

UNIVERSIDADE FEDERAL DE SÃO CARLOS
PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIA,
TECNOLOGIA E SOCIEDADE

Renato Aparecido Terezan de Moura

**ANÁLISE MULTIVARIADA DE CONTEÚDOS DA
CULTURA CIENTÍFICA VIA CIÊNCIA DE
DADOS SOCIAIS E APRENDIZADO DE
MÁQUINA RAG-LLM (Geração Aumentada por
Recuperação para Modelos Linguísticos de
Grande Dimensão)**

São Carlos-SP

2025

Renato Aparecido Terezan de Moura

**ANÁLISE MULTIVARIADA DE CONTEÚDOS DA
CULTURA CIENTÍFICA VIA CIÊNCIA DE DADOS
SOCIAIS E APRENDIZADO DE MÁQUINA RAG-LLM
(Geração Aumentada por Recuperação para Modelos
Linguísticos de Grande Dimensão)**

Tese apresentada ao Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade do Centro de Educação e Ciências Humanas, como requisito de defesa do título de Doutor em Ciência, Tecnologia e Sociedade.

Orientação do Prof. Dr. Cidival Morais de Sousa

São Carlos-SP
2025

de Moura, Renato Aparecido Terezan

Análise Multivariada de Conteúdos da Cultura Científica
via Ciência de Dados Sociais e Aprendizado de Máquina
RAG-LLM (Geração Aumentada por Recuperação para
Modelos Linguísticos de Grande Dimensão) / Renato
Aparecido Terezan de Moura -- 2025.
334f.

Tese de Doutorado - Universidade Federal de São Carlos,
campus São Carlos, São Carlos

Orientador (a): Cidoval Moraes de Sousa

Banca Examinadora: Ariadne Chloe Mary Furnival,
Cidoval Moraes de Sousa, Fernando Firmino da Silva,
Francisco Rolfsen Belda e Maria Teresa Miceli Kerbauy
Bibliografia

1. Ciência de dados sociais, análise de conteúdos e
aprendizagem de máquinas. I. de Moura, Renato
Aparecido Terezan. II. Título.

Ficha catalográfica desenvolvida pela Secretaria Geral de Informática
(SIn)

DADOS FORNECIDOS PELO AUTOR

Bibliotecário responsável: Arildo Martins - CRB/8 7180



UNIVERSIDADE FEDERAL DE SÃO CARLOS
Centro de Educação e Ciências Humanas
Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade

Folha de Aprovação

Defesa de Tese de Doutorado do candidato Renato Aparecido Terezan de Moura, realizada em 28/03/2025.

Comissão Julgadora:

Prof. Dr. Cidoval Moraes de Sousa (UEPB)

Profa. Dra. Ariadne Chloe Mary Furnival (UFSCar)

Profa. Dra. Maria Teresa Miceli Kerbauy (UFSCar)

Prof. Dr. Francisco Rolfsen Belda (UNESP)

Prof. Dr. Fernando Firmino da Silva (UEPB)

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade.

Dedico este trabalho à minha esposa e companheira Márcia Adriana Pierini, minha mãe Luzia Terezan de Moura e meu pai Lázaro Leite de Moura (ambos "In Memoriam") e meu filho Caio Pierini Coelho.

Agradecimentos

Inicialmente agradeço de maneira verdadeiramente profunda à minha mulher-esposa-companheira e musa Adriana Pierini, que me inspirou e cuidou do meu ser em diversos sentidos, durante os cinco anos do doutorado e nas dinâmicas da pesquisa, vivendo cada etapa e momento importante.

Parafraseando duas músicas conhecidas, reafirmo: “ Ela une todas as coisas e eu não sei viver sem ela”. Te amo Adri !

Ao meu filho Caio Pierini Coelho que me ouviu muito e suportou minhas reclamações, além de me ajudar a testar o TeseBot antes da defesa. Obrigado filho, vocês me inspiram a seguir pesquisando e estudando.

Nesta seara da luta pela ciência e comunicação pública de qualidade, agradeço imensamente mesmo, e de maneira efusiva, ao meu orientador Cidoval Moraes de Souza, o qual me possibilitou enxergar as ciências e a sociedade com novos olhares humanistas e desideologizados, mas nunca despolitizados. O “camarada” Cidoval me fez acreditar novamente na possibilidade de aproximar a sociedade das universidades pelo caminho da troca de saberes genuínos em busca das soluções viáveis e sustentáveis, econômica e ambientalmente.

A presença do professor Cidoval na trajetória foi essencial e determinante na objetividade científica e epistemológica da pesquisa, sem a qual não alcançaria as premissas de que a tese fosse original, inédita e até mesmo inovadora. Serei sempre grato pela sua acolhida e amizade únicas. Saúde e paz !

Agradeço ao sistema público de educação brasileiro, que entrega ensino de qualidade, realiza pesquisa básica e avançada e oferece ações de extensão eficazes que alcança comunidades, atendendo premissas de letramento e difusão científica ao povo brasileiro. O serviço educacional prestado pelo estado é comparável, em termos de alcance e universalidade, ao Sistema Único de Saúde (SUS), embora estejam em diferentes categorias, possuindo muitos problemas e contradições.

O cidadão brasileiro que obteve acesso ao ensino público fundamental, pode estudar gratuitamente até a pós-graduação, como foi o meu caso, em universidades mais inclusivas e com programas de permanência estudantil que viabilizam a vida acadêmica de milhões de brasileiros.

Também agradeço à Universidade Estadual Paulista (UNESP) que possibilitou, através de seu regimento, que eu me afastasse do trabalho por alguns dias da semana sem prejuízo de vencimentos, o que viabilizou o desenvolvimento da pesquisa e da tese, me levando a acreditar que as universidades públicas brasileiras são instituições que promovem ações de igualdade e inclusão inimagináveis, até mesmo em países economicamente mais desenvolvidos.

Por fim, agradeço imensamente meus irmãos Sérgio Domingos de Oliveira, Edna Aparecida de Moura Arcuri, Marco Antonio de Moura e Maria Chrystina Moura Andrade que me apoiaram na fase de defesa e após a mesma. Obrigado

“Estamos apenas no começo da ciência e muito distantes do tempo em que se acreditava possível descrever todo o universo nos termos de poucas leis fundamentais... Cabe às futuras gerações construir uma nova ciência que incorpore todos os aspectos do macro e do microcosmo, pois por enquanto, a ciência continua em sua infância.”

(Ilya Prigogine, in Ciência, Razão e Paixão)

Resumo

Orientados por enfoques epistemológicos interdisciplinares do campo CTS (Ciência Tecnologia e Sociedade), investigamos dinâmicas comunicacionais em dois projetos de uma universidade pública, utilizando uma abordagem híbrida que combina Análise de Conteúdo com Aprendizado de Máquina e Ciência de Dados Sociais. Coletamos e analisamos acervos de sites e mídias digitais, incluindo interações em chats, comentários e transmissões de eventos. Nossa pesquisa investigou a eficácia das dinâmicas de cultura científica, o alcance público dos projetos e revelou idiosincrasias dos usuários por meio da correlação de variáveis quantitativas e qualitativas. Os dados foram coletados e minerados com técnicas estatísticas e de Processamento de Linguagem Natural (PLN) no MAXQDA, e complementados por análises multivariadas. Para analisar conteúdos audiovisuais e textuais dos canais de streaming, desenvolvemos pipelines de aprendizado de máquina com RAGs-LLMs (Geração Aumentada por Recuperação para Modelos Linguísticos de Grande Dimensão). Os resultados da tese são apresentados em painéis visuais (dashboards) analíticos no site tecnociencia.net.br, onde também implementamos um robô conversacional (Tesebot), treinado com o conteúdo da pesquisa para responder perguntas. Se verifica na conclusão da tese, que sua corroboração é inevitável e carregada de certezas estatísticas e inferenciais que refletem e legitimam a complexidade da pesquisa na era da ciência pós-acadêmica, das big-techs, das redes sociais e da inteligência artificial. As estratégias metodológicas possibilitaram compreender as variáveis e instâncias estatísticas e relacionais envolvidas na observação dinâmica da cultura científica nos projetos de difusão pesquisados.

Palavras-chave: ciência pós-acadêmica, tecnociência, ciência de dados sociais, não linearidade, redes complexas, aprendizagem de máquinas, análise de conteúdos, máquinas de Boltzmann, grafos de conhecimento

Abstract

Guided by interdisciplinary epistemological approaches from the field of Science, Technology, and Society (CTS), we investigated communication dynamics in two projects at a public university, using a hybrid approach that combines Content Analysis with Machine Learning and Social Data Science. We collected and analyzed collections from websites and digital media, including interactions in chats, comments, and event broadcasts. Our research demonstrated the effectiveness of scientific culture dynamics, the public reach of the projects, and revealed user idiosyncrasies through the correlation of quantitative and qualitative variables. The data was collected and mined using statistical and Natural Language Processing (NLP) techniques in MAXQDA software, and complemented by multivariate analyses. To analyze audiovisual and textual content from streaming channels, we developed machine learning pipelines with RAGs-LLMs (Retrieval-Augmented Generation for Large Language Models). The results of the thesis are presented in visual and analytical dashboards on the website tecnociencia.net.br, where we also implemented a conversational robot (Tesebot), trained with the research content to answer questions. The conclusion of the thesis shows that its corroboration is inevitable and loaded with statistical and inferential certainties, which reflect and legitimize the complexity of research in the era of post-academic science, big tech, social networks, and artificial intelligence. The methodological strategies made it possible to understand all the variables and statistical and relational instances involved in the dynamic observation of scientific culture in the dissemination projects researched.

Keywords: post-academic science, technoscience, social data science, nonlinearity, complex networks, machine learning, content analysis, Boltzmann machines, knowledge graphs

Lista de figuras

Figura 1 - Comparativo de remuneração de ADS a sites de desinformação	17
Figura 2 - Índice de confiança comparativo antes e depois da pandemia	24
Figura 3 - Importância da ciência - Índice do Estado da Ciência	25
Figura 4 - Índice de confiança por fonte de informação (IC) de 2019 e 2023	25
Figura 5 - Percentual citando instituição de pesquisa e cientista brasileiro importante	26
Figura 6 - Níveis de dificuldade de contextualizar o conhecimento científico.....	26
Figura 7 - Frequência das fontes que os entrevistados buscam informações de C&T	27
Figura 8 - Fontes de informação sobre ciência e tecnologia no Brasil	28
Figura 9 - Nível de letramento científico por escolaridade no Brasil.....	30
Figura 10 - Fluxograma dos ciclos de Ciência dos Dados	39
Figura 11 - Hierarquia DIKW, do inglês Data-Information-Knowledge-Wisdom.....	50
Figura 12 - Síntese dos Modelos de Comunicação Pública da Ciência.....	52
Figura 13 - Conceitos sobre tecnociência solidária de Renato Dagnino	56
Figura 14 - Rede de citações em ciências sociais computacionais	65
Figura 15 - Gráfico de taxas falsas de correspondência em reconhecimento facial ..	73
Figura 16 - Amostra e vieses publicadas em matéria sobre reconhecimento facial ...	76
Figura 17 - Indicadores da pobreza e representatividade da população negra	81
Figura 18 - Neuroanatomia e conectividade do verme <i>C. elegans</i>	208
Figura 19 - Controlando a rede neural de <i>C. elegans</i>	209
Figura 20 - Ilustração sobre o “Demônio de Maxwell”	212
Figura 21 - Separação linear em 2 classes	206
Figura 22 - Plotagem de Um Gradiente Descendente	216
Figura 23 - Representação gráfica do problema XOR no Perceptron linear	218
Figura 24 - Simulação de crescimento populacional por função de mapa logístico .	222
Figura 25 - Programa Python com mapa logístico indicando caos a partir de 3,56 ..	223
Figura 26 - Comparativos de grafos de algoritmos de aprendizado de máquina	227
Figura 27 - Perceptron multicamada e resolução MLP no plano cartesiano	228
Figura 28 - Resolução do problema XOR no Perceptron Multicamada	228
Figura 29 - Comparativo entre Modelos de abordagem para redes neurais	229
Figura 30 - Grafo de Circuito Euleriano	230
Figura 31 - Grafo e fórmula do Número de Erdős	231
Figura 32 - Diferenças entre tipos de redes.....	232
Figura 33 - Tela do site com a transcrição do vídeo analisado e incorporado.....	253
Figura 34 - Tela do site com nuvem e tabela de frequência de palavras	253
Figura 35 - Figura 35 - Tela do site com análises de IA sobre a transcrição	254
Figura 36 - Tela do site com nuvem de palavras e análises de IA sobre comentários	255
Figura 37 - Tela do site com Análise de sentimentos dos comentários	255
Figura 38 - Tela do site com nuvem de palavras e tabela de frequências do chat ...	256
Figura 39 - Tela do site com Análise de sentimentos e de discurso do chat.....	258
Figura 40 - Captura do Google Trends sobre Deepseek em 01/2025.....	295

Lista de quadros

Quadro 1 - Cinco notícias falsas mais disseminadas sobre as vacinas	19
Quadro 2 - Bases de dados estatísticas fundamentais	34
Quadro 3 - Slides com fotogramas retirados do documentário da Padoka	60
Quadro 4 - Compara aspectos das Ciências Sociais Computacionais e de Dados	62
Quadro 5 - Níveis de aplicação em projetos de CSC	68
Quadro 6 - Definições de características das origens de inscrição (CSD)	97
Quadro 7 - Representação estatística de campos dos dataframes (CSD)	97
Quadro 8 - Variáveis analisadas, correlações de forças (CSD)	108
Quadro 9 - Correlações entre idade, visualizações e tempo de exibição	118
Quadro 10 - Plataformas da RAPLE e finalidades	128
Quadro 11 - Tipos de correlação das variáveis por idioma	137
Quadro 12 - Fonte primárias de multicolinearidade	141
Quadro 13 - Equações dos modelos de regressão	142
Quadro 14 - Diferenças entre modelos de regressão	166
Quadro 15 - Diferenças entre modelos de regressão logística alternativos	167
Quadro 16 - Descrição das fórmulas estatísticas descritivas	168
Quadro 17 - Comparativo básico entre Constante de Boltzmann e Teoria da Informação	213
Quadro 18 - Pioneiros redes neurais e armazenamento de informação em redes dinâmicas	219
Quadro 19 - Premissas das redes	232
Quadro 20 - Diferenças conceituais e variações entre as topologias das redes	233
Quadro 21 - Características das redes complexas	234
Quadro 22 - Fluxo de dados em um pipeline RAG básico	238
Quadro 23 - Três requisitos base da análise de conteúdo categorial	240
Quadro 24 - Fluxograma de análise de conteúdo proposto por Bardin	240
Quadro 25 - Detalhamento das etapas de AC em 3 etapas	244
Quadro 26 - Comparativo entre Bardin e Berelson (o autor)	245
Quadro 27 - Interface visual do MAXQDA exibindo codificação dos títulos audiovisuais ...	247
Quadro 28 - Livro de Códigos do Canal Ciências Sociais em Diálogo	248
Quadro 29 - Resumo de IA(LLM/RAG) do Livro de Códigos no MAXQDA (Bedrock Claude)	251
Quadro 30 - Respostas comparadas de RAGs (Base de Conhecimento própria + LLMs)	259
Quadro 31 - Resumo detalhado comparado de RAGs (Base de Conhecimento própria+LLM)	260
Quadro 32 - Pergunta sobre papel dos militares via RAGs + LLMs	260
Quadro 33 - Pergunta sobre a guerra via RAGs + LLMs	261
Quadro 34 - Resumo dos Comentários via RAGs + LLMs	261
Quadro 35 - Pergunta sobre a guerra via RAGs + LLMs	261
Quadro 36 - Resumo do chat durante a live via RAGs + LLMs	262
Quadro 37 - Análise de sentimentos comparadas do chat via RAGs + LLMs	262
Quadro 38 - Análise do discurso comparadas do chat via RAGs + LLMs	263
Quadro 39 - Respostas comparadas sobre os assuntos do texto de RAGs + LLMs	265
Quadro 40 - Respostas comparadas sobre resumo do texto de Abrucio RAGs + LLMs	267
Quadro 41 - Resposta sobre o estado de bem estar social de RAGs + LLMs	268
Quadro 42 - Respostas comparadas de resumos detalhados do texto Abrucio RAGs + LLM:	268
Quadro 43 - Comparativo análises de sentimento dos comentários on demand Abrucio	269
Quadro 44 - Comparativo resumo do chat na live de Abrucio	270
Quadro 45 - Comparativo análise de sentimento do chat na live de Abrucio - Gemini/Claude	271
Quadro 46 - Comparativo análise de discurso do chat na live de Abrucio - Gemini/Claude	272

Lista de tabelas

Tabela 1 - Serviços multiplataforma de difusão e cultura científica da FCLAr.....	29
Tabela 2 - Distribuição de visualização das variáveis do canal CSD	92
Tabela 3 - Variância entre médias de Visualizações (CSD).....	91
Tabela 4 - Retenção de inscritos por origem da inscrição do canal CSD	107
Tabela 5 - Mapa de calor dos 22 vídeos mais visualizados (CSD)	105
Tabela 6 - Análises Descritivas de Visualizações, Usuários ativos e Contagem de eventos	133
Tabela 7 - Valores de Aplicação do fator de inflação da variância (VIF) (RAPLE).....	142
Tabela 8 - Análise descritiva sobre a proporção dos tipos de usuários (RAPLE).....	168
Tabela 9 - Índice do uso de navegadores no site das RAPLE	150
Tabela 10 - Fabricantes e Usuários Ativos.....	151
Tabela 11 - Registros excluídos da amostra de distribuição (RAPLE)	152
Tabela 12 - Métricas descritivas das variáveis analisadas (RAPLE)	155
Tabela 13 - Tabela de variáveis por tipo de acesso (RAPLE)	158
Tabela 15 - Mapa de Calor dos usuários ativos por idioma e país (RAPLE)	161
Tabela 16 - Países com língua portuguesa como língua oficial (RAPLE)	162
Tabela 17 - Capilaridade dos idiomas dos usuários que acessam (RAPLE)	163
Tabela 18 - Equações dos modelos de regressão (RAPLE).....	165
Tabela 19 - Métricas descritivas de variáveis correlacionadas (RAPLE)	169
Tabela 20 - Matriz de correlação de variáveis de engajamento (RAPLE)	172
Tabela 21 - Tabela com métricas sobre estados, regiões, distritos ou províncias (RAPLE) .	174
Tabela 22 - Comparativo de engajamentos e tempo das cidades com maior acesso (RAPLE)	183
Tabela 23 - Métricas comparativas entre as cidades que mais acessam a RAPLE	184
Tabela 24 - Métricas e mapa de calor por sistema operacional (RAPLE)	193
Tabela 25 - Recorte temporal da coleta de dados nos projetos de extensão analisados .	275
Tabela 26 - Métodos alternativos de regressão logística e acurácia	280

Lista de gráficos

Gráfico 1 - Visualização de conteúdos do canal Ciências Sociais em Diálogo (CSD)	84
Gráfico 2 - Comparativo de acessos entre inscritos e não inscritos do canal (CSD)	85
Gráfico 3 - Visualizações por data e inscrição no Youtube (CSD)	86
Gráfico 4 - Visualizações entre inscritos, não inscritos e usuários recorrentes (CSD)	87
Gráfico 5 - Dispersão da correlação de Pearson entre inscritos e visualizações CSD	88
Gráfico 6 - Série temporal dos resíduos do canal CSD	89
Gráfico 7 - Dispersão da relação entre inscritos e resíduos do canal CSD	90
Gráfico 8 - Comparativo de visualizações X origem do tráfego do canal CSD	91
Gráfico 9 - Comparativo de tempo de exibição X origem do tráfego do canal CSD.....	92
Gráfico 10 - Origem do tráfego de visualizações do canal CSD	92
Gráfico 11 - Visualizações por origem de tráfego em série temporal (CSD)	93
Gráfico 12 - Visualizações por status de inscrição em série temporal (CSD).....	93
Gráfico 13 - Histograma de Visualizações por status de inscrição (CSD)	94
Gráfico 14 - Histograma de Visualizações por origem de tráfego (CSD)	94
Gráfico 15 - Retenção e perdas de inscritos por modalidade de origem (CSD)	96
Gráfico 16 - Correlação das visualizações por espectadores recorrentes (CSD)	98
Gráfico 17 - Dispersão das visualizações por espectadores recorrentes (CSD)	98
Gráfico 18 - Visualizações em série temporal (CSD)	99
Gráfico 19 - Série temporal sobre espectadores recorrentes (CSD)	99
Gráfico 20 - Total de inscrições obtidas e visualizações por modalidade de exibição (CSD)	100
Gráfico 21 - Percentual por modalidade de exibição (CSD)	101
Gráfico 22 - Origens de visualização mais frequentes (CSD)	102
Gráfico 23 - Rosca sobre plataformas e sites que mais originam visualizações (CSD)	102
Gráfico 24 - Decomposição da série Temporal de Compartilhamentos (CSD)	103
Gráfico 25 - Decomposição da série Temporal de Visualizações (CSD)	104
Gráfico 26 - Dispersão dos dados com a linha de regressão polinomial (CSD)	105
Gráfico 27 - Gráfico de dispersão de visualizações por demanda e live (CSD)	106
Gráfico 28 - Rosca sobre visualizações por disponibilidade (CSD)	107
Gráfico 29 - Histograma de diff sobre as médias de visualizações (CSD)	107
Gráfico 30 - Mapa de calor das Visualizações, Duração Média e Tempo de Exibição (CSD)	108
Gráfico 31 - Rosca sobre Visualizações sob demanda (on demand) e ao vivo (live) (CSD).	110
Gráfico 32 - Visualizações, Tempo de Exibição e Duração Média por Dispositivo (CSD) ..	111
Gráfico 33 - Linhas empilhadas em série temporal de visualizações por dispositivo (CSD)	112
Gráfico 34 - Linha de regressão indicando inviabilidade do modelo de regressão linear (CS	112
Gráfico 35 - Linha de regressão exponencial (CSD)	113
Gráfico 36 - Proporção de visualizações e tempo de exibição por dispositivo (CSD)	114
Gráfico 37 - Acessos do canal Ciências Sociais em Diálogo por idade (CSD)	115
Gráfico 38 - Dispersão das Visualizações por Tempo de Exibição (CSD)	116
Gráfico 39 - Correlação entre variáveis dependentes e Tempo de Exibição (CSD)	117
Gráfico 40 - Percentual de Visualizações por Gênero e Idade (CSD)	118
Gráfico 41 - Dispersão entre idade, gênero e visualizações (CSD)	120
Gráfico 42 - Rosca com termos mais buscados no Youtube (CSD)	121
Gráfico 43 - Dispersão das correlações entre variáveis (CSD)	123
Gráfico 44 - Modelo regressivo preditivo x Visualizações reais (CSD)	123
Gráfico 45 - Visualizações e tempo de exibição das playlists (CSD)	124
Gráfico 46 - Série temporal sobre desempenho anual do canal (CSD)	126
Gráfico 47 - Visualizações x Usuários Ativos (CSD)	130
Gráfico 48 - Visualizações x Contagem de Eventos (RAPLE).....	131
Gráfico 49 - Usuários Ativos x Contagem de Eventos (RAPLE)	131
Gráfico 50 - Regressão Linear entre Visualizações e Contagem de Eventos (RAPLE) ...	132
Gráfico 51 - Regressão não Linear entre Usuários Ativos e Contagem de Eventos (RAPLE)	133
Gráfico 52 - Histograma de visualizações das páginas do site (RAPLE)	134
Gráfico 53 - Histograma de usuários ativos das páginas do site (RAPLE)	134
Gráfico 54 - Histograma de eventos dos usuários nas páginas do site (RAPLE)	135
Gráfico 55 - Boxplot de Contagem de Eventos por idioma dos Títulos (RAPLE)	137

Gráfico 56 - Boxplot de Visualizações de Usuário Ativo por Idioma (RAPLE)	138
Gráfico 57 - Boxplot de Contagem de Usuários Ativos por idioma (RAPLE)	139
Gráfico 58 - Dispersão da regressão linear Visualizações e Usuários Ativos (RAPLE)	140
Gráfico 59 - Dispersão da regressão não linear (RAPLE)	140
Gráfico 60 - As 6 principais origens de tráfego (RAPLE)	144
Gráfico 61 - Rosca das principais origens de tráfego (RAPLE)	145
Gráfico 62 - Mapa de calor e correlação das variáveis da tabela (RAPLE)	146
Gráfico 63 - Porcentagem de Tráfego por Plataforma (RAPLE)	147
Gráfico 64 - Comparativo entre navegadores web na Raple e no mundo (RAPLE)	149
Gráfico 65 - Rosca sobre os Modelos de dispositivos mais usados (RAPLE)	151
Gráfico 66 - Distribuição por Modelos de dispositivos mais usados (RAPLE)	151
Gráfico 67 - Mapa de Calor da Matriz de Correlação das variáveis (RAPLE)	153
Gráfico 68 - Usuários ativos e sessões por plataforma (RAPLE)	155
Gráfico 69 - Média/Mediana de Usuários Ativos,engajamentos, Sessões e Eventos (RAPLE)	156
Gráfico 70 - Usuários ativos por idioma (RAPLE)	159
Gráfico 71 - Mapa de Usuários ativos por país e idioma (RAPLE)	160
Gráfico 72 - Usuários ativos e sessões por plataforma (RAPLE)	162
Gráfico 73 - Mapa de calor e correlações entre variáveis geográficas e idiomáticas (RAPLE)	163
Gráfico 74 - Mapa mundi com Usuários ativos por região, estado ou província (RAPLE) .	168
Gráfico 75 - Usuários ativos por sessões engajadas (RAPLE)	170
Gráfico 76 - Dispersão de Usuários ativos por contagem de eventos (RAPLE)	170
Gráfico 77 - Dispersão de Engajamento por contagem de eventos (RAPLE)	170
Gráfico 78 - Dispersão de Usuários ativos por Taxa de Engajamento (RAPLE)	171
Gráfico 79 - Dispersão de Sessões engajadas por contagem de eventos (RAPLE)	171
Gráfico 80 - Dispersão de Sessões engajadas por taxa de engajamento (RAPLE)	172
Gráfico 81 - Histograma das quatro variáveis relacionadas a engajamento (RAPLE)	173
Gráfico 82 - Rosca de usuários ativos por estado, região ou província (RAPLE)	174
Gráfico 83 - Mapa mundial com Contagem de Eventos por Região (RAPLE)	180
Gráfico 84 - Mapa sobre Taxa de engajamento por cidade (RAPLE)	185
Gráfico 85 - Mapa sobre o Tempo médio de engajamento por cidade (RAPLE)	186
Gráfico 86 - Minutos em sessões engajadas por cidade (RAPLE)	187
Gráfico 87 - Mapa sobre Sessões engajadas por cidade (RAPLE)	188
Gráfico 88 - Distribuição da Taxa de Engajamento (binned) (RAPLE)	189
Gráfico 89 - Mapa sobre Usuários ativos no mundo (RAPLE)	190
Gráfico 90 - Usuários ativos na América Latina (RAPLE)	191
Gráfico 91 - Engajamentos, Tempo Médio e eventos por usuários ativo (RAPLE)	192
Gráfico 92 - Métricas comparativas entre sistemas operacionais mais usados (RAPLE)	193
Gráfico 93 - Eventos e usuários por dispositivo (RAPLE)	194
Gráfico 94 - Eventos e usuários por tipo de interação (RAPLE)	195
Gráfico 95 - Roscas de Eventos e usuários por tipo de interação (RAPLE)	197
Gráfico 96 - Nuvem de palavras dos títulos de transmissões do canal	250
Gráfico 97: Tendências de palavras dos títulos de transmissões do canal	252

Sumário

1 FUNDAMENTOS E ESTRUTURA DA TESE

1.1 Introdução	17
1.2 Justificativa	25
1.3 Pergunta da pesquisa	31
1.4 Objetivos gerais e específicos	32
1.5 Ambiente da pesquisa	32
1.6 Plano Metodológico	33
1.7 Metodologia de Ciência de Dados Sociais.....	36
1.8 Estrutura do trabalho	40

2 CIÊNCIA PÓS-ACADÊMICA

42

2.1 A Ciência Pós-Acadêmica de Ziman e sua concepção	43
2.2 A <i>Práxis</i> Científica	45
2.3 Expertise e Controvérsias em Collins e Evans.....	46
2.4 O nó górdio: Modernidade, conhecimento e política	48
2.5 Do que é feita a tecnociência para Latour ?	50
2.6 Conhecimento e sabedoria popular	51
2.7 A apropriação social da ciência e da tecnologia	53
2.8 A tecnociência solidária de Renato Dagnino	55
2.9 Tecnologia social ou tecnociência solidária ?	58

3 CIÊNCIA DE DADOS SOCIAIS E ANÁLISES DESCRITIVAS

63

3.1 A origem na Ciência Social Computacional (CSC).....	65
3.2 Ciências Sociais Computacionais na onda positivista.....	67
3.3 A multidisciplinaridade da CSC.....	69
3.4 Um breve resumo sobre algoritmos.....	72
3.5 Aspectos fundamentais dos algoritmos de Aprendizado de Máquina	73
3.6 Vieses racistas e sexistas em algoritmos de reconhecimento facial	74
3.7 Contextos que são diretamente afetados pela Inteligência Artificial	77
3.8 Impactos sociopolíticos de algoritmos no cotidiano dos brasileiros	80
3.9 Tecno-solucionismo	82
3.10 Fundamentos da Ciência de Dados Sociais	82
3.11 Análises Descritivas CSD	87
3.11.1 - Projeto do canal de streaming Ciências Sociais em Diálogo	87
3.11.2 - Correlações entre engajamento e origem do tráfego	94
3.11.3 - Perda e obtenção de inscritos	98
3.11.4 - Relação entre Visualizações e Espectadores Recorrentes	100
3.11.5 - Inscrições e visualizações obtidas por modalidade de acesso ..	103
3.11.6 - Análise de correlações entre visualizações ao vivo e sob demanda	109
3.11.7 - Análise de acessos por plataforma e dispositivo	114
3.11.8 - Análise de Correlações entre Visualizações e Idade	118
3.11.9 - Analisando as correlações entre Idade, Gênero e Visualizações	122
3.11.10 - Termos mais buscados no Youtube	123
3.11.11 - Visualizações e Tempo de exibição (horas) das Playlists	124
3.11.12 - Visualizações e Horas de Exibição por ano	126
3.12 Análises Descritivas RAPLE	129
3.12.1 - Projeto da Rede de Apoio aos Professores de Língua Estrangeira	129
3.12.2 - Análises e Correlações	130
3.12.3 - Correlações de variáveis dos dados de acesso	131
3.12.4 - Análise Descritiva por Idioma	137
3.12.5 - Origens do Tráfego do site da RAPLE	145
3.12.6 - Proporcionalidade de Recorrência dos Usuários	148
3.12.7 - Usuários ativos por navegador web	150

3.12.8 - Sessões por origem de tráfego	154
3.12.9 - Métricas de Usuários Ativos por Geolocalização e Idioma	159
3.12.10 - Análise por Estado, Província ou Região	168
3.12.11 - Eventos por Região	181
3.12.12 - Análise de Distribuição por cidades	182
3.12.13 - Padrões de Engajamento nas Cidades Analisadas	183
3.12.14 - Minutos Totais em Sessões Engajadas	187
3.12.15 - Usuários Ativos	189
3.12.16 - Análise da Popularidade e Engajamento por Sistema Operacional	192
3.12.17 - Análises de Tipo de eventos por dispositivo	195

4 ANÁLISE DE CONTEÚDOS EM REDES COMPLEXAS POR APRENDIZAGEM DE MÁQUINAS RAG-LLM

200

4.1 Ciências da complexidade	203
4.2 Ciência das Redes	206
4.3 Complexidade, clusters e hubs	207
4.4 Redes neurais reais orgânicas (conectomas)	208
4.5 O Conexionismo (conectoma) e o Simbolismo (perceptrons)	211
4.6 A máquina de Boltzmann é um modelo estocástico gerador de grafos ..	231
4.7 Análises de conteúdos auxiliado por Aprendizado de Máquina RAG/LLM	253
4.8 Análises de AM/RAG da Live de Piero Leirner.....	258
4.9 Análises de AM/RAG da Live de Fernando Abrucio	265

5 ESTRATÉGIAS METODOLÓGICAS E RESULTADOS

5 As demandas de difusão e letramento científico	274
5.1 Estratégias metodológicas e resultados	233
5.2 Variáveis e métricas analisadas	276
5.3 - Análises Comparativas e Regressões.....	280
5.4 - Análise de Séries Temporais	282
5.5 - Análises Descritivas	282
5.6 - Análises Qualitativas e de Conteúdo	

6 CONCLUSÃO DA TESE

287

6.1 - Considerações Finais	292
----------------------------------	-----

7 REFERÊNCIAS BIBLIOGRÁFICAS

7 Referências Bibliográficas de todos os capítulos	297
--	-----

8 APÊNDICES

1 FUNDAMENTOS E ESTRUTURA DA TESE

1.1 Introdução

Após mais de 2 anos e meio de isolamento social imposto pela pandemia Covid-19, pesquisadores de diferentes campos da ciência puderam analisar os impactos das rápidas mudanças e disrupturas ocasionadas durante e após esse evento em nível global, que afetou bilhões de pessoas no mundo todo. Os pesquisadores dos estudos sociais da ciência e da tecnologia obtiveram uma oportunidade ímpar de observar as dinâmicas de relacionamento e níveis de confiança entre a sociedade e as instituições de pesquisa científica, além de analisar o impacto da emergência de novos fatores como a desinformação remunerada, a flexibilização da presença física e a consolidação das tecnologias informacionais como ferramentas de mediação do trabalho, da saúde, da cidadania e até mesmo das relações socioafetivas.

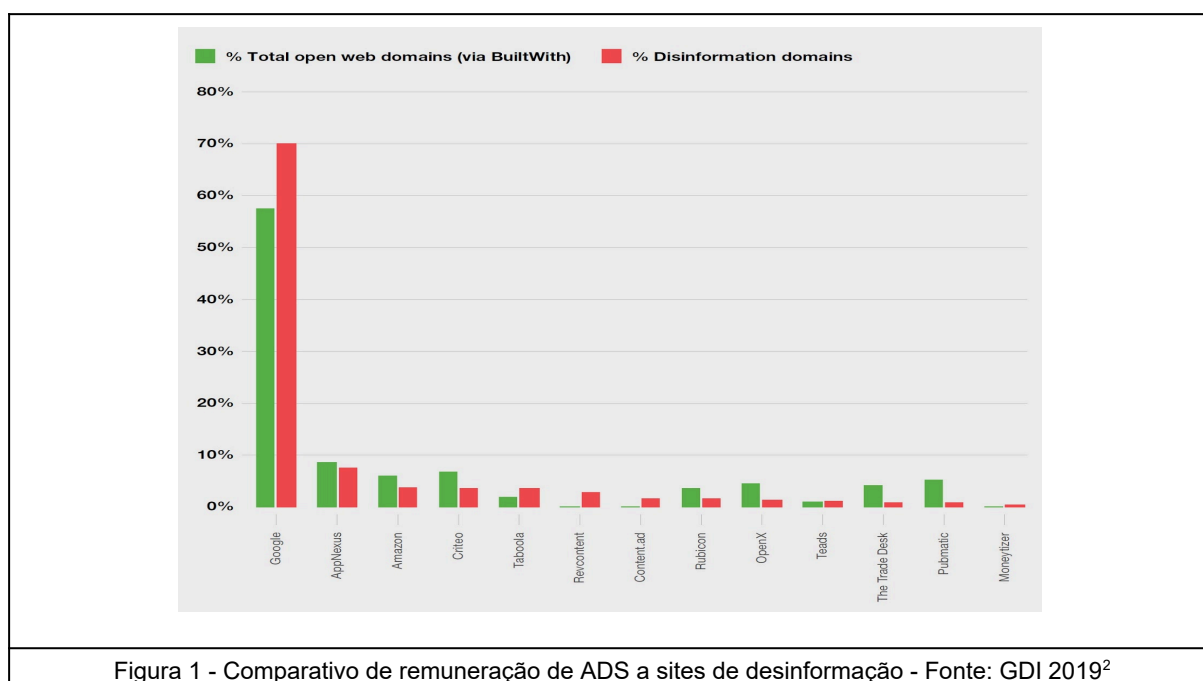
O impacto da flexibilização de tempo e lugar durante a pandemia, alterou rapidamente as concepções comunicacionais da sociedade globalizada (IANNI, 1994, 1998) conectada em redes informacionais (CASTELLS, 1999), com a imposição da presença remota em todas as instâncias da vida civil, como trabalho, escola, consultas, audiências, defesas de teses e mesmo nos relacionamentos interpessoais afetivos. Num espaço relativamente curto de tempo, todas as faixas etárias tiveram que se adaptar ao uso intensivo e/ou extensivo de tecnologias de informação e comunicação para se relacionarem com seus familiares, amigos, empregadores, fornecedores e o Estado, em suas modalidades de atendimento e prestação de serviço ao cidadão.

A nova realidade, imposta pelas modalidades virtuais de atendimento presencial remoto e sua flexibilização nas relações interpessoais e interinstitucionais, se acentuaram após a pandemia. Elas resultaram no que podemos inferir como prova empírica da emergência da cultura algorítmica, ao concebermos que algoritmos rodando em redes informacionais se envolvem ou dominam tarefas consideradas tradicionalmente culturais que incluem escolhas, buscas, classificação e hierarquização de objetos, pessoas, ideias e lugares (ZUIN e ZUIN, 2018; BEZERRA, 2017; JUST, N. & LATZER, 2016; GILLESPIE, 2011).

A migração do espaço vivencial para a virtualidade das redes informacionais

acelerou a circulação de desinformação¹ e a transformou em uma camada do modelo de negócios das *big techs*, através de algoritmos de publicações de anúncios que remuneram páginas e sites que produzem e difundem desinformação (ORESQUES e CONWAY, 2010). As big techs não checam o conteúdo e propósito das páginas, apenas consideram quantas exibições e acessos resultaram daquele anúncio e remuneram, indiscriminadamente, a produção reiterada, sistemática e objetiva da agnotologia, quando a produção de desinformação é deliberada, sistemática e remunerada (PROCTOR e SCHIEBINGER, 2008).

Segundo o relatório Índice Global de Desinformação, as empresas Google (70%), Taboola (4%), Revcontent (3%), Content Ad (2%), Teads (1%) e Monetizer (1%), dominam juntas 81% do mercado de anúncios e remuneram valores significativamente maiores aos sites e páginas de desinformação quando comparados com os todos os outros sites classificados na comparação.



Muitas empresas transnacionais frequentemente tem seus anúncios exibidos junto a conteúdos que geram desinformação ou disseminam notícias falsas, como pode ser observado no relatório de 2019 do Global Disinformation Index (GDI, 2019 updated 2021). Somente naquele ano foram distribuídos US\$ 235.000.000,00

¹<https://www.disinformationindex.org/research/2019-9-1-the-quarter-billion-dollar-question-how-is-disinformation-gaming-ad-tech/>

² <https://www.unesco.org/en/world-media-trends/global-disinformation-index-gdi>

(duzentos e trinta e cinco milhões de dólares) pagos anualmente a mais de 20.000 sites de desinformação por anúncios de produtos globais e locais³.

Até mesmo o site de jornalismo da gigante de mídia estadunidense ABC News, mundialmente conhecido, foi duplicado durante meses por uma versão falsa com o mesmo design, logotipos e até uma URL muito parecida (ABCnews.com.co), que misturava notícias verdadeiras publicadas no site original com notícias falsas e desinformação sistematizada⁴. O proprietário do domínio que hospedava o site alegou que ganhava mais de US\$10.000 (dez mil dólares) por mês em publicidade. O site foi retirado do ar, mas seus últimos conteúdos falsos ainda podem ser vistos em <https://archive.ph/bv4yc>.

Ainda mais inusitada é a forma como as *big techs* tratam esse mercado obscuro como negócios dentro da normalidade, pois embora saibam que estimulam crimes de diversas naturezas, garantem a impunidade pelo sigilo das operações de inserção e remuneração de anúncios no ambiente da plataforma. Como são automatizadas por algoritmos, ignoram a legislação e criam uma espécie de vácuo institucional, onde a desinformação é muito bem remunerada e estimulada pelos maiores *players* do mercado de anúncios da internet (CASSINO et al., 2021).

Podemos inferir a emergência de uma face da cultura algorítmica que cria e mantém bolhas de desinformação em nível global, com empresas que atuam em mercados bem definidos e atende instituições públicas e particulares, governos democraticamente eleitos, partidos políticos e grupos religiosos que fazem uso efetivo e frequente de notícias falsas, desinformação e manipulação algorítmica (EMPOLI, 2022; O'NEIL, 2020; D'ANCONA, 2018).

As vacinas são um dos principais alvos da desinformação em nível global e a parcela da população mundial que negligencia a proteção preventiva, oferecida gratuitamente pelos estados-nações, cresce em ritmo acelerado e ocasiona o retorno de doenças já erradicadas e mortes por sarampo, difteria e coqueluche, além do aumento de casos de paralisia infantil.

Em 2021, no segundo ano da pandemia Covid-19, houve cerca de 9 milhões

³<https://www.disinformationindex.org/research/2019-9-1-the-quarter-billion-dollar-question-how-is-disinformation-gaming-ad-tech/>

⁴ <https://dbpedia.org/page/ABCnews.com.co> ABCnews.com.co era um site de notícias falsas que imitava a URL, o design e o logotipo do site da ABC News

de casos e 128 mil mortes por sarampo em todo o mundo, com surtos em 22 países. Os casos de sarampo no mundo aumentaram 79% nos primeiros dois meses de 2022, em comparação com o mesmo período de 2021 e mais de 40 milhões de crianças não vacinadas estão suscetíveis à crescente ameaça da doença, segundo relatório da Organização Mundial da Saúde (OMS) e do Centro de Controle e Prevenção de Doenças dos Estados Unidos (CDC)⁵.

Durante a pandemia global da Covid-19, as vacinas foram atacadas sistematicamente sem nenhuma prova concreta ou sequer indicação de estudos que pudessem comprovar a ineficácia ou efeitos adversos atribuídos à vacinação (CASSINO et al., 2021). Houve também a disseminação de mentiras absolutamente fantasiosas sobre a inoculação de outros patógenos e chips eletrônicos nas vacinas, chegando-se ao extremo de divulgarem amplamente que a vacina transformaria uma pessoa em jacaré, como dito pelo ex-presidente Jair Bolsonaro entre 2020 e 2021⁶.

A Agência Lupa⁷, de checagem de fatos, listou 5 das notícias falsas mais disseminadas sobre as vacinas, com argumentos infundados entrelaçados a termos científicos para confundir segmentos da população que não obtiveram letramento científico suficiente para detectar a desinformação. A agência cita estudos de psicologia que usam o termo “viés de confirmação”, para explicar como as pessoas tendem a acreditar mais em notícias que correspondem ao que elas pensam.

1 - A vacina contra a Covid-19 vai modificar o DNA dos seres humanos
2 - A vacina contra a Covid-19 tem chip líquido e inteligência artificial para controle populacional
3 - Imunizantes contra Covid-19 estão relacionados à transmissão de HIV
4 - Vacinas contra Covid-19 criam campo magnético no corpo de quem é imunizado
5 - CoronaVac não tem comprovação científica
Quadro 1: Cinco notícias falsas mais disseminadas sobre as vacinas: Fonte: Lupa

O letramento científico rudimentar da população brasileira colabora, em alguma medida (IBLC, 2017), para o êxito da desinformação em praticamente todos os segmentos da sociedade. Fica evidente diante desses fatos e proposições que a comunicação, o letramento e a difusão de conteúdos e conhecimentos científicos

⁵ <https://news.un.org/pt/story/2022/11/1805892> - Relatório da OMS e CDC

⁶ <https://twitter.com/folha/status/1370052452146352134>

⁷ <https://www.ufsm.br/midias/experimental/agencia-da-hora/2021/11/11/top-5-fake-news-mais-absurdas-sobre-a-vacina>

podem colaborar com o fortalecimento da cultura científica na sociedade e fazer frente aos desafios cognitivos na era das redes informacionais, na qual, ironicamente, estamos imersos em muita desinformação.

O conhecimento e a cultura científica podem estar presentes cotidianamente na vida das pessoas (VOGT, 2016) agregando ferramentas de difusão, divulgação e letramento científicos, entendendo-as como necessariamente complementares e desenvolvendo um olhar crítico, ao mesmo tempo, para a ciência e para a mídia. (PEZZO, in VOGT, GOMES, MUNIZ, 2018, p. 87 a 95).

Os desafios da cultura científica em redes informacionais são amplamente complexos, pois ninguém está imune aos *walled gardens*⁸ controlados por algoritmos e robôs que filtram informações e manipulam redes, direcionando os usuários aos conteúdos monetizados por anúncios e coleta de metadados com fins comerciais, comumente chamados de bolhas⁹ (BARRETO JÚNIOR e PELIZARI, 2019). Eles são realimentados por sites de notícias falsas (*fake news*), veiculadas sistematicamente por robôs em fazendas de celulares com perfis falsos que manipulam até eleições presidenciais¹⁰.

A pandemia impactou fortemente outros aspectos da vida em nosso país e alterou hábitos, conceitos e a cultura como um todo. No socioeconômico, as empresas prestadoras de serviço e comércio de todas as dimensões foram obrigadas a se adaptar rapidamente à nova economia informacional e à virtualização de atendimento para manterem seus fluxos operacionais. Essa mudança incrementou substancialmente a demanda pelos serviços terceirizados de entregas e acabaram oferecendo uma ocupação remunerada a muitas pessoas que perderam seus empregos formais durante o período, embora o Instituto de Pesquisas Econômicas Avançadas (IPEA) tenha detectado apenas uma continuação do crescimento acentuado dessas ocupações, que foi de 979,8% entre 2016 e 2019¹¹.

⁸ Jardim murado (*walled garden*) ambiente controlado a conteúdos e serviços baseados em rede do usuário, direcionando a navegação a uma seleção de materiais ou impedir o acesso a outros. <https://www.techtarget.com/searchsecurity/definition/walled-garden>

⁹ <https://www.indexlaw.org/index.php/revistadgnt/article/view/5856>

¹⁰ <https://pp.nexojournal.com.br/ponto-de-vista/2021/O-que-s%C3%A3o-plataformas-de-fazendas-de-cli-que-e-por-que-elas-importam>

¹¹ <https://www.ipea.gov.br/cartadeconjuntura/index.php/tag/entregadores-de-mercadoria/>

Essa mesma pesquisa mostra que 31% dos trabalhadores em atividade no setor de transporte de passageiros e de mercadorias no Brasil fazem parte do que foi convencionado como *Gig Economy*¹², que se refere a contratos de serviços informais e esporádicos que funcionam por meio de aplicativos ou faz uso deles em alguma escala (GÓES et al. in IPEA, 2023).

As relações socioeconômicas resultantes dessa aparente reacomodação de ramo ou modalidade de trabalho, se tornam realidades recorrentes no mundo todo e caracterizam uma disruptura na forma de mediar o trabalho, que passa a ser realizada exclusivamente de forma indireta e despersonalizada através de plataformas e aplicativos de empresas transnacionais de serviços informacionais, as chamadas *big techs*.

Toda a relação de trabalho ou serviço entre os profissionais (considerados pelas *big techs* como autônomos) e os fornecedores que contratam o atendimento ou entrega, são mediados por empresas, que em sua maioria, não regulam essa relação pela legislação civil ou trabalhista brasileira, embora estejam sujeitos às mesmas, como qualquer outra.

De fato, emergiu uma modalidade de serviço realizado de forma pessoal, indireta e precária que resulta da inação do estado e da sociedade em reconhecer as novas dinâmicas relacionais da economia informacional, pois se faz necessário encarar, com seriedade, a realidade algorítmica nas relações de trabalho, serviços e consumo.

Na cadeia produtiva da *Gig Economy*, os consumidores, os fornecedores e os prestadores de serviço são atendidos por algoritmos de empresas transnacionais que não apenas mediam as relações de compra, pagamento e entrega entre eles, mas também recrutam e selecionam esses profissionais e estabelecem níveis diferentes de relacionamento na mesma plataforma. É como se criassem uma espécie de vácuo institucional onde atuam, sem se incomodarem ou serem incomodados com o tipo de relação socioeconômica excludente e precária que estabelecem com os entregadores (GÓES et al. in IPEA, 2023).

As novas dinâmicas que a economia informacional consolidou, como previsto por Castells há mais de vinte anos em seu livro sobre a sociedade em rede,

¹² Termo sem tradução direta que pode significar o que no Brasil se chama de bico (gig work)

deslocaram a centralidade da mediação nas relações de trabalho em algumas categorias, do estado de direito e suas leis trabalhistas para o ambiente virtualizado, desterritorializado e ubíquo das redes informacionais.

As redes informacionais se entrelaçaram às redes reais de relacionamentos, serviços e comércio em novos arranjos sociotécnicos que recriam as relações humanas pela economia informacional e algorítmica (CASTELLS, 1999; MOROZOV, 2018, AMADEU DA SILVEIRA in Cassino et al., 2021; PRZEYBILOVICZ, 2019). Elas não apenas flexibilizam a presença física dos seres humanos, mas também mediam muitas de suas relações com as instituições e o estado, que dependem exclusivamente de diversas *big techs* para prestar seus serviços, coletar impostos, burocratizar as relações (PRZEYBILOVICZ, 2019) e vigiar o cidadão (BAUMAN, 2014).

Um exemplo prático dessa dependência e do deslocamento da centralidade do estado na mediação de direitos e acesso à justiça em 2019, foi a tentativa do Tribunal de Justiça do Estado de São Paulo de implantar uma plataforma digital de serviços na nuvem da *big tech* transnacional Microsoft, para atender as demandas de registro, tramitação e arquivamento de todas as ações do maior tribunal do país. O contrato que previa investimentos de mais de um bilhão de reais não considerava possíveis riscos institucionais e socioeconômicos de vazamentos de dados estratégicos e secretos de pessoas, empresas e instituições públicas e privadas brasileiras, atendidas pelos serviços jurisdicionais do tribunal (AMADEU DA SILVEIRA in Cassino et al., 2021).

A suspensão do contrato¹³ ocorreu somente após ação do Conselho Nacional de Justiça, que o considerou um risco enorme à segurança nacional e à privacidade, contrapondo frontalmente a Lei Geral de Proteção de Dados (Lei 13.709/18), em suas diretrizes de segurança da informação. A soberania nacional estava sendo entregue ingenuamente à uma empresa estadunidense, que teria total acesso a dados sensíveis de cidadãos ou instituições brasileiras, pela simples requisição de um juiz de qualquer instância daquele país, conforme está previsto em sua legislação.

¹³ <https://www.conjur.com.br/2020-mai-20/tj-sp-rescinde-contrato-13-bilhao-microsoft>

Esse exemplo do tribunal demonstra a dificuldade das instituições brasileiras em racionalizar o uso de tecnologias na prestação de serviços sensíveis e estratégicos, apenas em troca de simples economia financeira, ao não precisar manter e administrar *data centers*¹⁴, como alegado pelo desembargador que tentou implantar o projeto e não via problemas na flexibilização de segurança.

A cultura algorítmica torna-se altamente estratégica ao influenciar fortemente as bases da mediação humana em seus relacionamentos com o estado, as instituições privadas e também familiares e amigos. Na massificação da economia informacional é necessário compreender, dominar e contextualizar o uso de algoritmos, tornando-os uma questão central em todos os aspectos das relações humanas mediadas por ferramentas digitais, que filtram, analisam e classificam quais informações se pode acessar e quais dados serão coletados para realimentar essa cultura.

O negacionismo científico dinamiza o impacto das falsas notícias porque busca legitimar, falsamente, meras discordâncias de opinião sem correspondência com qualquer fundamento ou método científico. Eventualmente são tratadas como verdades inconvenientes de conspirações globais de manipulação da realidade ou como elementos de profecias heréticas, mas há uma produção em nível industrial de criação e distribuição de desinformação, que realimenta ciclicamente essa rede multidimensional de financiamento e faturamento da agnotologia como atividade regulamentada, regulada e lucrativa (D'ANCONA, 2018) .

A cultura científica (VOGT, 2016) disseminada como tecnociência (LATOUR, 2000) em uma concepção sociotécnica das redes complexas de produção e difusão científica (LATOUR, 2012) como proposta nesta tese, também pode aproximar as instituições científicas e universidades da sociedade brasileira, que admira e confia nos cientistas e professores em níveis superiores a jornalistas, líderes religiosos e escritores (IBLC, 2017; CGEE, 2023).

A visibilidade da produção científica e sua difusão pelos canais institucionais de comunicação social das universidades se torna essencialmente estratégico na aproximação com a população e enfrentamento do negacionismo científico. A cultura

¹⁴ centro de processamento de dados, também conhecido como data center, é um local onde estão concentrados os sistemas computacionais de uma empresa ou organização

científica desmitifica a ciência infalível e culturalmente distante da era mertoniana (ZIMAN, 2000), que passa a ser compreendida como atividade socioeconômica com dinâmicas e idiossincrasias naturais e humanas, sujeitas a reveses, crises, eleições, políticas e extrema competitividade, características das sociedades industrializadas e democracias capitalistas.

A democracia é originalmente tensa e recria suas estruturas em bases socialmente construídas de relacionamento humano. Na tecnociência, os grupos de expertise, reconhecidamente proficientes no método científico de cada campo, estabelecem os limites do questionamento legítimo além da mera controvérsia. A ciência não abre suas caixas-pretas (LATOUR, 2000) por mera discordância infundada, mas recria o método e restabelece suas bases, sempre que emerge uma verdadeira superação, ruptura ou tese comprovável que a legitime.

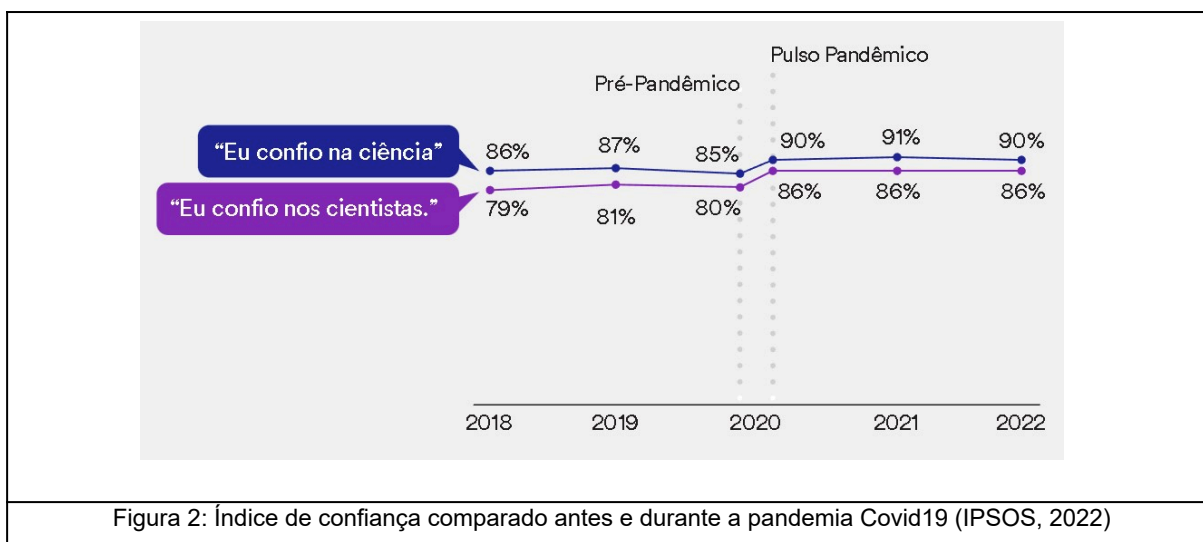
O rigor do método científico e sua reprodutibilidade, somente é possível pela divulgação aberta dos resultados em formato colaborativo entre os cientistas e laboratórios. Ela está fundamentada na legitimidade do empirismo ao coletar dados em campo, além da premissa de divulgação dos resultados em revistas ranqueadas e classificadas pela relevância e rigorosidade da sua revisão por pares.

Difícilmente, outro segmento da sociedade civil seria mais qualificado e adequado em contrapor a desinformação e o negacionismo científico do que a própria comunidade que produz ciência, comumente chamada como academia.

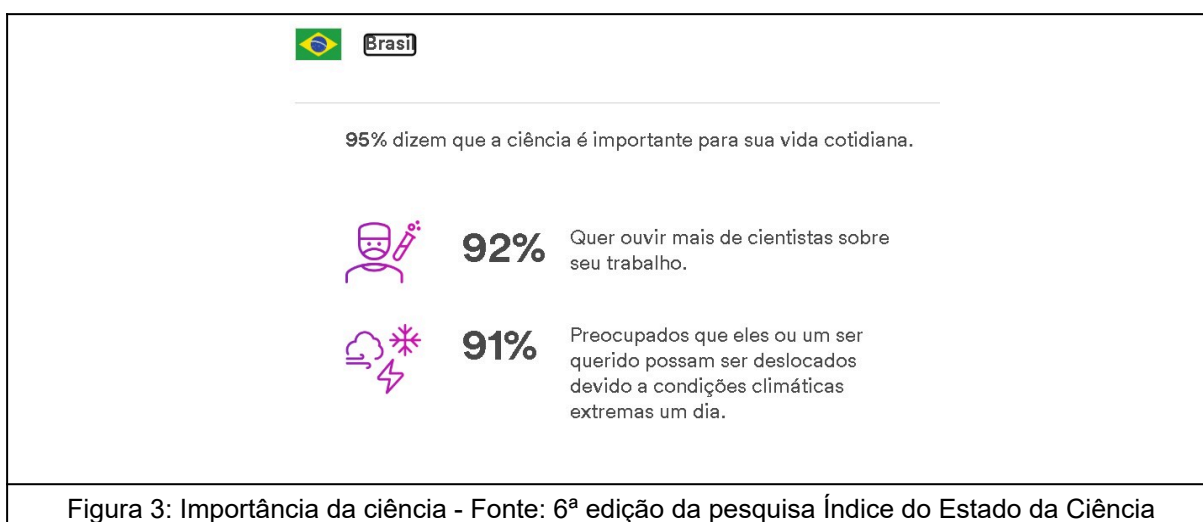
1.2 - Justificativa

A pesquisa ‘Confiabilidade Global’ do Ipsos¹⁵ de 2022, mediu a percepção de cidadãos brasileiros sobre a confiança nas principais profissões e 64% apontaram os professores como os mais confiáveis, seguidos dos cientistas (61%) e médicos (59%) . Na amostra de 28 países, o Brasil ocupa o segundo lugar no ranking mundial dos que mais confiam na categoria dos professores, sendo superado apenas pela China, que apresentou índice de confiança de 66%.

¹⁵<https://www.ipsos.com/sites/default/files/ct/news/documents/2022-07/Global%20trustworthiness%202022%20Report.pdf>



Em 2022, a 6ª edição da pesquisa Índice do Estado da Ciência, da transnacional 3M, revelou que 92% dos brasileiros confiam na ciência. Porém, mais de um terço dos brasileiros (38%) concordam ser céticos em relação à ciência, um aumento de 5 pontos desde o ápice da pandemia Covid19 em 2020.

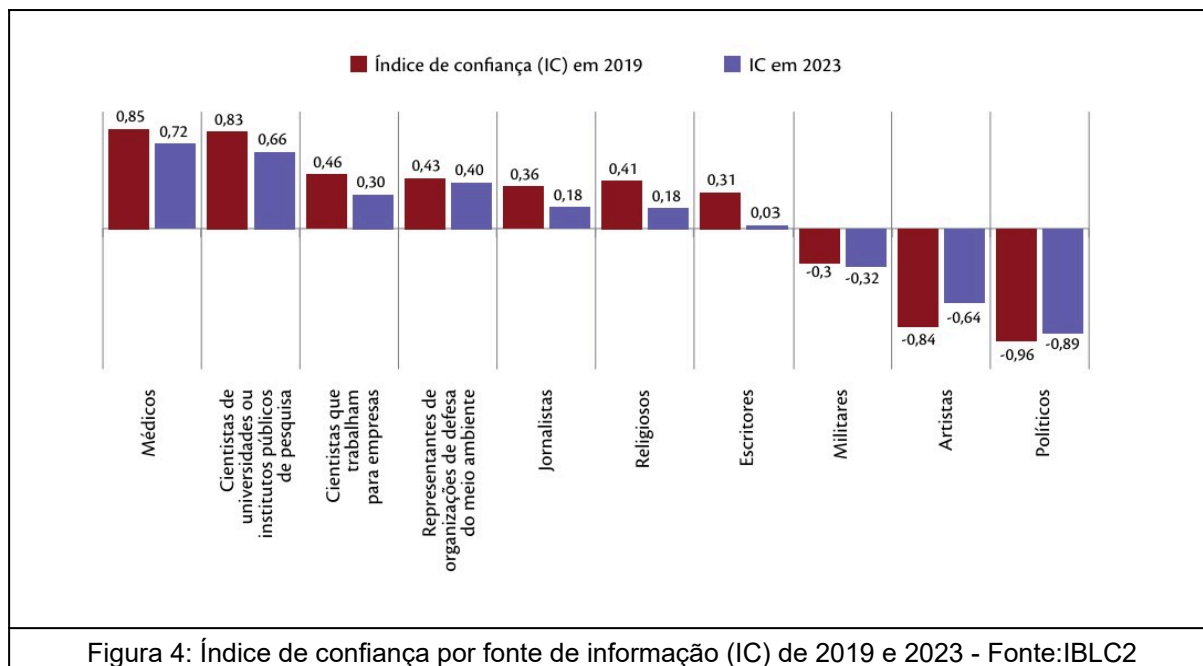


A escolha do canal de informação sobre ciência influencia na confiança do brasileiro, pois 77% dizem confiar mais em fatos científicos publicados na mídia tradicional, contra 51% que confiam em fatos científicos divulgados nas mídias sociais (contra 44% globalmente)¹⁶.

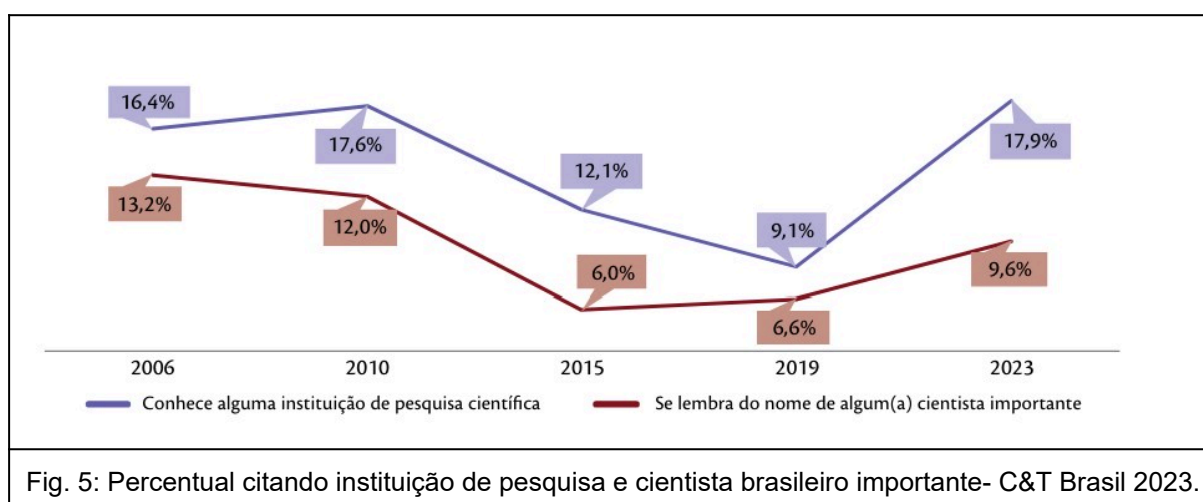
Pesquisa sobre percepção pública da ciência e tecnologia (CGEE, 2023), revelou que os cientistas brasileiros são bem avaliados, qualitativamente, com o segundo maior índice de confiança por fontes de informação da população brasileira

¹⁶ https://www.3m.com.br/3M/pt_BR/pesquisa-do-estado-da-ciencia/#

¹⁷, sendo os que atuam em instituições públicas considerados ainda mais confiáveis.



A amostra de 2019 demonstrou que mais de 90% dos brasileiros não se lembravam ou não sabiam apontar um cientista do país e 93% não se lembravam ou não sabiam indicar instituições do setor. (CGEE, 2019). Em 2023 esse índice praticamente dobrou e pode ser atribuído ao contexto da pandemia de 2020, pois, durante esse período, as instituições de pesquisa envolvidas no desenvolvimento de vacinas receberam ampla cobertura midiática (CGEE,2023).



Mesmo assim, o percentual de pessoas que consegue lembrar o nome de alguma instituição de pesquisa científica ou de algum cientista do País é muito

¹⁷ <https://www.cgge.org.br/web/percepcao>

baixo, estando entre os menores da América Latina.

Proporção de pessoas que “fariam com dificuldade” ou “não seriam capazes de fazer”					
Pessoas de 15 a 40 anos residentes nas 9 regiões metropolitanas + DF com pelo menos 4 anos de escolaridade					
	Total	Nível 1	Nível 2	Nível 3	Nível 4
Conferir a conta de consumo de energia elétrica.	26%	40%	28%	18%	8%
Ler manuais para instalar aparelhos domésticos.	26%	39%	27%	20%	18%
Combater um pequeno incêndio seguindo as instruções dos equipamentos contra fogo.	36%	45%	40%	27%	27%
Consultar dados sobre saúde e medicamentos na internet.	37%	58%	41%	24%	16%
Entender gráficos e tabelas inseridas em matérias de jornais, revistas etc.	43%	63%	47%	30%	19%
Estimar o consumo de energia de aparelhos elétricos a partir de suas especificações técnicas.	48%	63%	50%	39%	38%
Interpretar dados científicos incluídos nos rótulos de produtos alimentares (ex.: tabela nutricional, composição, etc.).	48%	60%	54%	37%	35%
Interpretar os resultados de um exame de sangue a partir dos valores de referência fornecidos pelo laboratório.	55%	67%	61%	44%	35%

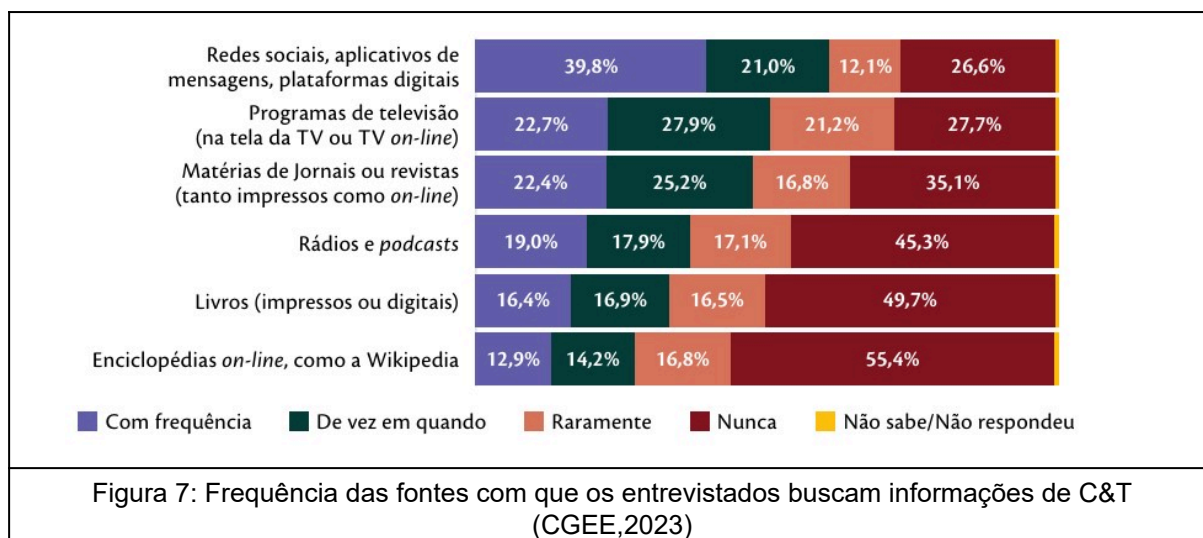
Figura 6 - Níveis de dificuldade de contextualizar o conhecimento científico. Fonte: IBLC (2017)

Mas em contraste ao nível de confiança, mais de 62% dos brasileiros não obteve em seu letramento, conhecimentos científicos suficientes para compreender dados e resolver problemas cotidianos, como analisar o consumo e tarifação de energia elétrica, interpretar dados dos rótulos dos alimentos ou compreender manuais técnicos de aparelhos domésticos.

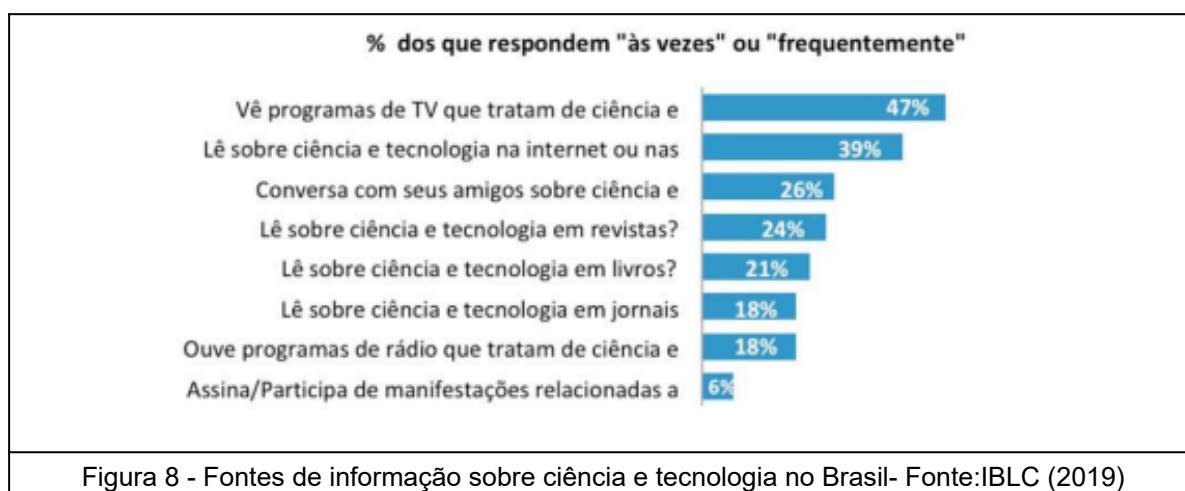
A popularização da ciência interessa à sociedade, sendo objetivo comum entre pesquisadores, população e comunicadores. A difusão e divulgação científica de pesquisas e indicadores torna-se uma premissa estratégica de aproximação com a sociedade, complementando o letramento científico da população brasileira. As universidades, instituições de pesquisa e agências de fomento encontram dificuldades em se comunicar com a diversidade de perfis da sociedade brasileira (CARIBÉ, 2015).

Mesmo possuindo baixo índice de letramento científico, a população brasileira considera que a pesquisa das universidades públicas produzem mais benefícios para a sociedade e que o investimento público em educação superior e pesquisa científica deve aumentar. A internet e a TV são os meios mais utilizados pela população para adquirir conhecimento científico e a busca e/ou acesso a

informações são realizados por redes sociais, aplicativos de mensagens e plataformas digitais chega a quase 40%, seguido pela TV (22,7%).



Os dados também indicam que a maioria dos cidadãos assistem conteúdos científicos na tv aberta e cabo, bem como no youtube, pois as gerações de nativos digitais¹⁸ (PRENSKY, 2001) já consideram a plataforma como televisão. Isso indica a predominância da linguagem audiovisual como opção e indica que infográficos, documentários e animações são bem aceitos como linguagem de representação do conhecimento, entre todas as gerações.



As mídias digitais e a imprensa, bem como a academia, não superaram as dificuldades intrínsecas ao processo de decodificação do discurso científico à linguagem usual de comunidades tão diferenciadas de nosso país continental,

¹⁸ <https://www.tdics.epsjv.fiocruz.br/glossario/nativo-digital-x-imigrante-digital>

dificultando o entendimento, mesmo das populações dos grandes centros, mais familiarizadas com a ciência e a tecnologia. (CARIBÉ, 2015). A grande imprensa não compreende bem a dinâmica da academia e tem menor interesse pelas ciências básicas ou pesquisa pura, priorizando áreas com maior apelo econômico ou aplicação industrial. (VOGT, GOMES, MUNIZ, 2018)

Segundo Cidoval Morais de Sousa, harmonizar a divulgação científica e o ensino de ciências se torna outra estratégia de aproximação das comunidades aos pesquisadores e instituições, criando empatia com a população de forma crítica, ensinando ciências como ferramenta de transformação social.

Adota-se, como suporte teórico, os enfoques CTS, que entendem a ciência como uma construção social, a comunicação pública da ciência como intrínseca ao fazer científico, o ensino de ciências como possibilidade de transformação social, e a sustentabilidade como ideal a ser perseguido pela tecnociência, agora comprometida com um mundo diferente, mas não desigual. “ (SOUSA, 2017)

Na Europa, desde os anos 1990, existem políticas públicas educacionais em nível básico de popularização da ciência, oferecendo disciplinas de educação científica, tanto para estimular o conhecimento desde as séries iniciais para os alunos com vocação para a pesquisa, quanto para melhorar os conhecimentos da população em geral, pois, segundo estudos, a cultura científica chega pelos alunos aos pais, aumentando o alcance dessas ações (CAIRES, 2019).

A ausência de uma cultura de comunicação e mídia nos centros produtores de conhecimento e pesquisa no Brasil, certamente colaborou para o distanciamento da população e influenciou a relação, nem sempre harmoniosa, entre divulgadores da ciência na imprensa, direcionados pela linha editorial de caráter privado e lucrativo das emissoras e veículos. Por outro lado, os pesquisadores praticamente só conhecem o financiamento público das pesquisas e sempre trabalharam em entidades públicas, onde mais de 90% delas são realizadas. (CARIBE, 2015; PORTO, 2009)

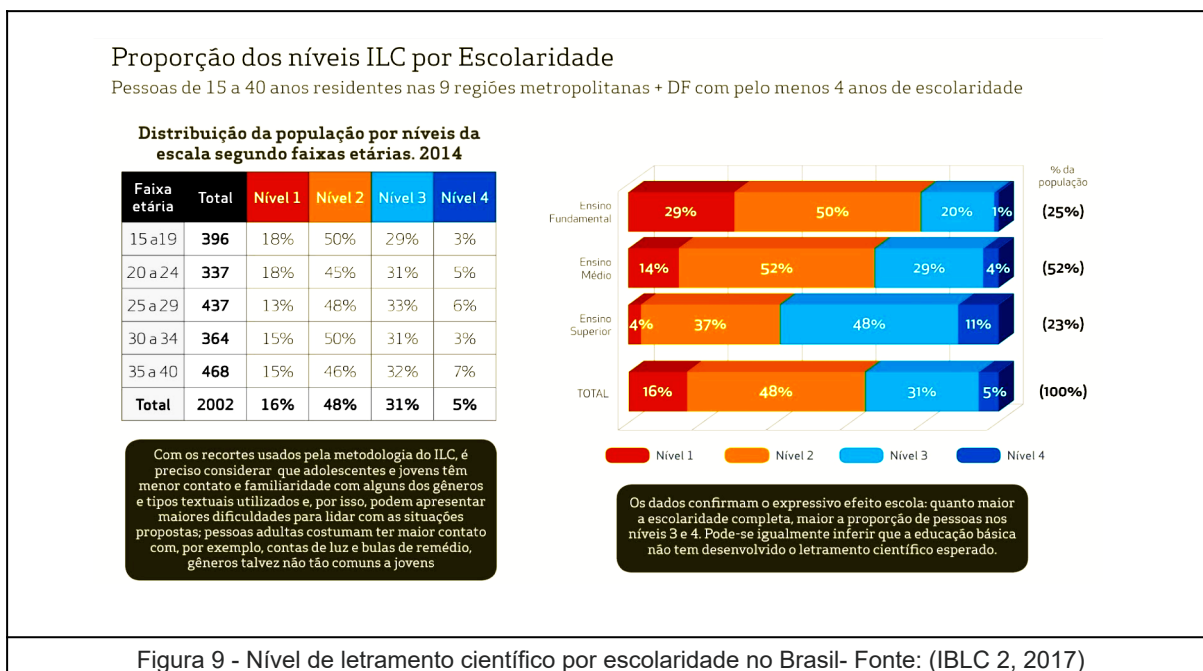
Contextualizando conceitos da convergência (BLACKMAN, 1998; JENKINS, 2003), da hipermídia (SANTAELLA, 2007), da cibercultura (PORTO, 2018; LÉVY, 2010) e da cultura da mídia (KELLNER, 2001) nos conteúdos que as instituições científicas e universidades já produzem e divulgam em redes sociais e plataformas de streaming, a cultura científica pode ser útil no enfrentamento da desinformação e

do negacionismo científico.

1.2.1 - Alfabetização científica ou letramento científico

Outra influência determinante é o chamado analfabetismo científico¹⁹ ou letramento científico rudimentar²⁰, que resulta das dificuldades de compreensão dos conteúdos pelas pessoas sem recursos, gerando uma mitificação²¹ da ciência que distancia o cientista e a academia da sociedade (SBPC, 2013). Isso também contribui para aumentar a falta de sensibilidade dos empresários e editores de veículos de comunicação para a importância da circulação de informações em Ciência, Tecnologia e Inovação. (PORTO, 2009).

A Base Nacional Comum Curricular (BNCC) estabelece que o letramento científico deve ser desenvolvido ao longo do Ensino Fundamental, permitindo aos alunos, de forma autônoma, compreender, interpretar e formular ideias científicas em uma variedade de situações, inclusive, as do cotidiano, acessando saberes comuns e conhecimentos científicos produzidos ao longo da história.



A pesquisa do Instituto Brasileiro de Letramento Científico indica ser ainda muito pequeno o percentual de alunos que possuem nível básico (20%) e proficiente (1%) no ensino fundamental, pois mesmo considerando as maiores dificuldades das

¹⁹ <http://portal.sbpcnet.org.br/noticias/pisaquebrao-analfabetismo-cientifico-no-brasil/>

²⁰ <http://iblc.org.br/wp-content/uploads/2018/01/3-publicacao-ilc.pdf>

²¹ mitificar - Conferir a alguém ou a algo qualidades ou características incomuns, não correspondentes à realidade. Fonte <http://michaelis.uol.com.br/busca?id=e3W1V>

séries iniciais, deveria haver um equilíbrio resultante das maiores habilidades dos alunos das séries finais do ciclo.

A pesquisa também demonstra que a população reconhece a importância do letramento científico, tanto para compreender o mundo, quanto para obter melhores oportunidades de trabalho. Existe forte contraste entre os que reconhecem que a formação científica permitiria melhores oportunidades no mercado de trabalho (62%) com uma proporção bastante significativa (46%) dos que não manifestam interesse por uma profissão na área de ciências. Alunos que declaram vocação para a ciência também têm percentual reduzido, indicando menos de 20% dos entrevistados.

1.3 Pergunta da pesquisa

Como analisar a eficácia de dinâmicas de cultura científica utilizando dados de acesso e interações em plataformas digitais de projetos de difusão científica e/ou extensão universitária ?

1.4. Objetivo geral

Investigar acervos digitais e mídias sociais para coletar, classificar e analisar séries de dados sobre o alcance público e os usuários, correlacionando dados multivariados através de Análise de Conteúdo, Aprendizado de Máquina e Ciência de Dados Sociais.

1.4.1 - Objetivos Específicos:

- A. Publicar a tese e seus resultados em painéis visuais e analíticos (dashboards) no site tecnociencia.net.br, difundindo as análises e estratégias metodológicas utilizadas através de infográficos e planilhas.
- B. Desenvolver um robô conversacional (TeseBot) como Agente de Conhecimento (RAG/LLMs) treinado com os conteúdos da tese que responda perguntas sobre a pesquisa e seus resultados no mesmo site.

1.4.2 - Hipótese:

A utilização de plataformas digitais nos projetos de difusão científica "Ciências Sociais em Diálogo" e "Rede de Professores de Língua Estrangeira (RAPLE)",

promove eficazmente a cultura científica e o letramento científico ao democratizar o acesso ao conhecimento para diversos públicos.

1.5 Ambiente da Pesquisa

As pesquisas são realizadas através da atuação do autor da tese como Webmaster e Suporte Acadêmico em Difusão Científica da Faculdade de Ciências e Letras da Unesp de Araraquara (FCLAr), onde atua profissionalmente desde 2008 até 2024.

Serviço	Tipo de mídia ou plataforma	Média acesso s por mês
Streaming	Transmissão ao vivo pela internet de conferências e eventos científicos via Rede Nacional de Ensino e Pesquisa (RNP) e Youtube	mais de 3200
Acervo On demand	Canal de Mídias digitais na Rede Nacional de Ensino e Pesquisa (RNP) ²² e Youtube	mais de 4.600
Webmaster	Suporte ao site institucional e desenvolvimento do CMS do projeto da Rede de Apoio aos Professores de Língua Estrangeira (RAPLE)	mais de 4100
Suporte Acadêmico	Criação de canais, vinhetas, artes e programação visual dos projetos pesquisados (Ciências Sociais em Diálogo e RAPLE)	mais de 62.000
Tabela 1 - Serviços multiplataforma de difusão e cultura científica da FCLAr: Fonte: o autor		

Atualmente, são realizadas ações de cultura científica (difusão, divulgação, letramento) em diferentes modalidades de comunicação multiplataforma pela internet, que podem ser acessadas por desktops, tablets, notebooks, smartphones e smart tvs. O arranjo multiplataforma permite realizar ações que caracterizam a cultura científica para difundir e divulgar projetos acadêmicos e as mídias digitais audiovisuais são imprescindíveis para as ciências humanas, pois integram o repertório de ferramentas de mediação hipermidiática, ampliando o alcance dos eventos acadêmicos e demais atividades.

As análises das redes sociais e suas estruturas relacionais são investigadas em dois projetos de extensão e suas instâncias que demandam serviços de comunicação, difusão e letramento científico multiplataforma através de site multimídia e canais de streaming, que são:

²² Canal da FCLAr na RNP - <https://eduplay.rnp.br/portal/search?term=fclar>

1. **Ciências Sociais em Diálogo:** Canal no Youtube criado em abril de 2020 e se dedica, segundo dados do projeto, a compartilhar com a comunidade universitária e a sociedade em geral, debates semanais e abertos sobre temas atuais e que interessam a todos, permitindo a interação do público com os expositores. O canal já realizou 241 transmissões de conteúdos que alcançaram mais de 52.500 visualizações, possui 2872 inscritos e totalizou 10.626 horas de exibição de conteúdos entre abril de 2020 e maio de 2024. <https://www.youtube.com/c/Ci%C3%A7%C3%A9nciasSociaisemDi%C3%A1logo>
2. **Rede de Apoio aos Professores de Língua Estrangeira:** Projeto de Extensão que visa facilitar acesso de professores a material digital inovador, organizado por temáticas e planos de aula, além de debates, oficinas e dinâmicas de ensino por streaming, estimulando a troca de ideias e conhecimentos sobre o ensino de línguas através do acervo audiovisual em nosso canal. Também se comunicam com sua rede de professores, alunos e interessados pelo Facebook, Instagram e Spotify. <https://raple.fclar.unesp.br/>

1.6 Plano metodológico

A delimitação do tema objetivou uma redução realista do objeto de pesquisa e tornou viável sua produção dentro do prazo e cronograma estabelecidos, demonstrando sua importância social e aderência aos campos já pesquisados e exaustivamente teorizados. Foi necessário checar sua atualidade e relevância através de uma revisão bibliográfica, que possibilitou melhor conhecimento do objeto e aprofundamento dos temas, buscando uma abordagem realista e legítima.

Essa abordagem permitiu que as duas modalidades sejam complementares, não excludentes e possam ter a mesma relevância na pesquisa, pois a ciência, em sua essência, alterna objetividade (quantitativo) e subjetividade (qualitativo), possibilitando inferências e/ou observações que dependem tanto dos aspectos estatísticos e precisos dos dados, como das injunções e influências dos fatores sociais, econômicos, epistemológicos, humanos, culturais e imateriais.

A pesquisa qualitativa buscou observar amostras, mesmo reduzidas, que refletem significados de hábitos, ações, culturas, valores, tendências de comportamento e aspectos sociais, através de coletas virtuais de campo, oralidade ou ainda observações em ambiente não controlado.

A coleta dos dados estatísticos nos canais e redes sociais buscaram referenciar desde os períodos iniciais da criação de cada página ou perfil nas

plataformas até os dias atuais. Para os dados dos conteúdos foram coletados os diálogos do chat online durante a transmissão (*streaming*), bem como os comentários sobre o vídeo no acervo do canal (on demand). Os acessos de cada conteúdo totalizam ambas as modalidades e as plataformas não fornecem meios de separá-los, portanto as visualizações são sempre totais, portanto atemporais.

Na análise de conteúdos audiovisuais, o volume de dados que se manipula e analisa é imenso e eventualmente supera a capacidade de processamento de algumas ferramentas. A transcrição de uma palestra (live) de aproximadamente uma hora e meia, com abertura, perguntas e encerramento chega a dezenas de páginas de texto transcrito e em alguns casos, teve que ser dividido em duas ou mais partes, sendo tais textos identificados na página da análise no site.

1.6.1 - Formas e vozes

Analisando pela ótica dos estudos do discurso na intertextualidade de Bakhtin, como mencionados na Revista Bakhtiniana²³, nesta pesquisa científica foram utilizadas as formas textuais de citação direta, paráfrase e menção indireta, além de externas e origem própria como dados, tabelas, tabulações, análises, gráficos e de terceiros, através de obras audiovisuais.

No sentido ideológico de caráter de aproximação e explicitude ²⁴, eventualmente estarão presentes as vozes do pesquisador, dos pesquisadores entrevistados e/ou gravados em meio digital, dos estudantes, das pessoas envolvidas na produção dos conteúdos e da população, através da coleta de dados que realimentam a pesquisa.

1.6.2 Corpus

Além da revisão bibliográfica metodológica, foram coletadas e referenciadas por bases de dados estatísticas públicas e privadas, como aporte quantitativo de dados empíricos para formular inferências estatísticas e qualitativas.

Instituição	Nome e Tipo de Pesquisa	Anos
IBGE ²⁵	Pesquisa Nacional por Amostra de Domicílios - PNAD Contínua Pesquisa de Inovação (PINTEC) e Módulo TICs da PNAD Contínua	2018-2023

²³ <http://www.scielo.br/pdf/bak/v8n2/11.pdf>

²⁴ <http://periodicoscientificos.ufmt.br/ojs/index.php/revdia/article/view/5084/pdf>

²⁵ IBGE: Instituto Brasileiro de Geografia e Estatística

IBLC²⁶	Letramento científico: um indicador para o Brasil Relatório Técnico, do Instituto Brasileiro de Letramento Científico	2017
MCTIC²⁷	Pesquisa sobre a Percepção Pública da C&T no Brasil Indicadores Nacionais de Ciência Tecnologia & Inovação	2019/2022 2018/2022
CAPES²⁸	Relatório sobre a Avaliação da Pós-Graduação Stricto Sensu Relatório Research in Brazil for CAPES de Clarivate Analytics	2016 2018
CETIC²⁹	TIC para o desenvolvimento sustentável Pesquisa sobre o uso das TICs nas escolas brasileiras	2019/2022 2017/2022
CGEE³⁰	Percepção Pública da C&T no Brasil – 2023. Resumo Executivo. Brasília: Centro de Gestão e Estudos Estratégicos	2023
Quadro 2 - Bases de dados estatísticas fundamentais		

1.6.3 – A metodologia da pesquisa

A metodologia adotada é a da pesquisa-ação (LEWIN, 1947), que tem como uma de suas principais características, a ausência inicial de todos os elementos necessários para a delimitação do objetivo e gradual formulação dos meios adequados para obter resultados satisfatórios.

Associando os dados coletados aos conceitos articulados na observação do modelo experimental em suas dinâmicas, busca-se articular as seis fases específicas da metodologia (CHIZZOTTI, 2018), visando a superação gradativa e sequencial daquilo que se pretende resolver.

As fases metodológicas, segundo CHIZZOTTI, podem ser resumidas em cinco etapas em retroalimentação cíclica.

1. Planejar etapa
2. Coletar e minerar dados
3. Analisar dados e ações
4. Avaliação de métricas e conteúdos
5. Geração de percepções, inferências e resultados

Dificuldades e imprevistos

O cronograma de execução da tese foi fortemente impactado pelo isolamento social imposto na pandemia, resultando na impossibilidade de captações de imagens, entrevistas, enquetes e outras coletas de dados nos projetos de extensão.

²⁶ Instituto Brasileiro de Letramento Científico

²⁷ Ministério da Ciência, Tecnologia, Inovação e Comunicação

²⁸ Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

²⁹ Centro Regional de Estudos para o Desenvolvimento da Sociedade da Informação

³⁰ Centro de Gestão e Estudos Estratégicos (CGEE)

Outro fator impactante na pesquisa foi a abrupta e não justificada decisão da *big tech* Meta, de interromper definitivamente todas as formas de acesso autorizado aos dados do facebook e instagram em 2024. Cerca de 30% da coleta de dados desta pesquisa precisou ser abandonada e ignorada por não haver mais como desenvolver as análises sem acesso aos dados atualizados.

1.7 Metodologia de Ciência de Dados Sociais

Na metodologia de Ciência de Dados, foram mapeados os problemas a resolver ou apoiar a solução, analisar os dados e propor um modelo de compreensão que possibilita observar padrões e até mesmo prever resultados, numa abordagem chamada top-down (cima para baixo).

Essa metodologia deve ser eficaz em atender as demandas apresentadas sem limitações ou dependência direta de qualquer tecnologia, adequada ao volume de dados que serão processados e gerar uma resposta confiável que represente a análise quantitativa e/ou qualitativa dos dados (AMARAL, 2018).

Este projeto analisou os dados sobre alcance, engajamento, interações e outras métricas sobre a comunicação e difusão científica realizada pelas redes sociais oficiais da instituição e seus resultados na dinâmica de cultura científica com as comunidades interna e externa e se ocorreu letramento científico.

Além da análises das redes sociais e suas estruturas relacionais, também são investigados projetos e instâncias que demandam serviços de comunicação, difusão e letramento científico multiplataforma através de sites multimídia e canais de streaming.

Além dos resultados descritivos e estatísticos demonstrados, os dados sobre conteúdos coletados permitem analisar fenômenos relacionais, linguísticos e comunicacionais que ocorrem em determinado tempo (períodos analisados), local (redes complexas) e cultura científica (divulgação, difusão e letramento) .

As cinco etapas oferecem um acompanhamento constante do desenvolvimento do modelo e possibilitam que ajustes e correções de premissas ou objetivos possam ser realizados durante as dinâmicas de preparação e limpeza na mineração de dados brutos e possíveis transformações necessárias nos mesmos, otimizando e viabilizando as etapas posteriores. O método também deve permitir a

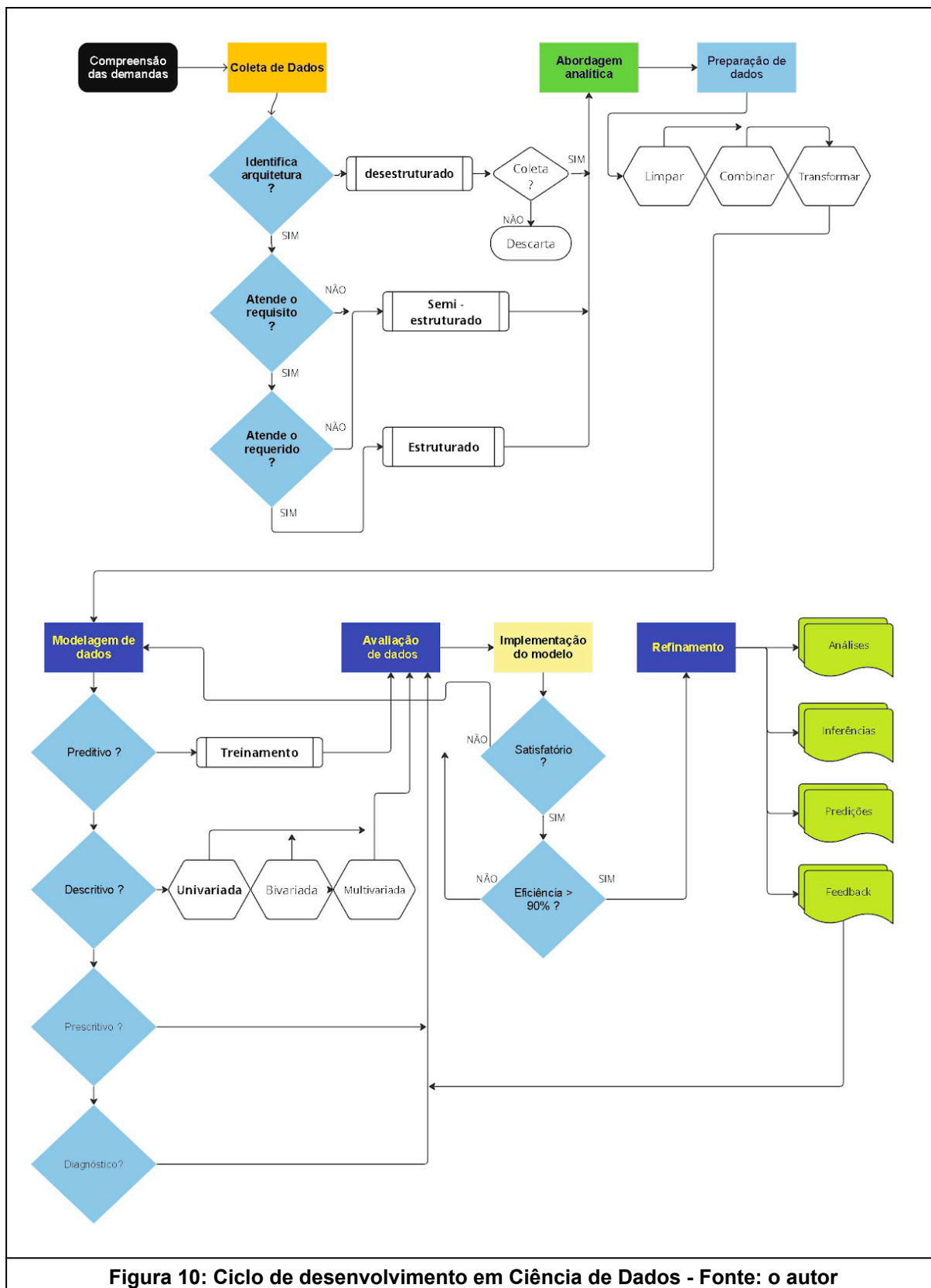
análise contínua dos dados em qualquer período de tempo em cinco etapas, não se limitando à sazonalidade de demandas (CASTRO e FERRARI, 2016). Os dados dos gráficos e tabelas podem ser atualizadas automaticamente nas planilhas ou por recursos JSON³¹

As fases da metodologia de coleta e análise de dados estão descritas:

1. Os dados são **coletados** em formato CSV nas plataformas e exportados para o Google Planilhas, onde são **analisados e limpos, rotulados e validados** também no Maxqda quanto aos dados dos conteúdos (**Preparação de Dados**).
2. Na segunda etapa (**Análise Exploratória dos Dados**) são descobertos padrões, distribuições e suas correlações para obter uma compreensão abrangente, **propondo inferências** quantitativas e/ou qualitativas que formulam hipóteses.
3. A terceira etapa consiste em utilizar métodos estatísticos como análises de variância, regressão linear e/ou logística, séries temporais e tabelas de frequências para **testar as hipóteses formuladas na etapa anterior** e criar tabelas dinâmicas que permitam implementar a **Modelagem de Dados**.
4. Após a **Modelagem Estatística**, os chats, comentários e vídeos são transcritos e analisados no Maxqda e submetidos a um **treinamento de aprendizado de máquina** em framework Python-NLTK de Processamento de Linguagem Natural (PLN) por métodos Naive-Bayes de probabilidades e Graph-RAG de vetorização+grafos de conhecimento para lidar com lematização, gerenciamento de derivações e redução de flexões.
5. Na última etapa são gerados os **gráficos das análises descritivas e os conteúdos submetidos à janelas de contexto** no Gemini Advanced (mais de 1.5 milhão de tokens) e de 64 mil palavras no Maxqda através de comandos diretamente no chat dessas ferramentas, quando são gerados os resumos, comentários e análises de sentimento e discurso. Os gráficos e tabelas são organizados em um site Wordpress com painéis de dados (dashboards)
6. Os vídeos transmitidos e armazenados no Youtube, seus comentários e chats são os conteúdos analisados qualitativamente, iniciando pela transcrição do mesmo em arquivo de texto PDF pela ferramenta Maxqda e disponibilizado juntamente com acesso ao audiovisual incorporado no site. Qualquer pessoa pode ler a transcrição e acessar o trecho do conteúdo no contexto da transmissão.

³¹ JSON, um acrônimo de JavaScript Object Notation, é um formato compacto, de padrão aberto independente, de troca de dados simples e rápida entre sistemas

1.7.1 - Fluxograma do ciclo de desenvolvimento em Ciência de Dados



1.9 - ESTRUTURA DO TRABALHO

1.9.1 - Ciência, Tecnologia e Sociedade

O capítulo referencia o campo dos estudos sociais da ciência e tecnologia (science studies) em breve resumo da concepção sobre ciência pós-acadêmica de Ziman. São contextualizadas as controvérsias em novas abordagens de autores como Knorr-Cetina; Latour E Woolgar, Sismondo, Collins, Dagnino e Sousa sobre a prática social da ciência e da tecnologia, a tecnociência e a visão de que as estratégias políticas são parte da conduta dos cientistas.

1.9.2 - Ciência de Dados Sociais e Análises Descritivas

Na essência do capítulo sobre a Ciência de Dados Sociais, a tecnologia informacional conectada pelas redes favoreceu novos métodos de pesquisa híbridos que oferecem recursos avançados de coleta, mineração e análises de dados que dinamizaram as análises sociológicas e de conteúdos. São descritos de forma resumida, os algoritmos, seus vieses e sua influência na vida cotidiana e os preconceitos raciais e misóginos no reconhecimento facial, para exemplificar a realidade da manipulação na cultura algorítmica.

1.9.3 - Análise de Conteúdos em Redes Complexas

São apresentadas diferenças entre redes neurais reais, orgânicas ou simuladas, o paradigma do conectoma, o connexionismo, a linearidade dos perceptrons, as visões de Leo Szilard sobre o de bit de informação, aspectos fundamentais da constante de Boltzmann que deu origem ao algoritmo mais utilizado em aprendizagem profunda, da teoria matemática da comunicação de Shannon, do Demônio de Maxwell, do pensamento linear no algoritmo Adaline. Também são analisadas a origem das Máquinas de Boltzmann Restritas (RBMs), Redes Neurais por Grafos (GNNs), o Modelo BARD e as pipelines LLMs/RAG.

1.9.4 - Resultados e conclusões

Os dados coletados sobre as dinâmicas de difusão e letramento científico são minerados, limpos e analisados por técnicas multivariadas, como as quantitativas lineares e não lineares, qualitativas e de conteúdo. Foram geradas dezenas de

percepções e inferências estatísticas/relacionais para validar a tese de eficácia das dinâmicas de divulgação, difusão e letramento (cultura científica) das redes complexas estudadas.

Também foi desenvolvido uma articulação de métodos de pesquisa e ferramentas computacionais que permitiram análises de conteúdos comunicacionais e audiovisuais de forma eficiente e documentada, cujas técnicas são debatidas nas considerações finais.

2.1 - A Ciência Pós-Acadêmica de Ziman e sua concepção

Em 1933, a tese de Robert K. Merton, cujo título, *Ciência, Tecnologia e Sociedade na Inglaterra do Século XVII*, já prenunciava a emergência dos estudos sociais da ciência ao avaliar a influência de fatores econômicos nas pesquisas. Somente após os anos 1930, Merton desenvolve e publica seus artigos sobre o ethos científico (MERTON, 1938), descrevendo as premissas institucionais que o definem: comunismo, universalismo, comunicação, desinteresse e ceticismo organizado. Evoluindo nesses estudos, o autor passa a reconhecer padrões de controle institucional em que a competição por prioridades tende a maior reconhecimento de cientistas consagrados (Efeito Matthew ou vantagem acumulada), enfatizando contradições entre metas, recursos e meios sociais na busca pelo sucesso pessoal e o próprio ethos (LATOUR, 1991).

Essa concepção clássica, quase idealista, evoluiu para novas abordagens críticas aos fatores sociais resultantes da ciência, principalmente no pós Segunda Guerra, com o advento de inúmeras catástrofes humanas e ambientais. Guerras com armas químicas e nucleares, trágicos acidentes nas indústrias com vazamentos e contaminações que mataram milhares e o impacto ecológico de extração mineral e petrolífera desenfreada e destruidora (COLLINS, 2010).

Sem dispensar o conceito de ethos acadêmico mertoniano, John Ziman desenvolve sua concepção de ciência pós-acadêmica, em que os valores sociais e epistêmicos do modelo consagrado nas universidades e centros de pesquisa, revistos pela cultura industrial e burocrática (ZIMAN, 2000, p. 82), se torna a confluência de ambas³².

Como afirma Ziman, a filosofia da ciência acadêmica integra sua cultura através dos princípios reguladores do seu ethos e uma imensa variedade de práticas, procedimentos e tradições epistêmicas, em correlação às normas mertonianas.

Em seu livro *Real Science*, ele dedica todo um capítulo aos novos modelos de produção de conhecimento em variados exemplos de que a atividade científica extrapolou os muros físicos e articulou uma abordagem que busca conciliar a visão

³² Ziman discordava da denominação Modo 1 e 2 de Gibbons et al.(1994), argumentando que seria então necessária a criação do modo 3, sendo justamente o somatório deles. (REIS,2013)

internalista da academia com as alterações ocorridas no modo de produção da ciência contemporânea e suas implicações filosóficas, históricas e sociais, que alteraram as relações entre ciência, tecnologia e sociedade (IBIDEM).

Por exemplo, a norma do 'comunalismo' está intimamente ligada ao princípio do empirismo - ou seja, a confiança nos resultados da observação e da experiência replicáveis. (ZIMAN, 2000, p.65)

Para Ziman, os princípios reguladores da ciência mudam com o tempo, assim como as práticas sociais que os conectam ao ethos acadêmico, mas que são quase os mesmos ideais filosóficos desde os tempos vitorianos. Sendo essencialmente preceitos e não regras imutáveis, alguns requisitos podem ser flexibilizados pela própria natureza específica da ciência em questão ou pela impossibilidade real de cumpri-los, como no exemplo dado por Ziman, de que a paleontologia não pode ser desmerecida, apenas porque não desenvolve pesquisa experimental (ZIMAN, 1979)

Crítico da imagem deturpada da ciência infalível que busca a verdade absolutamente perfeita da qual não se pode duvidar, nem mesmo nos aspectos em que ela não se aplica, também contrapõe o rebaixamento das questões complexas a mera relatividade de opiniões que não possam ser desafiadas por outro cientista. Mas tudo sempre dentro dos limites de como ela opera e regulada pelo tipo de informação, muitas vezes, extremamente especializada e endossada pelos grupos de expertise que a transformam em conhecimento explícito (ZIMAN, 2000, p. 58).

A visão predominante na era pré-mertoniana, da ciência pura e inquestionável, dá lugar a uma visão simples de tipo ideal de produção de conhecimento que esteja à altura do ethos declarado, moral ou intelectualmente. Ziman reconhece que não é suficiente apenas acreditar na ciência, limitando-se a análise de seus princípios, mas visualiza a emergência de novos modos de produção do conhecimento que superam a idealização da chamada ciência acadêmica. (IBIDEM, p. 59)

Sem sugerir que sejam abandonados os critérios tradicionais de validade científica e seus muitos aspectos filosóficos, éticos e epistemológicos, Ziman propõe superar o modelo acadêmico de mais de duzentos anos e aceitar o fato de que essa mesma superação já está em curso, diante de nossos olhos, uma transformação radical em que a ciência é regulada por um novo ethos e uma nova filosofia da natureza, tornando-se pós-acadêmica. Mudanças estruturais importantes advindas

com as novas formas de produzir conhecimento trazem questionamentos pertinentes e desafiadores ao ethos acadêmico e como ele pode ser interpretado.

A exploração comercial dos direitos de propriedade intelectual, por exemplo, pode ser conciliada com a norma do comunalismo? Pode a objetividade científica ser reivindicada para a investigação que é levada a cabo para resolver problemas sociais específicos? (ZIMAN, 2000, p. 61)

Com o advento da ciência pós-acadêmica proposto, ele analisa os efeitos das visíveis mudanças estruturais que ocorreram em universidades, institutos de pesquisa, laboratórios industriais e instituições governamentais de fomento, sem necessariamente ter que explicar porque elas ocorreram ou se as mesmas são temporárias ou permanentes. Sem repudiar os objetivos tradicionais da ciência, Ziman alega que o termo indica continuidade, mas também diferenças sutis em funções já desempenhadas nos mesmos espaços de atuação, nos quais as pressões políticas, econômicas e sociais são reconhecidas e tratadas com naturalidade e onde as ciências se fundem e se entrelaçam de maneira inovadora e multidisciplinar.

A ciência passou a ser vista como um sistema dinâmico que acumula êxitos e tensões ao responder demandas cada vez mais complexas e tecnológicas, sendo suscetível aos riscos históricos inerentes e flutuações de políticas públicas que ora dinamizam, ora atrasam o desenvolvimento. Segundo o autor, não há uma clara e inequívoca fase dessa transição que possa delinear historicamente essa passagem, porém ela pode ser visualizada por um fenômeno inegável, a coletivização da prática científica.

2.2 - A *Práxis* científica

A prática científica, tanto em relação à síntese do conhecimento como sua dinâmica nos laboratórios, configura-se como objeto e campo dos Estudos Sociais da Ciência e Tecnologia (CTS), que delinearam a emergência de concepções pós-estruturalistas da sociologia da ciência mertonianas (MERTON, 1973). As controvérsias nessas novas abordagens, evidenciadas nos debates sobre a ciência e a tecnologia como atividade formal e prática social, difundem conhecimentos, questionam e analisam o mundo natural, sendo pertinentes às sociedades industrializadas (KNORR-CETINA, 1992; LATOUR E WOOLGAR, 1991;

SISMONDO, 2004, COLLINS, 2010, ZIMAN 2000, 1987, 1979). Nesse sentido, a concepção da ciência como construção social de contextos científicos, permeados por uma diversidade de práticas e relações na qual se insere, seria como uma ponte entre o interior do laboratório e seu entorno (MONTEIRO, 2012).

Mesmo o construtivismo social, em sua concepção de que o ser humano constroi o conhecimento através de suas interações sociais ao invés de descobrir uma realidade objetiva e independente, teoriza, segundo o princípio do finitismo³³, que a aplicação de uma regra, de uma classificação ou de um termo referente a determinado fato ou fenômeno requer um julgamento de semelhança ou diferença, dependendo de seu uso e de sua aplicabilidade (SISMONDO, 1993).

Para KNORR-CETINA, os produtos da ciência são consequências de múltiplas decisões, impregnadas de outras, indicando outros problemas que predispoem novas soluções. Em muitos aspectos, a noção de um laboratório científico em sociologia da ciência representa o que, em história e metodologia de mesma, sempre foi a noção de "experimento".

O caráter "artificial" da ferramenta mais importante do cientista, o laboratório, reside no fato de que este não é mais que uma acumulação local de materializações de seleções anteriores. A seleção de investigações anteriores também afetam as seleções subsequentes, mobilizando as condições de novas decisões. (KNORR-CETINA, 1992)

Nessa concepção, laboratórios e experimentos combinam-se de maneiras diferentes em campos diferentes como processos alinhados à ordem natural e social, onde são criados, reconfigurados e viabilizados em um determinado tempo e lugar. A flexibilidade de comportamentos e possibilidades no laboratório seriam estratégicas, tanto na pesquisa de campo como na produção de conhecimento, tal qual um dispositivo sociotécnico. A ciência nos laboratórios pode otimizar aspectos de ordem natural e também social, ao considerar as estratégias políticas como parte da conduta dos cientistas (LATOUR, 2000; ZIMAN, 1987).

Mesmo para Merton, o desenvolvimento persistente e continuado da ciência somente ocorre em sociedades com certa ordem institucional, por pessoas ativamente interessadas e capacitadas nessas atividades, além de algumas garantias políticas (MERTON, 1973, pág 11).

³³ O uso, o significado da palavra estabeleceu-se a partir das funções que esta desempenha na atividade coletiva dos indivíduos. [Interfaces da Sociologia do Conhecimento de Karl Mannheim](#) (UFSCAR, 2015)

2.3 - Expertise e controvérsias em Collins e Evans

A afirmação de Collins sobre o fato de que mesmos métodos aceitos como científicos podem resultar em diferentes resultados até mesmo incompatíveis, não invalida sua concepção de que a ciência e a tecnologia sejam o melhor modo de extrair a experiência humana de um mundo incerto, mesmo diante de suas fraquezas epistemológicas (COLLINS, 2010, pág 4). As fragilidades e incoerências oriundas dessa dicotomia originaram uma crise de desconfiança, justificada também por falhas que resultaram em riscos à saúde humana e ao meio ambiente causados pelo progresso científico, resultando numa visão pós-moderna, crítica da expertise dos cientistas e típica de um populismo tecnológico (IBIDEM).

No livro *Repensando a Expertise*, Collins e Evans propõem a superação da concepção da mesma como competência individual típica de uma teoria relacional, argumentando que ela seria resultado da associação de grupos de experts, sendo sua aquisição um processo de socialização dentro das práticas dos mesmos, que pode ser perdida com o distanciamento do indivíduo desses grupos. Nesta abordagem, a atribuição da expertise, embora independa da aceitação explícita de quaisquer indivíduos ou mesmo grupos, demanda tempo e esforço na socialização de resultados com aqueles que atribuem esse status a outros, pois além de existirem diversas formas de expertise, a relação entre esses grupos também influencia (IBIDEM, pág 6).

A obra, não pretendendo analisar processos de atribuição do status de expert, trata inicialmente da simplificação do conceito de que especialistas que “sabem o que estão falando”, em última análise, teriam melhores condições de construir consensos e opinar sobre determinados assuntos, mesmo diante de ambiguidades sobre qual assunto, objetivamente, se está falando. O objetivo do livro, segundo os autores, seria de avaliar as probabilidades de que a posse de uma expertise real e substantiva elegeria naturalmente um expert, como questão de socialização dentro das práticas de um grupo deles.

Na visão dos autores, a sabedoria popular considera que a compreensão genuína envolve conhecimento tácito, havendo um conflito evidente entre elas. Os seres humanos compartilham conhecimento tácito em unidades complexas de grupos sociais, desenvolvendo e sustentando essa partilha, mesmo com novos

membros do grupo, por meios inexplicáveis por eles mesmos. Conhecimento tácito é a compreensão profunda que um indivíduo apenas pode obter por meio da imersão social em grupos que a possuem (COLLINS e EVANS, 2010).

Na visão pós-moderna do filósofo Jean François Lyotard, a ciência como um subconjunto do conhecimento, resultaria de enunciados nos quais os objetos sejam explicitamente observáveis e passíveis de pertencerem a um arcabouço (LYOTARD, 1988, p. 35), cuja linguagem seja considerada pertinente aos experts daquela especificidade.

Para Lyotard (IBIDEM), a didática assegura essa reprodução no ensino, sendo diferente da dialética na pesquisa. No ensino, o aluno assimila o conhecimento aceitando o status de expert atribuído ao professor pela instituição. Na ciência, os novos enunciados necessitam referenciar seus precedentes e reafirmar uma bibliografia considerada pertinente e institucionalmente aceita, mesmo nas teorias que emergem como contestação ou negação desses enunciados.

A verdade do enunciado e a competência do enunciador são assim submetidas ao assentimento da coletividade de iguais em competência. É preciso, portanto, formar iguais (IBIDEM, p. 46).

Neste sentido, a linguagem da ciência em seu formato narrativo prescinde dessa mesma legitimação quando comunicada aos leigos ou não especialistas. Não havendo paridade de conhecimento entre o comunicador e a sociedade, o status atribuído por seus pares do grupo de expertise legitima suas afirmações e corrobora seus intentos. Os autores Collins e Evans criaram uma tabela periódica de expertises, em que as pessoas podem se basear ao tomarem decisões técnicas.

2.4 - O nó górdio: Modernidade, conhecimento e política

Conceituar a modernidade ou o que pode ser considerado moderno é um enorme desafio que nos leva ao labirinto de abordagens filosóficas e etnográficas que Bruno Latour realizou em seu livro *Jamais Fomos Modernos*. Nele, o autor articula política e ciência, humano e não-humano, o socialmente construído e o naturalmente dado, para argumentar que a cultura intelectual não sabe bem como ou se deve conciliar e/ou separar esses conceitos.

Para exemplificar essa complexidade e a impotência em elaborar uma teoria que não recorra ao dualismo epistemológico homem/natureza, que ao mesmo tempo

separa conhecimento e política, Latour utiliza o termo nó górdio, não só em *Jamais Fomos Modernos*, mas também em *Políticas da Natureza*. Ora cortando, ora reatando o nó, utiliza essa metáfora de algo insolúvel, impossível de desatar, para criticar que os *Science Studies* não conseguem, nem mesmo agregando as teorias de análise do discurso, sistemas ou das redes, propor uma abordagem:

“mais flexível que a noção de sistema, mais histórica que a estrutura e mais empírica do que a complexidade, a rede é o fio de Ariadne destas histórias confusas ... Ou dizem respeito à natureza, ou à política ou ao discurso.”

Latour propõe o termo tripartição crítica para ilustrar o que ele chama de dilema que permite a etnografia no mundo moderno, pois argumenta que a antropologia trata, sem crises ou crítica, o tecido inteiriço das naturezas-culturas. É capaz de juntar em uma mesma monografia, os mitos, as etnociências, genealogias, formas políticas, técnicas, religiões, epopeias e ritos dos povos que estuda.

Nem um só elemento que não seja ao mesmo tempo real, social ou narrado (LATOUR, 1991, p. 16).

Mas se os antropólogos tradicionais não aceitam ou mal consideram uma antropologia do mundo moderno, é justamente nessa negação que reside a crítica revisionista de Latour, de que precisamos buscar uma definição mais ampla que apenas uma passagem de tempo, pois ao tentar definir modernidade, contrastamos esse conceito com algo arcaico e estável, superado ou mesmo ultrapassado, como se existisse um ponto de ruptura na passagem regular do tempo, reduzido a uma cronologia.

Ao buscar definir o que é moderno (LATOUR, 1991, p. 20), na hipótese de ser necessário abrigar as redes de ciência e técnicas, designa dois conjuntos de práticas diferentes e distintas. Um conjunto de misturas de seres completamente novos, híbridos de natureza e cultura, ao qual chamou de redes; e o segundo cria duas zonas ontológicas distintas entre humanos de um lado e não-humanos de outro, ao qual chamou de crítica.

Logo no início do livro, Latour cita, ao ler o jornal, diversas notícias em que são abordados, de forma híbrida, misturados mesmo, reações químicas dos aerossóis que afetam a camada de ozônio e as reações políticas de governos industrializados, de que não seria absolutamente aceitável culpar a poluição

industrial, mas também de empresas que tentam mudar suas cadeias de produção para eliminar o clorofluorcarbono.

Se naquela época do livro *Jamais Fomos Modernos*, em 1991, já se debatia jornalisticamente, essa dicotomia entre países poluidores e países que sequestram carbono, hoje em dia já se tornou prática comum, grandes empresas poluidoras comprarem títulos de sequestro de carbono de empresas agropecuárias brasileiras ou organizações não governamentais de preservação e sustentabilidade ambiental (MARTINS, 2004).

Atualmente, já não questionamos isso separando o que é política do que é ciência, são conceitos já híbridos e socialmente aceitos, traduzidos para toda a sociedade, o que pode indicar que já superamos a necessidade da simetria, que seria alcançada pelos conjuntos de práticas de purificação e tradução.

Ao afirmar a necessidade de uma nova constituição, moderna, em seu ponto de vista, Latour realiza uma analogia entre a Teoria da Separação dos Três Poderes e propõe que admitamos que jamais fomos modernos, pois a proliferação de híbridos, tais como “embriões congelados, sistemas especialistas, máquinas digitais, robôs munidos de sensores, milho híbrido, bancos de dados, psicotrópicos liberados de forma controlada, baleias equipadas com rádio-sondas, ... etc.”(p. 53), revolucionaram a forma como vemos e conceituamos as coisas de forma entrelaçada, impregnadas de ciência e cultura ao mesmo tempo.

Aqui o nó górdio se mostra possível de ser desatado, ao invés de cortado. Mas Latour, atualizando sua concepção, revisita e reanalisa algumas questões centrais de *Jamais Fomos Modernos* em sua obra de 1999, *Políticas da Natureza*, onde se propõe a associar as ciências à democracia.

2.5 - Do que é feita a tecnociência para Latour ?

O termo que motivou toda digressão realizada até aqui, pode agora ser melhor entendido, mesmo que haja uma imensa controvérsia sobre a origem ou significado real do mesmo. Considerando qual olhar ou campo teórico estamos pressupondo sobre o conceito, que é natural para os Estudos Sociais da Ciência e Tecnologia e para a sociedade capitalista, é ao mesmo tempo contraditório para os

marxistas que alegam uma distinção ideológica da primazia da teoria sobre a prática. (BUNGE, 2012, KOSLOWSKI, 2015)

Latour usa o termo em todo livro *Ciência em Ação* sem especificar claramente sua referência. No capítulo 4, chega a discorrer de que é feita a tecnociência, contrastando a ciência que é feita do lado de dentro, no laboratório, da realizada no imenso lado de fora, captando recursos para que a ciência pura possa ser executada. Ele provoca, questionando quem está realmente fazendo ciência ?

É justamente dessa concepção de que a ciência pura e a ciência aplicada são indissociáveis, mesmo quando diferenciadas em suas finalidades e *modus-operandi*, pois considera não haver necessidade de separar as contribuições da natureza e da técnica(ou tecnologia) na concepção de objeto ou fato científico (NORDMANN, 2011).

Afinal, o que pode ser considerado ciência ou política, ou ambos, quando um pesquisador chefe se dedica quase expressamente à captar recursos, conseguir reconhecimento estatal, angariar parceiros e colaboradores enquanto os demais cientistas estão no laboratório realizando os experimentos com o mesmo rigor e dedicação ?

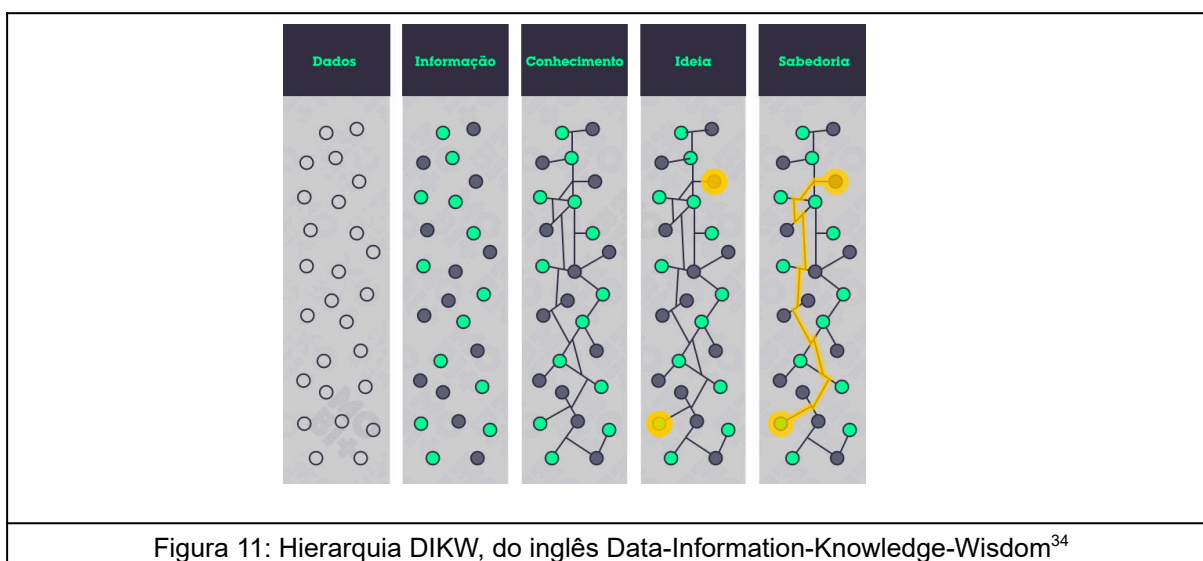
Para Latour, somente construímos a tecnociência quando observamos simultaneamente quem fica dentro dos laboratórios e quem fica fora deles, entendendo que a retroalimentação gerada pelo recurso obtido e aplicada na pesquisa resulta nos avanços ou resultados que legitimam a busca de mais financiamento. Seriam 3 ciclos circulares compostos pelos elementos: dinheiro e força de trabalho, instrumentos e objetos novos, argumentos e inovações.

Ele alega que a palavra tecnociência descreve todos os elementos conectados ao conteúdo científico, embora haja algum risco de se conferir créditos da ciência a todos os grupos de interesse. Porém argumenta que seja possível que todos os elementos, desde os ratos de laboratório, os dependentes de drogas, os chefes, assessores, ministros e congressistas se alinhem de tal modo que a tecnociência influencie as políticas públicas nacionais de saúde.

2.6 - Conhecimento e sabedoria popular

Todo aprendizado pode gerar conhecimentos, num processo que se inicia desde a gestação, quando aprendemos de forma inconsciente, a reagir aos estímulos externos, como quando ouvimos as vozes dos pais e aprendemos a identificar quem está próximo, sem nem haver nascido ou mesmo sem enxergar bem nos primeiros dias de vida. Aprender é base fundamental da vida, sendo um processo natural, contínuo e indispensável ao desenvolvimento dos seres vivos, pois vivemos aprendendo e acumulando conhecimento durante a vida toda.

O conhecimento não é uma competência desenvolvida somente através da acumulação, mas também pela articulação de seus elementos sobre o mundo físico, social, cultural e agora também digital, visando entender e explicar a realidade para contribuir com o desenvolvimento da sociedade (conforme a BNCC - MEC). Existem diversas abordagens em áreas distintas sobre esse processo, que ainda pode ser subdividido em conhecimento tácito ou explícito, como conceituado pela área da Gestão do Conhecimento, que os considera complementares e naturalmente conectados (NONAKA e TAKEUCHI , 2008). A era informacional também trouxe novas concepções sobre subdivisões do processo de aprendizagem, ao propor que a forma como se articula dados, que processados e classificados podem gerar informação, que organizados podem originar conhecimento e que articulados podem resultar em sabedoria. (SANTOS e SOUZA, 2010)



³⁴ SANTOS & SOUZA: A idéia de uma hierarquia entre esses termos já havia sido esboçada nas artes, em um poema de 1934 de Thomas Stearns Eliot (1888-1965) "*The Rock*" e na música de Frank Zappa

Articulando os conceitos de conhecimentos tácito e explícito ao da hierarquia DIKW, pode-se inferir que a ciência também sistematiza o conhecimento através da análise e formalização dessas mesmas conexões. Através do método, o levantamento de dados aliado à observação descritiva e consequente proposição de solução a dinâmica do conhecimento que capacita à ação ou realização de tarefa, indica que ele deve possuir tanto características de objeto, quanto de processo (SANTOS e SOUZA, 2010). Nesta concepção, a ciência também referencia os conhecimentos e métodos precedentes, ao reconhecer e reafirmar bibliografias consideradas pertinentes e institucionalmente aceitas.

A visão de Collins e Evans sobre a sabedoria popular considera que a compreensão genuína envolve conhecimento tácito e um conflito evidente entre elas.

Estranhamente, tudo isso vai de encontro a um outro tema central na análise social da ciência: a ideia de que a compreensão genuína envolve conhecimento tácito. Conhecimento tácito é a compreensão profunda que um indivíduo apenas pode obter por meio da imersão social em grupos que a possuem (COLLINS e EVANS, 2010).

Os seres humanos compartilham conhecimento tácito em unidades complexas de grupos sociais, desenvolvendo e sustentando essa partilha, mesmo com novos membros do grupo, por meios inexplicáveis por eles mesmos. A linguagem humana também é apreendida por imersão no grupo social ao qual ela pertence, mantendo e melhorando sua fluência através da interação social contínua com o mesmo.

Nenhuma entidade não humana conhecida, seja ela artificial ou biológica, possui essa capacidade que emerge espontaneamente na maioria dos seres humanos. Sendo assim, podemos inferir que a expertise é aprendida e mantida pela interação social.

2.7 - Apropriação Social da Ciência e Tecnologia

Redefinindo o papel dos cidadãos na apropriação social da ciência como problema de redistribuição dos saberes, emergem novas formas de gerar conhecimento científico e tecnológico, chamadas de desenvolvimento cooperativo (ALONSO, 2008). Neste sentido, a abordagem relacional sociotécnica da teoria ator-rede (LATOUR, 2012) avalia a flexibilidade como propriedade emergente das

redes envolvidas nestes arranjos.

O conceito de apropriação social do conhecimento científico e tecnológico neste trabalho, considera que, embora sejam inegáveis as transformações na vida da parcela da população que usufrui dos benefícios do desenvolvimento científico e tecnológico, isto não melhora o entendimento público da ciência e da tecnologia nem aproxima os cidadãos dos cientistas (HAYASHI, SOUSA, e ROTHBERG, 2011)

Embora tenha acesso a recursos tecnológicos³⁵, há um inegável distanciamento entre a população e as instituições de pesquisa e desenvolvimento científico, onde são desenvolvidas muitas das patentes que sustentam a própria infraestrutura da internet e da cibercultura (IBIDEM, 2011).

Algumas estratégias foram desenvolvidas por autores que se dedicaram a mapear a dinâmica do entendimento público da ciência e não apenas seu acesso, conhecidos como modelos de comunicação pública da ciência (BROSSARD; LEWENSTEIN, 2010; FARES; NAVAS; MARANDINO, 2007; KNORR-CETINA, 1999; LIMA; NEVES; DAGNINO, 2008 APUD ARRUDA, et al 2017).

³⁵ De acordo com o IBGE, em 2023, 92,5% dos domicílios brasileiros tinham acesso à internet, o que representa 72,5 milhões de residências, embora 22,4 milhões de pessoas fossem consideradas “excluídas digitais” pelo Instituto.

<https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/41024-internet-foi-acessada-em-72-5-milhoes-de-domicilios-do-pais-em-2023#:~:text=Destaques,%25%20para%2081%2C0%25>

Modelos	Objetivos	Visão da ciência	Conhecimento científico	Comunicação da ciência
Déficit de conhecimento científico	Mensagem tem que ser compreendida; limpar ruídos e otimizar neutralidade dos meios. Ciência para alguém (público leigo)	Conhecimento linear, independente do meio social	Analfabetismo científico; popularização e vulgarização da ciência	Unidirecional: do especialista para o público leigo
Déficit complexo		Forma de conhecimento verdadeira e superior	Maior entendimento e valorização de atividades científico-tecnológicas	Unidirecional, mas amplia a capacidade da participação pública
Experiência leiga	Mensagem tem que ser democratizada e apropriada socialmente; ruídos revelam natureza dos meios, mensagens	Mais próxima do público; conhecimento baseado em experiência cotidiana	Parcial, provisório e controverso; influência de reflexões promovidas pelos Estudos CTS	Bidirecional; busca promover e ampliar a participação direta de atores sociais em decisões e conflitos
Democrático ou participação pública		Mais próxima do público; compartilhamento de decisões no âmbito da C&T	Busca difusão de informações para maior participação pública da sociedade	Bidirecional; comunicação de equidade entre cientistas e não-cientista; valorização de várias formas de expertise
Apropriação Social da C&T	Equivaler as linguagens utilizadas por cientistas e não cientistas e tornar publicamente conhecidos os processos e conceitos relacionados à C&T (e inovação), gerando engajamento e participação efetivos	Não é o único conhecimento válido; não reificada, o conhecimento é fruto da integração entre cientistas e não cientistas	A ciência é um dentre vários conhecimentos válidos. Ruptura da visão de propriedade	Polifônica, em rede

Figura 12 - Síntese dos Modelos de Comunicação Pública da Ciência - ARRUDA, et al 2017

.A cultura científica está entrelaçada com a comunicação da ciência, pela premissa do debate público, por dinâmicas e idiossincrasias da difusão e da divulgação científica pelos cientistas ao público leigo.

Concebemos nesta pesquisa que essa comunicação se caracteriza por um caráter dialógico, ao considerar que o processo ativo de troca de saberes valoriza as experiências culturais e conhecimentos tácitos, ao contrário do modelo de déficit. (HAYASHI, SOUSA, e ROTHBERG, 2011).

O conhecimento não emerge de uma estrutura acadêmica neutra, sem interesses e auto-orientada, muito pelo contrário, como nos esclarece Ziman, a construção do conhecimento científico é resultado de diversas negociações e acordos entre os atores.

Os enunciados discordantes precisam ser analisados e eventuais discordâncias passam pelo crivo dos grupos de expertise ou associações de classe

que a legitimam ou refutam, sendo boa parte dessa dinâmica realizada pela publicação em revistas científicas, mas também pela comunicação da mesma à sociedade.

A informação não sai do laboratório direto para a publicação num determinado meio. O percurso é complexo, cheio de disputas, interesses que envolvem indústrias, agências de fomento, governos, veículos e o próprio público (IBIDEM, 2011).

A comunicação pública da ciência também é orientada pelo diálogo entre cientistas e não cientistas em suas dimensões literária, coletiva e epistêmica (KNORR-CETINA, 1999), sem a qual, não existiriam socialmente (VESSURI, 2002).

A apropriação social do conhecimento científico seria facilitada se os cientistas e os meios de comunicação estimulassem uma atitude crítica que evocasse os limites e as conquistas da ciência, bem como a natureza das importantes decisões políticas relacionadas com o campo científico (VESSURI, 2002)

A crítica de Ziman à imagem deturpada da ciência infalível que busca verdades absolutamente perfeitas, sem rebaixar tudo a uma mera relatividade de opiniões, corrobora uma visão humanista da ciência e dos cientistas, que os considera atores sociais que precisam convencer não apenas seus pares, mas também a sociedade através da cultura científica.

Do ensino de ciências, particularmente do que se alinha aos enfoques CTS, o jornalismo científico pode absorver, na prática, o rechaço ao determinismo científico e tecnológico, à ciência como produtora de verdades absolutas, ao desenvolvimento tecnológico como última palavra em salvação, ao cientista como homem acima de qualquer suspeita, aos conteúdos desvinculados da vida cotidiana, ao propalado desinteresse que sacraliza o ethos científico (SOUSA, 2017).

Neste contexto de apropriação social do conhecimento, a expertise e o conhecimento tácito se entrelaçam e refletem a real necessidade de maior interação entre academia e cidadãos. Os chamados “leigos” podem contribuir decisivamente para que a pesquisa consolide um modelo em que os saberes sejam valorizados e trocados para atender demandas específicas de sua cultura, identidade e território.

2.8 - A tecnociência solidária de Renato Dagnino

Ao utilizar a palavra tecnociência, estamos nos referindo a um conceito amplo, com inúmeras concepções e finalidades diferentes, mas que precisa ser

melhor especificado aqui, dirimindo dúvidas e evitando a dualidade em afirmações presentes no texto.

Essa especificidade fica mais evidente quando associamos a palavra “solidária”, que cria um termo que possui nuances originadas de sua derivação do termo “tecnologia social”, amplamente empregada no âmbito da economia solidária, representando um “conjunto de redes de produção e consumo baseadas na propriedade coletiva dos meios de produção e na autogestão” (DAGNINO, 2019-2), buscando obter meios de sustentabilidade ao enfrentar as dificuldades de adaptação à economia capitalista periférica. Dagnino ainda salienta que o termo tecnologia social estaria atrelado ao conceito de negação de tecnologia tradicional.

Nas duas obras do autor e no artigo aqui referenciado, fica evidente o caráter de seu pensamento latino-americano sobre ciência, com forte influência de valores do desenvolvimentismo da CEPAL e dos debates sobre imperialismo, dependência e modernização (BAUMGARTEN, 2008) ou subdesenvolvimento, dependência e desigualdade (DAGNINO, 2019 - 2, pág. 47)

Pelo arcabouço dos Estudos Sociais da Ciência e Tecnologia (ESCT), haveria duas categorias de abordagem, a primeira de C&T com foco na análise, com suas variantes de neutralidade (não influencia a sociedade) e do determinismo tecnológico e a segunda com foco na sociedade, com caráter socialmente determinado, que se subdivide em duas vertentes, a de que as características de C&T são socialmente determinadas (tese fraca da não neutralidade), contrapondo a noção de que em sua funcionalidade, a C&T inibe a mudança social (tese forte da não neutralidade). (BAUMGARTEN, 2008)

A visão de Dagnino se aproxima do conceito sobre Rede de Atores, neste aspecto da tecnociência, pois a ciência não consiste em teoria pura e nem a tecnologia seria apenas aplicação da ciência, pois todos os seus nós participam de sua elaboração. Se para Latour, o alinhamento dos envolvidos (os nós) na tecnociência poderia resultar na emergência natural de uma política pública de saúde, para Dagnino, ela seria determinante do contexto social, podendo, inclusive, coibir sua mudança. Ele considera que a tecnociência estaria alinhada ao preceito neoliberal de mercantilização e que as próprias instituições que produzem e

financiam ciência, sejam públicas ou privadas, geram soluções tecnológicas apropriadas aos mercados dominantes e inerentes ao capitalismo.

Ainda nesse sentido, Dagnino afirma que chegou a considerar em seus estudos a tecnociência como capitalista, porém percebeu que os processos das redes orientado pelos interesses e valores dos que nela trabalhavam, era anterior ao conceito de capitalismo, embora os interesses das minorias que detêm a hegemonia pelo poder e dinheiro, estejam entranhados nas políticas de C&T.

“... primeiro a ciência, sem valores e interesses, seria gerada e, depois, a tecnologia, que a aplicava, seria colocada a serviço do capital, como diria um crítico marxista. Como se o conhecimento científico originalmente neutro fosse posteriormente “contaminado” mediante o desenvolvimento da tecnologia com os valores e interesses capitalistas da propriedade privada e da exploração do trabalho humano e, introduzida no processo de trabalho, possibilitando a extração de mais-valia relativa” (DAGNINO, 2017-2, pág. 50) .

Ao contextualizar e integrar as realidades inerentes ao capitalismo periférico da América Latina, da necessidade de conscientização, mobilização, participação e empoderamento dos movimentos populares e políticas públicas, na geração de conhecimento, que dinamize e fomenta os empreendimentos solidários sustentáveis (ESS - ONU) em redes de economia solidária e interesse coletivo, Dagnino formulou a síntese:

“tecnociência solidária é a decorrência cognitiva da ação de um coletivo de produtores sobre um processo de trabalho que, em função de um contexto socioeconômico – que engendra a propriedade coletiva dos meios de produção – e de um acordo social –que legitima o associativismo–, os quais ensejam, no ambiente produtivo, um controle –autogestionário– e uma cooperação –de tipo voluntário e participativo–, provoca uma modificação no produto gerado cujo resultado material pode ser apropriado segundo a decisão do coletivo –empreendimento solidário.”

Evitando aderir à simples concepção da contradição, de que haveria uma ciência imbuída de consciência social e boas intenções e outra sem ética e orientada apenas para a exploração, procurou desenvolver e aperfeiçoar o conceito de tecnociência solidária, para substituir o termo tecnologia social em sua negação, pode-se dizer, ideológica, da tecnologia convencional e ao mesmo tempo contrapor a ideia de uma neutralidade, mesmo que reversa, de atores que não se contaminaram com os interesses privados da tecnociência capitalista.

O autor ainda esclarece que não teria abandonado o termo tecnologia social, apenas por ele ter sido adaptado ao ressignificar teorias e conceitos distintos aos

originais, mas também porque essas adaptações e usos poderiam inferir equívocos, como de uma releitura da tecnologia tradicional, que seria desenvolvida na interação com a sociedade e orientada para uma transformação social. Diante desses e outros fatos, considerou mais apropriado assumir os riscos de conceber um novo conceito, que pode ser melhor compreendido pelas expressões contidas no artigo, explicitadas pela tabela a seguir:

(1) modo	é um modo do original, peculiar, aberto, mutante e adaptativo baseado numa cultura, marco analítico-conceitual e em instrumentos metodológico-operacionais específicos: a tecnociência solidária é uma proposta em construção
(2) conhecimentos	de qualquer natureza –científico, tecnológico, religioso, ancestral...– e origem –academia, empresas, povos originários, movimentos populares, excluídos...– desde que coerentes com os valores e interesses característicos da economia solidária
(3) devem	a tecnociência solidária é uma proposta assumidamente normativa, utópica, um objetivo estratégico em processo, a ser alcançado; associado a um dever-ser
(4) empregados	o que envolve instrumentos metodológico-operacionais específicos, como o da adequação sociotécnica que propõe com suas sete modalidades o reprojeto da tecnociência capitalista
(5) produção	a partir de insumos naturais, produtos e rejeitos de empresas, produtos de redes de economia solidária, etc.
(6) consumo de bens e serviços	consumo final de bens e serviços pelas famílias, diretamente, mediante moeda social, etc., ou através das compras públicas: <i>merenda e transporte escolar</i> ; consumo produtivo de bens e serviços de empresas ou de redes de economia solidária; aquisição, pelo Estado, de bens para implementar políticas públicas não relacionadas ao consumo das famílias – <i>uniforme para presos</i> – ou de serviços para implementar aquelas não relacionadas ao consumo das famílias – <i>manutenção de equipamentos públicos, cursos d'água</i> –, intermediação monetária, creditícia e financeira
(7) redes de economia solidária	insistência numa associação exclusiva e excludente entre tecnociência solidária e economia solidária, caracterizada pela propriedade coletiva dos meios de produção, autogestão, relações horizontais, solidariedade, distribuição do excedente material ou econômico decidida pelos seus integrantes; viabilizada mediante estratégias de “trabalho e renda” elaboradas –formuladas, implementadas e avaliadas– pelo Estado
(8) interesses	processos de adequação sociotécnica deverão sempre levar em conta os valores –morais, culturais, econômicos, etc.– e o interesse de consolidação e expansão da economia solidária
(9) necessidades	referência à diferença entre necessidade e demanda –necessidade com poder de compra–: satisfação de necessidades mediante valores de uso –necessidades materiais e demandas cognitivas associadas
(10) coletivas	o foco da economia solidária e, por isto, da tecnociência solidária, deve ser em atividades que visem ao bem-estar de toda a sociedade –e da conservação ambiental–, em especial as levadas a cabo de maneira também coletiva e associadas à vida em comunidade

Figura 13 - Conceitos sobre tecnociência solidária de Renato Dagnino

2.9 - Tecnologia social ou tecnociência solidária ?

Os conceitos tecnologia social e tecnociência solidária vêm se consolidando, pelo menos nos últimos 10 anos, como elementos muito presentes e pertinentes aos debates sobre Políticas de CT&I, bem como nos campos da Economia Solidária e dos Estudos Sociais da Ciência e Tecnologia. Nos três segmentos, existem imbricações (GRANOVETTER, 1973); (SCHOR, 2007) que aproximam esses campos, porém também exibem aspectos diametralmente opostos, embora não sejam mutuamente excludentes.

Esta dicotomia pode significar simplesmente abordagens teóricas específicas de campos de pesquisa diferentes ou podem estar relacionadas à concepção, orientação ou até mesmo à ideologia política da pessoa e/ou entidade que aborda ou pesquisa temas relacionados, indicando a complexidade do debate acerca da neutralidade da ciência.

A linha tênue que separa os mesmos conceitos nos campos da Economia, da Sociologia, dos estudos CTS (Ciência Tecnologia e Sociedade) e mesmo do campo da Administração Pública e da Ciência Política, também motivam debates acadêmicos, que de certa forma, são positivos para a concepção de quais aspectos são unânimes, discordantes ou específicos em cada abordagem.

A difusão da produção científica compreendida como tecnociência, permite o letramento científico da população local, mesmo dos que não estejam em fase escolar, colaborando também com a democratização da ciência. Ao divulgar ações e resultados da produção acadêmica para a população, a universidade aproxima-se do cidadão para poder ouvi-lo e entender suas demandas através da troca de saberes culturais ou profissionais.

Uma experiência real nesse sentido pode ser observada no minidocumentário ³⁶ do autor desta tese, sobre o projeto de extensão desenvolvido pelo NEPESC (Núcleo de Extensão e Pesquisa em Economia Solidária, Criativa e Cidadania da UNESP/FCLAr) em parceria com a Associação das Mulheres Assentadas da Padoka no Assentamento Monte Alegre em Araraquara - SP.

Esse grupo de mulheres de um dos assentamentos rurais do município fundou, em 2008, a Associação de Mulheres Assentadas (AMA), objetivando criar

³⁶ <https://eduplay.rnp.br/porta1/video/padoka2>

um local para produzir e comercializar produtos com qualidade alimentar e nutricional, com matérias-primas naturais e sem conservantes, cujas técnicas de produção protejam o meio ambiente. Embora em 2008 ainda não houvesse consolidado, localmente, o conceito de Empreendimento Econômico Solidário (ESS), a iniciativa das mulheres já estava alinhada, há anos, com objetivos de desenvolvimento sustentável (OBS) 1, 2, 5, 11 e 12 da UNESCO.

O minidocumentário mostra essa experiência no ecossistema empreendedor do município, cujo perfil agroindustrial sucroalcooleiro e citrícola mecanizado, contrasta fortemente com a produção manual de alimentos que podem ser considerados orgânicos. O empreendimento de economia solidária atende requisitos de erradicação da pobreza, geração de emprego e renda e criação de um ecossistema empreendedor sustentável.

Considerando que a difusão do conhecimento científico pode resolver problemas reais da população como tecnociência solidária (DAGNINO, 2019), este estudo pode mensurar impactos das ações de difusão científica, extensão universitária e troca de saberes com comunidades locais, coletando dados sobre o alcance dessa difusão e do letramento científico ali oferecidos e nos canais de mídia digital das universidades.

Os saberes, acadêmicos ou não, tácitos ou explícitos, registrados por oralidade ou multimídia digital, permeiam as relações econômicas e sociais dos moradores do território do assentamento. No projeto de Empreendimento Econômico Solidário (ESS), a universidade troca conhecimentos e saberes com os assentados de maneira espontânea, observa sua atuação para compreender suas idiossincrasias e propõe soluções simples e acessíveis. O intuito é fortalecer a circulação da produção no território, mas também possibilitar que possam ser comercializados em seu entorno para complementar a renda e obter acesso ao que não se produz no assentamento.

A Padoka foi criada no Assentamento Monte Alegre, que abrange uma área de 1.3 mil hectares, onde os assentados produzem cultivos sazonais, como milho, mandioca e feijão, bem como culturas perenes, como laranja, limão, goiaba, manga e café e hortaliças.

Também atuam na bovinocultura do leite, ovinocultura e avicultura de corte e postura, além de formarem pequenas agroindústrias familiares para a produção de farinha de mandioca, polvilho, rapadura, açúcar mascavo, doces, pães, compotas, mel e queijos.

Esses produtos abastecem as famílias dos assentados e são trocados entre eles, bem como vendidos por meio de políticas públicas federais ou estaduais, com destaque para o Programa Aquisição de Alimentos (PAA), o Programa Nacional de Alimentação Escolar (PNAE) e o Programa Paulista da Agricultura de Interesse Social (PPAIS).

Padoka do Assentamento Monte Alegre - AMA

O Empreendimento Solidário Sustentável de produção alimentícia, gerido pela Associação das Mulheres Assentadas (AMA), conhecido como "Padoka", realiza atividades de geração de trabalho e renda por uma população de assentados em área de alta vulnerabilidade socioeconômica no município de Araraquara.



Padoka do Assentamento Monte Alegre - AMA

Utilizam hortifrutigranjeiros produzidos no local em estreita relação de colaboração, encurtando a cadeia de produção e consumo através da comercialização de produtos prontos (pães e similares) no próprio assentamento, mas também em feiras e mercados rurais e urbanos, eventos públicos e privados no entorno.



Padoka do Assentamento Monte Alegre - AMA

Essas ações transformaram a realidade das mulheres do assentamento, ao criar um local para produzir e comercializar produtos com qualidade alimentar e nutricional, matérias-primas naturais, sem conservantes e cujas técnicas de produção protejam o meio ambiente.



Padoka do Assentamento Monte Alegre - AMA

Embora em 2008 ainda não houvesse consolidado, localmente, o conceito de empreendimento econômico solidário (ESS), a iniciativa das mulheres estava alinhada com objetivos de desenvolvimento sustentável preconizados pela ONU, como meta para a Agenda 2030.



Padoka do Assentamento Monte Alegre - AMA

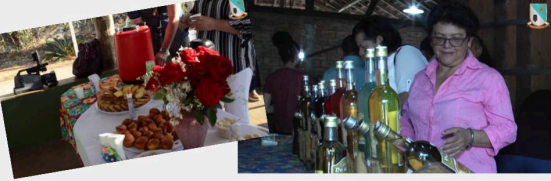
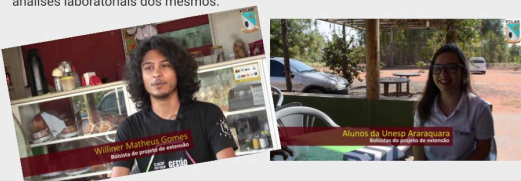
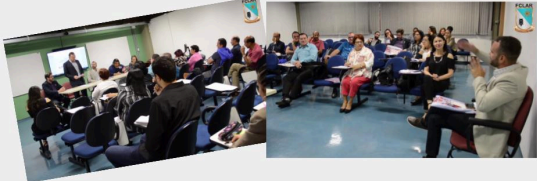
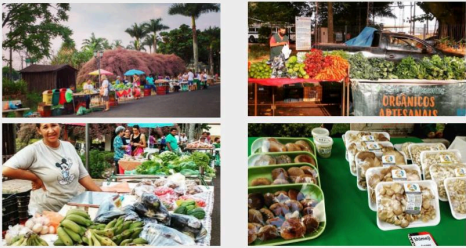
Tais metas buscam reduzir a extrema pobreza, erradicar a fome, buscar a igualdade de gênero, promover a produção e o consumo responsável em cidades e comunidades sustentáveis.



Padoka do Assentamento Monte Alegre - AMA

Nesse projeto que já dura 12 anos, houve investimento público através de ações da Unesp, da Coordenadoria de Economia Solidária, do Instituto de Terras de SP (ITESP - assentamento), de doações de entidades privadas e associações de classe e iniciativas da população que promove eventos que levam público ao local.



Padoka do Assentamento Monte Alegre - AMA Aos finais de semana há um grande afluxo de visitantes da região que apreciam muito os sucos, café, leite e lanches produzidos com matérias primas naturais do empreendimento solidário. A rota de estradas de terra é conhecida como trilha de "bikers" (grupos de ciclistas e familiares) do entorno (Araraquara, Matão, Bueno de Andrada, Rincão, etc ...) 	Padoka do Assentamento Monte Alegre - AMA O assentamento fica 15 quilômetros distante da cidade e abrange uma área de 1.3 mil hectares, onde os assentados produzem cultivos sazonais, como milho, mandioca e feijão, bem como culturas perenes, como laranja, limão, goiaba, manga e café e hortaliças. Também atuam na bovinocultura do leite, ovinocultura e avicultura de corte e postura, além de formarem pequenas agroindústrias familiares para a produção de farinha de mandioca, polvilho, rapadura, açúcar mascavo, doces, pães, compotas, mel e queijos. 
Padoka do Assentamento Monte Alegre - AMA Esses mesmos produtos abastecem as famílias dos assentados e são trocados entre eles, bem como vendidos por meio de políticas públicas federais ou estaduais, com destaque para o Programa Aquisição de Alimentos (PAA), o Programa Nacional de Alimentação Escolar (PNAE) e o Programa Paulista da Agricultura de Interesse Social (PPAIS). 	Padoka do Assentamento Monte Alegre - AMA As mulheres do empreendimento solidário sustentável recebem assessoria em tecnologias sociais dos alunos dos cursos de Economia, Administração Pública e Ciências Sociais em relação aos aspectos socio-econômicos e dos alunos dos cursos de Farmácia e Biotecnologia sobre a qualidade, conservação, rótulos dos alimentos e análises laboratoriais dos mesmos. 
Padoka do Assentamento Monte Alegre - AMA A mobilização das instâncias envolvidas (Universidade, Prefeitura e comunidades) conferem legitimidade aos projetos junto à população em situações de vulnerabilidade ou riscos (fome, sem teto) 	Núcleo de Extensão e Pesquisa em Economia Solidária, Criativa e Cidadania Feira de produtos orgânicos, artesanais e caseiros de produtores locais Assessoria às entidades e associações de produtores 
Quadro 3 - Slides com fotogramas retirados do documentário da Padoka - Fonte: o autor	

Neste capítulo são abordados fundamentos e origem da Ciência de Dados Sociais e sua aplicação nesta tese através de análises estatísticas, gráficos, tabelas, percepções e inferências quantitativas. Também são debatidos somente alguns aspectos fundamentais e sociopolíticos dos algoritmos e inteligência artificial neste item, porque serão extensivamente avaliados e analisados no próximo capítulo sobre redes complexas e Aprendizado de Máquina (AM).

As Ciências Sociais Computacionais praticamente emergiram das possibilidades do *big data* surgidas entre 2000 e 2008 e iniciaram um período comumente chamado de era dos dados (SANTOS 2, 2023). A natural transição da tecnologia da informação e da comunicação para a sociedade da informação e do conhecimento, desencadeou o pensar sobre a era dos dados, sendo inicialmente focada em desenvolver métodos computacionais com objetivo de obter avanços nos modelos de pesquisa em Ciências Sociais.

As práticas metodológicas de ambas se aproximam em diversos aspectos, porém há diferenças fundamentais em conceitos e finalidades (SOARES, 2021), embora a mais importante seja o fato de que as Ciências de Dados Sociais aplicam as metodologias, inclusive estatística e de Aprendizado de Máquina da Ciência de Dados, ao invés de apenas desenvolver métodos computacionais para atender demandas específicas. O quadro a seguir traça um comparativo entre quatro aspectos em comum das mesmas e esclarece diferenças e semelhanças entre si:

Aspecto	Ciência Social Computacional	Ciência de Dados Sociais
Foco	Desenvolvendo métodos computacionais	Aplicando métodos de ciência de dados
Objetivo	Avançar na pesquisa em ciências sociais	Resolver problemas sociais
Métodos	Modelagem baseada em agentes, aprendizado de máquina, mineração de dados	Aprendizado de máquina, estatística, visualização de dados
Aplicações típicas	Simulação de sistemas sociais complexos, análise de redes sociais	Previsão de resultados sociais, análise de dados de mídia social
Quadro 4 - Compara aspectos das Ciências Sociais Computacionais e de Dados - Fonte: o autor		

A tese aborda a concepção das Ciências Sociais na era dos dados, partindo do artigo que propõe a chamada Ciência Social Computacional em 2009.

3.1 - A origem na Ciência Social Computacional

A realidade interconectada da sociedade globalizada em redes informacionais (CASTELLS, 1999; BLACKMAM, 1998; JENKINS, 2003; LÉVY, 2000) ocasionou a emergência de compreendermos a complexidade de nosso comportamento social e comunicacional. Em apenas algumas décadas ela se alterou profundamente e desencadeou fenômenos culturais, econômicos e políticos que reafirmam a disruptura como estratégia de transformação digital (FERREIRA, 2016) .

Sugerindo que a emergência de novos sistemas comunicacionais e produtivos é correlata a processos sociais adaptativos (LUHMANN, 2005), a disparidade de velocidade entre a evolução cultural e evolução genética é um indicativo desta nova realidade, contrapondo a forma como ela é vista na teoria do processo evolucionário de dupla herança (BOYD, RICHERSON, 1985).

A tecnologia da informação apresenta oportunidades para o desenvolvimento econômico, mas também pode ser prejudicial e eliminar mercados existentes. Por exemplo, o streaming teve um impacto negativo sobre as locadoras de vídeo, mas fortaleceu o mercado de produção de filmes.

Embora não haja consenso acadêmico sobre essas emergências que funcionam como catalisadores de processos sociais adaptativos, os indivíduos conectados estão buscando se conscientizar das idiossincrasias das redes informacionais, adaptando-se à nova economia digital e acoplando suas expertises aos meios digitais de reprodução ampliada do capital tecnológico-informacional (ARAÚJO, MELO, 2007)

No artigo Manifesto de Ciência Social Computacional publicado no The European Physics Journal Special Topics (CONTE et al, 2013), doze pesquisadores da Europa e dos EUA afirmam que a crescente integração das tecnologias da informação globais na vida quotidiana está a gerar uma quantidade sem precedentes de circulação e tratamento de dados.

...cientistas sociais e comportamentais, cientistas cognitivos, teóricos de simulação baseada em agentes, cientistas da computação, matemáticos e físicos podem cooperar lado a lado para que se possa chegar a modelos inovadores e teoricamente fundamentados a partir de fenômenos em foco (CONTE et al, 2013, p. 5)

Para esse grupo de cientistas, essa realidade indicava a necessidade de compreensão da complexidade em uma nova disciplina chamada Ciência Social Computacional (LAZER et al., 2009) . Essa abordagem interdisciplinar pode indicar uma nova era de compreensão das estruturas e funções da sociedade em diferentes níveis através da colaboração entre cientistas no desenvolvimento de modelos que respondam aos desafios dessa era (IBIDEM).

A Ciência Social Computacional (CSC) pode oferecer os meios e ferramentas para lidar com o volume massivo de dados gerados pela sociedade em redes informacionais. A nova abordagem articula a pesquisa empírica das nossas pegadas digitais³⁷ com a teoria científica que modela técnicas computacionais de redes complexas em que estamos imersos.

O artigo de 2009 da Science, de Lazer et al., questiona as barreiras ao reconhecimento da Ciência Social Computacional como uma nova fronteira científica, argumentando que as tecnologias de informação globalizadas já permitiam a análise de estruturas e conteúdos em séries temporais, uma vez que esses dados já estariam disponíveis em repositórios de empresas e instituições (*big data*) .

O artigo discutiu o uso eleitoral de redes de informações e o impacto do Processamento de Linguagem Natural (PLN) na organização e análise de textos da Internet. No entanto, ele não previu o surgimento de fazendas de robôs virtuais influenciando as eleições ou o uso generalizado do IA generativa para recriação de conteúdo em todos os segmentos da sociedade.

Surgiram obstáculos institucionais ao avanço da Ciência Social Computacional (CSS) devido às leis de privacidade de dados e à necessidade de consentimento explícito para o uso de dados pessoais. Também mencionam um maior distanciamento físico e institucional entre departamentos de ciências sociais, engenharia e informática nas instituições de ensino superior e nas corporações privadas.

Na finalização, o artigo assume as restrições de acesso e o direito à privacidade como possíveis entraves aos métodos da Ciência Social Computacional

³⁷ Uma pegada digital, às vezes chamada de sombra digital ou pegada eletrônica, refere-se ao rastro de dados que você deixa ao usar a Internet. Inclui sites que você visita, e-mails que você envia e informações que você envia on-line. <https://www.kaspersky.com.br/resource-center/definitions/what-is-a-digital-footprint>

(CSC) e propõe um regime auto-regulador de procedimentos, tecnologias e regras que reduz estes riscos, mas preserve o potencial investigativo das pesquisas.

No encerramento os autores questionam se os investimentos aportados no desenvolvimento das ciências cognitivas, que já articulavam plenamente a atuação de biólogos, filósofos e cientistas computacionais trabalhando coletivamente, poderiam receber os mesmos incentivos e investimentos das instituições públicas ou privadas

3.2 - Ciências Sociais Computacionais na onda positivista

Em 2020, a Ciência Social Computacional (CSC) era um campo estabelecido globalmente com produção acadêmica relevante. Ela havia desenvolvido métodos híbridos para analisar bancos de dados grandes e complexos, e novas modalidades de pesquisa estavam surgindo em vários subcampos (Edelmann et al. 2020).

O artigo daquele ano analisa a evolução da ciência social computacional na sociologia por meio de um estudo bibliométrico e da análise de subcampos com produção relevante em novas modalidades de pesquisa.

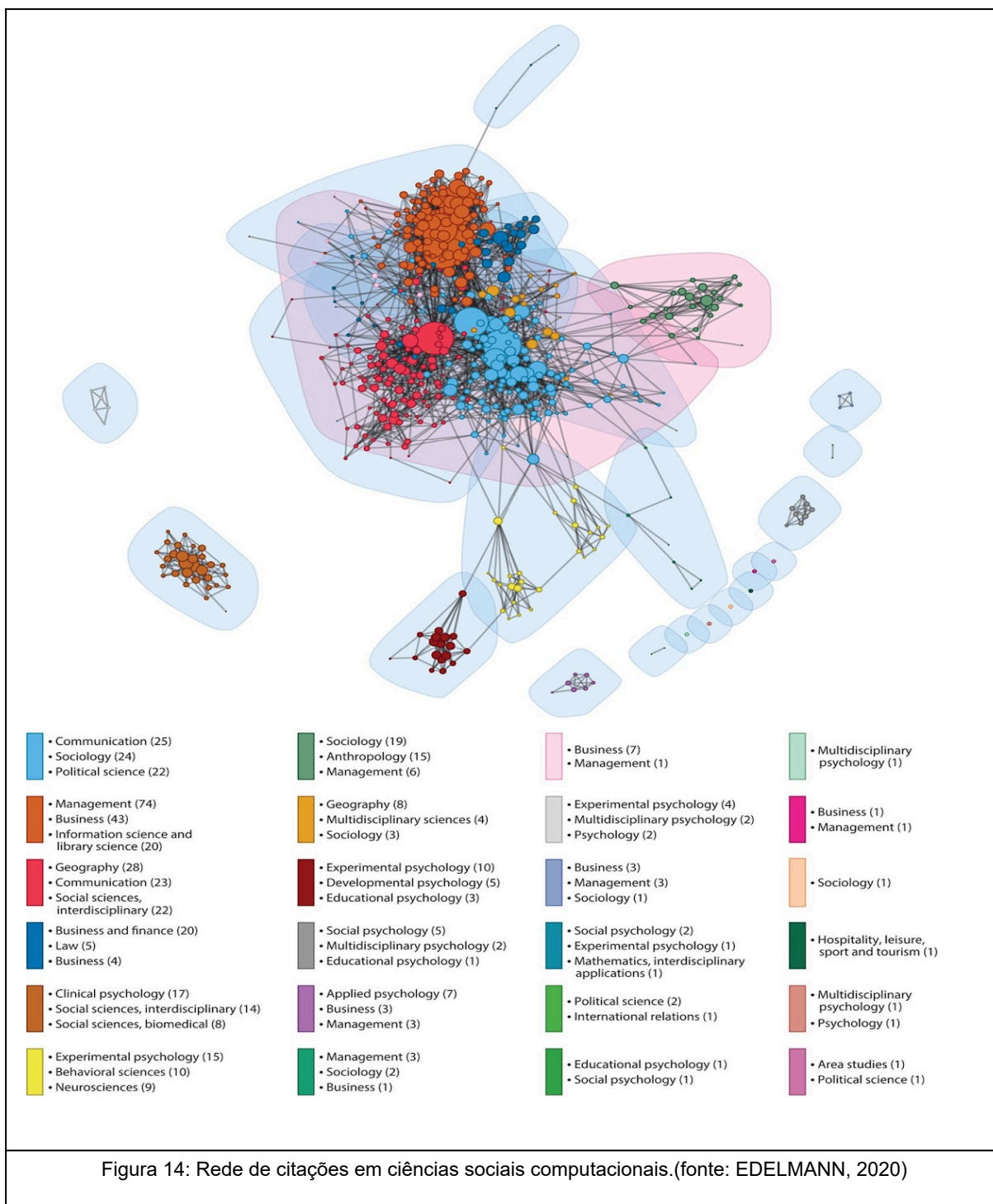
Ao contrário dos conjuntos de dados convencionais coletados por cientistas sociais, essas novas fontes digitais geralmente fornecem detalhes ricos sobre a evolução das relações sociais em grandes populações conforme elas acontecem (IBIDEM)

A revisão mostrou que as ciências sociais estariam na vanguarda da pesquisa sobre comportamento humano de longo prazo e espaços digitais. Inicialmente, o termo se referia a métodos computacionais de modelagem baseada em agentes ou simulações de comportamento humano.

Edelmann et al. (2020) também encontraram contradições no uso do termo “ciência social computacional”, principalmente no campo STEM, onde muitas vezes, interpretam erroneamente qualquer análise de dados em larga escala do comportamento humano como ciência social computacional, mesmo quando usam teorias de física e matemática sem metodologias de ciências sociais.

Alguns renomados pesquisadores de STEM trataram as ciências sociais como campo a ser colonizado e produziram verdadeiras aberrações metodológicas e confusões conceituais, como classificar essa nova ciência como forense (MCFARLAND et al. 2015) ou tentar chamá-la de física social (PENTLAND, 2015),

termo amplamente rejeitado pelos cientistas sociais desde a primeira tentativa positivista perpetrada por Augusto Comte em 1822³⁸.



A pesquisa usando “big data” e “ciência social computacional” (CSC) entre 2000 e 2016 mostrou um crescimento nas publicações de CSC em várias disciplinas

³⁸ Comte apresentou o termo “física social” em 1822 em um ensaio que marcou época, onde ele expôs o programa positivista. in LEVINE, D. Visões da tradição sociológica. Jorge Zahar Eds., 1997. 325p. - RJ

desde 2012. Uma análise de rede de citações identificou 24 áreas de conhecimento conectadas por meio de métodos e processos de CSC.

O grafo gerado a partir dessa rede de autores e citações mostrou um conglomerado de quatro comunidades no centro da rede. A primeira delas conecta comunicação, sociologia e ciência política. A segunda congrega trabalhos em geografia, comunicação e ciências sociais interdisciplinares, a terceira une um considerável número de papers sobre negócios e biblioteconomia e a quarta e última conecta negócios, finanças e direito.

Alguns desses estudos pesquisaram a difusão em redes sem escala e o controle de suas topologias, mostrando particularidades e semelhanças entre redes fortemente agrupadas e aleatórias. Também mostrou que a topologia de rede pode moldar a adoção de convenções sociais e que dados do facebook indicam haver sinergia entre a difusão de estilos comunicativos emocionais e racionais em uma grande rede de debates sobre saúde pública.

Juntos, alguns desses estudos mostram que as interações em micronível entre indivíduos podem gerar padrões de difusão em macronível antecipados pelos primeiros teóricos das ciências sociais, como Gabriel de Tarde (EDELMANN et. al., 2020).

Um estudo que abrangeu 56 milhões de usuários do Twitter e 58 milhões de assinantes de smartphones contradiz o conceito de que os laços de redes sociais próximas são mais fortes do que as conexões distantes. Estudando essa relação em 11 redes de grande escala populacional culturalmente diversas em quatro continentes, concluíram que fortes laços entre redes podem ser importantes para a propagação de ideias ou doenças, podendo influenciar também a difusão e integração social (PARK et. al, 2018)

Os autores concluem que a rápida expansão da ciência social computacional torna as revisões de literatura rapidamente obsoletas. Eles acreditam que pesquisas futuras influentes vincularão níveis macro (como mudança cultural) a processos de tomada de decisão micro. No entanto, alertam que isso exigirá métodos híbridos criativos, e que o aprendizado de máquina sozinho não explicará o comportamento coletivo.

3.3 - A multidisciplinaridade da CSC

A CSC como novo campo sem claras delimitações metodológicas ou mesmo epistemológicas, integra as ciências sociais com as áreas das ciências da computação, da ciência das redes e das ciências da complexidade.

As ciências computacionais aportam as ferramentas (softwares e hardwares) para a coleta, armazenamento, manipulação, consolidação, processamento, análise, classificação e visualização de dados, a ciência das redes oferece o recurso dos grafos e metodologias de percolação, redes pequeno mundo com hubs e clusters e as ciências da complexidade oferecem o arcabouço conceitual, matemático e estatístico de análise.

Diferentes pesquisadores definem a CSC com algumas particularidades e muitas similaridades, não havendo um pleno consenso metodológico ou conceitual, mas concebem a vida social como ancorada nas noções de emergência e complexidade (PAVESI e VALENTIM, 2021) para compreender a dinâmica social.

Níveis de aplicação em projetos de CSC	
Fase Empírica	Construção de Big Data: Bancos de Dados Públicos e privados, Dados coletados em Campo (automatizados, semi - automatizados e artesanais), Web Scraping ³⁹ , limpeza e preparação de dados;
Fase Analítica	Análise de Redes Sociais: construção de modelos por Machine Learning e Processamento de Linguagem Natural para categorizar dados e formular hipóteses;
Fase Teórica	Transformação do modelo: de Machine Learning em uma Simulação Computacional Baseada em Agentes para fazer experimentos, previsões, predições e formular teorias
Quadro 5: Níveis de aplicação em projetos de CSC (in PAVESI e VALENTIM, 2021)	

O crescimento exponencial dos dados na era da informação permitiu análises quantitativas e qualitativas e em 2022, o uso generalizado de smartphones transformou isso em um novo modelo de negócios para as grandes empresas de tecnologia (*big-techs*). As mídias sociais permitem a análise em tempo real dos processos sociológicos e apoia o monitoramento de políticas públicas, tornando os indicadores mais acessíveis aos movimentos sociais e ao público em geral.(BALTAR E BALTAR, 2013).

³⁹ A coleta de dados web, ou raspagem web, é uma forma de mineração que permite a extração de dados de sites da web convertendo-os em informação estruturada para posterior análise

Novos métodos de computação trazem novos conhecimentos e terminologias com os quais os cientistas sociais não estão familiarizados. Esses cientistas levantam a preocupação de que antigos debates, como o evolucionismo social, o determinismo biológico e o funcionalismo, estejam sendo revividos (IBIDEM, pág. 16). A evolução social, cultural e organizacional é inegável e vem sendo estudada há muito tempo. O tempo não é um fator fundamental na transformação social; a dinâmica histórica e a noção de emergência mostram que a mudança é constante.

“Os fenômenos sociais emergem das relações sociais e não da vontade individual, ainda que os indivíduos sejam agentes racionais dessas mudanças” (IBIDEM. p. 17)

Partindo da concepção de que os fenômenos sociais estão inter- relacionados entre si, a noção de emergência decorre da complexidade de mudanças que afetam estruturas compostas por diversas partes de um todo conectadas em redes informacionais.

As CSCs permitem análises quantitativas e qualitativas, modelagem e simulações para entender o surgimento de fenômenos problemáticos (por exemplo, migrações, desigualdades, desinformação) e fenômenos positivos (por exemplo, redes de solidariedade, cooperação científica) que desafiam os métodos tradicionais das ciências humanas.

Os diversos dossiês sobre a CSC no Brasil e exterior aqui citados indicam um caráter quase panfletário de reflexões que identificam a difícil tarefa de superar o artesanato intelectual (BALTAR E BALTAR, 2013) , desde sua provável consolidação como arcabouço conceitual de coleta, processamento, classificação, análise de dados e inferência estatística como método que requer infraestrutura, cooperação multidisciplinar e investimentos para enfrentar a multitude em escala local-global das relações humanas virtualizadas ou não.

Não há consenso nas ciências sociais sobre fenômenos simples que explicam a dinâmica de instituições e das superestruturas, das normas e da coerção social ou mesmo da noção de indivíduo ou do devir. Alguns estudos computacionais sobre a cooperação em sistemas sociais criaram modelos estatísticos que foram muito utilizados para explicar conceitos como cooperação, altruísmo ou racionalidade limitada através de simulações que usam analogias de bares (Modelo El Farol) e

prisões (Dilema do prisioneiro) e exemplificar a noção de emergência e também a chamada “Economia da Complexidade” (MITCHELL, 2009).

As Ciências de Dados Sociais já é vista e considerada uma condensação das modalidades de coleta, mineração e análises de dados aqui expostas, que descreveram algumas origens e idiosincrasias da CSC nas dinâmicas de pesquisa no mundo todo. A ciência de dados sociais é um campo multidisciplinar que usa técnicas e conhecimentos de estatística, matemática, ciência da computação e ciências sociais.

No momento da coleta e análise de dados desta pesquisa, optou-se por usar modelos algorítmicos para apoiar as tarefas e obter dataframes multidimensionais que foram processados, classificados, vetorizados e organizados manualmente, mas também com auxílio de recursos de aprendizado de máquina, comumente chamados de IA generativa.

3.4 - Um breve resumo sobre algoritmos

Em matemática e ciência da computação, um algoritmo é uma sequência finita de ações executáveis que visam obter uma solução para um determinado tipo de problema, sendo procedimentos precisos, não ambíguos, padronizados, eficientes e corretos. (DASGUPTA, PAPADIMITRIOU E VAZIRANI, 2008)

Os diagramas lógicos de execução de tarefas também representam uma estrutura de dados semânticos, filtros de imagem e de som que moldam a ordem e os interesses das pessoas, mesmo sem sua plena consciência ou consentimento. Eles também enviam anúncios personalizados com base nos metadados coletados, refletindo os hábitos, os interesses e até mesmo as doenças dos indivíduos. (LÉVY, 2000) (PARISER, 2012).

O algoritmo mais usado globalmente é o mecanismo de busca do Google. Inicialmente, ele se baseou na pesquisa do matemático italiano Massimo Marchiori, que criou o Hypersearch, um mecanismo de busca com foco nas relações entre páginas da Web. Ele inspirou os cofundadores do Google, Larry Page e Sergey Brin, que citaram o algoritmo ao introduzirem o PageRank⁴⁰ em 1997.

O algoritmo PageRank foi patenteado pela Universidade de Stanford e

⁴⁰ <http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>

licenciado exclusivamente para o Google. Stanford recebeu em troca 1,8 milhão de ações do Google, que foram vendidas em 2005 por US\$336 milhões. O atual algoritmo RankBrain evoluiu a partir do PageRank e agora usa redes neurais artificiais, além de gráficos matemáticos e topologias de rede de mundo pequeno.

Os algoritmos dos mecanismos de pesquisa, ativos valiosos para as empresas de tecnologia, são explorados pela priorização de anúncios pagos nos resultados de pesquisa. Esses algoritmos, como as caixas pretas (LATOUR, 2000) são inescrutáveis; suas entradas e saídas são conhecidas, mas seu funcionamento interno permanece secreto.

Quando usamos mecanismos de pesquisa, fornecemos metadados que são processados por algoritmos para fornecer resultados e anúncios adaptados aos nossos hábitos. Em 2022, o Google pode coletar dados em tempo real sobre nossas localizações e fornecer um mapa que rastreia nossos movimentos, incluindo detalhes como horário, rotas e compras.

Consentimos com esses mecanismos quando concordamos com os termos de uso, muitas vezes sem entender a profundidade da coleta de dados. A ausência de leis específicas que regem as operações algorítmicas, na maioria dos países, permite que as grandes empresas de tecnologia explorem nossos dados pessoais, transformando-nos em mercadorias para sua hiper infraestrutura global voltada para o lucro. (EMPOLI, 2022; O'NEIL, 2020; D'ANCONA, 2018)

Há diversas violações de nossa privacidade, um princípio fundamental do bem-estar social e consagrado em nossa Constituição. Essas empresas desconsideram nossa intimidade, vida privada, honra e imagem, e seus algoritmos perpetuam preconceitos racistas, sexistas e colonialistas. (AMADEU DA SILVEIRA in Cassino et al., 2021)

3.5 - Aspectos fundamentais dos algoritmos de Aprendizado de Máquina

O desenvolvimento de algoritmos levou ao conceito de inteligência artificial (IA), que, embora pareça comparável às capacidades intelectuais humanas, não se iguala ao poder de processamento, à conectividade e à plasticidade do cérebro humano. A IA está relacionada a emulações e simulações de redes neurais que podem aprender a detectar padrões e sinais em dados, estruturados ou não,

processados em neurônios artificiais (SEJNOWSKI, 2019).

Embora sejam inspirados em modelos de funcionamento de neurônios biológicos, os recursos da IA são extremamente limitados quando comparados à eficiência da mente humana, mas muito mais eficientes em outras tarefas como reconhecimento de padrões, percepção de anomalias ou agrupamentos e processamento paralelo massivo (BABINI & MARRANGHELLO, 2007).

Elas também podem aprender com as simulações e o chamado treinamento, quando os dados são apresentados sistematicamente aos neurônios artificiais que os processam em diferentes níveis e sistemas de pesos, simulando nosso processo de aprendizagem (IBIDEM).

Uma diferença fundamental entre os algoritmos é a linearidade das ações que os mais simples performam executando uma tarefa atrás da outra, como numa receita de bolo sendo executada passo a passo. Já os algoritmos de IA executam ações simultâneas diferenciadas em paralelo, além de poderem evoluir através do aprendizado resultante de processamentos anteriores.

Os seres humanos são especializados em reconhecer padrões e nosso cérebro funciona por reações que resultam de nossa percepção dos mesmos em eventos anteriores, ou seja, a memória de nossas experiências nos orienta sobre as decisões que tomamos em nossa rotina diária. O aprendizado é fruto da realimentação de nossas experiências no processamento de novas informações.

Uma instância de IA funciona de forma semelhante, definindo e analisando dados que são chamados de objetos (análise de objetos), extraindo as características que os diferenciam e delimitam (extrator de características), selecionando as mesmas (seletor) e implementando um classificador que reconhecerá padrões e poderá realizar inferências estatísticas sobre o conjunto de dados (SEJNOWSKI, 2019).

Os dados podem ser processados de forma Simbólica, onde cada elemento de informação possui um significado a ele associado e/ou de forma Sub-simbólica, onde dados numéricos são manipulados na busca de padrões que geram classificações dessas características (MINSKY, 1991); (SILVA et al., 2011). Essas duas modalidades de processamento não são antagônicas e nem auto-excludentes, podendo ser combinadas e até mesmo complementares.

O processo de composição de uma IA consiste em quatro estágios: ingestão, preparação, treinamento e inferência. Os dados processados são armazenados na nuvem e disponibilizados para aplicações que atendem às demandas dos usuários. Exemplos de aplicações cotidianas de IA incluem tradutores de idiomas e reconhecimento biométrico, embora este último tenha levantado debates sobre vieses racistas e sexistas, mesmo que não intencionalmente.

3.6 - Vieses racistas e sexistas em algoritmos de reconhecimento facial

O Instituto Nacional de Padrões e Tecnologia dos EUA (NIST) realiza testes em produtos tecnológicos para medir padrões, eficiência e métricas. O NIST é uma instituição pública que compartilha suas descobertas por meio de boletins e já empregou cinco ganhadores do Prêmio Nobel. Sua meta é impulsionar a inovação e a competitividade industrial nos EUA por meio de avanços na ciência, nos padrões e na tecnologia de testes e padronizações, com o objetivo de melhorar a segurança econômica e tecnológica.

A entidade submete as tecnologias biométricas a testes tecnocientíficos desde os anos 1960 e no caso do reconhecimento facial, desde 2000 oferece o Teste de Fornecedor de Reconhecimento Facial (FRVT)⁴¹, avaliando as capacidades desses algoritmos nas modalidades de reconhecimento de “um para muitos” quando utilizado para a identificação de indivíduos em bases de dados fotográficas e de “um para um” que são utilizados para verificação de identidades afirmadas.

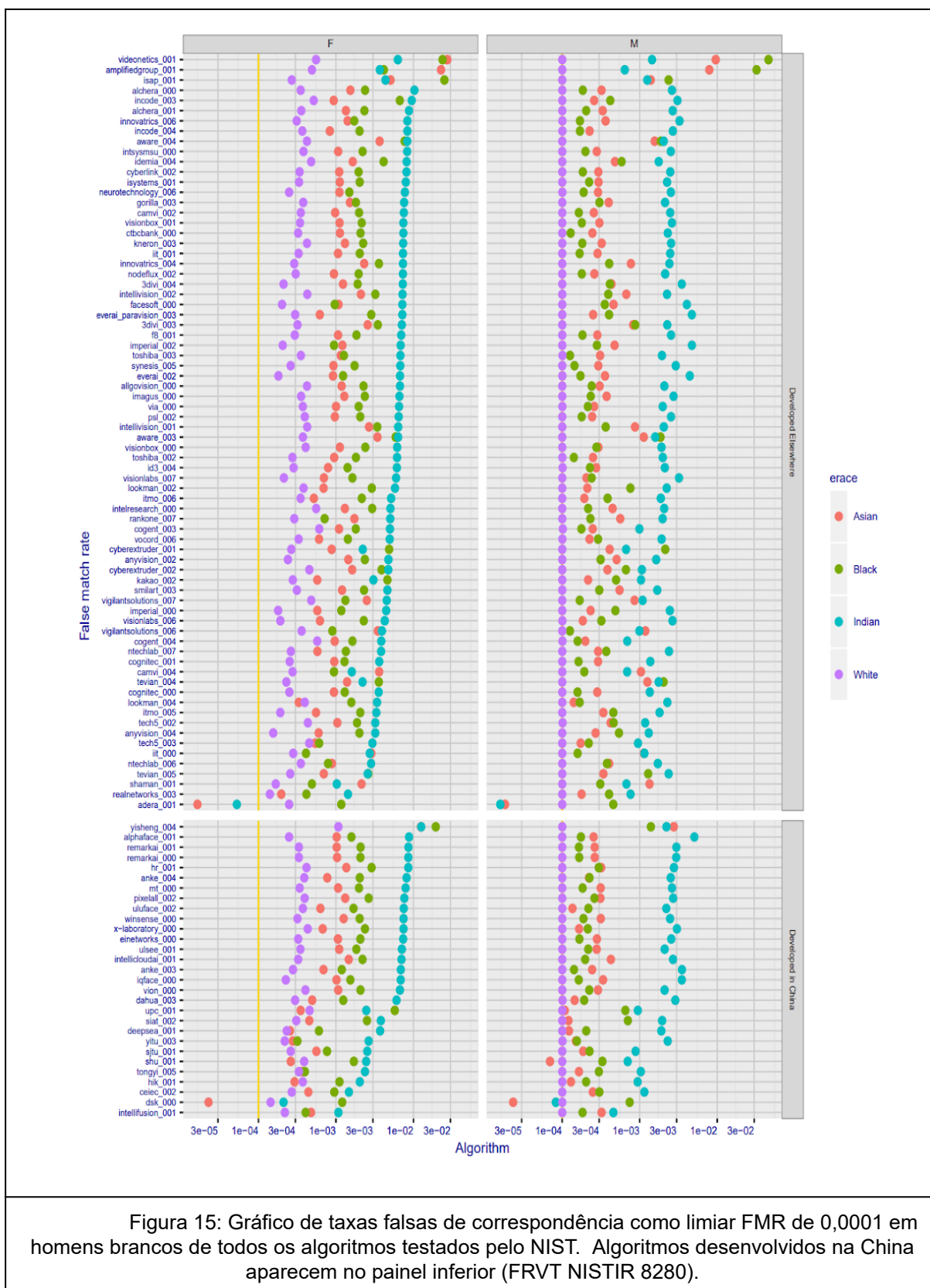
A precisão dos modelos de reconhecimento facial é avaliada com base em falsos positivos e falsos negativos. Um falso positivo ocorre quando o software identifica incorretamente dois indivíduos diferentes como a mesma pessoa. Um falso negativo ocorre quando o software não consegue reconhecer a mesma pessoa em duas fotos diferentes.

NIST realiza testes FVRT regulares e abertos, e em março de 2022, lançou a publicação SP 1270. Esta publicação oferece orientação sobre como gerenciar o viés na IA, explicando que o mesmo se origina não apenas dos dados e processos de aprendizado de máquina, mas também de fatores sociais e institucionais mais amplos que influenciam o desenvolvimento e a implantação da IA.

⁴¹ [Face Recognition Vendor Test \(FRVT\) | NIST](#)

O estudo NISTIR 8280 FRVT de 2019 encontrou evidências de viés demográfico em algoritmos de reconhecimento facial de 99 fornecedores, incluindo empresas, universidades e startups. O estudo usou 18,27 milhões de fotos de 8,49 milhões de pessoas, de bancos de dados operacionais fornecidos pelo Departamento de Estado dos EUA, Departamento de Segurança Interna e FBI.

As fotos foram coletadas de conjuntos de dados para viagens autorizadas, imigração ou aplicação da lei, e a análise foi classificada em 24 países de 7 regiões globais. Os testes mostraram diferenças demográficas nas taxas de falsos positivos e falsos negativos. Os diferenciais de falsos positivos eram muito maiores do que os diferenciais de falsos negativos, e nem todos os algoritmos exibiam os dois tipos. As taxas de falsos positivos variaram de 10 a mais de 100 vezes os resultados falsos, enquanto as taxas de falsos negativos eram menos de 3 vezes os resultados falsos.



A amostra FRVT NISTIR 8280 de 2019 mostrou que os algoritmos das empresas asiáticas, treinados em amostras étnicas mais diversas, tiveram um desempenho variado. Apesar da variação dos níveis de precisão, o viés favorece o

reconhecimento de homens brancos devido à predominância de dados de treinamento. O gráfico revela um enorme desequilíbrio nos dados de treinamento ou preconceito institucional, evidente nas taxas de erro entre os gêneros e em todos os 99 fornecedores.

Para cada algoritmo de verificação, os pontos representam taxas de falsa correspondência para comparações entre pessoas do mesmo sexo e da mesma raça. O limite, definido para cada algoritmo, produz FMR = 0,0001 para homens brancos. Os algoritmos são ordenados pelo pior caso de FMR, normalmente para mulheres indígenas americanas (FRVT NISTIR 8280).

Os falsos positivos na tecnologia de reconhecimento facial são significativamente maiores em fotografias de pessoas da África Ocidental e Oriental e do Leste Asiático, e menores em indivíduos do Leste Europeu, com diferenças de mais de 100 ocorrências entre os países. Também são numerosos entre os índios das Américas do Norte e do Sul, populações de descendência africana e asiática, especialmente mulheres, e também são recorrentes no reconhecimento de idosos e crianças.

Mesmo dentro de um parâmetro mínimo de erros em amostra de larga escala, são estatisticamente desfavoráveis às mulheres, principalmente as afrodescendentes, indígenas e orientais. Os falsos positivos podem desencadear sérios problemas na privacidade, direitos civis e liberdade das pessoas, chegando até mesmo a criminalização por acusações fundadas em sistemas de inteligência artificial.

3.7 - Contextos que são diretamente afetados pela Inteligência Artificial Há muitos casos em que a IA é usada para coletar dados, analisar dados biométricos e mapear emoções sem nosso conhecimento ou consentimento. Há um consenso geral de que a IA deve ser desenvolvida e aplicada de forma ética devido a preocupações com a perda de controle e a privacidade individual (KERR et al, 2021). A falta de transparência nas suposições subjacentes dos algoritmos e sua complexidade podem levar a uma desconfiança em relação à tecnologia de IA, especialmente porque ela pode afetar diretamente os cidadãos sem seu total consentimento. Isso pode resultar em sentimentos de estigmatização e causar

danos morais, sociais e psicológicos. Além disso, a manipulação e a classificação de grandes conjuntos de dados podem ampliar os preconceitos existentes, sejam eles intencionais ou não, e causar mais danos (IBIDEM).

A falta de transparência e a complexidade dos algoritmos de IA podem levar à desconfiança e afetar diretamente os cidadãos, resultar em sentimentos de estigmatização e causar danos morais, sociais e psicológicos. Além disso, a manipulação e a classificação de grandes conjuntos de dados podem amplificar os preconceitos existentes, intencionais ou não, e causar mais danos.

Os sistemas de recomendação em tempo real, embora lucrativos, podem parecer invasivos e diminuir a confiança devido à sua natureza insistente, levando a um efeito contraproducente no comportamento de compra.

A inteligência artificial não é imune à discriminação racial, social ou a práticas injustas. O NIST recomenda uma abordagem sociotécnica da IA, reconhecendo que a IA não é apenas constituída por processos matemáticos e construções computacionais, mas também por interações humanas.

O enviesamento nos sistemas técnicos é um fenômeno estatístico que distorce os erros aleatórios, o que pode afetar a manipulação dos dados. Até a Organização Internacional de Normalização (ISO) define o enviesamento como o grau em que um valor de referência se desvia da realidade, tornando o sistema enviesado e impreciso.

Os preconceitos enviesados na IA podem ser perpetuados por códigos de ética e legislação, apesar do tratamento de dados pessoais automatizado e sigiloso pelas grandes empresas de tecnologia. Sistemas de recomendação e ranqueamento, afetados por preconceitos cognitivos, podem perpetuar vieses e preconceitos sistêmicos e institucionais se não forem tratados.

O relatório do NIST confirmou os preconceitos da IA e propôs uma abordagem sociotécnica que combina técnicas matemáticas/computacionais com valores sociais. O relatório sublinhou o estabelecimento de protocolos públicos para a transparência, conjuntos de dados abertos e métricas T.E.V.V (Transparência, avaliação, validação, e verificação) e destacou a urgência de abordar as implicações sociais e os direitos relacionados com o desenvolvimento da IA.

O viés em IA não se originou na tecnologia da informação, mas o desenvolvimento e uso de IA criaram uma necessidade urgente de melhores práticas para lidar com as implicações sociais e os direitos relacionados. O viés da IA vem de várias fontes e os desenvolvedores precisam mitigá-lo em desafio contínuo, com implicações sociais crescentes que exigem práticas melhores e flexíveis, independentemente da política e da indústria (NIST, 2019).



Figura 16: Exemplos de amostra e vieses publicada em matéria sobre reconhecimento facial: [NYT](#)

O uso da tecnologia de reconhecimento facial pelo governo exige alta precisão devido à diversidade da população e à natureza sensível das informações utilizadas. Os erros nesses sistemas podem causar danos irreparáveis aos cidadãos e imagens de alta qualidade também são essenciais para o reconhecimento facial preciso.

Os sistemas de reconhecimento facial requerem imagens de alta qualidade para serem exatas e diferentes aplicações de IA têm diferentes requisitos de segurança, precisão e correspondência, dependendo do seu objetivo. As leis devem obrigar uma supervisão humana e critérios claros para buscas e detenções baseadas no reconhecimento facial.

O preconceito sistêmico pode resultar de procedimentos e práticas de instituições públicas e privadas que, involuntariamente, favorecem certos grupos

sociais em detrimento de outros, muitas vezes devido a costumes e práticas que não proporcionam acesso universal ou não acomodam a diversidade.

As tecnologias de IA podem ser tendenciosas, levando à exclusão social no Brasil, onde a fome sistêmica e o desemprego afetam desproporcionalmente as mulheres negras e os pobres, que podem ser ainda mais marginalizados pelos sistemas de candidatura a emprego baseados em dados na nossa sociedade orientada por algoritmos.

3.8 - Impactos sociopolíticos de algoritmos no cotidiano dos brasileiros

Atualmente no Brasil, cerca de 56% da população brasileira é negra e as mulheres representam 51,8% dos habitantes, segundo o PNAD do IBGE, porém, no que se refere à inclusão social e igualdade de condições e oportunidades, essa proporcionalidade não favorece essas duas maiorias (IBGE, 2019).

Segundo o mesmo relatório, as mulheres estudam mais, trabalham mais e ganham menos que os homens, pois ocupam 37,4% dos cargos gerenciais e recebem 77,7% do rendimento dos homens⁴² e a população negra representa até 41% das pessoas abaixo da linha da pobreza (vivem com menos de US\$5,50 por dia)⁴³.

Ainda há um agravamento dessas dificuldades para as mulheres negras, que embora representem 47,8% da população feminina, seus rendimentos alcançam apenas 44% em relação aos salários dos homens brancos.

Também há uma discrepância na proporcionalidade de representação política, onde 75,6% dos deputados e deputadas federais eleitos em 2018 são pessoas brancas, contra apenas 24,4% de pessoas negras e pardas⁴⁴.

A sociedade brasileira é predominantemente liderada por homens brancos, o que dificulta a inclusão social de pessoas negras e a igualdade de oportunidades para as mulheres. Essa desigualdade pode ser agravada pelo controle das empresas “big tech” sobre o acesso à informação por meio de algoritmos. Os algoritmos de aprendizado de máquinas não são regionalizados e geralmente refletem vieses racistas e/ou machistas de seus países de origem, no

⁴²<https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/20234-mulher-estuda-mais-trabalha-mais-e-ganha-menos-do-que-o-homem>

⁴³ https://biblioteca.ibge.gov.br/visualizacao/livros/liv101681_informativo.pdf

⁴⁴ <https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2101681>

hemisfério norte ocidental. Já é de conhecimento público que o reconhecimento facial executado por esses algoritmos não efetivam ou erram a análise de rostos de mulheres negras⁴⁵, criando constrangimento e realimentando sua exclusão.

Estes preconceitos reforçam uma visão colonialista e excludente, especialmente contra as mulheres negras, que sistematicamente não têm acesso igual ao trabalho e ao rendimento. Isto é evidente nas notícias recorrentes de mulheres negras que são libertadas da escravidão moderna⁴⁶ de forma recorrente.

O acesso à internet é essencial para que os brasileiros exerçam a cidadania, mas esse acesso não é gratuito. Isto cria uma exclusão sistemática dos pobres, especialmente das mulheres negras, que estão desproporcionalmente representadas abaixo do limiar da pobreza.

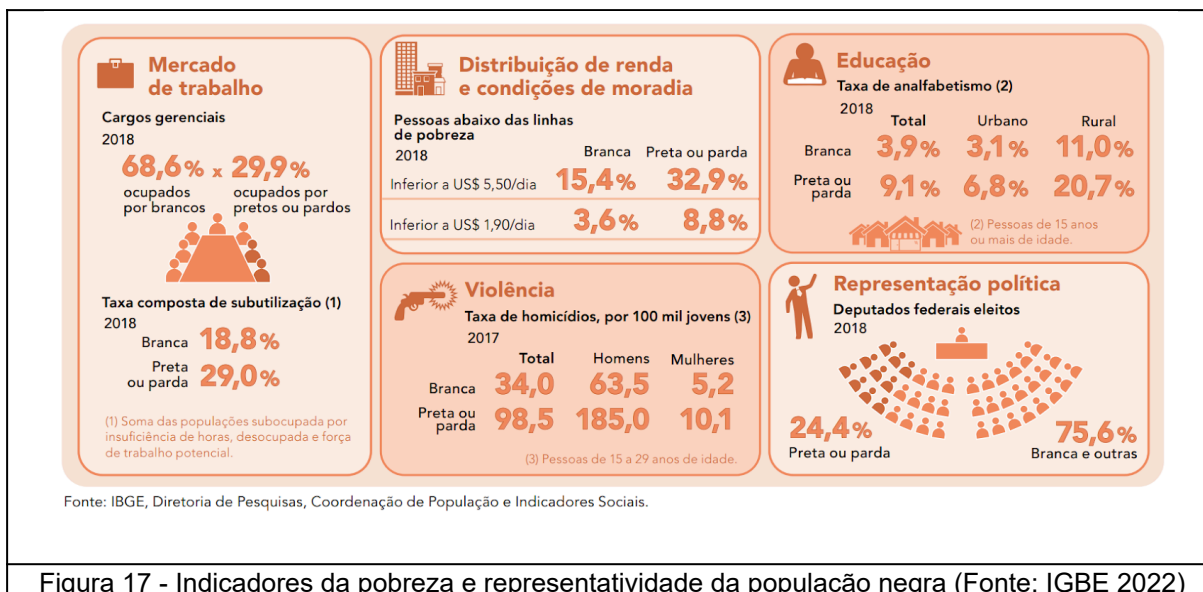


Figura 17 - Indicadores da pobreza e representatividade da população negra (Fonte: IGBE 2022)

O PL 21/2020 visa estabelecer um marco legal para a IA no Brasil, mas é criticado por ser superficial e uma comissão do Senado analisou o projeto e propôs um substitutivo que pode influenciar a lei final para abordar as complexas implicações da IA. O reconhecimento facial não é amplamente utilizado no Brasil, mas o governo recentemente coletou dados biométricos da população.

Os agentes públicos brasileiros precisam regulamentar o aprendizado de máquina e suas aplicações em sistemas de reconhecimento que envolvam pessoas, privacidade e informações sensíveis.

⁴⁵ <https://jornal.usp.br/ciencias/qual-o-impacto-da-tecnologia-de-reconhecimento-facial-na-populacao-negra/>

⁴⁶ <https://www1.folha.uol.com.br/mercado/2022/04/mulher-passa-50-anos-em-situacao-de-escravidao-em-bairro-nobre-de-santos.shtml>

Embora nem todos os algoritmos sejam utilizados em aplicações críticas, eles precisam de ser altamente precisos no nosso mundo diversificado, onde o reconhecimento facial já é utilizado em tempo real e pode reforçar preconceitos e exclusão.

3.9 - Tecno-solucionismo

As possibilidades subjetivas e empíricas de vieses étnicos e de gênero na aplicação da tecnologia de aprendizado de máquina em IA resultam de amostras que não representam a população alvo da aplicação. Também podem surgir em conjuntos de dados durante processos algorítmicos ou de treinamento que não extrapolam a heterogeneidade em suas representações simplificadas matematicamente.

As empresas tecnológicas, movidas pelo tecno-solucionismo, acreditam que a tecnologia pode resolver questões sociopolíticas e éticas complexas. Esta abordagem visa resolver preconceitos humanos e proteger os direitos digitais através do aperfeiçoamento da tecnologia, mas é criticada por ser simplista e autocentrada.

Os sistemas de reconhecimento facial de IA têm desempenho semelhante ou pior do que os especialistas humanos, sugerindo que a combinação das modalidades humana e de IA pode fornecer resultados confiáveis e atenuar o preconceito racial e de gênero.

O uso de IA em algoritmos que auxiliam na tomada de decisões judiciais e em processos seletivos de emprego podem perpetuar e amplificar injustiças sociais. No primeiro caso, o reconhecimento facial, sujeito a erros, pode levar à prisão injusta de suspeitos, com maior impacto em grupos minoritários. No segundo, critérios subjetivos de seleção podem perpetuar preconceitos e discriminação contra grupos já marginalizados.

A vida no mundo virtual reflete as mesmas distorções, racismos e preconceitos que a sociedade impõe aos grupos étnicos e ao gênero que, embora possam representar a maioria da população, como no caso do Brasil, não conseguem sequer propor uma mudança estrutural que humanize as relações e mitiguem a fome e a miséria profunda.

3.10 - Fundamentos da Ciência de Dados Sociais

A Ciência de Dados Sociais pode analisar variáveis e conteúdos sociais usando algoritmos e técnicas de aprendizado de máquina para processar grandes conjuntos de dados em tempo real e identificar tendências emergentes. Isso pode ser usado no desenvolvimento de políticas públicas, permitindo que governos e gestores tomem decisões orientadas por dados. Estes dados sociais podem auxiliar a identificar regiões que precisam de investimentos em infraestrutura ou prever surtos de doenças em uma comunidade.

As plataformas de mídia social também geram grandes quantidades de dados que podem ser analisados para identificar padrões e tendências no discurso do usuário, o que ajuda a entender as percepções da sociedade. A análise de sentimentos e opiniões também é um aspecto importante da ciência de dados sociais que permitem uma observação mais detalhada das dinâmicas conversacionais da população, cada vez mais atuante nos fóruns e grupos de discussão (SANTOS 2, 2023) .

A análise de dados pode ser uma ferramenta poderosa para o estado de bem estar social, ao interpretar os dados e traduzir as percepções em ações que promovam mudanças sociais positivas, com um objetivo claro e foco na geração das mesmas. Esta tese propõe novas formas de realizar pesquisa social digital ao expandir o repertório de métodos e ferramentas em pesquisa sociológica como resposta às novas dinâmicas da sociedade digital interconectada.

Para Deborah Lupton, uma das pesquisadoras mais citadas em artigos sobre sociologia digital, a coleta dos dados digitais é uma atividade natural diante da gigantesca massa de dados digitais agregados e quantificáveis da internet, chamados de transacionais (LUPTON, 2015). Ela é gerada de forma discreta como parte de outras atividades rotineiras em dispositivos diferentes de casa, do trabalho, da rua ou no lazer.

A pesquisadora também afirma que desde 2013 houve uma distinção entre objetos de dados digitalizados, quando um material anteriormente analógico é digitalizado armazenado para exibição na web e objetos de dados nativamente digitais, quando os mesmos são produzidos a partir das propriedades da internet e /ou formuladas especificamente como parte de sua operação, sendo mais atraentes

para os pesquisadores por oferecerem veracidade e validade que os dados digitalizados pelos pesquisadores (ROGERS, 2013 apud LUPTON, 2015).

Os cientistas de dados sociais enfrentam desafios metodológicos devido à escala sem precedentes e à produção contínua de dados digitais. Embora ofereça muitas oportunidades, o grande volume de dados também gera preocupações quanto à definição dos limites da pesquisa. Também começaram a lidar com o fato de outros atores e linhas de pesquisa disputarem posições de especialistas em pesquisa social com os mesmos, tanto pelas exigências de conhecimento computacional ao coletar e analisar dados quanto pela aplicação de técnicas estatísticas, com forte influência das ciências das redes e métodos quantitativos em estudos de mídia e comunicação (IBIDEM)

Outro autor (ACKLAND, 2013) detalhou minuciosamente os métodos de coleta e análise e também descreve como os dados da internet gerados pelas nossas pegadas digitais estão revolucionando as redes sociais e fornecendo novas questões de pesquisa, novos métodos e novos insights. Para ele, a revolução digital transformou a maneira como as pessoas buscam informações, trabalham, interagem e consomem bens e serviços, permitindo que os cientistas sociais revisitem questões tradicionais de pesquisa, como influência social, coletividade e engajamento político.

Em seu livro *“Web Social Science: Concepts, Data and Tools for Social Scientists in the Digital Age”*⁴⁷, Ackland explica que a correlação entre duas variáveis não implica em causalidade, e a causalidade pode ser direta ou indireta. Para o autor, embora existam métodos estatísticos de pesquisa que podem identificar e medir os efeitos causais, é importante considerar os efeitos da endogeneidade⁴⁸ e a seleção da amostra ou a direção dos efeitos causais ao estudar essas questões.

Seu livro traduz o conceito de Web Social Science como uma abordagem interdisciplinar e quantitativa que se concentra nos impactos sociais, econômicos e políticos da Internet e nas percepções fornecidas pelos dados. O autor enfatiza que o comportamento on-line e as redes sociais podem ter efeitos no mundo real e

⁴⁷ <https://sk.sagepub.com/book/mono/preview/web-social-science.pdf>

⁴⁸ Endogeneidade: decorre normalmente da existência de variáveis omitidas, erros de mensuração das variáveis incluídas no modelo e/ou simultaneidade entre variáveis dependentes e independentes. (fonte: <https://hedibert.org/wp-content/uploads/2016/05/endogeneidade3-testes.pdf>)

analisa as ferramentas e os métodos de pesquisa da ciência de dados sociais, com foco na análise de redes sociais e hiperlinks (IBIDEM).

A visão da sociologia digital nas palavras de Lupton, reafirma que prática sociológica precisa ser repensada se quisermos entender a vida social, econômica e política de hoje. Isso ocorre porque essas vidas são cada vez mais vividas em e por meio de tecnologias digitais (Ruppert et al. 2013 apud LUPTON).

Apesar de não exigirem técnicas avançadas de ciência de dados ou habilidades de codificação, essas novas abordagens sociológicas aproveitam várias tecnologias digitais para gerar, analisar e visualizar dados sociais. Os métodos tradicionais de pesquisa social precisam ser reavaliados à luz das possibilidades oferecidas pelos objetos de dados nativamente digitais e das ferramentas que podem ser usadas para configurá-los, analisá-los e visualizá-los (LUPTON, 2015).

Para Latour e os outros autores do artigo intitulado “O Todo é sempre menor que suas partes: Teste digital das mônadas de Gabriel Tarde”, que não usava conceitos como “indivíduo” ou “sociedade” em sua teoria, agora pode ser reexaminada devido a novos conjuntos de dados digitais que permitem uma navegação mais fácil das conexões individuais, tornando a ideia de “mônadas” mais plausível. Essa nova maneira de navegar pelos dados poderia mudar a teoria social (LATOUR et al. 2012).

O artigo afirma que os dados criados pelas interações dos usuários com as tecnologias digitais oferecem uma oportunidade para repensar a teoria social. Ele afirma que na era digital, em que as informações sobre as pessoas podem ser encontradas por meio de mecanismos de busca, os atores agora são definidos por suas redes digitais:

“Quanto mais você quiser identificar um ator, mais terá de mapear sua rede de atores” (IBIDEM).

Ainda segundo LUPTON, as tecnologias digitais estão profundamente arraigadas na vida contemporânea, moldando indivíduos, relacionamentos e a sociedade. Seu impacto é estruturado por categorias sociais e incorporado às instituições sociais (LUPTON, 2015).

As tecnologias digitais são de interesse crescente para os sociólogos devido ao seu impacto na vida cotidiana e nos conceitos de identidade, relações sociais e

interações. Essas tecnologias também deram origem a novas percepções sociológicas, pois são artefatos socioculturais que transformaram a economia do conhecimento e reconfiguraram as relações de poder (IBIDEM).

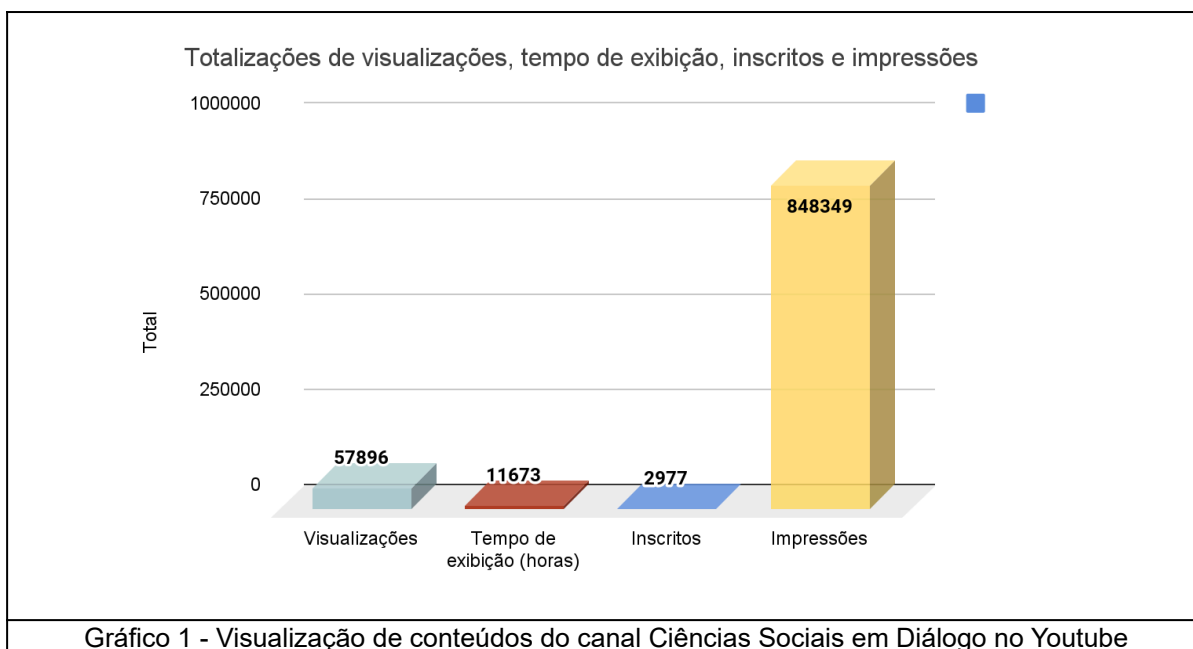
As tecnologias digitais aprimoraram a pesquisa sociológica ao fornecer novas ferramentas para análise e interpretação de dados, garantindo que as Ciências Sociais permaneçam relevantes num mundo em constante mudança.

3.11 - ANÁLISES DESCRITIVAS

Coletamos dados das planilhas dos canais de Youtube de ambos os projetos de extensão universitária e o site de apenas um deles, que foram analisados estatisticamente com técnicas lineares (BERTARELLI, 2016) e não lineares (HOFFMANN, 2016, ZEVIANI et al. 2013) nesta pesquisa .

3.11.1 - Projeto do canal de streaming Ciências Sociais em Diálogo

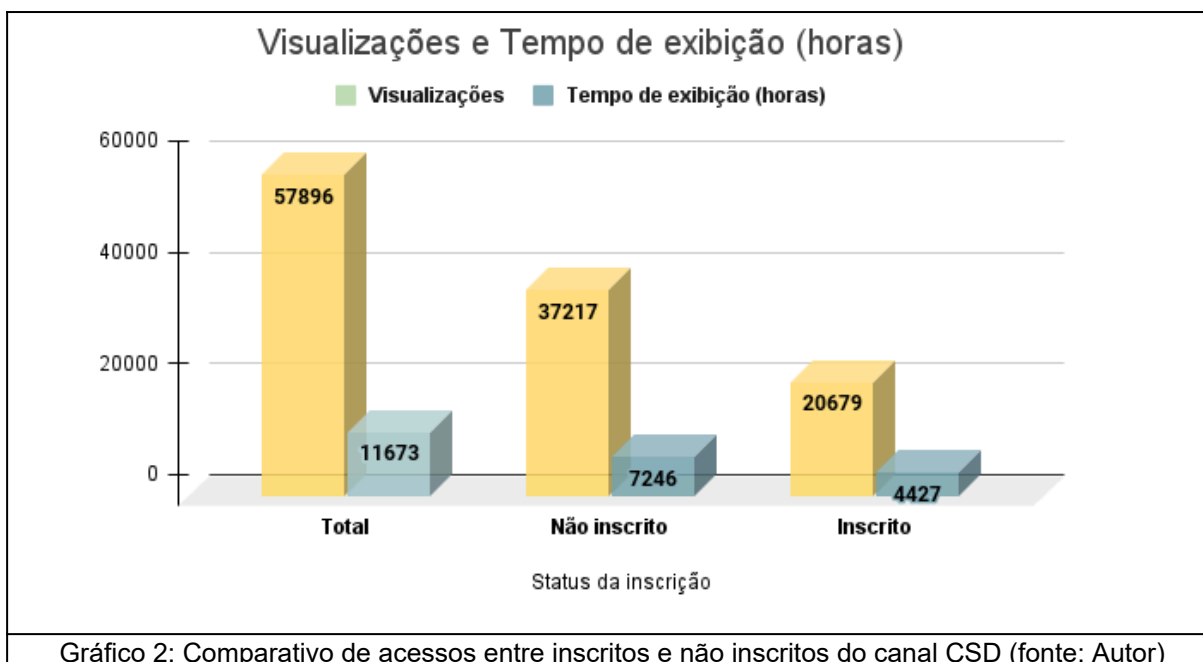
No início do isolamento imposto pela pandemia de Covid-19, o Departamento de Ciências Sociais da FCLAr Unesp de Araraquara, demandou a criação de um canal de streaming que iniciou suas atividades em abril de 2020 e se dedica, segundo dados do projeto, a compartilhar com a comunidade universitária e a sociedade em geral, debates semanais e abertos sobre temas atuais e que interessa a todos, permitindo a interação do público com os expositores. Foram produzidas vinhetas de abertura e encerramento, a programação visual e criação do canal, além de edições compactas, coleções e cortes de resumo para motivar a aderência ao canal, fidelizar a audiência e incentivar o engajamento dos inscritos.



Para demonstrar o modelo de análises descritivas, seguem gráficos e inferências sobre dados coletados desde a criação do canal em abril de 2020 até setembro de 2024. Para exemplificar e demonstrar a análise de conteúdos do canal e ciência de dados sociais, utilizamos uma amostra de 10 audiovisuais com o segundo maior número de acessos on-line e on demand do acervo. O site da tese terá o acesso liberado após a defesa, em dezembro de 2024.

O canal exibiu 270 transmissões online (streaming) que já ultrapassaram 57.890 visualizações em 11.674 horas de exibição no final de setembro de 2024, obtendo mais de 2970 inscritos e alcançando mais de 848.349 impressões. Cada vídeo foi exibido em média 215 vezes por streaming e/ou demanda, sendo que o tempo médio de exibição é de 43 horas por conteúdo e de 19 visualizações para cada usuário inscrito.

Cada usuário inscrito visualizou 4 horas em média e houve 3.142 impressões para cada um dos conteúdos, sendo que os usuários não inscritos superam os inscritos em 28,56% com 16.538 usuários a mais.



O engajamento é um conceito relacionado ao empenho dos inscritos em consumir conteúdos do canal e também uma estratégia de estimular o interesse deles em conhecer, comentar e compartilhar sobre o seu conteúdo. As pegadas digitais, comentários e seus metadados são recursos relevantes para os algoritmos dos buscadores e os novos recursos de IA generativa (Perplexity, GPTs, Genesis). O usuário inscrito pode influenciar positivamente outros usuários ao promover o conteúdo do canal e atrair novos inscritos, realimentando o engajamento.

O cálculo, geralmente, é realizado a partir da divisão entre o número de interações (curtidas, descurtidas e comentários) pelo número de seguidores, e o resultado multiplicado por 1.000, segundo métrica indicada pelo Buzzmonitor⁴⁹, utilizando dados fornecidos pelo Youtube.

Os não inscritos, em sua maioria, acessam por recomendações na barra lateral e sugestões ao final de outras reproduções, então é recomendável promover outros vídeos do canal na tela final, inserir marca d'água para fixar logotipo do canal, moderar e responder todos os comentários e fornecer uma boa descrição de cada conteúdo e do canal.

⁴⁹<https://buzzmonitor.com.br/blog/como-o-engajamento-e-calculado-para-cada-rede-social/#:~:text=M%C3%A1ximo%20de%20engajamento,-Representa%20o%20dia&text=o%20per%C3%ADodo%20selecionado%5D,-A%20conta%20%C3%A9%20feita%20a%20partir%20da%20divis%C3%A3o%20entre%20o,utilizando%20dados%20fornecidos%20pelo%20YouTub>

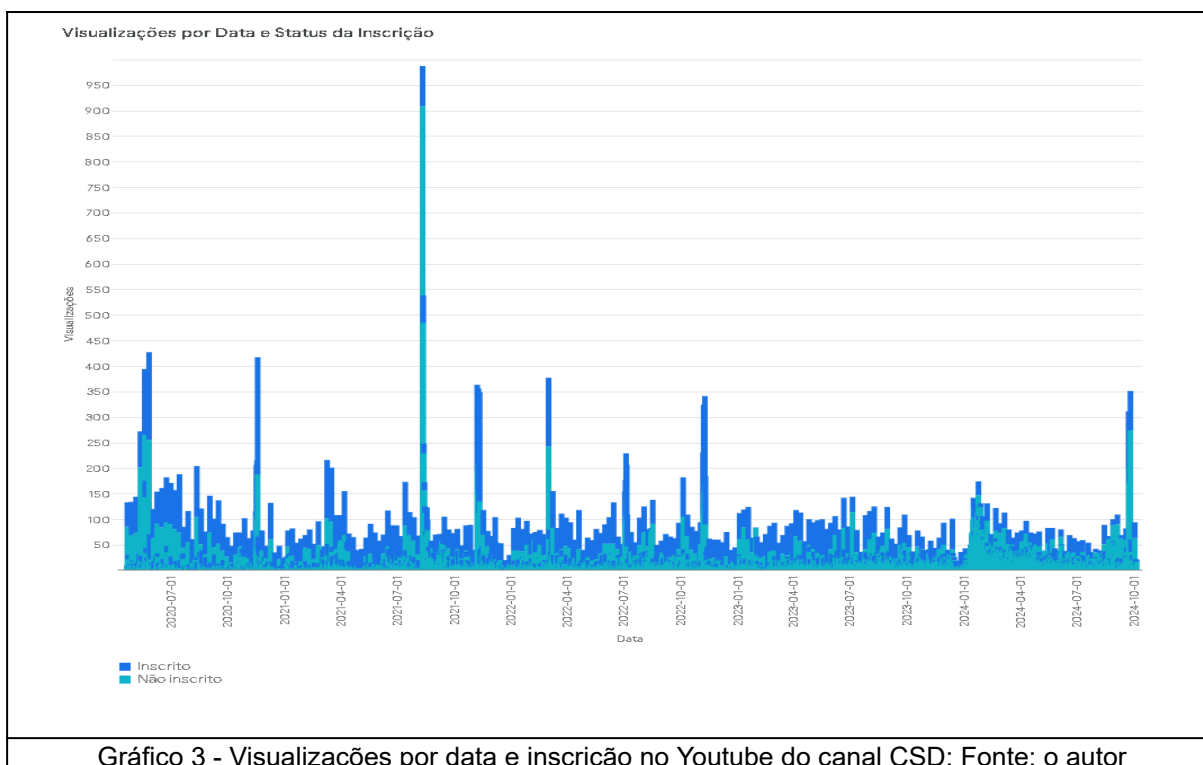


Gráfico 3 - Visualizações por data e inscrição no Youtube do canal CSD: Fonte: o autor

Percepções:

- O número de visualizações para usuários inscritos e não inscritos varia bastante ao longo do tempo.
- Usuários inscritos e não inscritos tiveram visualizações semelhantes ao longo do período, mas algumas datas apresentaram grande diferença.
- Em certos períodos, os usuários inscritos tiveram mais visualizações do que os não inscritos, como em agosto de 2020.
- Em outros, os usuários não inscritos tiveram mais visualizações do que os inscritos, como em janeiro de 2021.
- A variabilidade nas visualizações dos usuários parece diminuir com o tempo.

Observando um gráfico das visualizações de conteúdos por semana paralelamente ao gráfico de espectadores recorrentes por semana, pode-se inferir que haja uma correlação entre ambos, indicando que as lives mais assistidas agregam concentradamente, novos inscritos no canal.

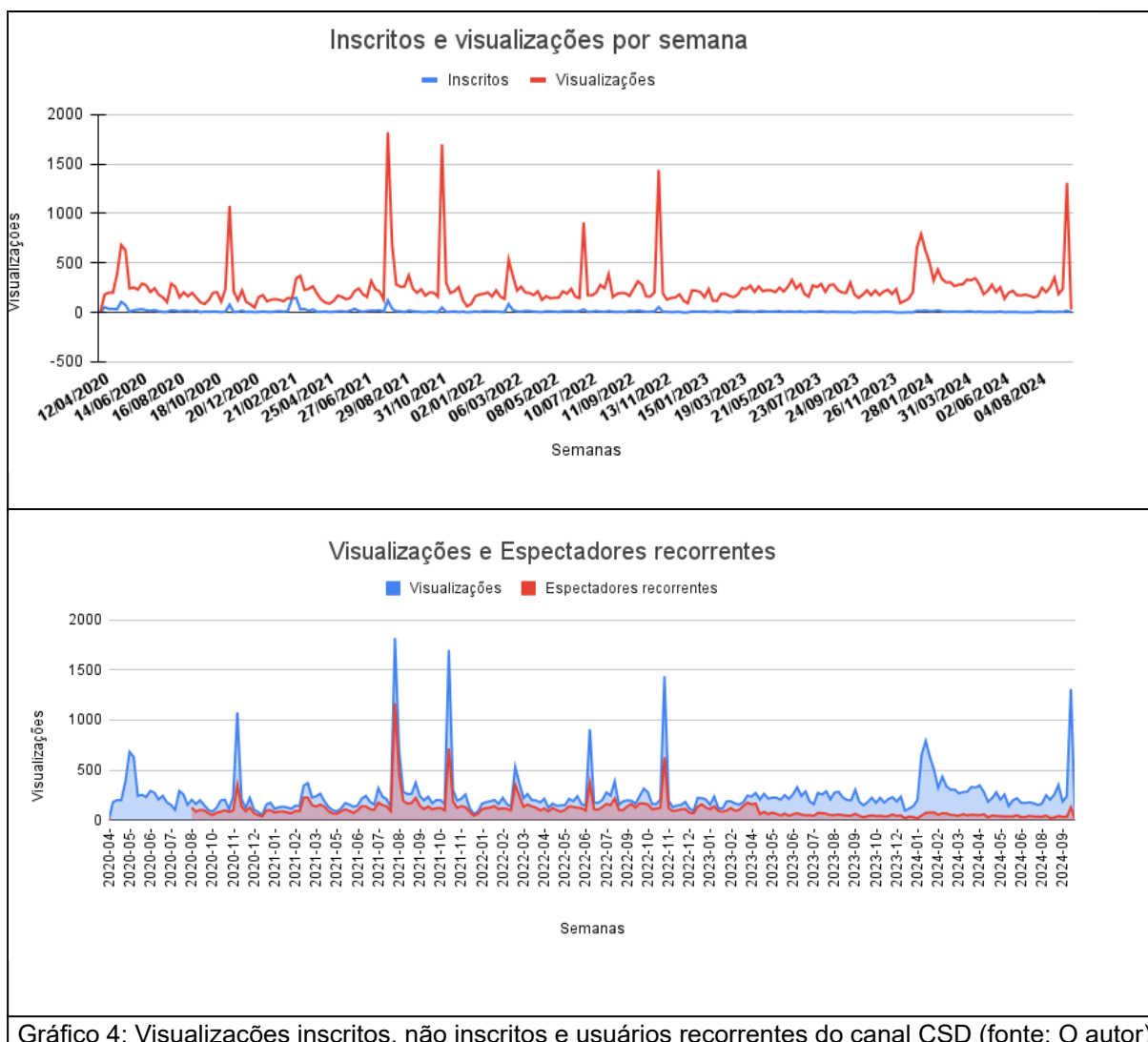
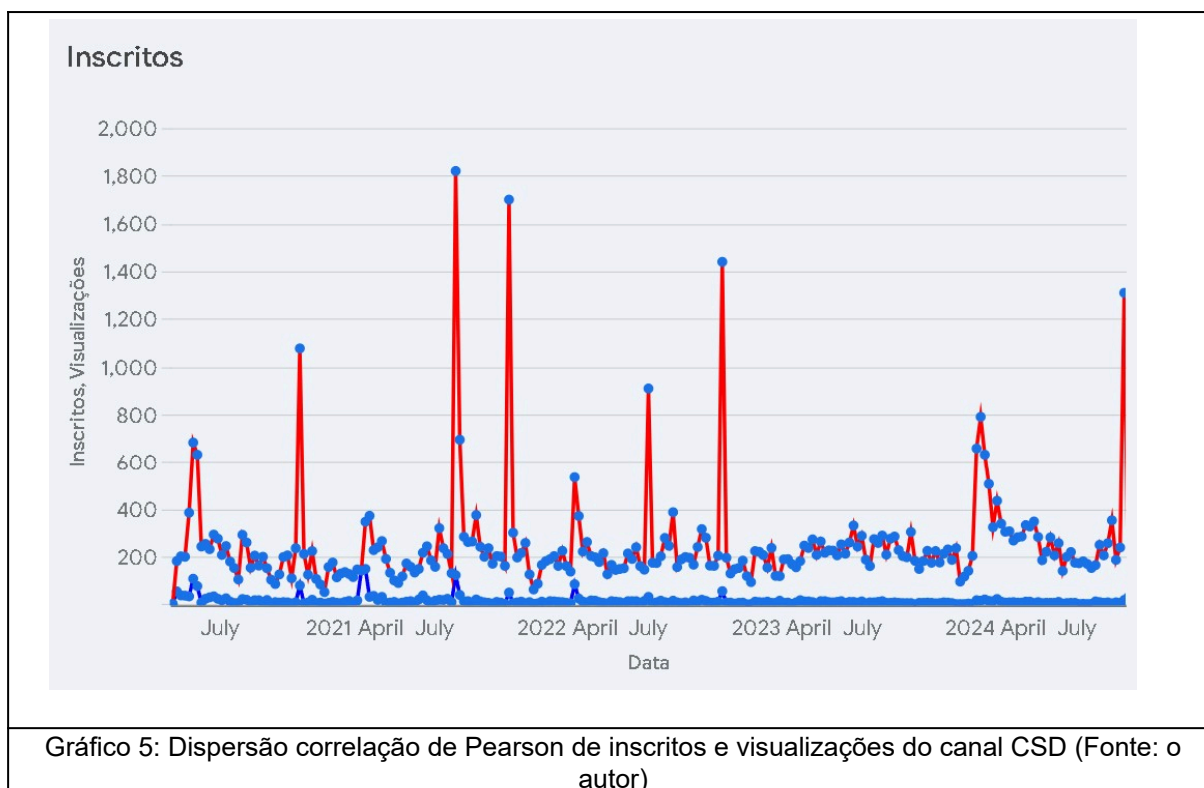


Gráfico 4: Visualizações inscritos, não inscritos e usuários recorrentes do canal CSD (fonte: O autor)

As duas tabelas possuem a mesma estrutura com as colunas Data, Status da inscrição e Visualizações, sendo esta última do tipo numérico (int64), portanto, nenhuma conversão adicional é necessária para realizar os cálculos. Como as tabelas possuem a coluna “Data” em comum e queremos analisar a correlação entre as visualizações dos inscritos e não inscritos ao longo do tempo, vamos combinar as duas tabelas usando a coluna Data como chave.

Ao realizar uma análise de correlação, construiremos um modelo de regressão linear para entender a relação entre as visualizações dos dois grupos e identificar visualmente em um gráfico com o espaço entre os eixos de visualizações dos dois grupos. Um coeficiente de Correlação de Pearson = 0.526 indica uma correlação moderada e positiva entre as duas variáveis. Isso sugere que, em geral, quando o número de inscritos aumenta, o número de visualizações também tende a

aumentar, e vice-versa, mas a relação não é muito forte e outros fatores podem estar influenciando o número de visualizações.

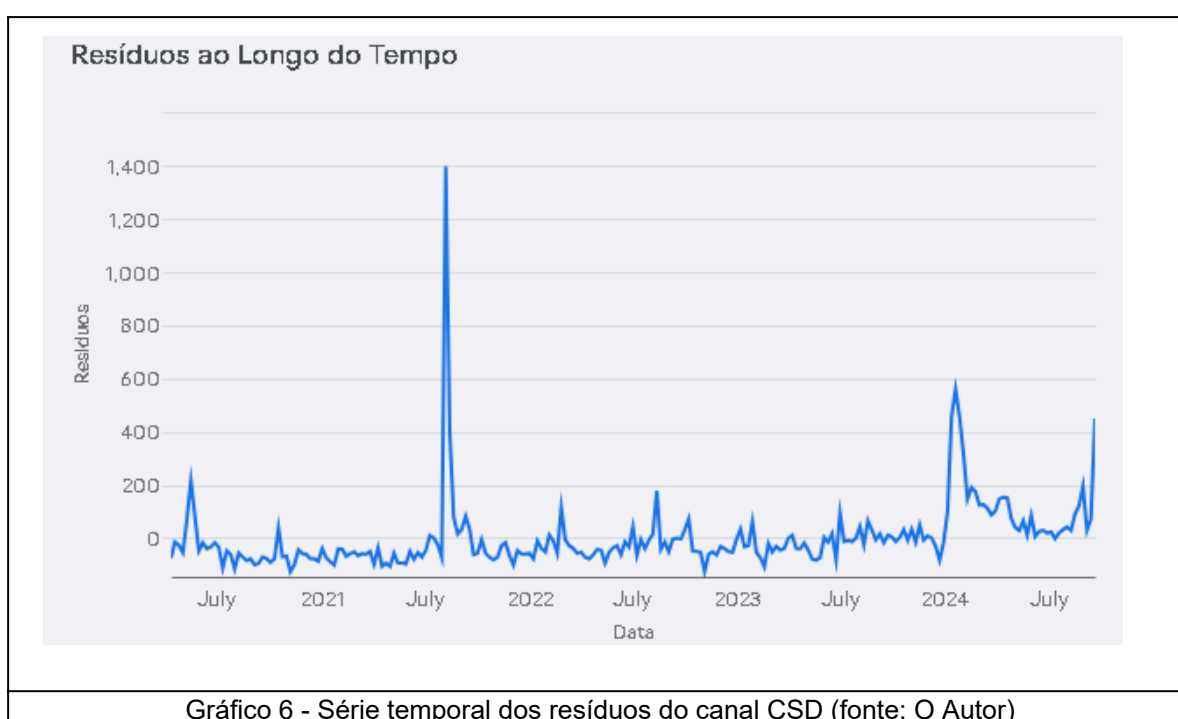


Dado que o coeficiente de correlação é maior que 0.5 (em termos absolutos), sugerimos uma regressão linear para aprofundar essa relação. O modelo resultante tem um R^2 de aproximadamente 0.27, o que indica que cerca de 27% da variação nas visualizações de não inscritos pode ser explicada pelas visualizações de inscritos. A inclinação da linha de regressão é de aproximadamente 0.93, sugerindo que, para cada aumento de uma visualização entre os inscritos, podemos esperar um aumento de 0.93 visualizações entre os não inscritos. O intercepto de 77.12 indica o número previsto de visualizações de não inscritos quando as visualizações de inscritos são zero.

O gráfico acima demonstra essa relação, com a linha de regressão ilustrando a tendência geral. É importante notar que, embora haja uma correlação e a regressão linear forneça algumas percepções, outros fatores além das visualizações de inscritos provavelmente influenciam as visualizações de não inscritos, dados o coeficiente de correlação moderado e o R^2 relativamente baixo. Para aprofundar a análise e identificar outros fatores que possam influenciar as visualizações de não

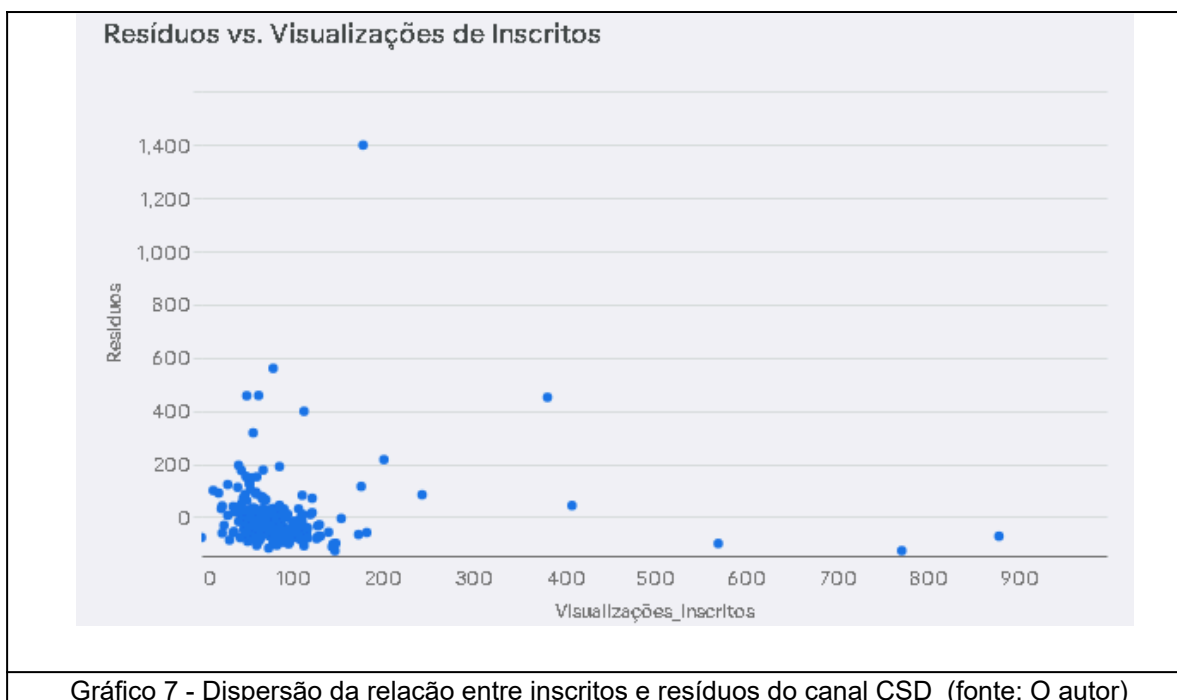
inscritos, além das visualizações de inscritos, podemos realizar uma análise de resíduos do modelo de regressão linear.

Os resíduos representam a diferença entre os valores observados e os valores previstos pelo modelo, e podem nos ajudar a identificar padrões ou tendências não capturadas pela regressão linear simples. Calculando os resíduos, podemos plotá-los em relação às visualizações de inscritos e ao tempo (data), buscando por quaisquer padrões ou relações que possam indicar a presença de outros fatores influentes.



Resíduos ao Longo do Tempo: Este gráfico de linha ilustra como os resíduos evoluem ao longo do tempo. Se houver algum padrão temporal dos resíduos, como uma tendência crescente ou decrescente, ou sazonalidade, isso pode indicar a presença de fatores externos que variam com o tempo e afetam as visualizações de não inscritos, mas não foram incluídos no modelo.

Havendo uma tendência clara nos resíduos ao longo do tempo, isso pode indicar a existência de fatores sazonais ou outros eventos externos que influenciam as visualizações.



Resíduos versus Visualizações de Inscritos: O gráfico de dispersão acima, mostra como os resíduos (a diferença entre as visualizações reais de não inscritos e as previstas pelo modelo) se comportam em relação às visualizações de inscritos. Idealmente, os resíduos devem estar distribuídos aleatoriamente em torno de zero, sem nenhum padrão claro. Qualquer padrão visível pode indicar que o modelo de regressão linear não está capturando completamente a relação entre as variáveis, sugerindo a presença de outros fatores influentes.

No caso de os resíduos apresentarem um padrão não linear em relação às visualizações de inscritos, isso pode sugerir que a relação entre as variáveis não é adequadamente descrita por um modelo linear simples. Se o gráfico de dispersão exibir um padrão cônico, com os resíduos aumentando ou diminuindo em magnitude à medida que as visualizações de inscritos aumentam, pode ser um sinal de heterocedasticidade⁵⁰. Em outras palavras, a variância dos erros não é constante.

Ao reconhecer essas tendências, podemos propor teorias sobre quais outros elementos podem afetar as visualizações de não inscritos e, posteriormente, reunir dados adicionais ou aprimorar o modelo para incluir esses elementos, aprimorando,

⁵⁰ A heterocedasticidade pode significar que o modelo de regressão não é confiável. Isso porque, quando as amostras não respeitam a homocedasticidade, os resíduos não são minimizados, e a reta não é um bom estimador para as observações

dessa forma, a capacidade de explicar e antecipar as visualizações. Convém esclarecer que a análise de resíduos é um processo iterativo e exploratório que possibilita experimentar diferentes visualizações e técnicas estatísticas para identificar os padrões relevantes e formular hipóteses sobre os fatores influentes.

3.11.2 - Correlações entre engajamento e origem do tráfego

Ao analisarmos os dados correlacionando as variáveis de visualizações às diferentes origens do tráfego e se o usuário é inscrito ou não, podemos adequar um modelo de regressão pelo fato da variável “Visualizações” não ser categórica e inconstante ao longo do tempo. Nesse caso é viável utilizar um modelo de regressão linear para correlacionar as variáveis “Origem do tráfego” e “Status da inscrição” à variável “Visualizações”.

Inicialmente podemos aplicar o método ANOVA bidirecional (LARS & WOLD, 1989), usado quando há uma variável quantitativa e duas variáveis nominais. Ele permite investigar o efeito simultâneo de duas variáveis independentes na variável dependente, sendo que o teste F para a não normalidade (análise de variações) foi estudada desde a década de 1930, produzindo resultados polêmicos com evidências a favor e contra sua robustez (BLANCA MENA, 2017).

O teste consiste em analisar as diferenças entre as médias das “Visualizações” para cada categoria de “Origem de Tráfego” e “Tempo de Exibição”. As análises ANOVA indicaram diferenças significativas nas médias de Visualizações entre as diferentes categorias:

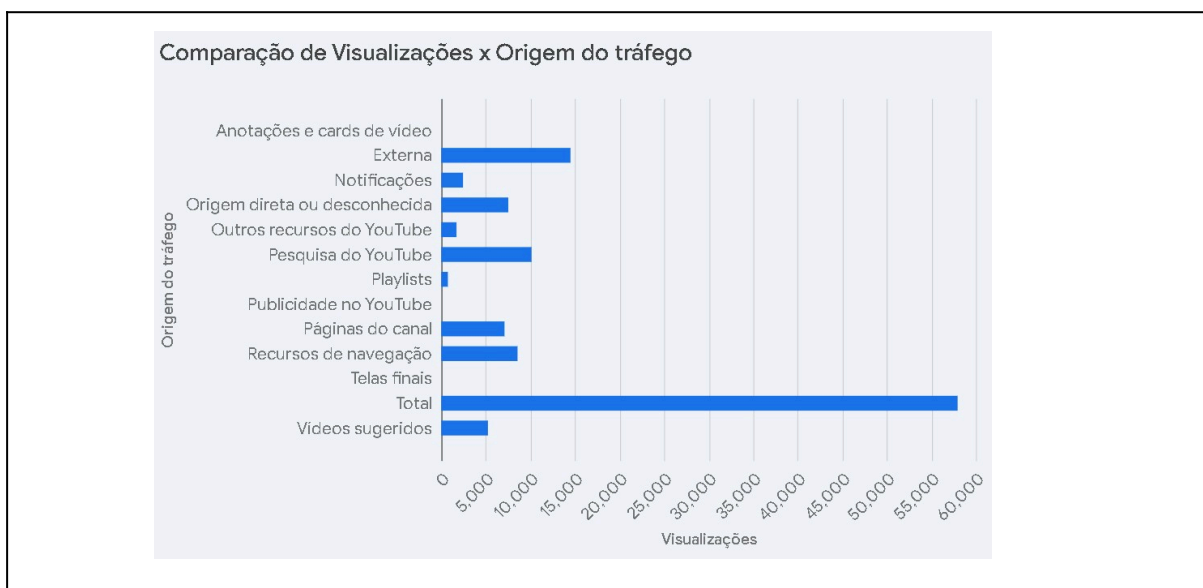


Gráfico 8: Comparativo de visualizações X origem do tráfego do canal CSD (fonte: O autor)

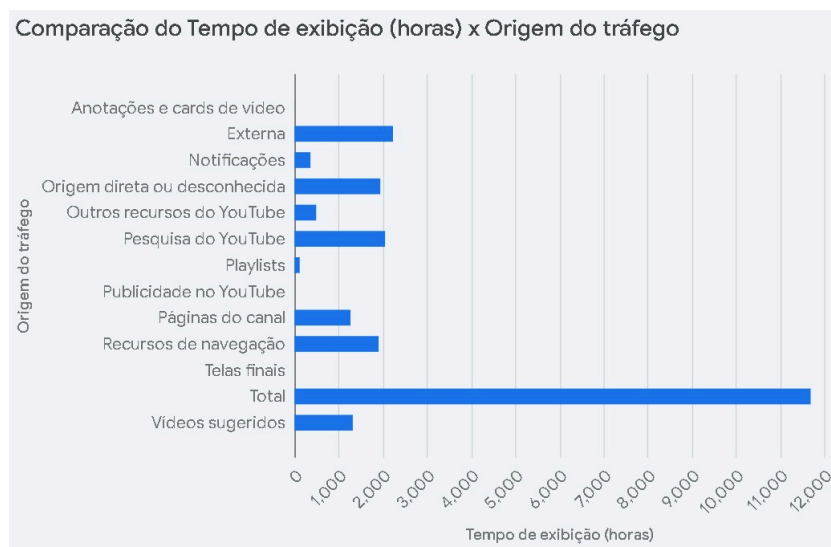


Gráfico 9: Comparativo de tempo de exibição X origem do tráfego do canal CSD (fonte: O autor)

Os resultados da análise descritiva, da regressão linear e dos dados articulados sugerem que as variáveis “Origem do tráfego” e “Status da Inscrição” exercem uma influência significativa em “Visualizações”. Entretanto, é essencial notar que a distribuição de “Visualizações” não é normal, o que pode limitar a adequação de um modelo de regressão linear.

Origem Tráfego	Status da Inscrição
F-statistic = 79,25	F-statistic = 90,76
p-value = 0,000	p-value = 0,000

Tabela 2 - Distribuição de visualização das variáveis do canal CSD

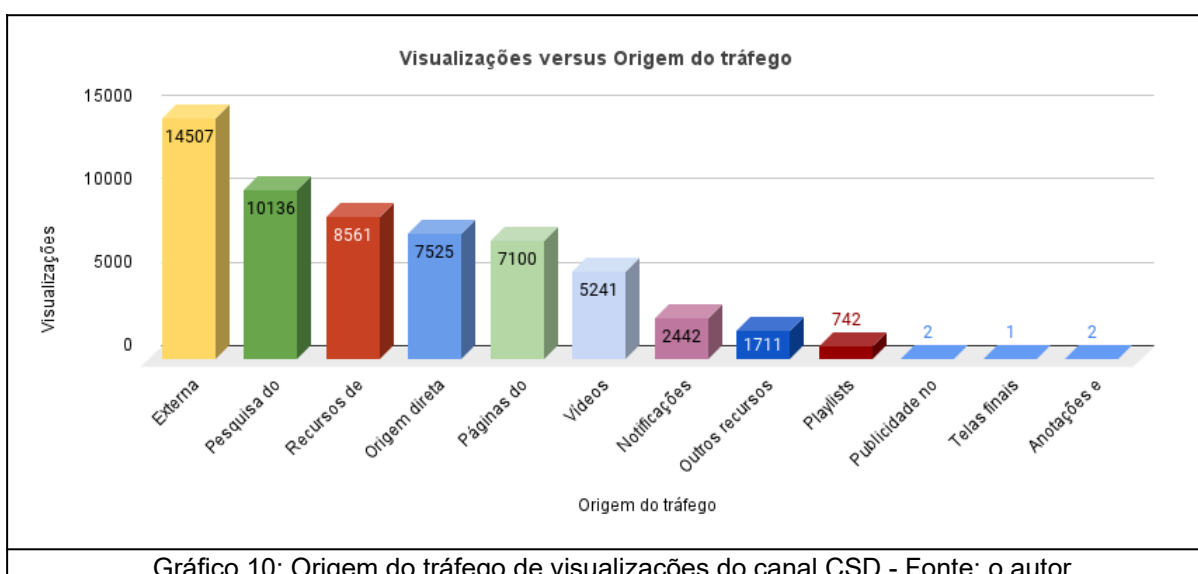


Gráfico 10: Origem do tráfego de visualizações do canal CSD - Fonte: o autor

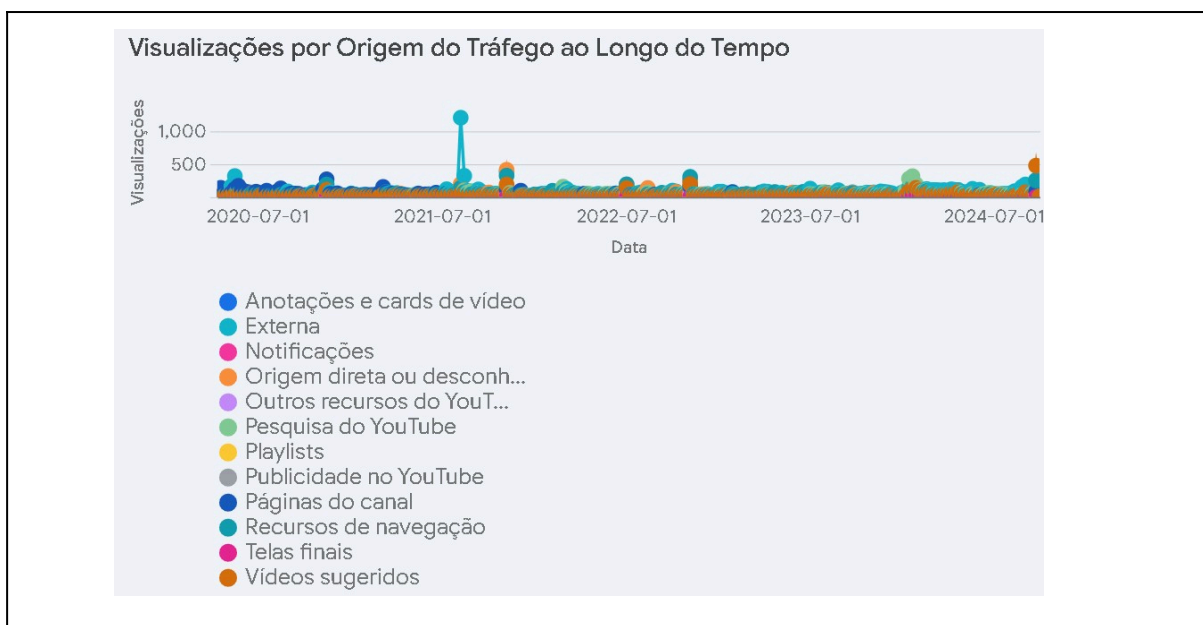


Gráfico 11: Visualizações por origem de tráfego em série temporal do canal CSD (fonte: o autor)

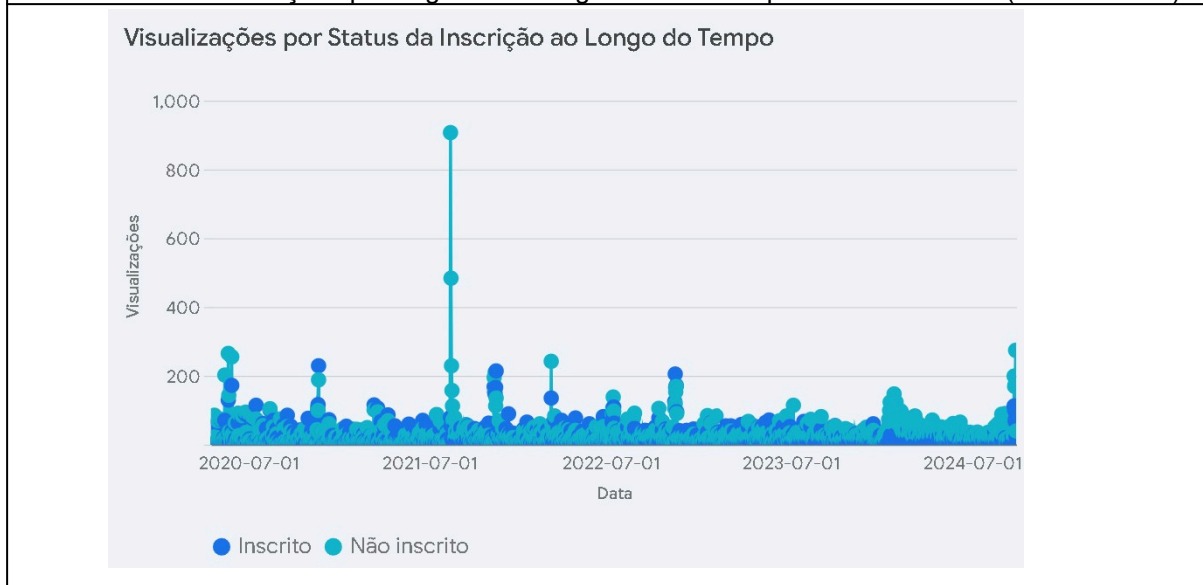


Gráfico 12: Visualizações por status de inscrição em série temporal do canal CSD(Fonte: o autor)

No gráfico por origem de tráfego, o valor de R ao quadrado de 0,495 indica que o modelo explica 49,5% da variância nas visualizações. Todos os níveis de origem do tráfego são preditores significativos de visualizações, exceto 'Anotações e cards de vídeo', 'Playlists', 'Publicidade no YouTube' e 'Telas finais', pois possuem valores de p superiores a 0,05.

No gráfico Visualizações por status de inscrição, o valor de R ao quadrado de 0,311 indica que o modelo explica 31,1% da variância nas visualizações. Ambos os níveis de status de inscrição são preditores significativos de visualizações, com valores de p inferiores a 0,05. Analisando os mesmos dados por histograma:

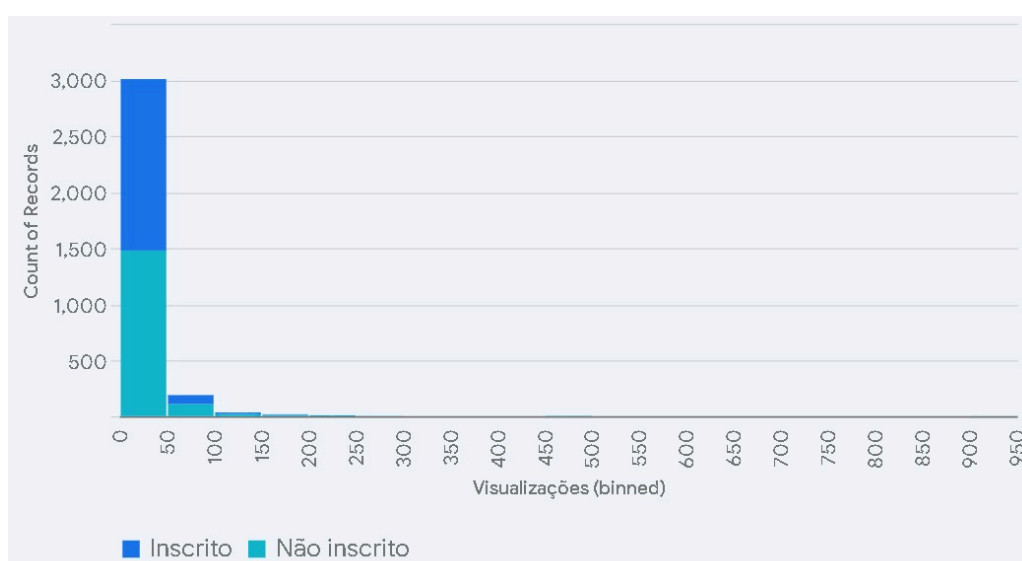


Gráfico 13: Histograma de Visualizações por status de inscrição do canal CSD (Fonte: o autor)

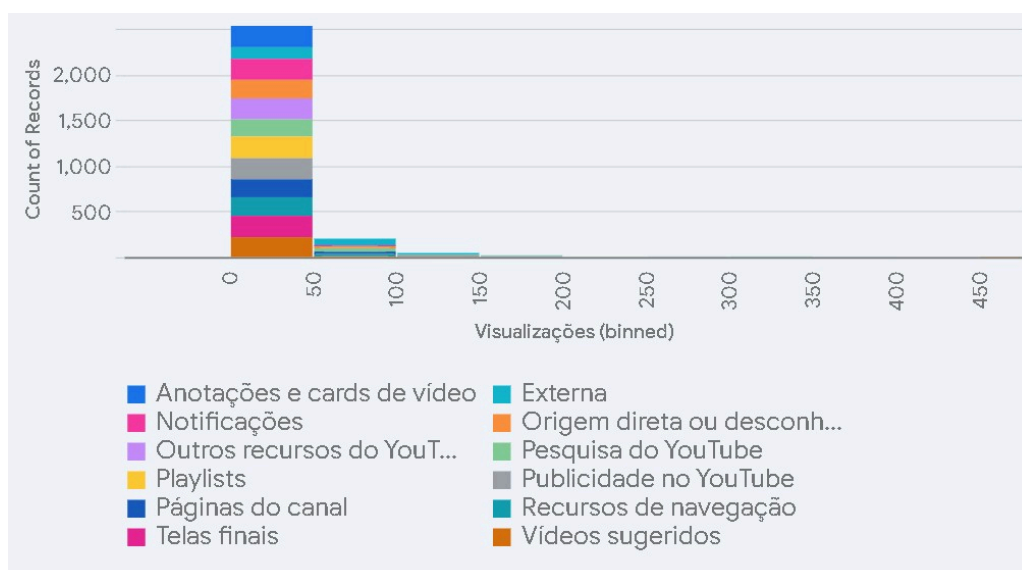


Gráfico 14: Histograma de Visualizações por origem de tráfego do canal CSD (Fonte: o autor)

Os campos “Origem do Tráfego” e “Status da Inscrição” influenciam a variável “Visualizações”, que na média, varia de 0,00854701 para 'Anotações e cards de vídeo' a 61,9957 para 'Origem Externa', sendo um desvio padrão elevado para a maioria dos níveis de “Origem do Tráfego”, indicando extensa variação nos dados.

Variância média mínima e máxima de “Visualizações”	
Anotações e cards de vídeo	Origem Externa
0,00854701	61,995

Tabela 3: Variância entre médias de Visualizações

Em relação à média de “Visualizações”, os 'Inscritos' somam 12,7177 e de 22,8887 para 'Não inscritos', sendo um desvio padrão também elevado, sugerindo alta variabilidade para o “Status da Inscrição”:

- O modelo para “Origem do Tráfego” explica 49,5% da variância em “Visualizações”.
- O modelo para “Status da Inscrição” explica 31,1% da variância em Visualizações.
- “Anotações e cards de vídeo”, “Playlists”, “Publicidade no YouTube” e “Telas finais” têm valores-p maiores que 0,05, indicando que não são estatisticamente significativos.
- Pode-se inferir que nem todos os níveis de “Origem de Tráfego” são preditores significativos de “Visualizações”.

Desta vez os histogramas também demonstram que a variável “Visualizações” não segue uma distribuição normal e uma regressão linear pode não ser a abordagem mais adequada. Não é possível aplicar uma regressão logística para analisar as correlações entre as colunas das tabelas fornecidas, pois a regressão logística é usada para prever uma variável de resultado categórica e a variável “Visualizações” é uma variável contínua.

3.11.3 - Perda e obtenção de inscritos

O Youtube disponibiliza outra tabela de dados quantitativos sobre inscrições obtidas e perdidas em série temporal de todo o período do canal, desde abril de 2020 a setembro de 2024. Os três primeiros indicadores são todos complementos da plataforma que podem ser melhor aproveitados para promover o canal aos não inscritos. Convém destacar que a “Pesquisa do Youtube” de busca na plataforma indica uma oportunidade de otimizar os rótulos e metadados dos conteúdos para melhorar vetores de busca semântica, pois se tornam ainda mais relevantes com o advento das IAs generativas⁵¹.

Em todo o período de 4 anos e 5 meses desde sua criação, o canal possui resiliência em manter o número de inscritos em crescimento linear contínuo. Devido à especificidade de ser um canal de ciências sociais e políticas, com público

⁵¹ A IA generativa é um subcampo da inteligência artificial centrada no desenvolvimento de algoritmos que geram novos dados ou conteúdos a partir de uma análise consistente dos mesmos.

relativamente restrito aos meios acadêmicos e educacionais com um total de 2981 inscritos, o canal obteve e mantém uma taxa de retenção média de 89,51% em relação à todas inscrições obtidas, tendo perdido, em média 10,49% dos inscritos ao longo do período.

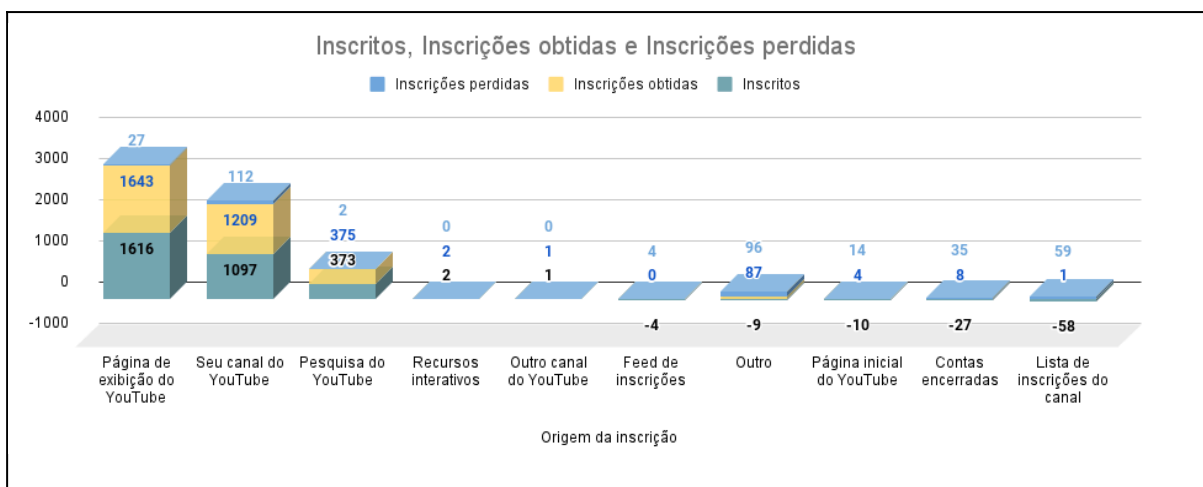


Gráfico 15: Retenção e perdas de inscritos por modalidade de origem do canal CSD (Fonte: o autor)

As perdas de inscrição relacionadas às “Pesquisas do Youtube” representam apenas 0,54% , indicando um nível de retenção de 99,40% das inscrições vindas de resultados de pesquisa com base nessas consultas. As “Inscrições” que são feitas por meio do botão de inscrição na página do seu canal, chamadas de “Seu canal de Youtube”, representam 32,95% do total delas, com perdas de 9,26% das obtidas. “Inscrições” feitas por meio do botão de inscrição nas páginas de exibição dos vídeos são chamados de “Página de exibição YouTube” e obteve a retenção de 98% das inscrições obtidas, considerando perdas irrisórias de 1,64% dos inscritos.

Origem da inscrição	Inscritos	Inscrições obtidas	Inscrições perdidas
Total	2981	3330	349
Página de exibição YouTube	1616	1643	27
Seu canal do YouTube	1097	1209	112
Pesquisa do YouTube	373	375	2
Recursos interativos	2	2	0
Outro canal do YouTube	1	1	0
Feed de inscrições	-4	0	4
Outro	-9	87	96
Página inicial do YouTube	-10	4	14
Contas encerradas	-27	8	35
Lista de inscrições do canal	-58	1	59

Tabela 4: Retenção de inscritos por origem da inscrição do canal CSD - Fonte: o autor

As Origens de inscrição “Recursos interativos” (0,1%), “Outro canal do YouTube” (0,0%), “Feed de inscrições” (-0,1%), “Outro” (-0,3%), “Página inicial do YouTube” (-0,3%), “Contas encerradas” (0,9%) e “Lista de inscrições do canal” (2,0%), por não obtiveram índices de inscrição maiores que 0,1%, não são relevantes para a análise e não afetará os resultados se desprezadas, pois somadas, continuariam negativas.

Origem da inscrição	Definições fornecidas pelo Youtube
Recursos interativos	Inscrições de recursos interativos em vídeos como telas finais, marca d'água de branding ou anotações.
Outro canal do YouTube	Inscrições de outros canais que recomendaram seu canal ou seus vídeos ou de outros canais relacionados ao seu canal.
Feed de inscrições	Inscrições do feed de inscrições.
Outro	Inscrições de outras fontes.
Página inicial do YouTube	Inscrições da página inicial do YouTube.
Contas encerradas	Inscrições perdidas por contas encerradas de usuários
Lista de inscrições do canal	Inscrições da lista de inscrição

Quadro 6: Definições de características das origens de inscrição do canal CSD - Fonte: o autor

Embora os valores dessas origens sejam numericamente insignificantes, é importante descrever cada uma delas, em definições fornecidas pelo próprio Youtube, conforme a tabela acima, pois podem se tornar relevantes ao se definirem estratégias de campanhas para engajamento e inscrições no canal.

3.11.4 - Relação entre Visualizações e Espectadores Recorrentes

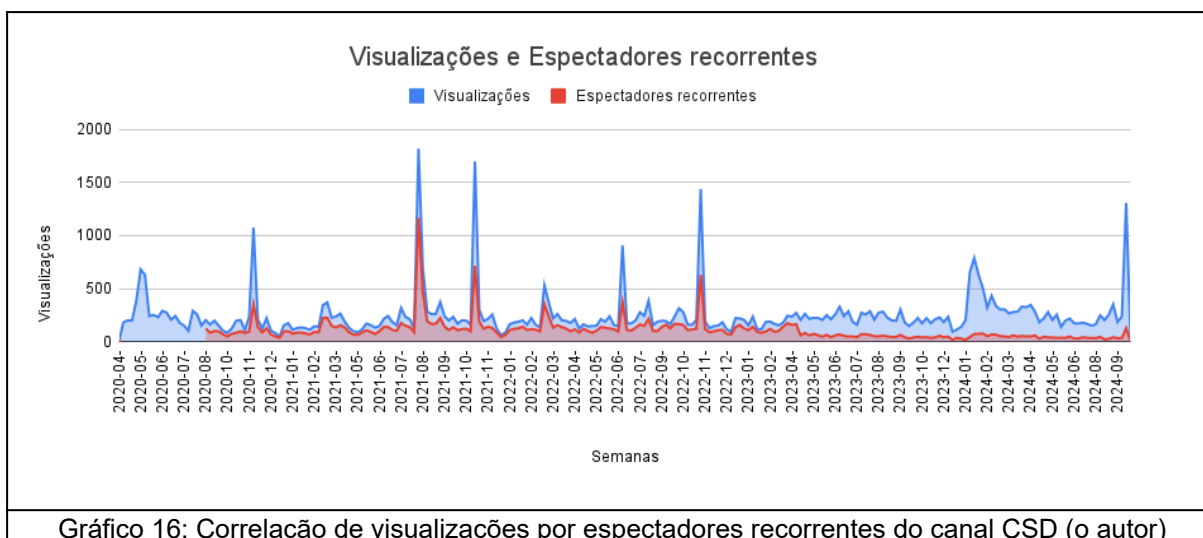
Outro conjunto de dados que pode conter informações sobre os espectadores e seu engajamento com o conteúdo do YouTube é o que relaciona as “Visualizações” e “Espectadores Recorrentes”.

A tabela de totalizações consiste em 3 campos:

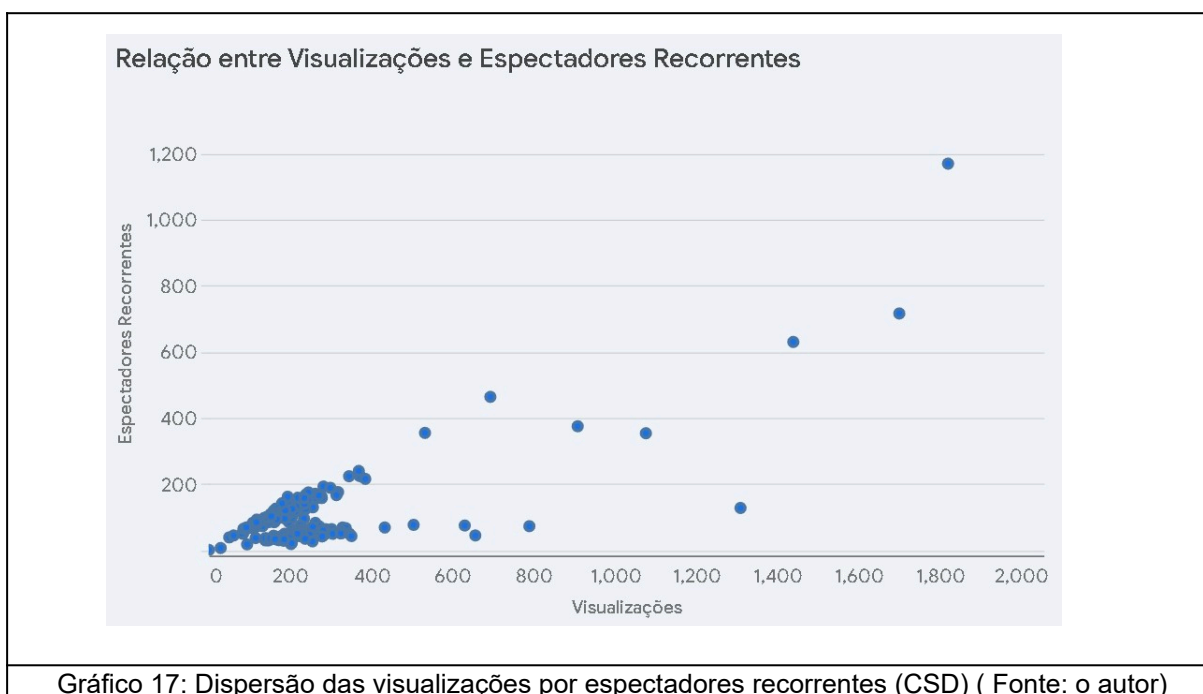
Campo do dataframe	Representação estatística
“Data”	Representa a data de coleta de dados ou observação.
“Visualizações”	Alcance de visualizações de vídeos na semana
“Espectadores Recorrentes”	Espectadores que assistem repetidamente ao conteúdo do canal.

Quadro 7: Representação estatística de campos dos dataframes do canal CSD - Fonte: o autor

A visualização sobreposta de dados de acesso de “Espectadores recorrentes” correlacionados com os totais desse alcance são indexados por uma série temporal para obter percepções sobre as fontes de assinatura e seu impacto no número de assinantes. Embora a correlação não pareça óbvia, o número de “Espectadores Recorrentes” é relativamente alto em relação aos totais, pois mesmo não sendo “Inscritos”, representam a maioria dos usuários, como pode ser observado no gráfico unificado de linhas comparativo entre essas variáveis:



Os registros de “Espectadores Recorrentes” se iniciam somente em agosto de 2020 e seguem com valores proporcionalmente equivalentes aos totais até abril de 2023, quando a correlação diminui consideravelmente até os dias atuais.



Essa visível alteração na variável pode ser estudada através da articulação entre as análises descritivas e as de conteúdo e serve para obter percepções sobre as fontes de recorrência e seu impacto no número de seguidores.

No gráfico de 'Visualizações ao longo do tempo' é possível observar um pico de visualizações próximo de 2000 no dia 28 de maio de 2023, com outro pico um pouco menor, com cerca de 1400 visualizações, em 17 de setembro de 2023. No geral, o gráfico mostra picos de visualizações pontuais, mas não constantes.

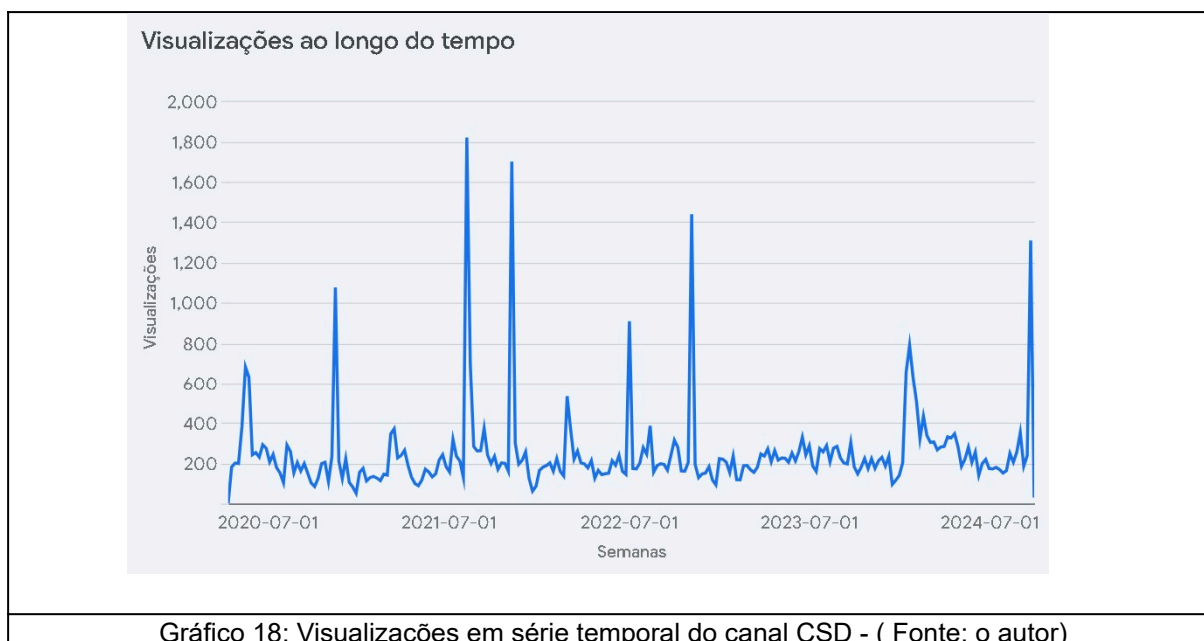


Gráfico 18: Visualizações em série temporal do canal CSD - (Fonte: o autor)

O gráfico de 'Espectadores recorrentes ao longo do tempo' também mostra picos pontuais, sendo o maior deles próximo de 200, em 28 de maio de 2023, coincidindo com o pico de visualizações, e outro um pouco menor, próximo de 150, em 17 de setembro de 2023, coincidindo também com o pico de visualizações nesta data. No geral, o gráfico mostra que geralmente o número de espectadores recorrentes fica abaixo de 50.



O número de recorrentes pode ser um indicativo de um amplo público que assiste conteúdos do canal de forma contínua e usual, mas há picos de acesso.

Percepções

- No gráfico de 'Relação entre Visualizações e Espectadores Recorrentes' é possível observar uma correlação positiva entre as duas variáveis, ou seja, quanto maior o número de visualizações, maior o número de espectadores recorrentes.
- O gráfico de 'Distribuição de Inscritos, Inscrições Obtidas e Inscrições Perdidas por Origem da Inscrição' mostra que a principal origem de inscrições é a 'Página de exibição do YouTube', seguida por 'Seu canal do YouTube' e 'Pesquisa do YouTube'.
- As inscrições obtidas e perdidas seguem a mesma proporção, sendo que a principal origem de inscrições, 'Página de exibição do YouTube', é também a que mais perde e que mais ganha inscritos.
- Juntos, os gráficos permitem observar que as visualizações e os espectadores recorrentes têm picos pontuais e não constantes, sendo que geralmente o número de visualizações fica abaixo de 200 e o número de espectadores recorrentes abaixo de 50.

3.11.5 - Inscrições e visualizações obtidas por modalidade de acesso

Ao analisarmos os dados de inscrições obtidas em relação ao número de visualizações, fica evidente uma discrepância dos dados de conversão de obtenção por modalidade de exibição, conforme o gráfico de barras que segue:

Nota-se claramente que as inscrições obtidas pela modalidade “Outros” obtêm 4,92 inscrições para cada visualização, enquanto as relacionadas às “Transmissão ao vivo” obtêm apenas 0,03 inscrições para cada visualização.

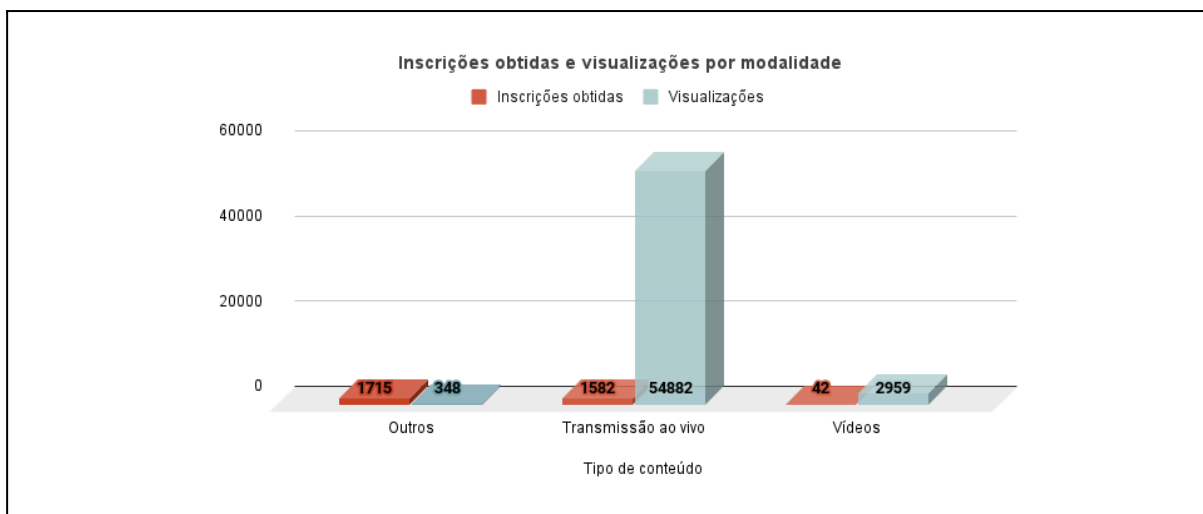


Gráfico 20: Inscrições obtidas e visualizações por modalidade exibição do canal (CSD) - (O autor)

Calculando o número de inscritos, visualizações e tempo total de exibição para cada tipo de conteúdo, exibimos os resultados em um gráfico de barras. Ele permite visualizar o tamanho relativo de cada tipo de exibição em termos de número de inscritos, visualizações e tempo total de exibição.

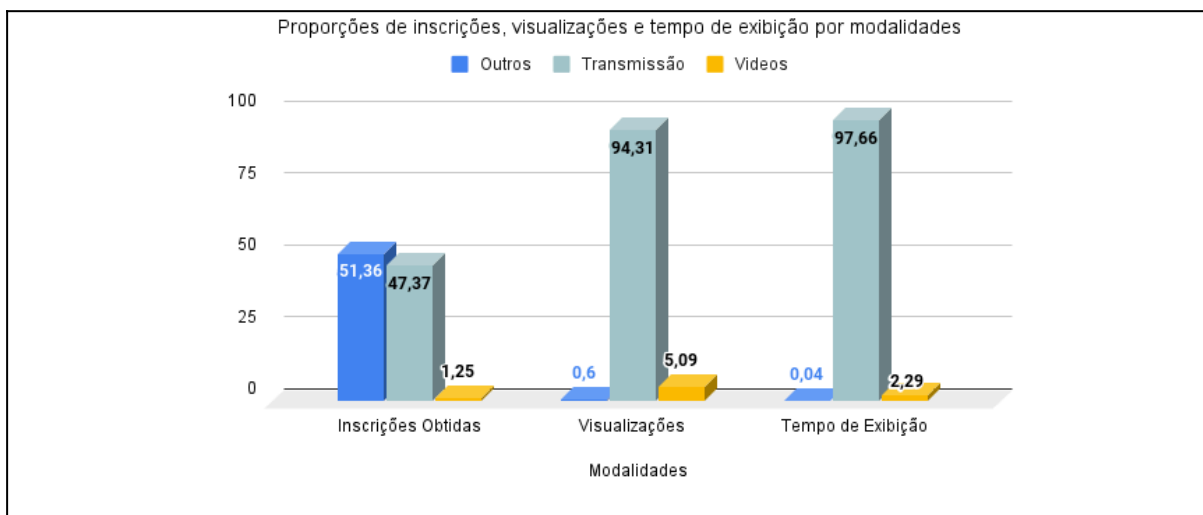


Gráfico 21: Percentual por modalidade de exibição (CSD) - (Fonte: O autor)

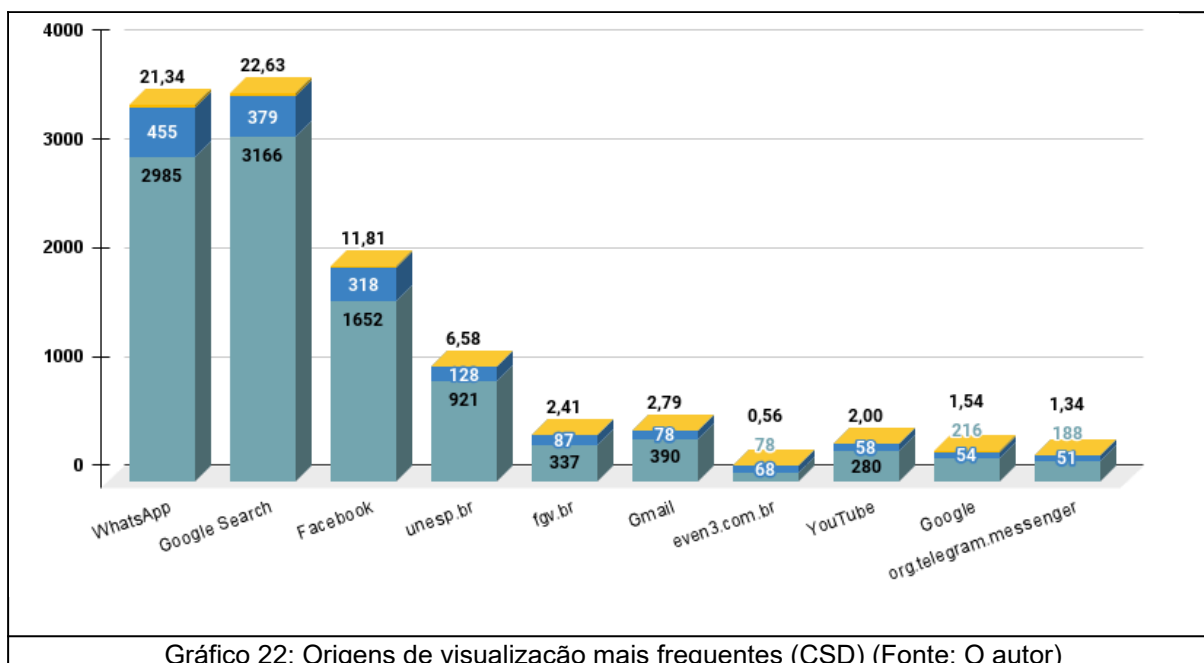
As transmissões ao vivo têm o maior número de inscritos, visualizações e tempo total de exibição, representando 94% do total de visualizações e 97% do tempo total de exibição. Por outro lado, o conteúdo 'Outros', que tem o segundo

maior número de inscritos, representa apenas 0,6% do total de visualizações e 0,04% do tempo total de exibição.

Considerando os percentuais das visualizações em relação às inscrições obtidas, observa-se um valor muito baixo, referenciando um percentual de mais da metade de inscrições obtidas. O que ocorre é que o Youtube não totaliza o que foi originado de Playlist separado de outros acessos, fazendo com essa contagem não represente adequadamente as inscrições que estariam relacionadas a mais de uma origem, não somente de Playlists.

Como as playlists são apenas listas com diferentes classificações dos mais de 240 conteúdos transmitidos e já estão contabilizadas no total de "Transmissão", consequentemente, as inscrições obtidas também não corresponderiam a valores tão ínfimos de "Visualizações" e "Tempo de Exibição".

Para melhorar a análise, obtivemos dados sobre quais as plataformas e sites que originam as visualizações, duração média da visualização, percentual médio e tempo de exibição em horas. Os 10 principais sites e/ou plataformas agregam 75% da origem de tráfego, totalizando 10.375 visualizações, também correspondem às maiores médias de Tempo de Exibição.



Podemos aplicar uma análise de séries temporais de ambos os data frames para identificar tendências, sazonalidade e padrões ao longo do tempo.

Decompomos as séries temporais, separamos os componentes de tendência, sazonalidade e residual. Testamos as hipóteses para comparar as médias de diferentes grupos nos dados.

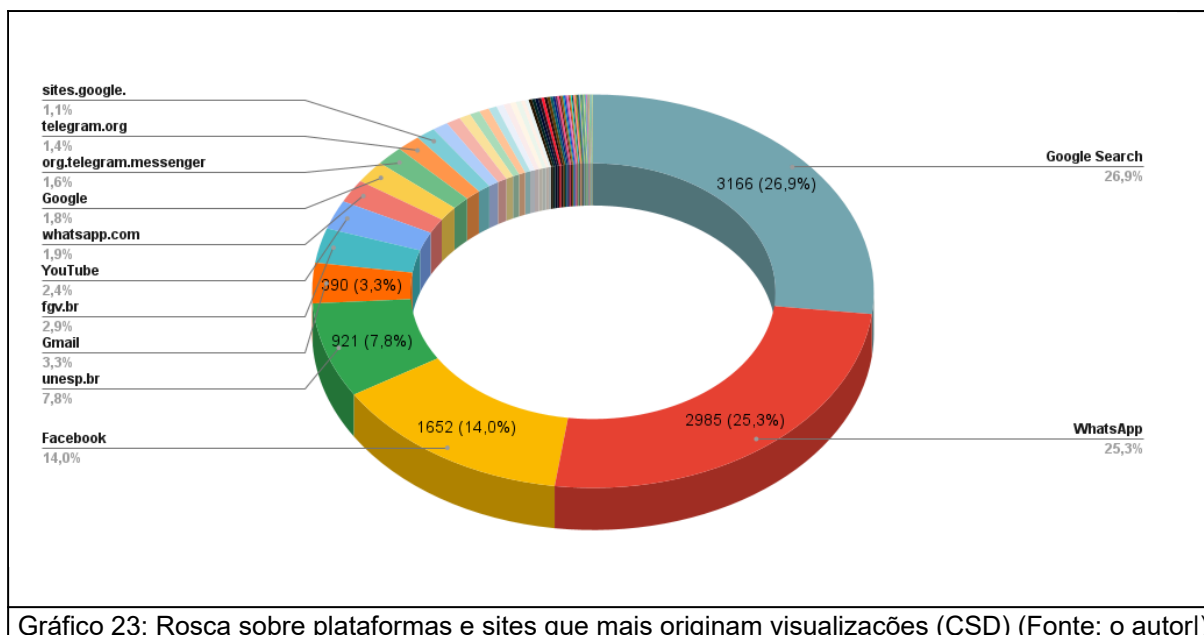


Gráfico 23: Rosca sobre plataformas e sites que mais originam visualizações (CSD) (Fonte: o autor)

Para uma análise preditiva simples usaremos modelos ARIMA, buscando antecipar valores futuros com base em dados anteriores. Faremos a decomposição dos dados definindo a coluna Data como o índice do Data Frame gerado a partir das duas tabelas de dados do Youtube. O Modelo Autorregressivo Integrado de Média Móvel (ARIMA) é uma técnica estatística amplamente utilizada para previsão de séries temporais. Seu nome decorre da combinação de três componentes:

1. **Autorregressivo (AR):** Modela a relação entre uma observação atual e um número pré-determinado de observações anteriores (atrasos).
2. **Integrado (I):** Diferencia os dados da série temporal para torná-los estacionários. Estacionariedade implica que as propriedades estatísticas da série temporal (média, variância, autocorrelação) permaneçam constantes ao longo do tempo, o que é um pré-requisito para a modelagem ARIMA.
3. **Média Móvel (MA):** Modela a relação entre uma observação atual e um número especificado de erros de previsão anteriores.

Esses três componentes em conjunto capturam os padrões subjacentes nos dados da série temporal e produzem previsões.

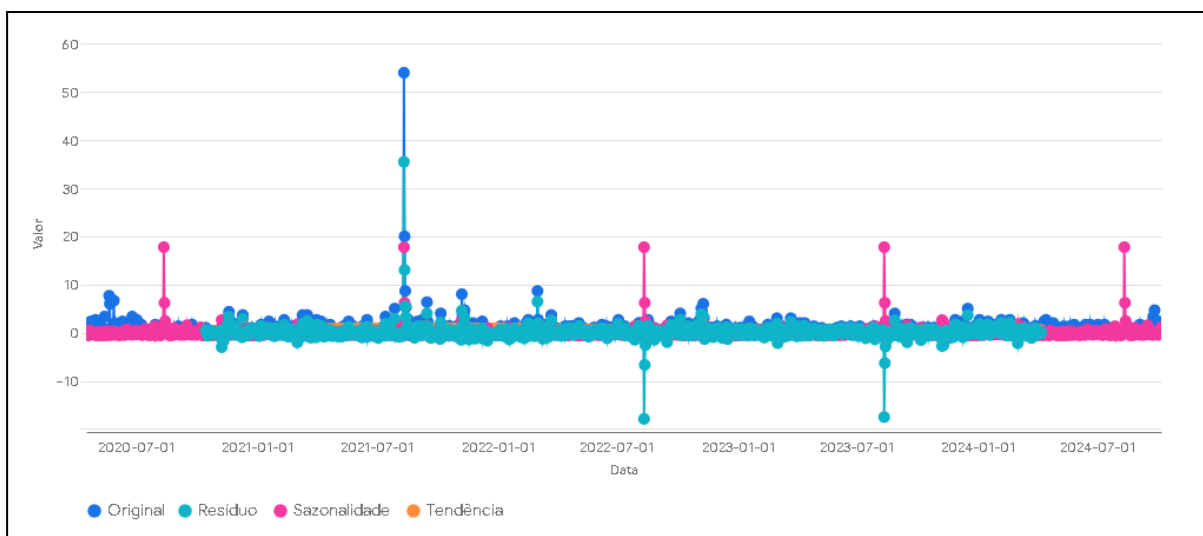


Gráfico 24: Decomposição da série Temporal de Compartilhamentos (CSD) (Fonte: O autor)

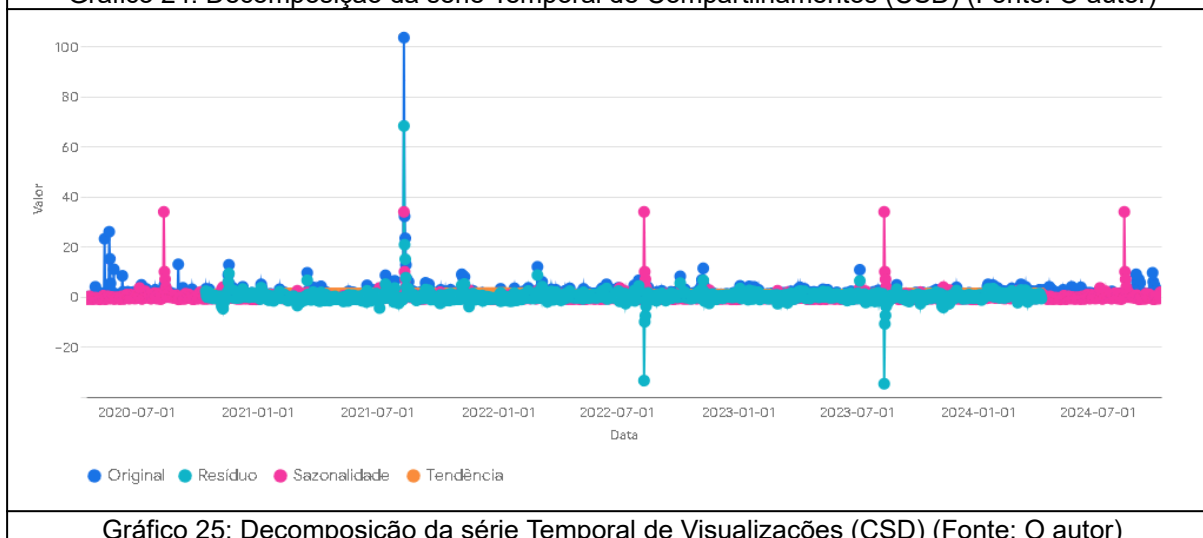


Gráfico 25: Decomposição da série Temporal de Visualizações (CSD) (Fonte: O autor)

O modelo requer dados estacionários, o que significa que a média, variância e autocorrelação devem ser constantes ao longo do tempo. Isso pode ser alcançado pela diferenciação dos dados. Ele pode ser ajustado aos dados estacionários, estimando os parâmetros dos componentes AR, I e MA.

Após o ajuste, o modelo pode fazer previsões futuras com base em dados históricos e parâmetros estimados. No caso destes dados, o modelo ARIMA foi capaz de prever com precisão os valores futuros das variáveis “Visualizações” e “Compartilhamentos”. Isso sugere que o modelo é um bom ajuste para os dados e que pode ser usado para fazer previsões confiáveis sobre o futuro.

Inferências

A análise de séries temporais de visualizações e compartilhamentos revelou uma tendência ascendente ao longo do tempo, com picos nos meses de verão, o

que sugere a presença de sazonalidade nos dados.

A decomposição das séries temporais demonstrou que a tendência é o componente mais significativo tanto para visualizações quanto para compartilhamentos. A sazonalidade também é relevante, mas o componente residual é relativamente pequeno.

O teste **T**, aplicado para comparar as médias de visualizações e compartilhamentos, indicou uma diferença estatisticamente significativa entre as médias dessas duas variáveis.

Os modelos autorregressivos integrados de médias móveis (ARIMA) foram capazes de prever com precisão os valores futuros de visualizações e compartilhamentos, evidenciando haver forte relação entre essas variáveis e sugerindo que essa relação pode persistir no futuro.

Percepções

A correlação entre o número de compartilhamentos e visualizações é de 0,51. Isso indica uma relação positiva moderada entre as duas variáveis.

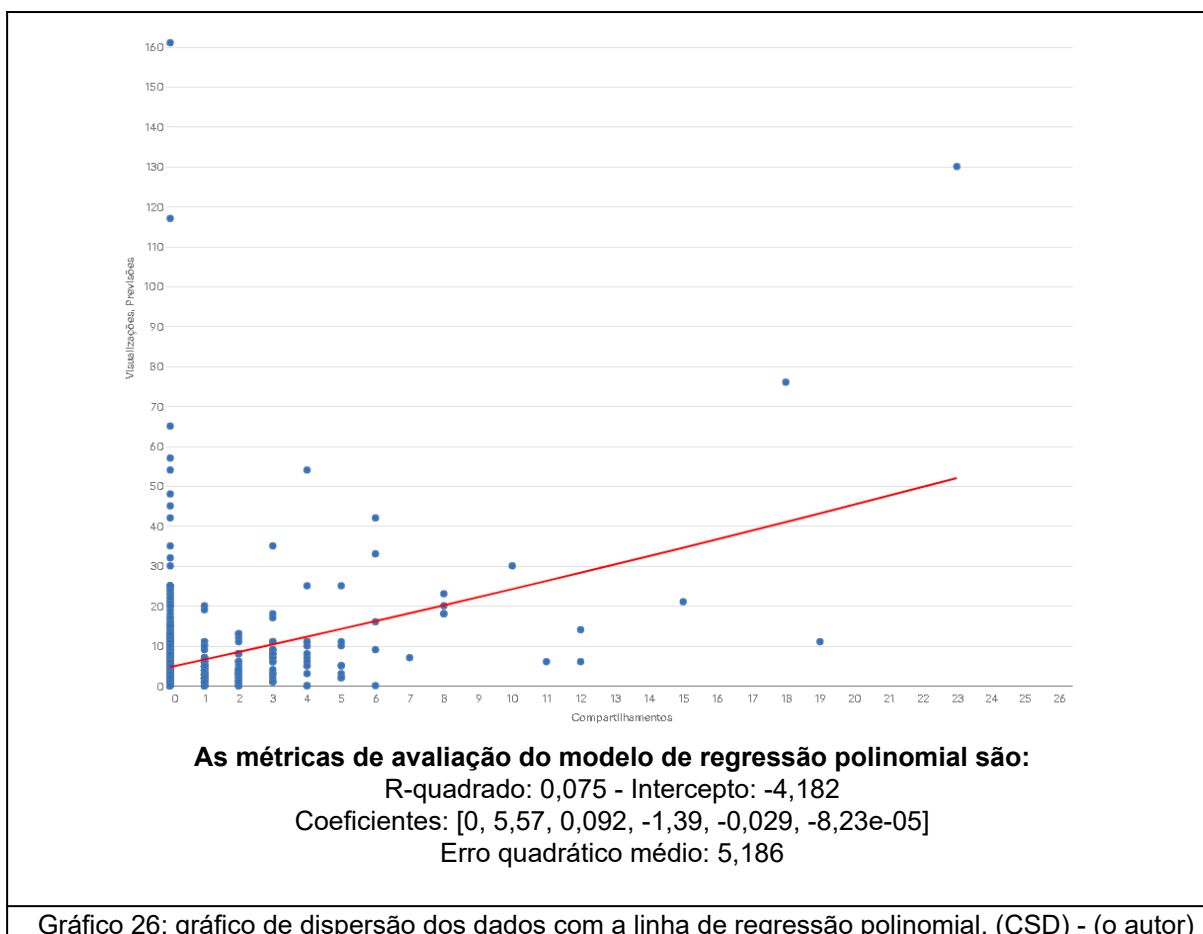
O modelo de regressão linear indica que:

- Um aumento de 1 compartilhamento está associado a um aumento de 2,82 visualizações em média.
- O valor de R^2 de 0,112 sugere que o número de compartilhamentos explica 11,2% da variância nas visualizações.

Embora a correlação seja moderada, o modelo de regressão linear não parece ser um bom ajuste para os dados, conforme indicado pelo baixo valor de R^2 . Pode haver outros fatores que estão contribuindo para o número de visualizações.

Regressão Polinomial

Entre modelos não lineares que podem ser usados para analisar esses dados, a regressão polinomial é uma forma de análise de regressão na qual a relação entre a variável independente X e a variável dependente Y é modelada como um polinômio de grau N .



O modelo de regressão polinomial se ajusta aos dados um pouco melhor do que o modelo de regressão linear, conforme indicado pelo valor de R ao quadrado de 0,133. No entanto, o modelo ainda não explica uma grande quantidade da variação nos dados. Pode haver outros fatores que estejam contribuindo para o número de visualizações, além do número de compartilhamentos.

O valor de R-quadrado de 0,075 indica que o modelo explica 7,5% da variância dos dados. Isso sugere que o modelo pode não ser um bom ajuste para os dados. O erro quadrático médio de 5,186 indica que o modelo tem um erro médio de 5,186 unidades ao prever os valores de Compartilhamento.

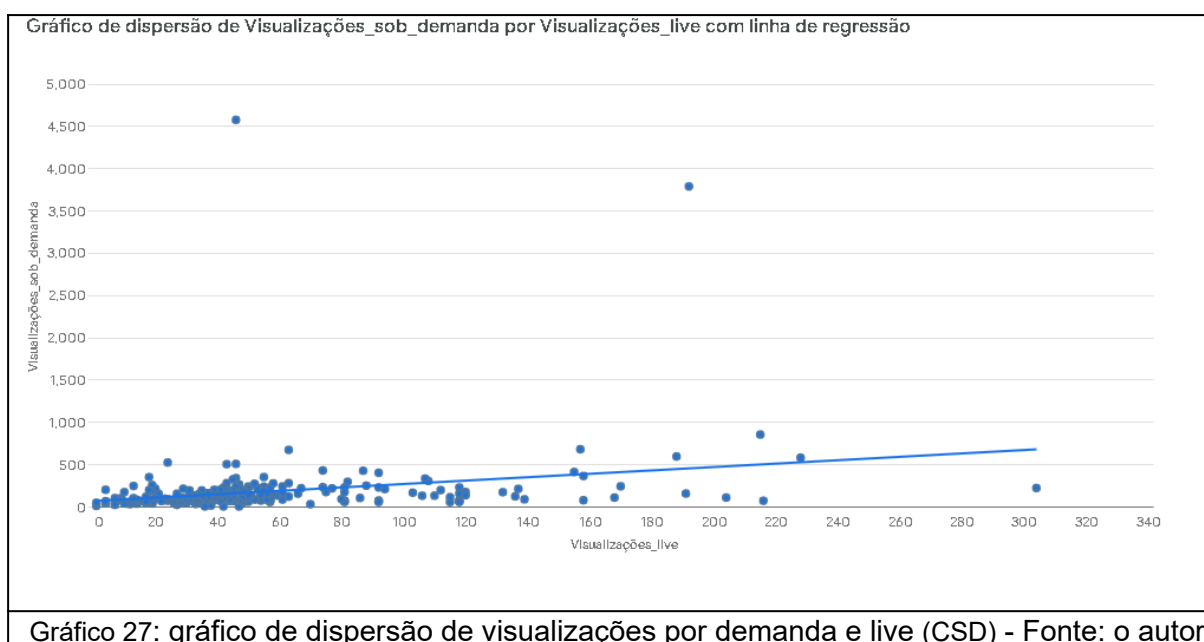
3.11.6 - Análise de correlações entre visualizações ao vivo e sob demanda

Ao analisar os dados, percebe-se que as duas tabelas podem ser vinculadas através da coluna "Título Do Vídeo". Após a limpeza dos dados, foi possível buscar correlações entre as tabelas. Observa-se um coeficiente de correlação de 0,24, um número positivo, o que sugere que a correlação seja fraca.

Um aumento de 1 unidade em 'Visualizações_live' está associado a um aumento de 2,01 unidades em 'Visualizações_sob_demanda'.

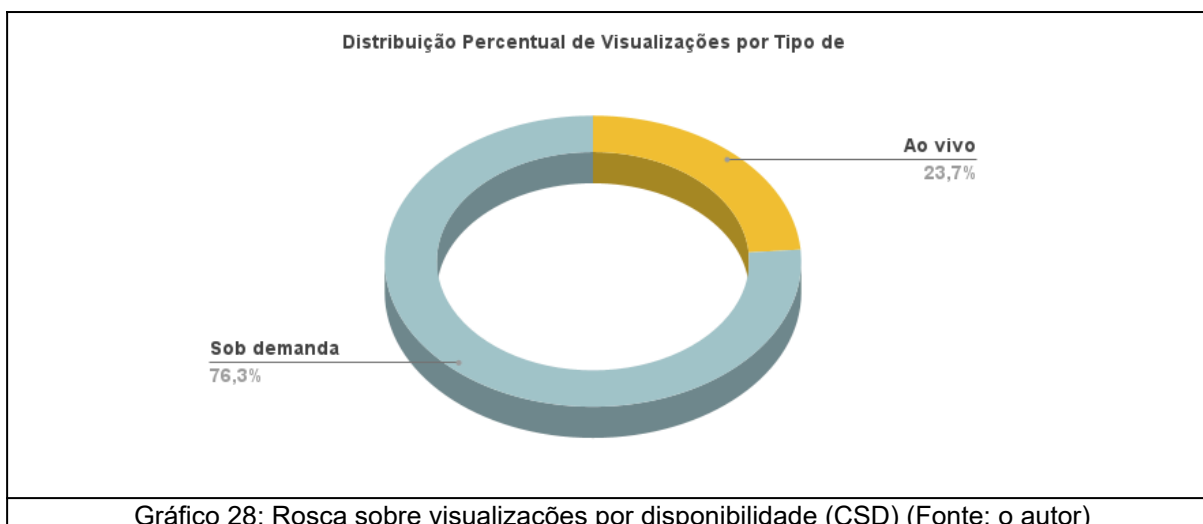
O valor de R ao quadrado de 0,058 indica que o modelo explica apenas 5,8% da variabilidade em 'Visualizações_sob_demanda'.

O baixo valor de R ao quadrado sugere que pode haver outros fatores importantes que influenciam 'Visualizações_sob_demanda' que não foram incluídos no modelo.

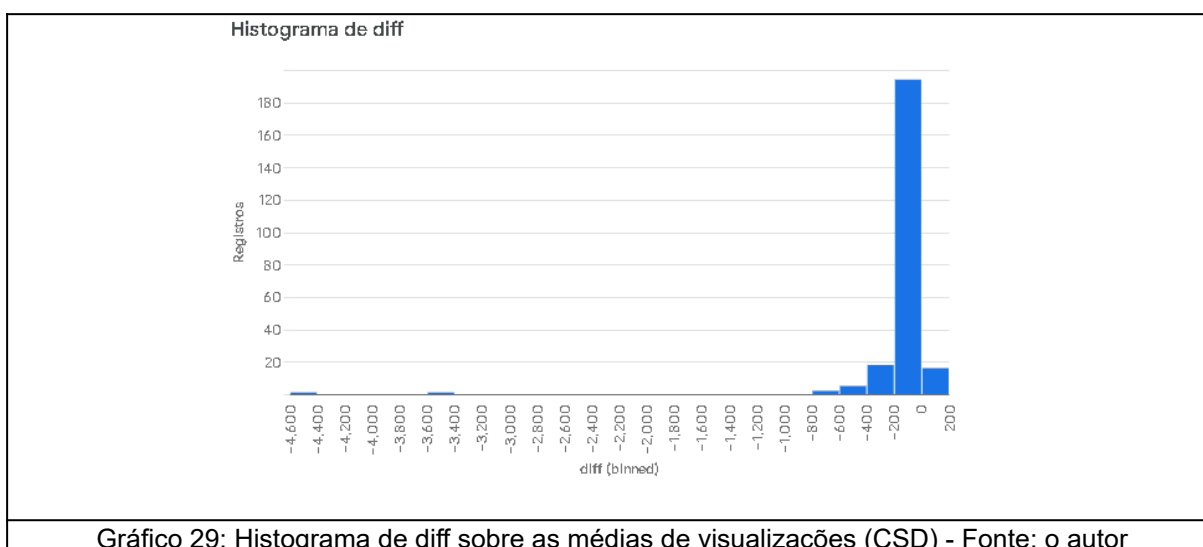


O gráfico de dispersão mostra que há uma relação positiva entre as visualizações ao vivo e sob demanda, mas não é muito forte. Isso é consistente com a análise de correlação, que encontrou uma correlação fraca entre as duas variáveis.

Verificamos se há alguma diferença significativa entre as visualizações ao vivo e sob demanda usando um teste T, cujo valor de $6.075501775251668e-07$, é menor que o nível de significância de 0,05. Portanto, podemos rejeitar a hipótese nula de que não há diferença entre as visualizações ao vivo e sob demanda.



A média de visualizações ao vivo é 108,65 e o desvio padrão é 41,32, sendo que a média de visualizações sob demanda é 231,96 e o desvio padrão é 448,17, mostrando haver uma diferença significativa entre as visualizações ao vivo e sob demanda e que há uma relação positiva entre as duas variáveis, mas não é muito forte.

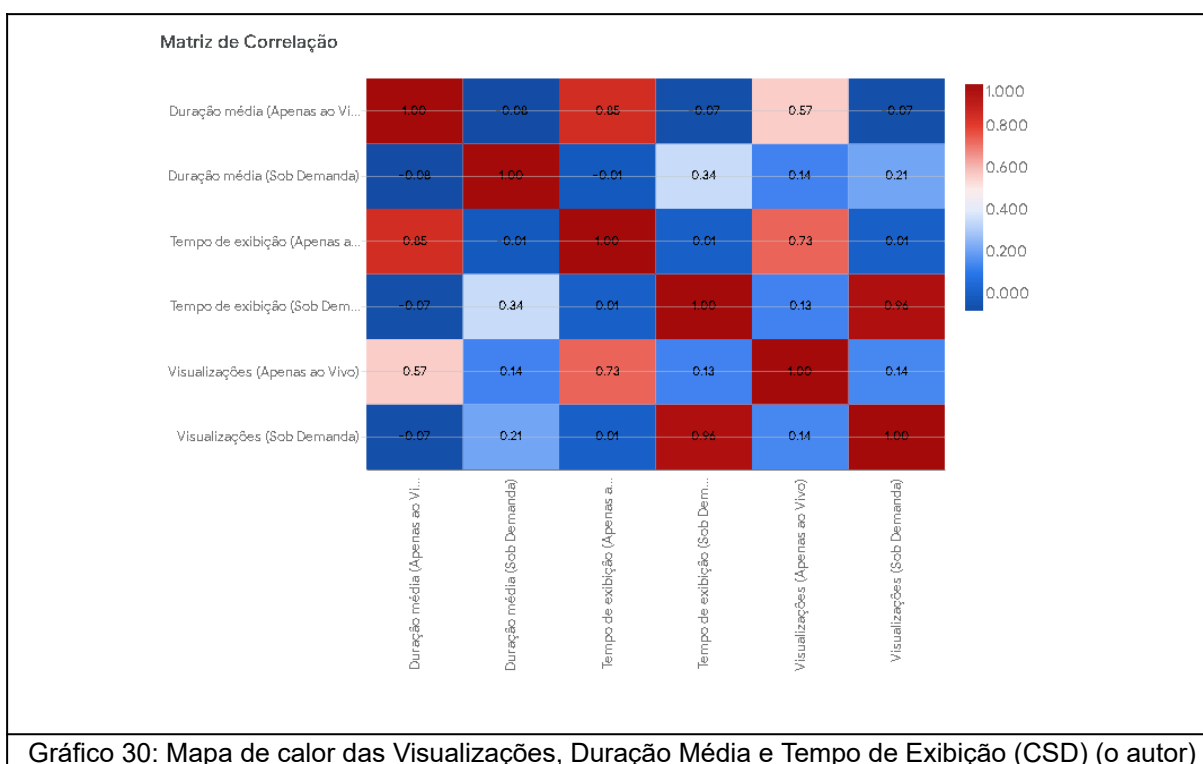


Podemos encontrar correlações entre colunas “Visualizações”, “Data”, “Tempo de exibição” e “Duração média da visualização”, criando um mapa de calor que demonstre as correlações entre essas variáveis. A matriz de correlação exibe as relações entre as variáveis:

Variáveis	Correlações e suas forças
Visualizações (Sob Demanda) e Visualizações (Apenas ao Vivo)	Correlação positiva fraca (0.240), indicando que vídeos com mais visualizações sob demanda tendem a ter mais visualizações apenas ao vivo, mas a relação não é muito forte.

Tempo de Exibição (Sob Demanda) e Tempo de Exibição (Apenas ao Vivo)	Correlação positiva moderada (0.439), sugerindo que vídeos com maior tempo de exibição sob demanda geralmente têm maior tempo de exibição apenas ao vivo.
Duração média (Sob Demanda) e Duração média (Apenas ao Vivo)	Correlação positiva forte (0.717), indicando que vídeos com maior duração média sob demanda tendem a ter maior duração média apenas ao vivo.
Visualizações (Sob Demanda) e Tempo de exibição (Sob Demanda)	Correlação positiva forte (0.817), mostrando que vídeos com mais visualizações sob demanda também têm maior tempo de exibição sob demanda.
Visualizações (Apenas ao Vivo) e Tempo de exibição (Apenas ao Vivo)	Correlação positiva quase perfeita (0.989), indicando uma relação muito forte entre visualizações e tempo de exibição para vídeos apenas ao vivo
Quadro 8: Variáveis analisadas, correlações de forças (CSD) - Fonte: o autor	

A correlação entre as visualizações “Sob Demanda” e “Apenas ao Vivo” é de 0.240, o que indica uma correlação positiva fraca. Isso sugere que, em geral, vídeos com mais visualizações “Sob Demanda” tendem a ter mais visualizações “Apenas ao Vivo”, mas essa relação não é muito forte.



Percepções:

As visualizações sob demanda e apenas ao vivo têm correlação positiva, mas fraca, o que sugere que o público que assiste sob demanda pode ser diferente do público que assiste apenas ao vivo.

O tempo de exibição é um indicador importante tanto para vídeos sob demanda quanto para vídeos apenas ao vivo, e tem correlação positiva com as visualizações.

A duração média da visualização é consistente entre vídeos sob demanda e apenas ao vivo, o que pode indicar que o tipo de conteúdo influencia a duração média da visualização.

As correlações mais fortes encontradas foram entre visualizações e tempo de exibição, o que sugere que o tempo de exibição pode ser um indicador importante do interesse do público em um vídeo.

É importante lembrar que a correlação não implica causalidade e que outras análises poderiam expor as causas por trás dessas correlações. O mapa de calor e a Tabela de Correlações oferecem mais percepções que permitem analisar relações entre algumas variáveis do modelo, como nos exemplos abaixo:

Relação entre tempo de exibição e duração média: O mapa de calor não mostra uma correlação forte entre a duração média da visualização e o tempo de exibição. Isso sugere que vídeos mais longos não necessariamente levam a um maior tempo de exibição. É possível que outros fatores, como o tema do vídeo e a qualidade do conteúdo, influenciem mais o tempo de exibição do que a duração do vídeo em si.

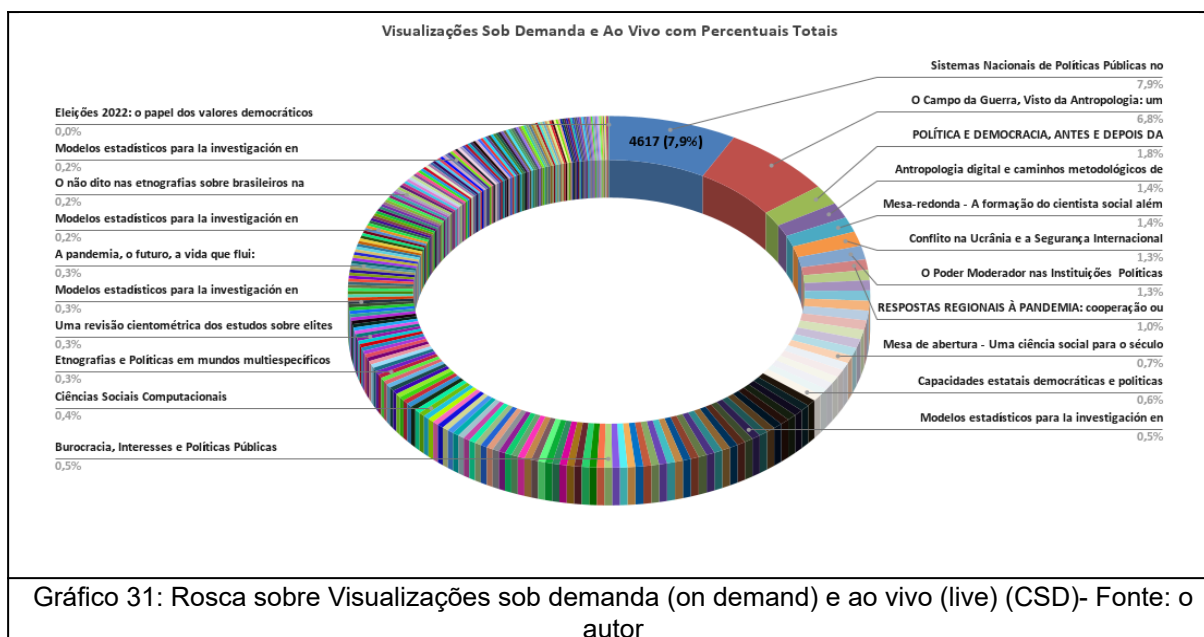
Comparação entre vídeos sob demanda e apenas ao vivo: As correlações entre as visualizações, tempo de exibição e duração média são geralmente mais fortes para vídeos apenas ao vivo do que para vídeos sob demanda. Isso pode indicar que o público que assiste a vídeos apenas ao vivo é mais homogêneo em termos de comportamento de visualização, enquanto o público sob demanda é mais diverso.

Análise individual de vídeos: O mapa de calor fornece uma visão geral das correlações entre as variáveis, mas também pode ser usado para analisar vídeos individualmente. Por exemplo, você pode identificar vídeos com alto número de visualizações e baixo tempo de exibição, o que pode indicar que o público está perdendo o interesse no vídeo rapidamente.

Identificação de outliers: É possível identificar outliers no mapa de calor, ou seja, vídeos que se desviam significativamente do padrão geral. Esses outliers podem ser interessantes para uma análise mais aprofundada, pois podem revelar percepções sobre vídeos com características únicas.

O gráfico rosca abaixo uma distribuição ponderada dos conteúdos mais assistidos ao vivo e sob demanda. É importante verificar que a análise demonstrou claramente que as visualizações obtidas na modalidade "Ao vivo" são mais consistentes e com maior "Duração de visualização", o que pode nos levar a inferir

que os conteúdos permanecem ativos e agregando até 3 vezes mais visualizações, porém com uma redução considerável nos minutos visualizados.



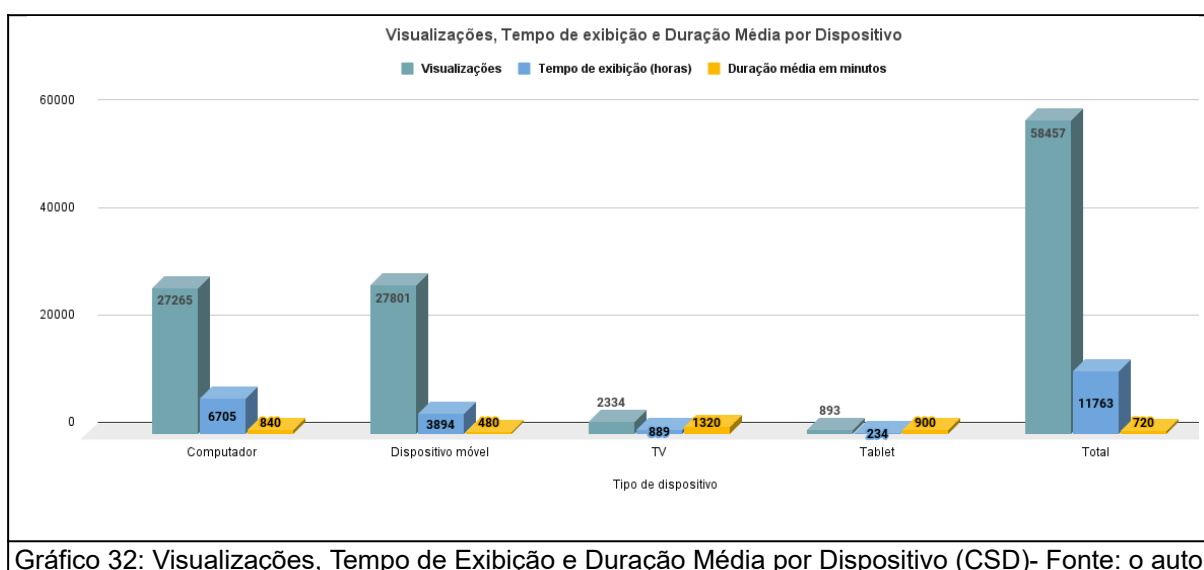
Uma visualização direta do mapa de calor dos 22 mais visualizados na própria tabela também pode oferecer uma abstração concreta entre o peso de cada modalidade no total visualizado.

Título do vídeo	Sob Demanda	Ao Vivo	Total
Sistemas Nacionais de Políticas Públicas no Federalismo Brasileiro	4571	46	4617
O Campo da Guerra, Visto da Antropologia: um sobrevoo no Brasil atual	3784	192	3976
Política e Democracia, Antes e Depois da Covid-19	850	215	1065
Antropologia digital e caminhos metodológicos de pesquisas etnográficas online e offline	677	157	834
Mesa-redonda - A formação do cientista social além do "cânone"	577	228	805
Conflito na Ucrânia e a Segurança Internacional	591	188	779
O Poder Moderador nas Instituições Políticas Brasileiras: o pretérito e o presente	669	63	732
Respostas Regionais à Pandemia: cooperação ou nacionalismo?	408	155	563
Um novo olhar para as Relações Institucionais e Governamentais	504	46	550
A governança regional da Amazônia no governo Bolsonaro	520	24	544
Diplomacia Parlamentar	500	43	543

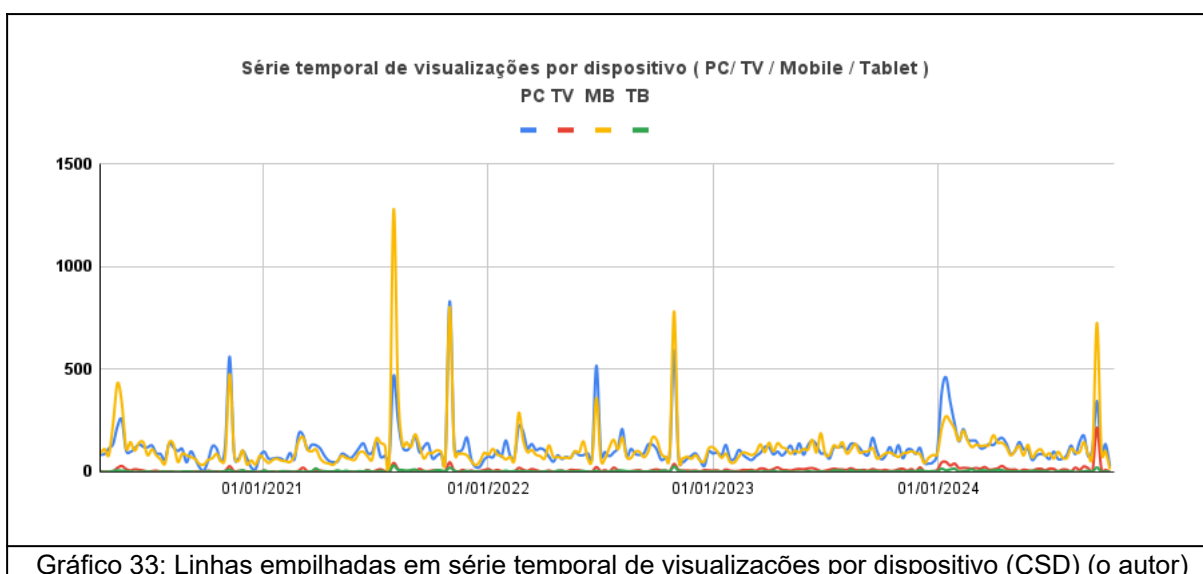
Conferência de abertura - O cenário brasileiro pós-pandemia	218	304	522
População Negra, Racismos e a Covid-19 no Brasil	360	158	518
O trabalho sem fim e sem direitos: o empreendedorismo e os ataques à regulação do emprego	423	87	510
Política e Classes Sociais na atual conjuntura brasileira	426	74	500
Populismo	398	92	490
Direitos Humanos em Tempos de Intolerância	330	107	437
Mesa de abertura - Uma ciência social para o século XXI: o que somos e o que podemos ser?	302	108	410
Conferência de abertura - A importância do movimento indígena contemporâneo	238	170	408
Federalismo, relações intergovernamentais e políticas sociais no contexto da Pandemia	348	55	403
Desafios sociais na Era dos Desastres	336	46	382
Tabela 5: Mapa de calor dos 22 videos mais visualizados (CSD) - (o autor)			

3.11.7 - Análise de acessos por plataforma e dispositivo

Os dados obtidos nessa análise mostram a relevância de conhecer como os usuários acessam e por quais dispositivos para obter percepções que podem influenciar a decisão de adequar conteúdos aos diferentes tipos de telas em suas disponibilidades. O canal multiplataforma é acessado por dispositivos móveis (47,5%), computadores (46,64%), televisores (3,99%) e tablets (1,53%), sendo praticamente equivalente no percentual em relação aos dois primeiros.



A maior duração média de visualizações é nas TVs (22m50s), Tablets (15m42s), computadores (14m45s) e a menor em dispositivos móveis (8m24s), possivelmente pelo alto consumo de franquia de dados móveis e/ou fadiga visual ocasionado pela pequena tela dos mesmos. Totalizando então 58.457 visualizações em 11.763 horas de exibição com duração média total 12m04s.



Uma correlação positiva significativa (0,916) existe entre as colunas Visualizações e Tempo de exibição (horas). Isso significa que, à medida que as visualizações aumentam, o tempo de exibição também tende a aumentar. O gráfico de dispersão gerado visualiza essa correlação, mostrando como as visualizações e o tempo de exibição tendem a aumentar simultaneamente.

Porém, a reta de regressão linear sobreposta mostra que com apenas uma reta, não separamos os pontos adequadamente, portanto, faz sentido tentar ajustar um modelo não linear aos dados.

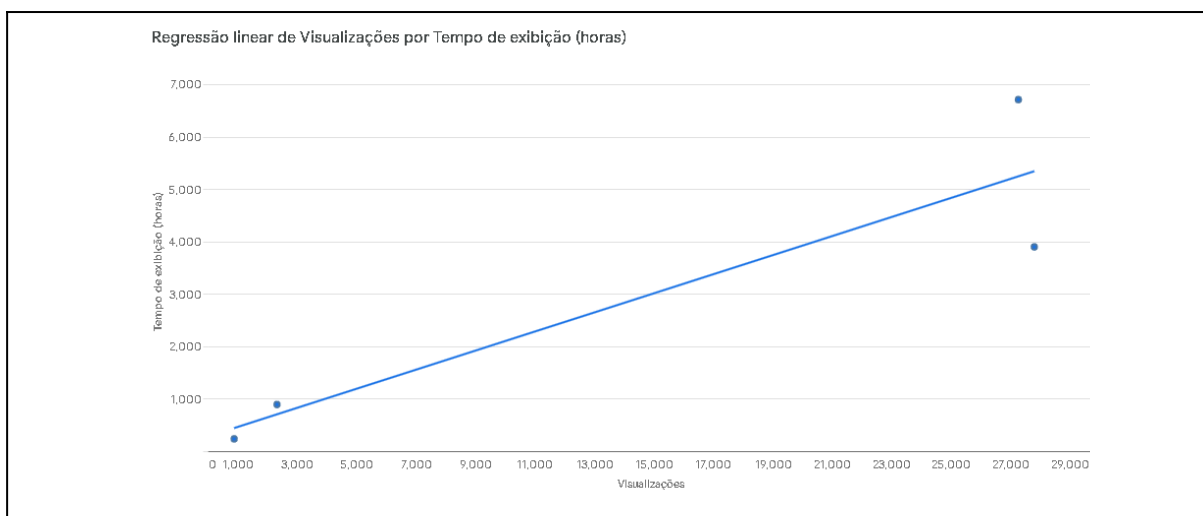


Gráfico 34: Linha de regressão indicando inviabilidade do modelo de regressão linear (CSD) (o autor)

Podemos ajustar um modelo não linear aos dados usando a função exponencial. Para o valor de R-quadrado ser calculado, corrigimos a escala dos dados para traçar um gráfico de dispersão dos dados com a curva ajustada sobreposta para visualizar o ajuste.

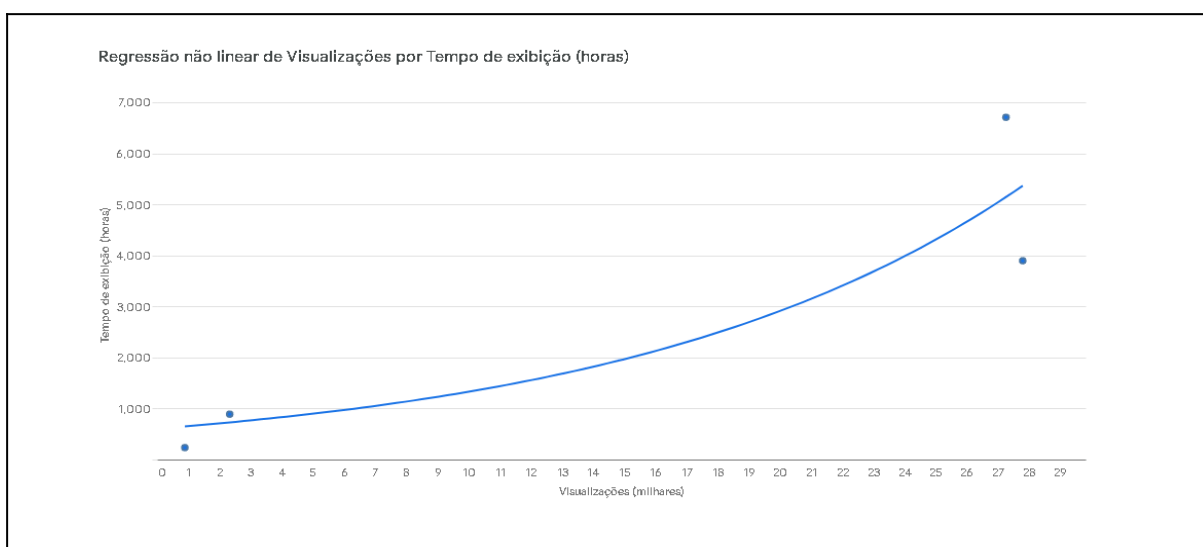


Gráfico 35: Linha de regressão exponencial (CSD) (Fonte: o autor)

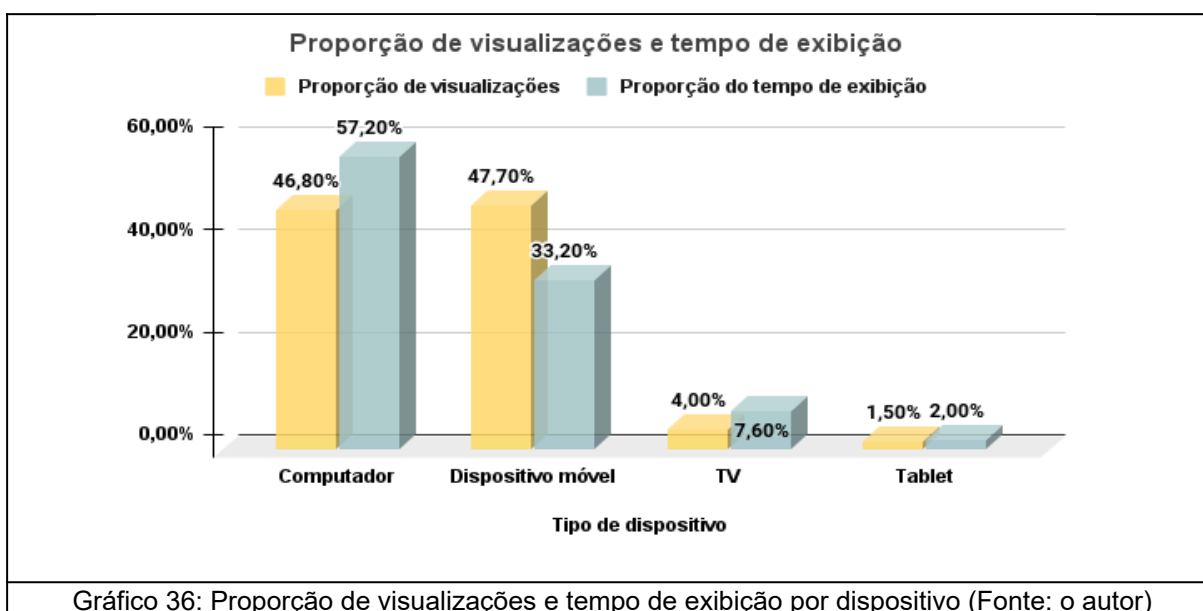
Uma função exponencial foi usada para ajustar uma regressão não linear aos dados. Os parâmetros ajustados foram $a = 606,30$ e $b = 0,078$. Com um R-quadrado de 0,820, o modelo explica 82% da variabilidade dos dados.

O R-quadrado é uma medida estatística que determina a proporção da variância na variável dependente que é explicada pela variável independente. No contexto do conjunto de dados fornecido, um R-quadrado de 0,82 sugere que 82% da variação no tempo de exibição é atribuída ao número de visualizações.

O modelo exponencial ajustado aos dados explica 82% da variação no tempo de exibição. Isso significa que o modelo captura a maioria dos fatores que influenciam o tempo de exibição. No entanto, os 18% restantes da variabilidade não são explicados por ele. Isso pode ser devido a uma série de fatores, como o tipo de conteúdo assistido, a hora do dia em que as visualizações ocorreram ou outros fatores não incluídos nesta análise.

O coeficiente de determinação (R^2) não é um indicador da qualidade de um modelo, mas apenas da proporção da variabilidade explicada. Um alto R^2 pode ser obtido mesmo com um modelo que não se ajusta bem aos dados, enquanto um baixo R^2 pode ser obtido com um modelo que se ajusta bem, mas que explica apenas uma pequena parte da variabilidade total.

O valor de R-quadrado de 0,82 para este conjunto de dados é relativamente alto, o que indica que o modelo exponencial se ajusta bem aos dados. Podemos realizar uma análise a partir de uma tabela com agregação proporcional dos dados de “Visualizações” e “Tempo de Exibição”.



A média da proporção de visualizações é de cerca de 0,25, com um desvio padrão de 0,26. Os valores variam de 0,015 a 0,477, sendo que a proporção média do tempo de exibição é de cerca de 0,25, com um desvio padrão de 0,25, variando de 0,02 a 0,572.

Os computadores desktop e notebook apresentam a maior proporção de tempo de exibição (57,20%), enquanto a proporção de visualizações (46,80%) é ligeiramente menor que a do dispositivo móvel. Isso sugere que as pessoas tendem a assistir a vídeos por mais tempo no computador, mesmo que acessem com um pouco menos de frequência em comparação aos dispositivos móveis.

Já os Dispositivos Móveis têm a maior proporção de visualizações (47,70%), mas um tempo de exibição significativamente menor (33,20%) em comparação ao computador. Isso indica que as pessoas usam mais o celular para acessar os vídeos, mas assistem por períodos mais curtos.

As Smart TVs possuem baixa proporção de visualizações (4%), mas o tempo de exibição (7,60%) é proporcionalmente quase o dobro das visualizações. Os Tablets apresentam as menores proporções, tanto de visualizações (1,50%) quanto de tempo de exibição (2%), sendo o menos utilizado para assistir aos vídeos.

Percepções:

Os dispositivos preferidos para assistir a vídeos variam conforme o tipo de conteúdo, público-alvo e outros fatores. No entanto, alguns padrões podem ser observados.

O computador é o dispositivo mais indicado para assistir a vídeos com mais atenção e por períodos mais longos. Isso se deve à sua tela maior e à possibilidade de usar fones de ouvido ou alto-falantes de alta qualidade. O dispositivo móvel é o mais popular para acesso rápido aos vídeos, mas com menor tempo de exibição. Isso se deve ao seu tamanho compacto e à conveniência de poder assistir a vídeos em qualquer lugar.

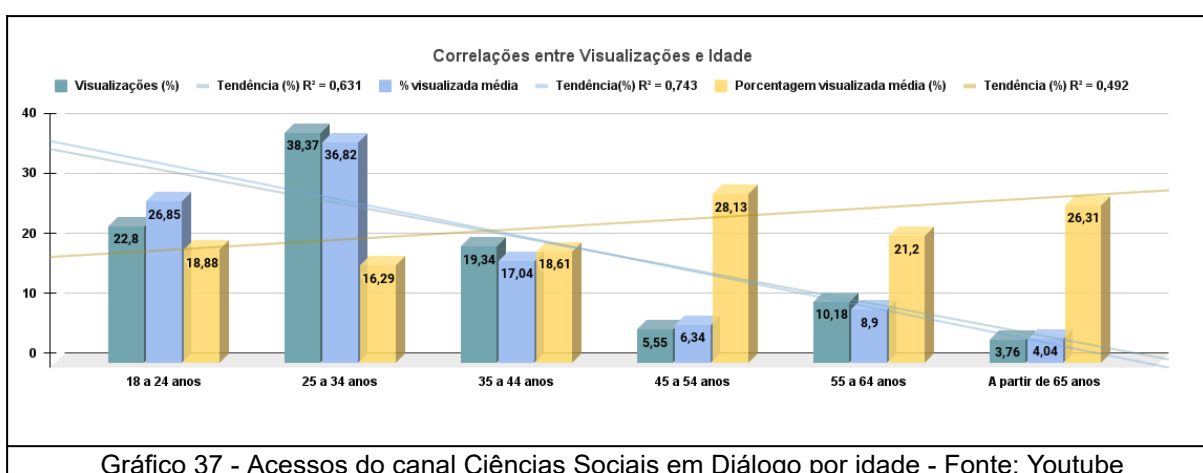
A TV e o tablet são dispositivos menos relevantes para o consumo de vídeos. Isso se deve à sua tela menor e à falta de fones de ouvido ou alto-falantes de alta qualidade. Pode ser estratégico investir em criar formatos de vídeo mais longos e com maior qualidade para o computador e conteúdos mais curtos e dinâmicos para dispositivos móveis, ideais para consumo rápido.

Vale ressaltar que esses dados representam apenas uma fatia do consumo de vídeos e podem variar dependendo do tipo de conteúdo, público-alvo e outros

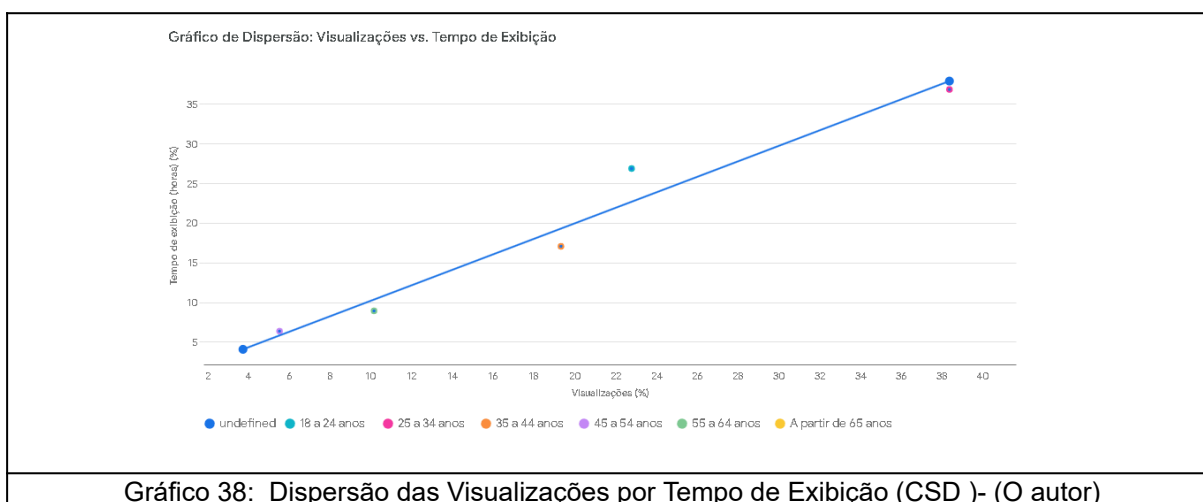
fatores. Análises mais aprofundadas, com dados demográficos e de comportamento do usuário, podem fornecer percepções mais completas e estratégicas.

3.11.8 - Análise de Correlações entre Visualizações e Idade

Em relação ao alcance dos conteúdos por idade, predomina a quantidade de acessos dos usuários (78,95%) e maior tempo de exibição entre 18 e 44 anos (78,95%), porém o percentual médio de visualizações é maior pelos usuários entre 45 e mais de 65 anos.



Para compreender melhor a correlação entre a idade do espectador e as métricas de visualização, realizamos uma análise detalhada dos dados. Além disso, utilizamos técnicas de correlação e regressão para determinar e medir numericamente a relação entre a idade do espectador e o tempo gasto assistindo ao conteúdo..



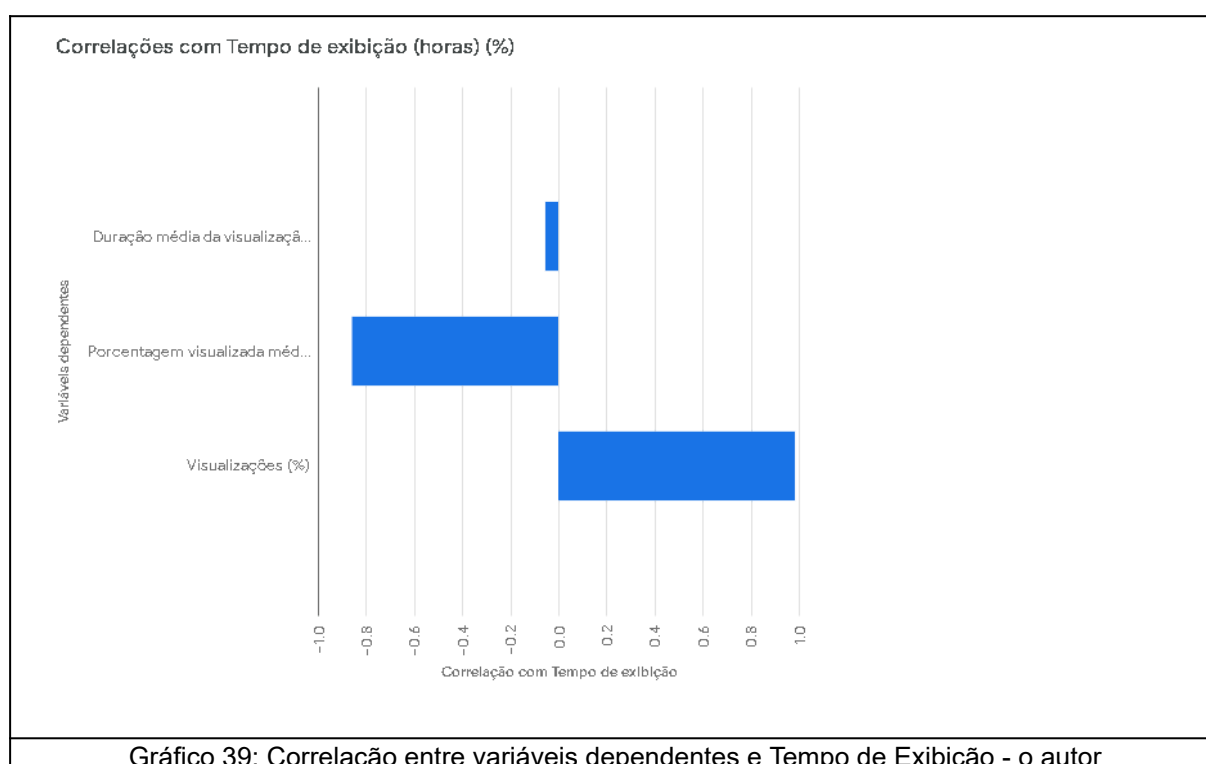
Existe uma forte correlação positiva entre o número de visualizações e o tempo total de exibição. Isso indica que os espectadores que assistem a mais vídeos tendem a gerar um tempo de exibição maior.

Existe uma correlação negativa moderada entre a porcentagem média visualizada e o tempo total de exibição. Isso sugere que, embora os espectadores que assistem a partes maiores de cada vídeo possam contribuir para um tempo de exibição maior por vídeo, eles podem não gerar tanto tempo de exibição total quanto aqueles que assistem a vários vídeos.

A duração média de visualização não está fortemente correlacionada com o tempo total de exibição. Isso pode indicar que a duração média de visualização não é um fator determinante no tempo total de exibição.

A análise de regressão ilustra ainda mais a forte relação linear positiva entre o número de visualizações e o tempo total de exibição. Se o número de visualizações é o preditor mais significativo do tempo total de exibição, aumentar o número de visualizações terá um impacto direto no aumento do tempo total de exibição.

Seria interessante incentivar os espectadores a assistirem a vários vídeos, recomendando conteúdo semelhante e criando playlists, criando engajamento e melhorando os tempos de visualização.



Avaliando o Tempo de Exibição médio, encontramos uma possibilidade plausível de que o percentual das visualizações “Sob Demanda” refletem o fato de que um conteúdo pode ser visto em várias partes, com diversos acessos em datas diferentes e granular demais a estatística não impulsiona a análise.

Metas que ensejam a conclusão dos vídeos na íntegra são irreais e não refletem a possibilidade de flexibilização de tempo que a internet permite e a disponibilidade quase ubíqua deles na nuvem. Também é inegável que transitamos entre telas de diferentes dimensões, dia e noite, em casa, na escola, no trabalho e na rua. Em cada uma delas há prós e contras sobre resolução, responsividade, colunas e plugins que atendem os usuários em vários dispositivos (Smart TV, Smartphones Tablet e Videogames).

Oferecer uma visualização agradável e limpa pode aumentar o número de visualizações e ser estratégia mais eficaz para incrementar o tempo total de exibição. Ao atrair mais espectadores, aumentam-se as probabilidades de que assistam a uma porção maior do conteúdo, o que pode levar a um envolvimento e engajamento mais significativos.

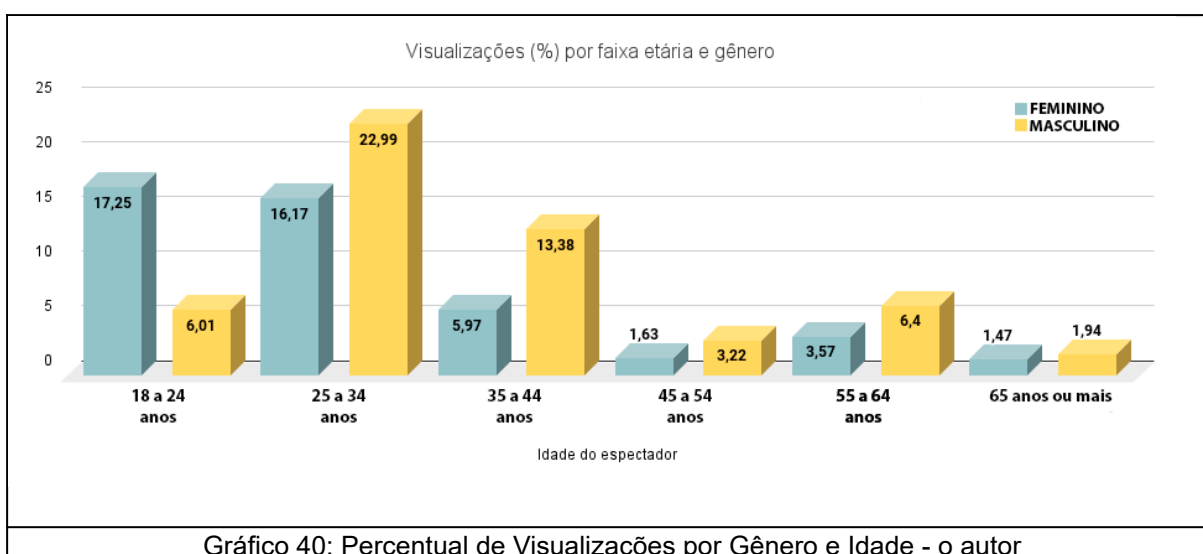
Percepções:

A análise dos dados por faixa etária revela percepções interessantes sobre o comportamento do espectador em relação ao tempo de exibição.

Faixa etária mais jovem (18-24 e 25-34 anos):	Contribui com a maior parte das visualizações e tempo de exibição, com a faixa etária de 25-34 anos liderando em ambos os aspectos.
	Apesar de visualizações mais longas, a porcentagem média visualizada é menor em comparação com faixas etárias mais velhas, particularmente 45-54 e 55-64 anos.
Faixa etária mais velha (45-54 e 55-64 anos):	Mostra maior tempo de exibição por vídeo, com pico na faixa etária de 45-54 anos.
Faixa etária mais velha (45-54 e 55-64 anos):	Apesar do menor número de visualizações, o tempo de exibição por vídeo é alto, indicando engajamento mais profundo com o conteúdo assistido.
Faixa etária de 35-44 anos	Comportamento intermediário entre as demais faixas

Faixa etária de 65+ anos	Menor tempo de exibição por vídeo, sugerindo preferência por conteúdo mais curto ou menos engajamento.
Quadro 9: Correlações entre idade, visualizações e tempo de exibição - o autor	

A análise por faixa etária do comportamento de visualização revelou diferenças significativas entre os grupos etários. Os grupos mais jovens geram um maior volume de visualizações, enquanto os mais velhos demonstram maior engajamento com o conteúdo assistido. Essas descobertas oferecem percepções valiosas para otimizar o conteúdo direcionado a públicos específicos de diferentes idades.



O gráfico de barras representa visualmente a distribuição percentual de visualizações dos conteúdos online de acordo com a faixa etária e o gênero dos espectadores.

Dados:

- Público maior: 25 a 34 anos, principalmente feminino (22,99%).
- Segundo maior público: 18 a 24 anos, ambos os gêneros com porcentagens próximas.
- Menor público: acima de 45 anos, com pouca representatividade.
- Em 65 anos ou mais, a diferença é pequena, com leve vantagem para o público masculino.
- Nas demais faixas etárias, a proporção entre os gêneros é relativamente equilibrada.

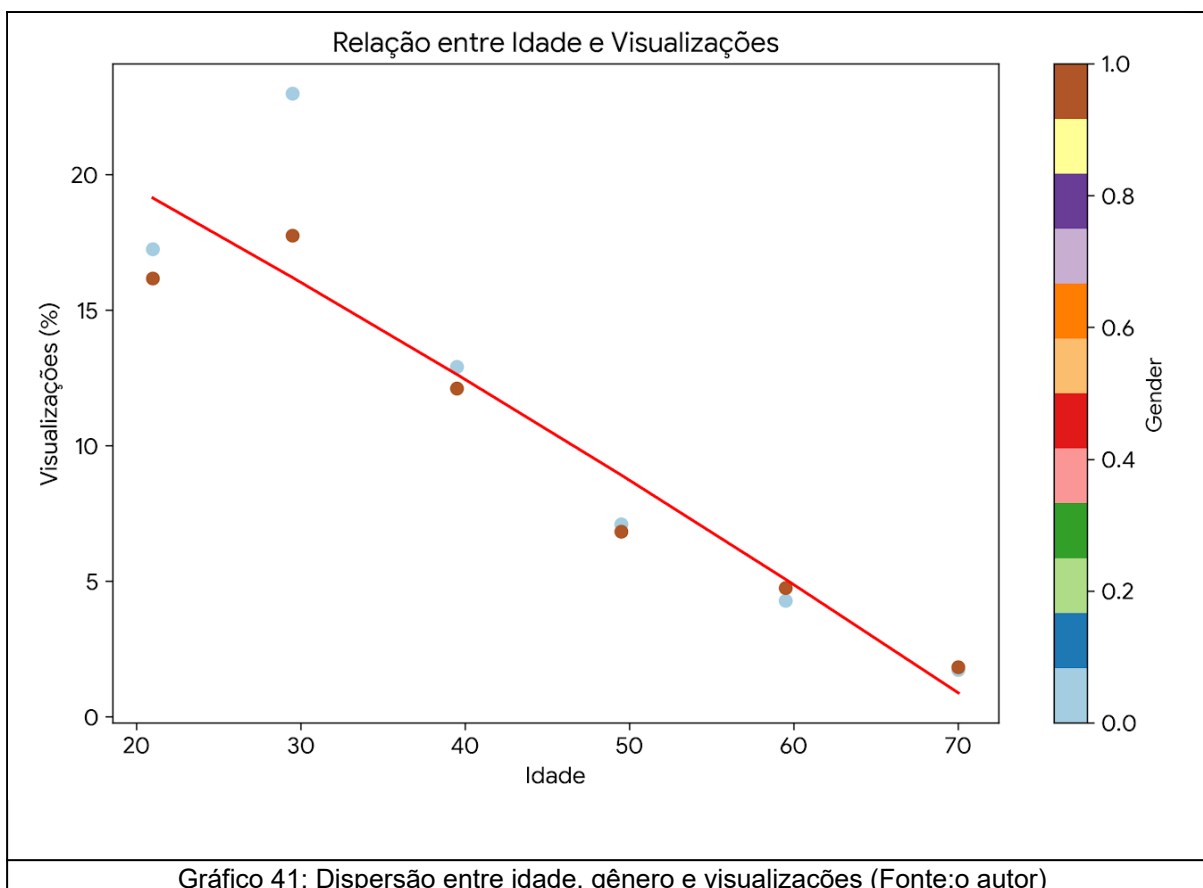
Percepções:

- Os conteúdos são mais populares entre jovens adultos, especialmente mulheres entre 25 e 34 anos.
- Há baixa representatividade do público acima de 45 anos, o que pode ter relação com dificuldades de acesso ou falta de familiaridade com a plataforma, além da probabilidade dos conteúdos não serem atrativos para essa faixa etária.
- Essas diferenças em algumas faixas etárias pode ser indicativo de se analisar a adaptação da linguagem ou formato do conteúdo para diferentes públicos.

3.11.9 - Analisando as correlações entre Idade, Gênero e Visualizações

A análise estatística dos dados revelou uma forte correlação negativa entre a idade e a porcentagem de visualizações (coeficiente de correlação de Pearson de -0.932, $p\text{-value} < 0.001$). Isso significa que, em geral, as faixas etárias mais jovens tendem a ter uma porcentagem maior de visualizações.

Devido a essa forte correlação, realizamos uma regressão polinomial para modelar a relação entre idade e visualizações. O modelo resultante teve um R-quadrado de 0.869, indicando que explica uma grande parte da variabilidade nos dados.



O gráfico de dispersão polinomial gerado mostra a relação entre idade e visualizações, com a linha de regressão polinomial destacada. Você pode observar visualmente como o modelo se ajusta aos seus dados.

Em resumo, a análise confirma que a idade é um fator importante na determinação da porcentagem de visualizações, com os mais jovens sendo o público dominante. O modelo de regressão pode ser usado para prever visualizações com base na idade, embora seja importante lembrar que outros fatores também podem influenciar as visualizações.

3.11.10 - Termos mais buscados no Youtube

Através da soma das visualizações de cada título, calculamos a porcentagem de visualizações em relação ao total delas, permitindo identificar os principais assuntos, ordenando os títulos por número de visualizações e porcentagem. No gráfico abaixo, o termo de busca “Ciências sociais em diálogo (40,64% das

visualizações)” foi suprimido para obter a visualização de termos de fato e não somente buscas com o nome do canal.



Seguem algumas percepções quantitativas e qualitativas de alguns assuntos mais buscados) excluindo “**Ciências sociais em diálogo**” (40,64% das visualizações):

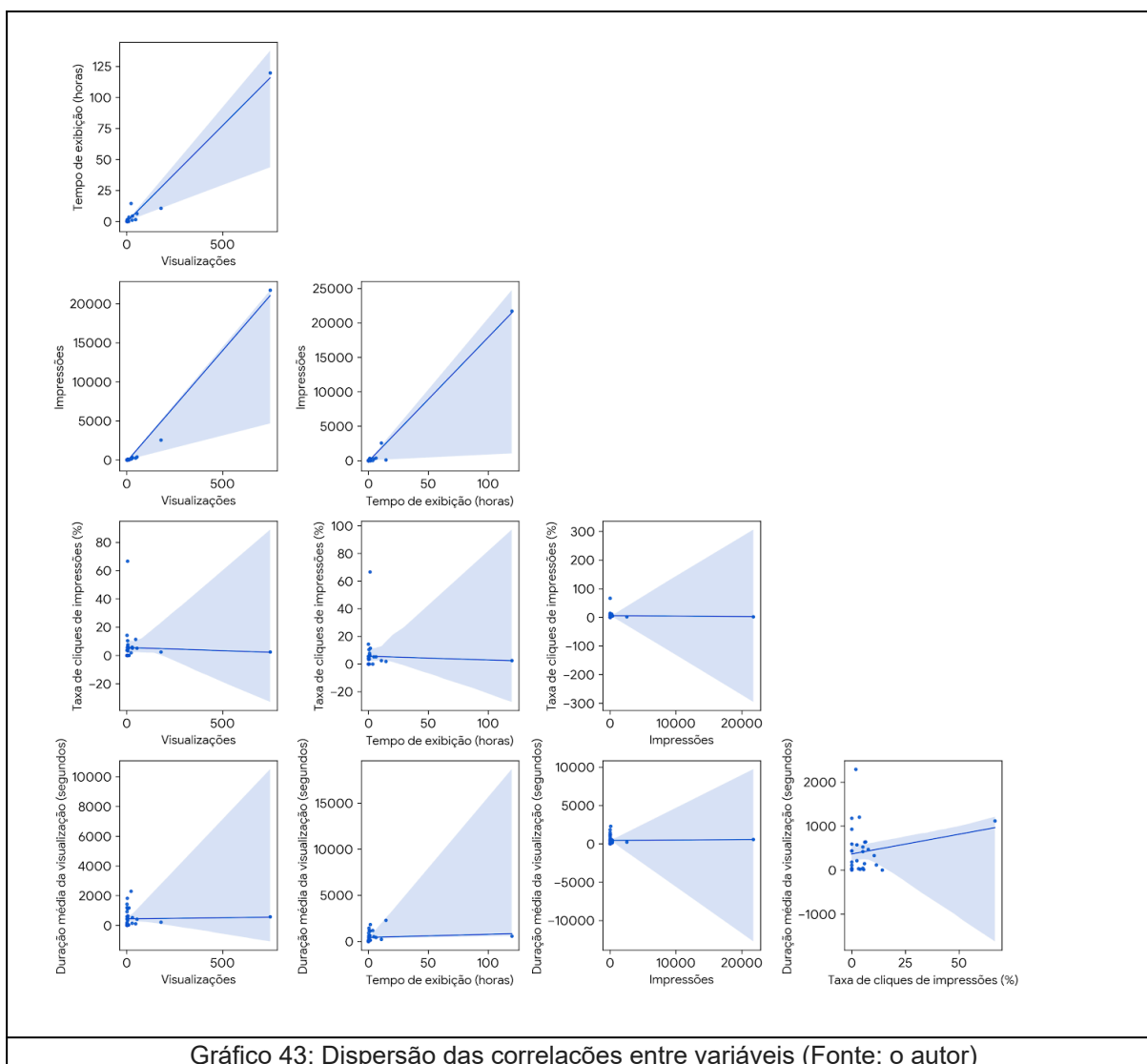
1. **Federalismo e descentralização de políticas públicas no Brasil** (22,9% das visualizações - três registros na tabela com títulos ligeiramente diferentes.)
2. **Piero Leirner** (18,9% das visualizações)
3. **Organização e funcionamento dos sistemas de programas nacionais** (11,13% das visualizações- dois registros na tabela, com títulos ligeiramente diferentes.)

3.11.11 - Visualizações e Tempo de exibição (horas) das Playlists

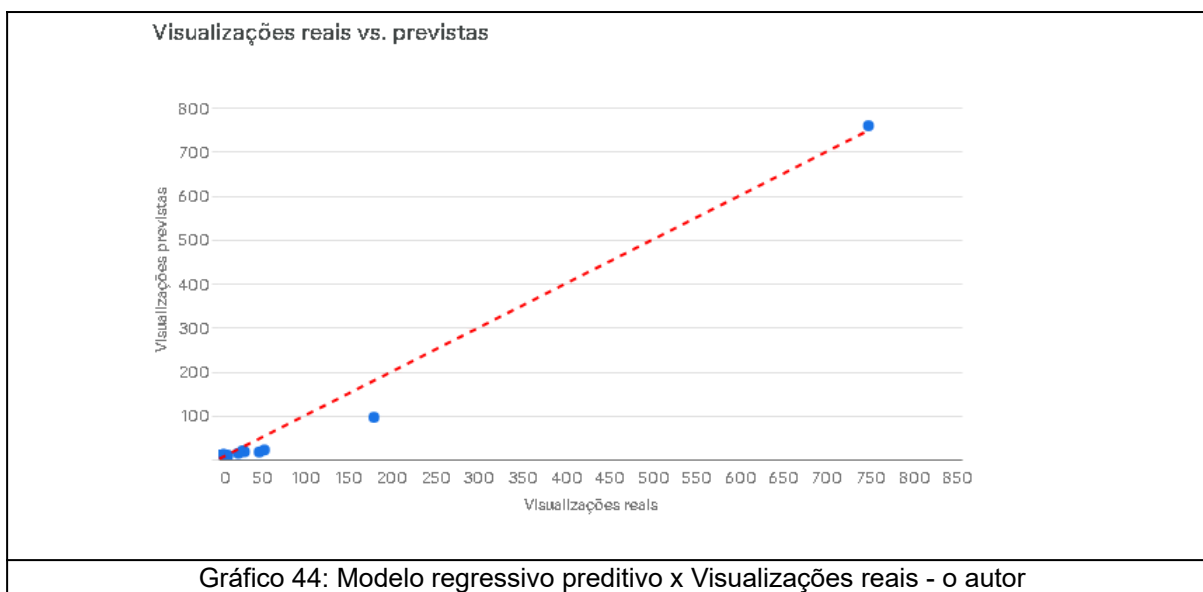
Analisamos as “Visualizações’ como variável dependente e ‘Tempo_de_exibicao_horas’ e ‘Impressões’ como variáveis independentes. Observamos resultados da regressão em que o R-quadrado do modelo é 0.981, o que significa que o modelo explica 98.1% da variância nos dados.

Além disso, o valor de p para ‘Impressões’ é 0.000, o que significa que ele é estatisticamente significativo. Por outro lado, o valor de p para ‘Tempo De Exibição’ é 0.894, tornando-o estatisticamente insignificante.

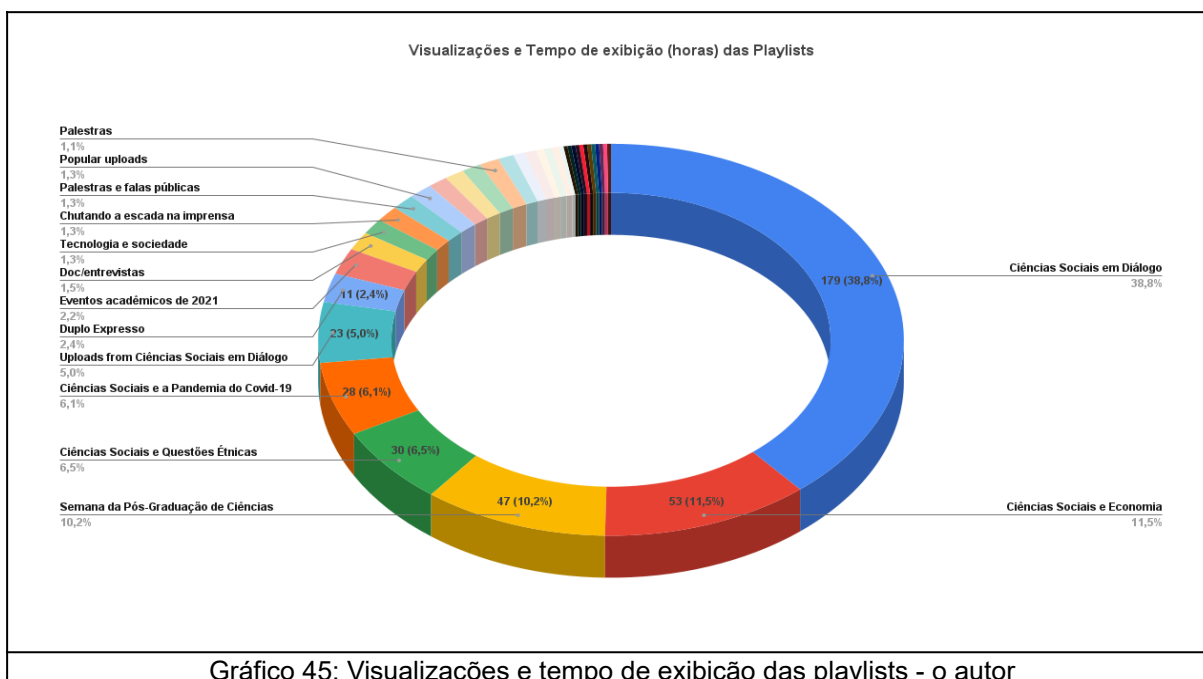
Para visualizar a relação entre cada par de variáveis, gráficos de dispersão foram criados para todas as possíveis combinações de pares de ‘Visualizações’, ‘Tempo de exibição (horas)’, ‘Impressões’, ‘Taxa de cliques de impressões (%)’ e ‘Duração média da visualização (segundos)’.



O modelo de regressão que melhor representa os dados é $\text{Visualizações} = 8.1784 + 0.1538 * \text{Tempo_de_exibicao} + 0.0337 * \text{Impressões}$. Este modelo explica 98,1% da variação nas 'Visualizações' dos dados.



Ao analisar os gráficos de dispersão, observamos uma correlação positiva entre "Visualizações", "Tempo de exibição (horas)" e "Impressões". Isso significa que, quando uma dessas variáveis aumenta, as outras também tendem a aumentar. Por outro lado, "Taxa de cliques de impressões (%)" não parece ter uma correlação linear clara com nenhuma das outras variáveis.



"Duração média da visualização (segundos)" parece ter uma correlação positiva fraca com "Visualizações", "Tempo de exibição (horas)" e "Impressões", mas a correlação não é tão forte quanto entre as três primeiras variáveis.

3.11.12 - Visualizações e Horas de Exibição por ano

Ao longo dos anos, a dinâmica de visualizações e tempos de exibição dos conteúdos no YouTube apresentou uma tendência de queda nas visualizações e crescimento no tempo de exibição, conforme demonstrado pelas linhas de tendência. Os dados são apresentados em um período de cinco anos, de 2020 a 2024, permitindo uma análise da evolução das métricas ao longo do tempo. Isso inclui tanto o alcance do conteúdo (visualizações) quanto o engajamento do público (tempo de exibição).

Esta análise fornece percepções valiosas sobre as mudanças no comportamento do público e as estratégias de conteúdo que podem ser adotadas para atender às necessidades dos espectadores.

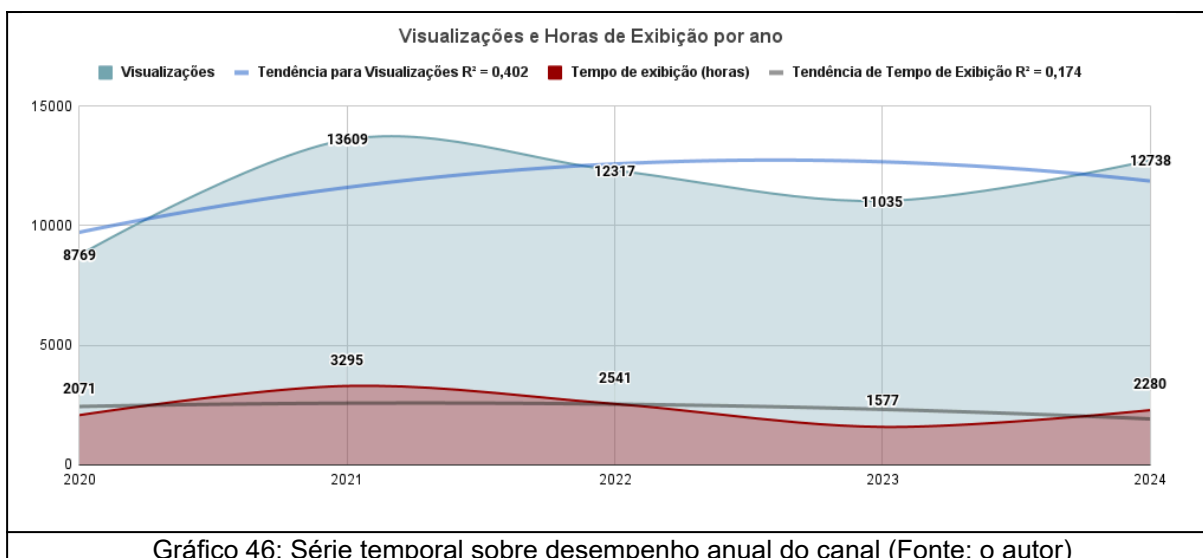
O gráfico mostra a evolução das visualizações e do tempo de exibição (em horas) de vídeos no YouTube entre 2020 e 2024.

Visualizações:

- **Crescimento geral:** Houve um aumento no número de visualizações ao longo do período, com um pico em 2021 e uma leve queda em 2022, seguida de nova alta em 2024.
- **Tendência de crescimento:** Apesar das oscilações, a tendência geral de visualizações é de crescimento, indicada pela linha azul no gráfico. O valor de $R^2 = 0,402$ indica que a linha de tendência explica 40,2% da variabilidade nos dados de visualizações.

Tempo de Exibição:

- **Variação:** O tempo de exibição teve um pico em 2021, seguido de uma queda até 2023 e um leve aumento em 2024, porém as oscilações são suaves e mudam gradativamente ao longo dos anos. O padrão senoidal reflete a regressão não linear de caráter exponencial e pode representar melhor as oscilações ao longo dos quase 5 anos do canal.
- **Tendência de estabilidade:** A linha cinza que indica a tendência do tempo de exibição é praticamente horizontal, revelando que, apesar das oscilações anuais, o tempo de exibição permaneceu relativamente estável ao longo do período. O valor de $R^2 = 0,174$ sugere que a linha de tendência explica apenas 17,4% da variabilidade nos dados de tempo de exibição, o que confirma a fraca correlação entre os dados e a tendência.



Percepções:

Engajamento: Embora o número de visualizações tenha aumentado, o tempo de exibição não acompanhou esse crescimento. Isso sugere que, apesar de mais pessoas estarem clicando nos vídeos, elas não estão assistindo por tanto tempo quanto antes.

Conteúdo: O tempo de exibição pode ter sido afetado por mudanças no tipo de conteúdo ou na duração dos vídeos ao longo do tempo. Vídeos mais curtos ou menos envolventes podem ter reduzido o tempo de exibição, apesar do aumento no número de visualizações.

Retenção de público: A queda no tempo de exibição pode ser um sinal de que o canal precisa otimizar suas estratégias para manter os espectadores engajados e assistir aos vídeos por mais tempo.

Recomendações:

- **Analisar o conteúdo:** Identifique os vídeos com maior e menor tempo de exibição para compreender o que mantém a atenção do público.
- **Otimizar a duração dos vídeos:** Experimente diferentes durações para encontrar o formato ideal que maximize o envolvimento.
- **Incluir chamadas para ação:** Incentive o público a assistir a outros vídeos, curtir, comentar e se inscrever no canal.
- **Utilizar técnicas de storytelling:** Crie narrativas envolventes que prendam a atenção do público do início ao fim.

As estatísticas descritivas são acessíveis por gráfico interativo e planilha de dados dinâmica em HTML no site <http://tecnociencia.net.br//elementor-383/>, que permite classificar e filtrar os dados para avaliações por assunto ou palavra chave, visualizações, horas de exibição, número de inscritos, impressões e cliques.

3.12 - ANÁLISES DESCRITIVAS DA RAPLE

3.12.1 - Projeto da Rede de Apoio aos Professores de Língua Estrangeira

O projeto de extensão, aprovado pelo Núcleo de Ensino da Faculdade de Ciências e Letras da Unesp em Araraquara, teve como propósito inicial a criação e customização de uma plataforma (REDE) de apoio a professores de línguas estrangeiras, que consistiu em um repositório de Objetos de Ensino-Aprendizagem (OA) e material digital para auxiliar professores, de diferentes idiomas, inseridos em contextos diversos, com sugestão de materiais relevantes .

A Rede de Apoio a Professores de Línguas Estrangeiras foi organizada como uma plataforma (ou repositório de OA) de materiais digitais devidamente ordenados e classificados como planos de aula, coletâneas de músicas, filmes, links e programas de autoria para uso do professor de línguas de diferentes contextos.

A plataforma atende docentes e alunos em dezenas de países em todos os continentes, mas principalmente no Brasil e outros países de língua portuguesa, ao disponibilizar materiais originais relevantes e significativos para os alunos de línguas da contemporaneidade,

Foram adotadas ferramentas de código aberto, livre e gratuito, de forma a não exigir investimentos financeiros em servidores físicos ou licenças pagas de software. Optou-se pelo desenvolvimento por meio do Content Management System (CMS) Wordpress Multisite em servidor Linux virtualizado na universidade. O projeto da REDE foi concebido a partir de uma abordagem que opta pelo desenvolvimento descentralizado, sustentável e colaborativo do site, pela construção coletiva de conteúdos e pelo uso de ferramentas de código aberto, gratuitas e bastante disseminadas entre os usuários e colaboradores que o constroem.

Atualmente a Rede de Apoio a Professores de Línguas Estrangeiras (RAPLE), está vinculada ao Centro de Línguas e Desenvolvimento de Professores (CLDP), no Campus da Faculdade de Ciências e Letras de Araraquara (FCLAr), tem como principal objetivo facilitar o acesso de professores a materiais didáticos digitais inovadores, autênticos e culturais. O projeto é organizado por línguas estrangeiras e temáticas acerca do ensino-aprendizagem de idiomas e fornece subsídios como planos de aula, indicações pedagógicas e conteúdos transmitidos por streaming.

3.12.2 - Análises e correlações

O Projeto de Extensão foi idealizado em 2013 e começou a tomar forma a partir de 2015 com a criação da primeira versão do site, agregando outras plataformas de mídia e redes sociais no decorrer dos anos. O site raple.fclar.unesp.br foi desenvolvido na plataforma Wordpress desde o início, sendo o eixo central dos recursos da rede e a maior fonte de materiais didáticos e conteúdos multimídia do projeto. Na segunda etapa, foram criadas as redes sociais da RAPLE para promover e divulgar ações e conteúdos do site, além de popularizar a rede e obter inscritos e engajamento.

Plataforma	Finalidade
Youtube	Transmissão de debates, oficinas e dinâmicas de ensino por streaming, estimulando a troca de ideias e conhecimentos sobre o ensino de línguas através do acervo audiovisual do canal.
Facebook	Comunicação e interação da rede de professores, alunos e interessados pelo facebook, divulgando conteúdos do site e atividades da comunidade.
Instagram	Divulgam eventos, disponibilizam conteúdos e se comunicam também pelo perfil no instagram.
Spotify	Playlists de diversos gêneros musicais em diferentes idiomas para auxiliar professores e estudantes de língua estrangeira.
Quadro 10: Plataformas da RAPLE e finalidades - o autor	

As redes foram implantadas ao longo dos anos, atendendo as demandas que novos estagiários e colaboradores traziam e forneceram meios de otimizar a comunicação social e midiática da rede. Diante desses fatos, é importante salientar que a coleta de dados será realizada no site e nas plataformas e/ou programas que complementam o repositório.

No caso do Youtube, há geração de novos conteúdos e ações de difusão e divulgação científica que permitem coletar dados e obter percepções sobre as métricas de visualização, engajamento e retenção de inscritos no canal, além de análises qualitativas sobre a transcrição, os comentários e chat das lives.

No caso do Facebook e Instagram, a finalidade sempre foi de comunicação com a comunidade que desenvolve e dissemina as redes, então somente será

realizada a análise descritiva de ambas, mas sempre descrevendo de forma clara, sobre qual delas são as métricas, análises e resultados.

3.12.3 - Site da RAPLE - Correlações de variáveis dos dados de acesso

A visualização de páginas é a principal variável dependente do estudo porque pode ser incluída em praticamente qualquer planilha de extração de dados do Youtube, que embora possibilite customizar os campos das tabelas do Analytics, não permite incluir todas as variáveis na coleta de dados. O Google Analytics do site da RAPLE permite articular as variáveis com mais possibilidades de mineração cruzada e favorece a análise multivariada.

Visualizações e Usuários Ativos por página do site

Inicialmente, a matriz de correlação é calculada, a qual mede a correlação linear entre cada par de colunas em um conjunto de dados. Ela revela a força e a direção da relação linear entre cada par de variáveis. Seus valores variam de -1 a 1, em que 1 indica uma correlação positiva perfeita, -1 uma correlação negativa perfeita e 0 ausência de correlação linear. As correlações indicaram:

- **Visualizações x Usuários Ativos: 0.85 (forte correlação positiva)** - Sugere que o aumento das visualizações da página está fortemente associado ao aumento do número de usuários ativos.
- **Visualizações x Contagem de Eventos: 0.998 (muito forte correlação positiva)** - Indica uma relação quase linear entre o número de visualizações e a contagem de eventos.
- **Usuários Ativos x Contagem de Eventos: 0.87 (forte correlação positiva)** - Sugere que um maior número de usuários ativos está geralmente associado a uma contagem maior de eventos.

Uma regressão linear modela cada par de variáveis com forte correlação:

Regressão linear entre Usuários ativos e Contagem de eventos

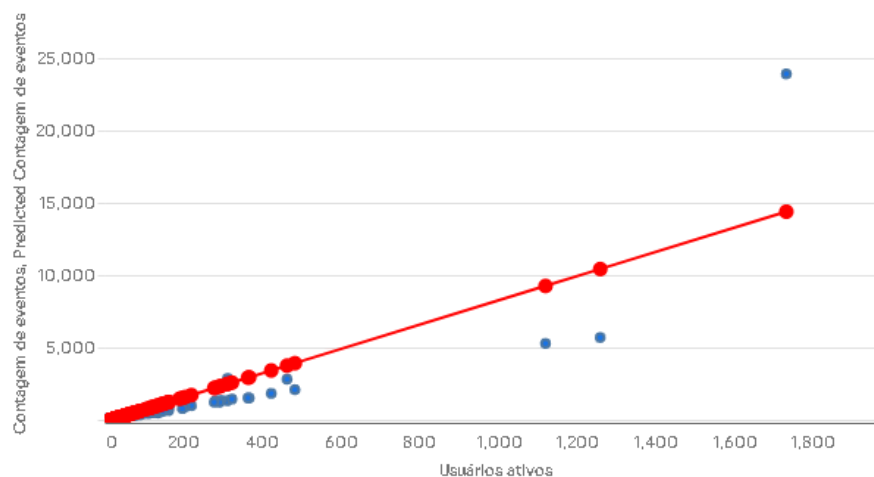


Gráfico 47: Visualizações x Usuários Ativos (RAPLE): Equação da linha de regressão: Usuários ativos = $0.000004 * \text{Visualizações} + 23.58$ - Coeficiente de determinação R^2 : 0.72

Regressão linear entre Visualizações e Contagem de eventos

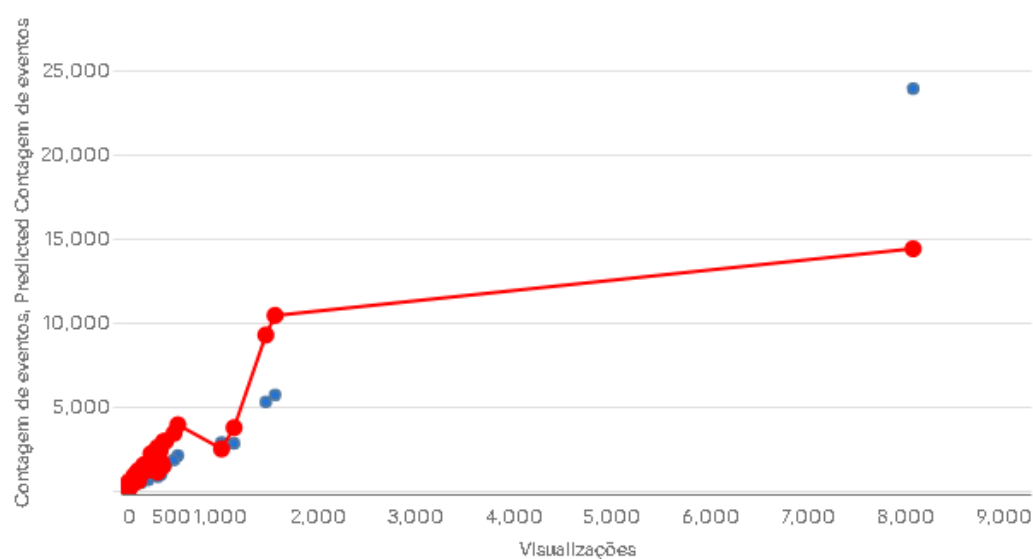
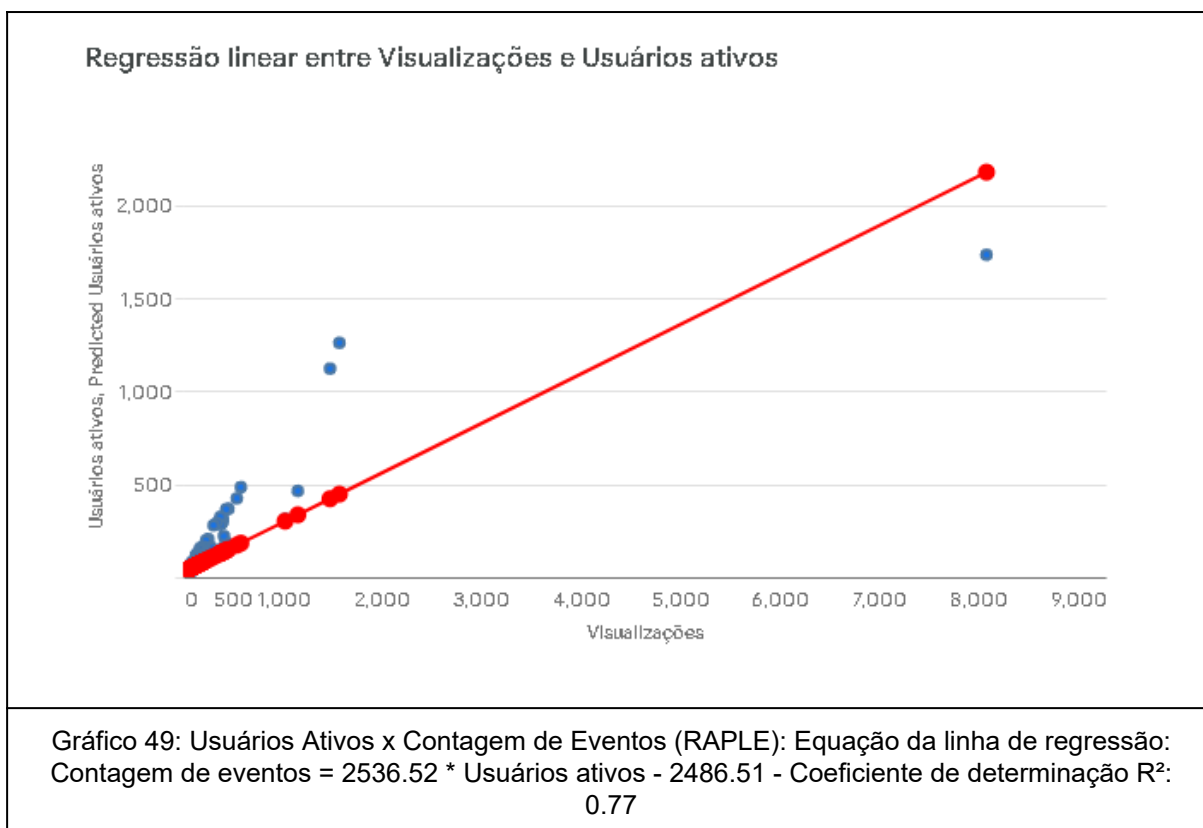


Gráfico 48: Visualizações x Contagem de Eventos (RAPLE): Equação da linha de regressão: Contagem de eventos = $1.00 * \text{Visualizações} + 11.73$ - Coeficiente de determinação R^2 : 1.00



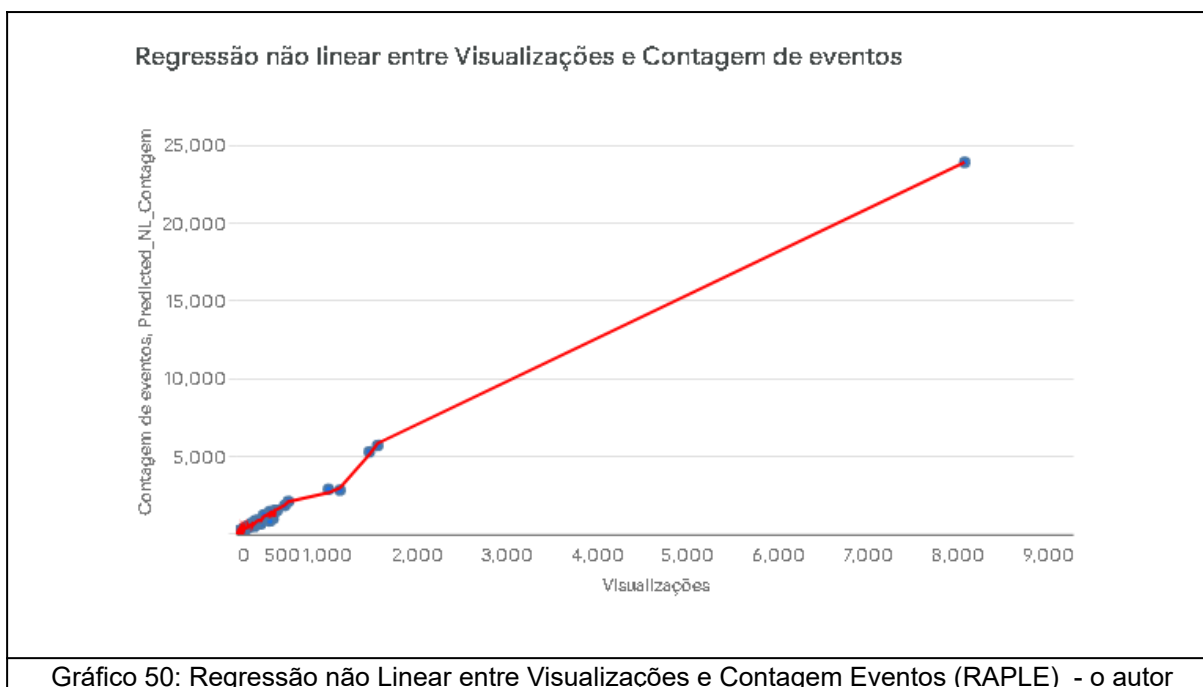
Conclusões da regressão: As análises apresentaram relações lineares significativas entre as variáveis estudadas. A quantidade de visualizações de páginas mostrou alta correlação tanto com o número de usuários ativos quanto com a contagem de eventos. Além disso, o número de usuários ativos também apresentou forte associação com a contagem de eventos.

Modelo de regressão não linear

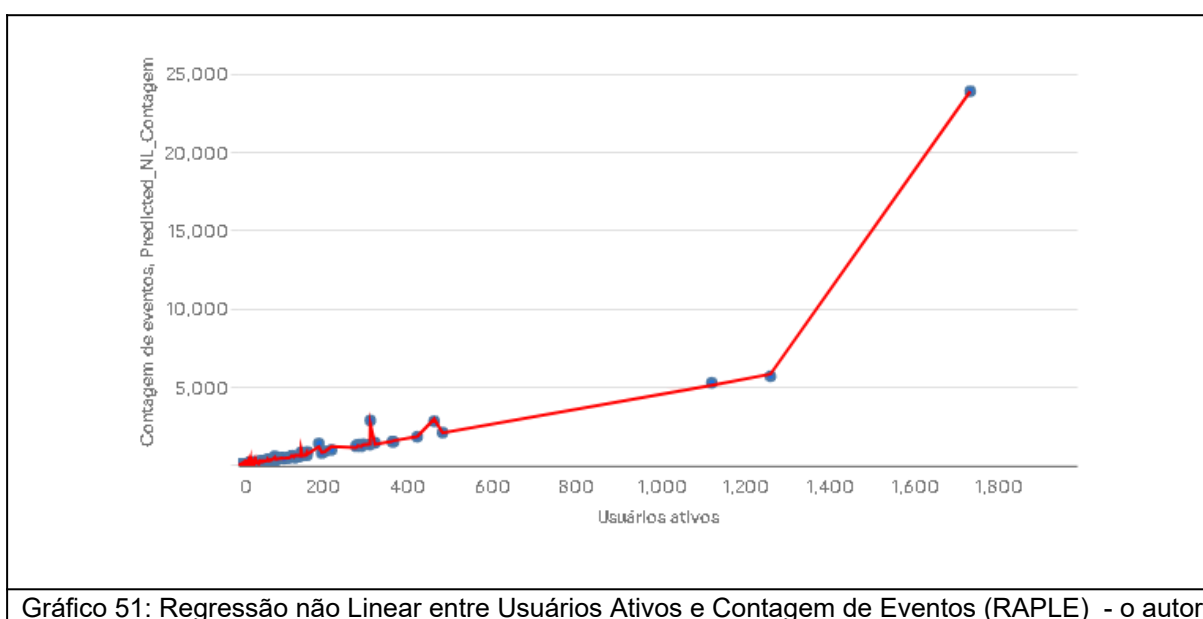
Ao aplicar um modelo de regressão não linear usando um modelo polinomial de grau 2, podemos verificar se há um melhor ajuste em relação à regressão linear. Vamos usar as variáveis 'Visualizações' e 'Usuários ativos' como preditoras e 'Contagem de eventos' como variável resposta.

Os gráficos gerados mostram a relação entre as variáveis e a qualidade do ajuste da regressão não linear. No gráfico 'Visualizações' x 'Contagem de eventos', podemos ver que a curva de regressão não linear se ajusta perfeitamente aos dados. No gráfico 'Usuários ativos' x 'Contagem de eventos', a curva de regressão não linear também se ajusta muito bem aos dados.

A regressão não linear (polinomial de grau 2) resultou em um R^2 de 1.00, o que indica um ajuste perfeito aos dados.



Ao comparar os coeficientes de determinação (R^2) dos modelos de regressão linear e não linear, podemos observar que ambos os modelos apresentam um bom ajuste aos dados. No entanto, o modelo de regressão não linear ($R^2 = 0.999$) apresenta um ajuste ligeiramente melhor do que o modelo de regressão linear ($R^2 = 0.77$). Isso sugere que o modelo não linear consegue capturar melhor a relação entre as variáveis 'Visualizações', 'Usuários ativos' e 'Contagem de eventos'.



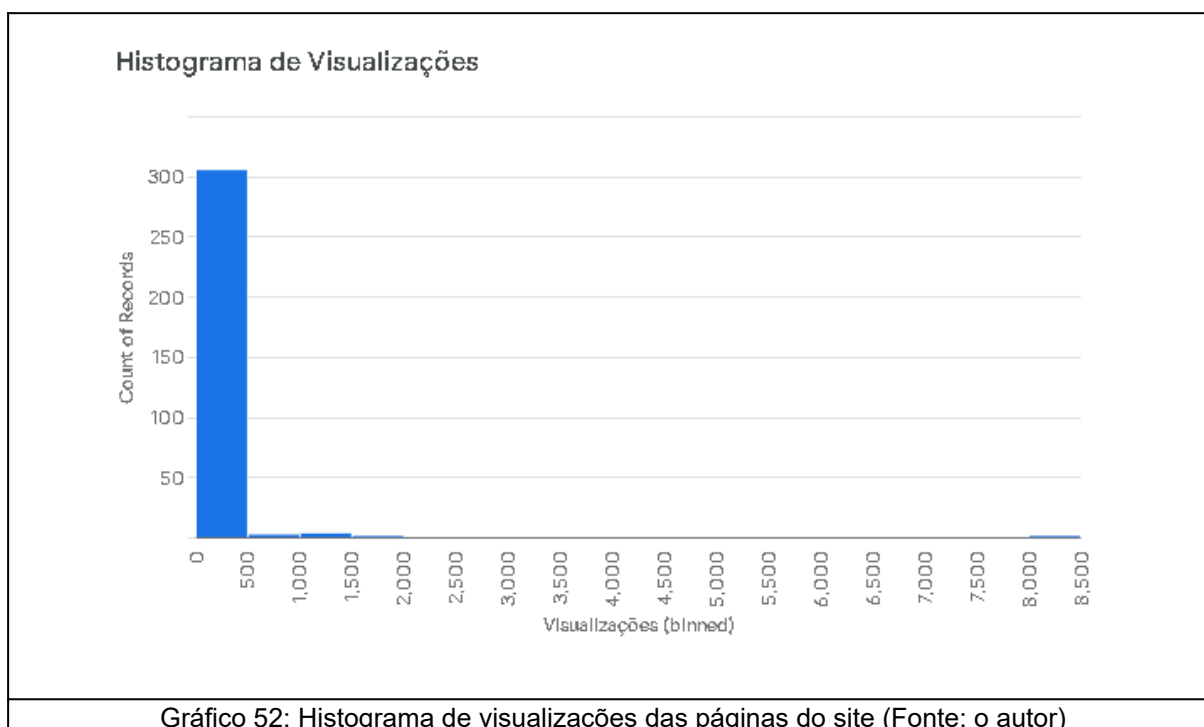
Apesar da pequena diferença entre os valores de R^2 , podemos considerar o modelo de regressão não linear como mais adequado para descrever a relação entre as variáveis. Ao selecionar o modelo mais apropriado para uma situação específica, é fundamental considerar fatores adicionais, como a complexidade do modelo e a interpretabilidade dos resultados.

Histogramas das Análises Descritivas Fundamentais

Ainda realizamos estatísticas descritivas como média, mediana, desvio padrão, mínimo, máximo e quartis para cada variável numérica, e plotamos histogramas para visualizar suas distribuições.

Origem	Média	Desvio Padrão	Mediana	Mínimo	Máximo
Visualizações Totais	94	484	22	0	8080
Usuários Ativos	53	152	17	1	1733
Contagem de eventos	279	1452	54	1	23876

Tabela 6 - Análises Descritivas de Visualizações, Usuários ativos e Contagem de eventos - o autor



Estatisticamente, o histograma de visualizações mostra que a maioria dos títulos de página possui uma contagem de visualizações relativamente baixa, com

um número pequeno de títulos de página apresentando uma contagem significativamente alta. Essa distribuição sugere que as visualizações podem ser influenciadas por alguns poucos títulos de página com alto tráfego ou engajamento, enquanto a maioria dos títulos de página tem um impacto menor na contagem geral.

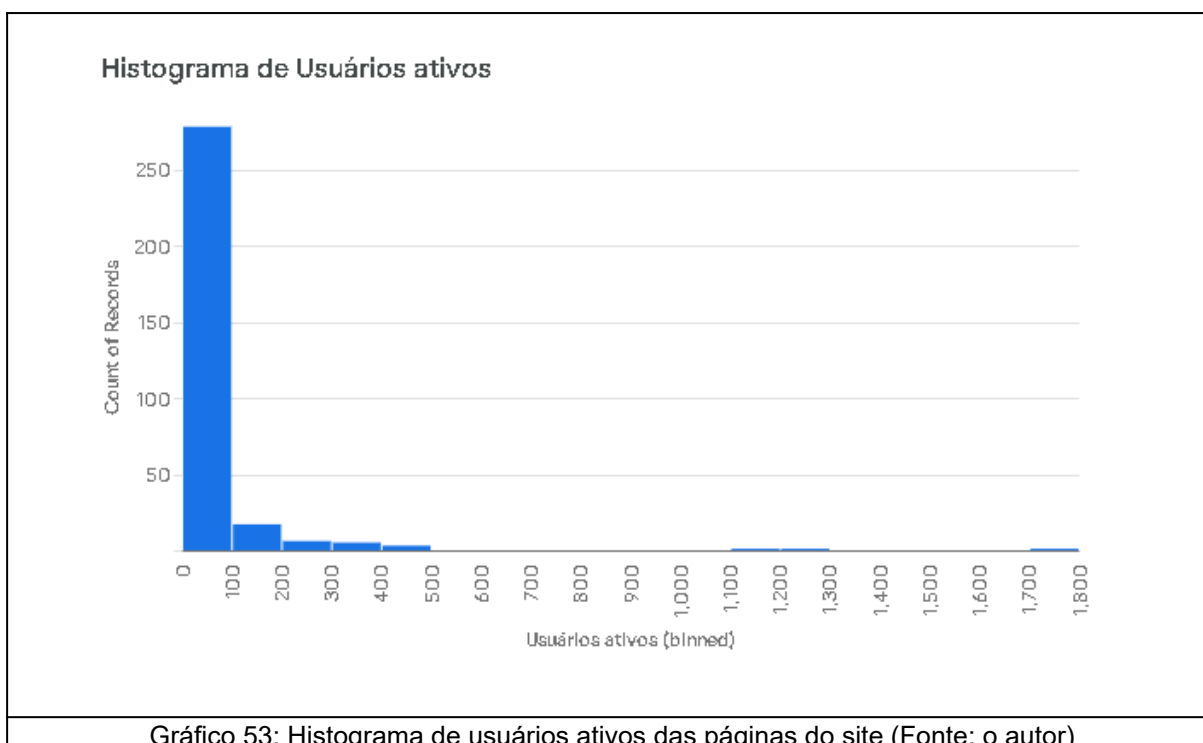
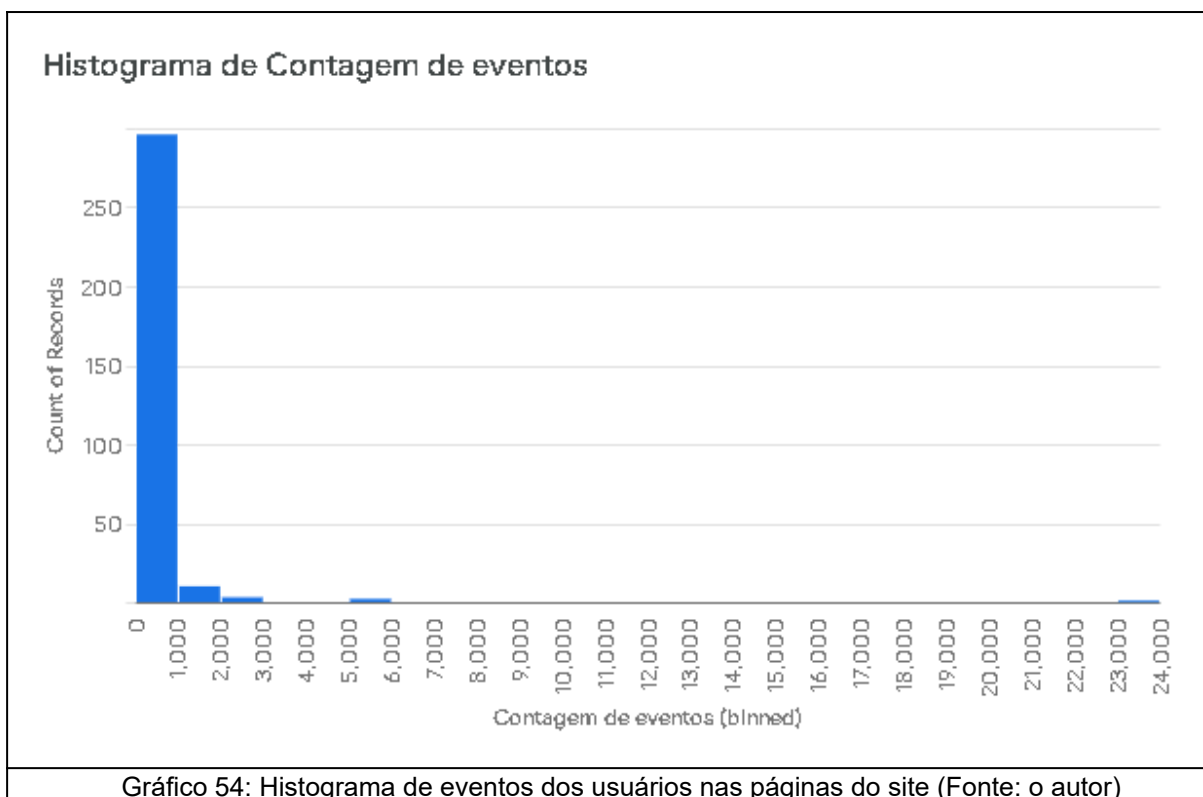


Gráfico 53: Histograma de usuários ativos das páginas do site (Fonte: o autor)

O histograma de Usuários Ativos mostra uma distribuição assimétrica positiva, com a maioria dos títulos de página tendo um número relativamente baixo de usuários ativos e alguns poucos títulos de página com um número muito alto de usuários ativos.

Os títulos de página com baixo número de usuários ativos podem representar uma oportunidade de otimização. É recomendável investigar os motivos pelos quais essas páginas não atraem tantos usuários e realizar melhorias para aumentar o alcance e engajamento.

A análise da distribuição dos usuários ativos pode auxiliar na segmentação de público e na personalização do conteúdo para diferentes grupos de usuários. Ao se deparar com possíveis outliers, como um número de usuários ativos muito acima da média, pode-se compreender o que torna alguns títulos de página tão populares e se é possível obter melhores resultados com outros títulos de página.



A maioria dos eventos está agrupada em um número limitado de títulos de página, que podem ser vistos como "hotspots" de atividade. Esses títulos de página exigem atenção especial para garantir que a experiência do usuário seja otimizada e que os eventos registrados sejam relevantes para os objetivos do canal. Uma baixa contagem inesperada de eventos em certas páginas pode indicar a existência de gargalos ou desafios de usabilidade que dificultam o acesso dos usuários.

Percepções

Os três histogramas exibem uma distribuição semelhante, caracterizada por uma assimetria positiva. Essa assimetria sugere que a maioria dos títulos de página tem um desempenho modesto em termos de visualizações, usuários ativos e contagem de eventos, enquanto um pequeno número de títulos de página apresenta um desempenho excepcionalmente alto.

Essa similaridade na distribuição das três métricas indica uma forte correlação entre elas. Ou seja, títulos de página com muitas visualizações tendem a ter também um grande número de usuários ativos e uma alta contagem de eventos. Essa correlação pode ser explicada pelo fato de que as três métricas estão relacionadas ao engajamento do usuário com o conteúdo do site.

No entanto, é importante observar que a correlação não implica causalidade. Ou seja, o fato de um título de página ter muitas visualizações não significa necessariamente que ele terá também um grande número de usuários ativos e uma alta contagem de eventos. Outros fatores, como o tipo de conteúdo, o design da página e as estratégias de divulgação, também podem influenciar essas métricas.

A análise comparativa dos histogramas permite identificar os títulos de página que se destacam em cada métrica e aqueles que apresentam um desempenho abaixo do esperado. Essa informação pode ser utilizada para otimizar o conteúdo do site, melhorar a experiência do usuário e direcionar as estratégias de difusão de forma mais eficiente.

3.12.4 - Análise Descritiva por Idioma

A tabela “Visualizações por Título da página e usuário ativo” possui 5 colunas, cujas variáveis numéricas se correlacionam conforme o grau de influência entre si e outras colunas.

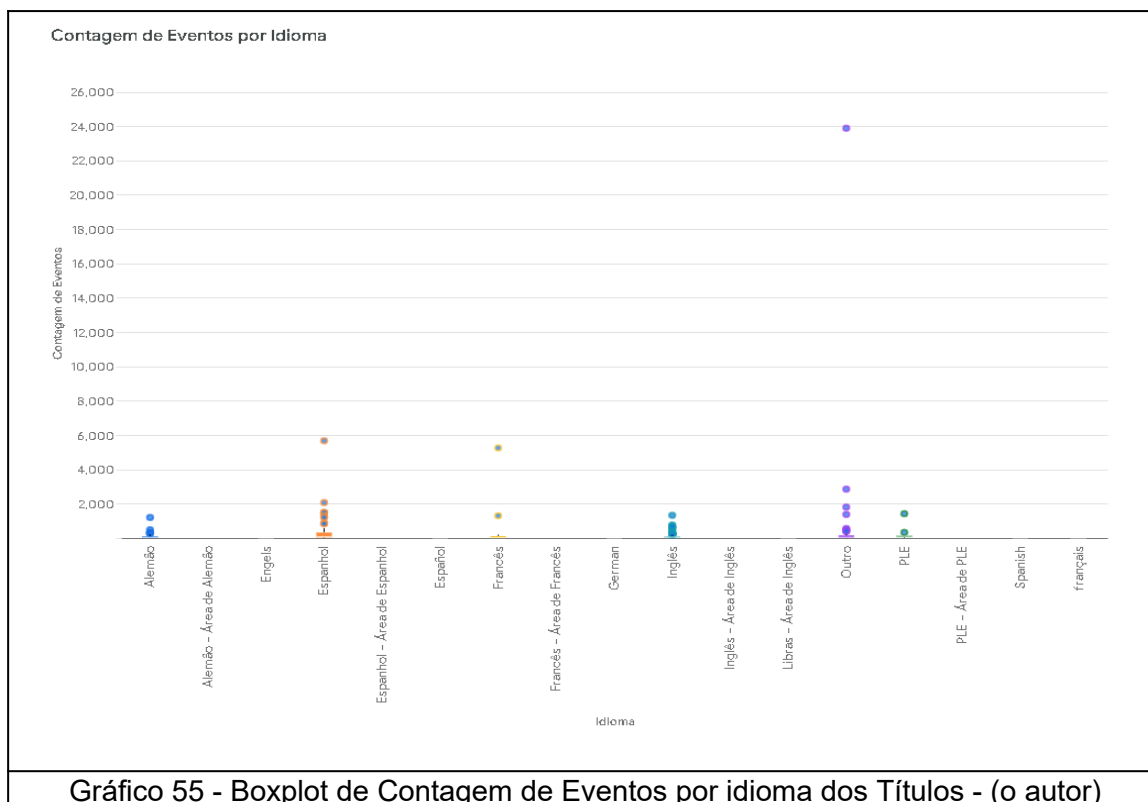
Algumas estatísticas descritivas mostram:

- A média de Visualizações é de 94 e o número médio de Usuários Ativos é de 53.
- Cada página do site alcança, em média, 1,3 Visualizações por Usuário Ativo.
- A contagem média de eventos é de 279.
- O desvio padrão para Visualizações, Usuários ativos e Contagem de eventos é alto, indicando ampla dispersão dos dados em torno da média. Os valores máximo e mínimo confirmam essa dispersão, evidenciando uma ampla gama.
- As medianas para Visualizações, Usuários ativos, Visualizações por usuário ativo e Contagem de eventos são inferiores às suas médias correspondentes, o que indica que as distribuições estão enviesadas positivamente (para a direita).
- “Visualizações”, “Usuários ativos” e “Contagem de eventos” estão altamente e positivamente correlacionados entre si. Isso significa que quando um aumenta, os outros também aumentam, e vice-versa.
- Visualizações por usuário ativo tem uma correlação moderada com Visualizações e Contagem de eventos, e uma correlação baixa com Usuários ativos.

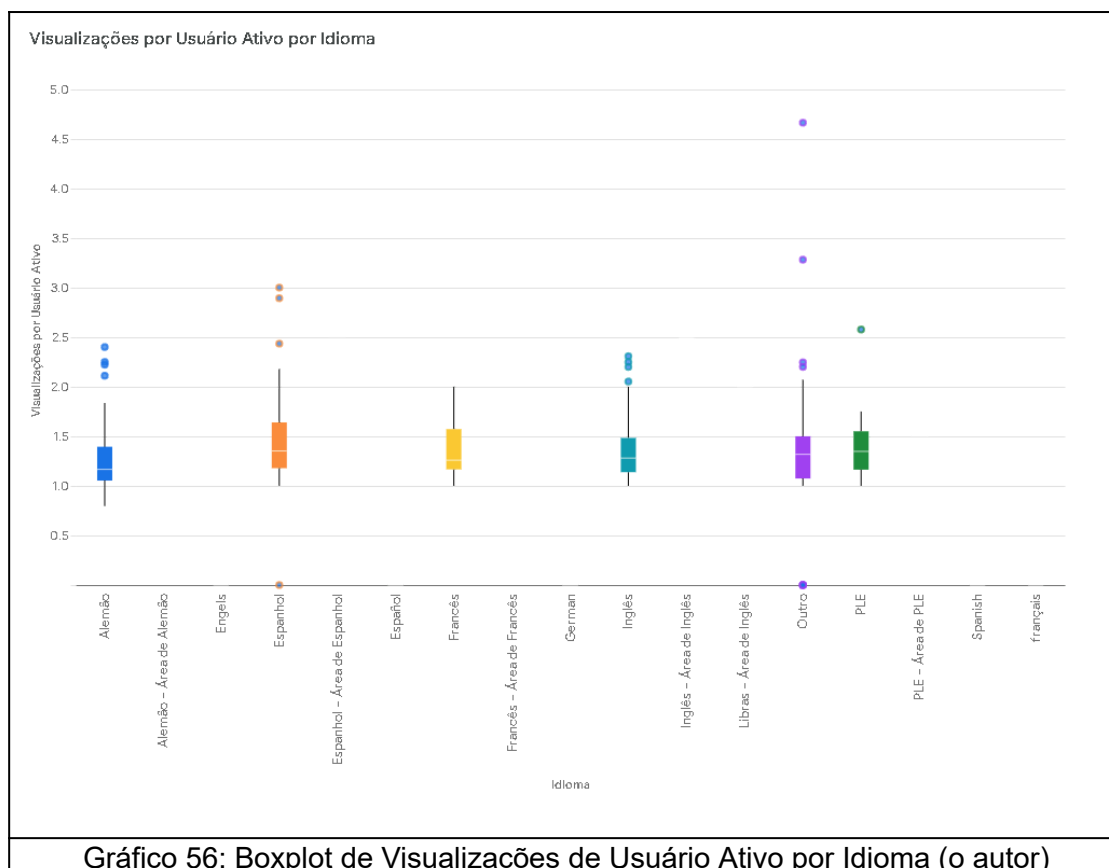
Colunas	Tipo de correlação		
Variáveis	FORTE	MODERADA	BAIXA
Visualizações,			
Usuários ativos			
Visualizações por usuário ativo			
Contagem de eventos			
Quadro 11 - Tipos de correlação das variáveis por idioma - o autor			

Quando referenciados pela coluna “Título da página”, encontramos correlações entre os títulos que podem ser classificados e analisados quanto ao idioma e palavras chaves empregadas e obter percepções que são interessantes para um site relacionado ao ensino de idiomas.

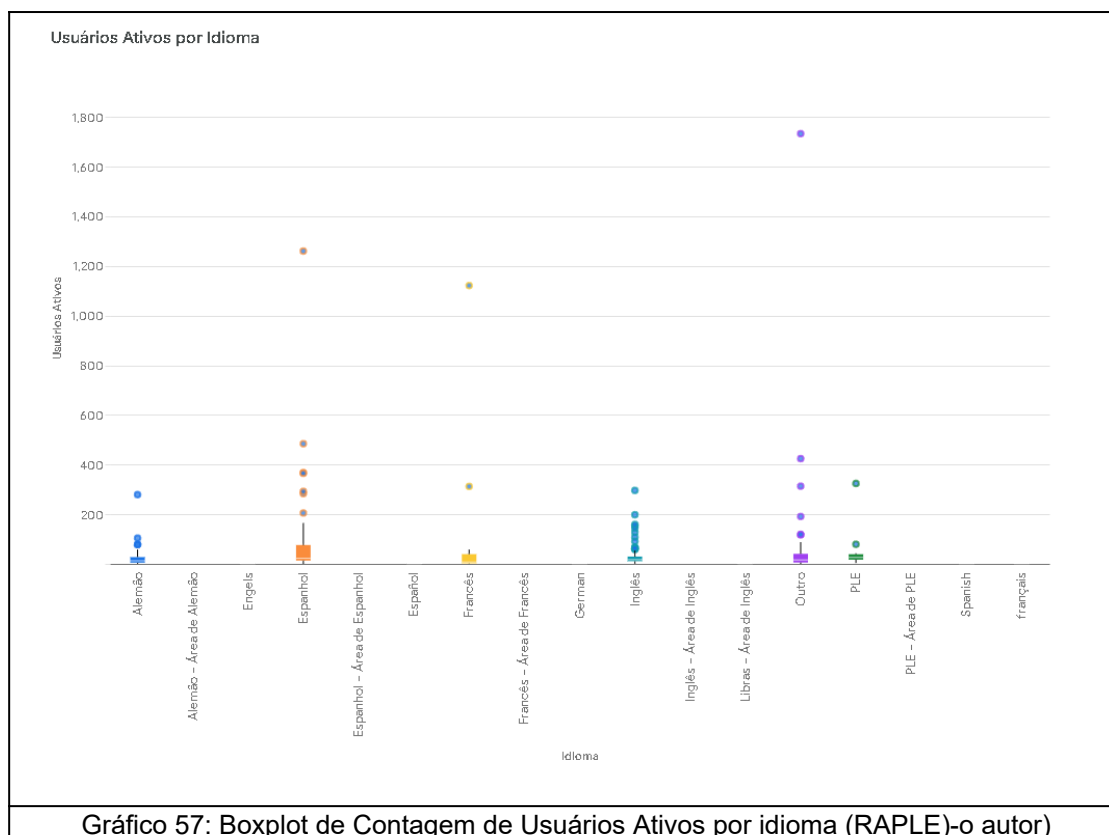
Analisamos e calculamos a distribuição de cada métrica para cada idioma e propomos as seguintes inferências:



- As páginas da Web em alemão têm, em média, 29 visualizações, com desvio padrão de 44, o que indica que o número de visualizações varia de página para página.
- As páginas da Web em espanhol têm, em média, 122 visualizações, com um desvio padrão maior de 247, o que mostra uma variação ainda maior no número de visualizações.
- As páginas da Web em francês têm, em média, 104 visualizações, com um desvio padrão de 319 visualizações, o que também indica uma grande variação no número de visualizações.



- As páginas da Web em inglês têm, em média, 40 visualizações, com um desvio padrão de 55, o que indica variações moderadas no número de visualizações.
- As páginas da Web em 'Outro' têm, em média, 288 visualizações, com um enorme desvio padrão de 1.295 visualizações, o que indica uma variação muito grande no número de visualizações.
- Páginas da Web em espanhol têm a maior média de usuários ativos, com 94 usuários por página.
- Páginas da Web em 'Outro' têm a segunda maior média de usuários ativos, com 90 usuários por página.
- Páginas da Web em francês têm a terceira maior média de usuários ativos, com 79 usuários por página.
- Páginas da Web em inglês têm a quarta maior média de usuários ativos, com 30 usuários por página.
- Páginas da Web em alemão têm a menor média de usuários ativos, com 23 usuários por página.
- As páginas da Web em inglês têm a maior média de visualizações por usuário ativo, com 1,34.
- As páginas da Web em alemão, espanhol e francês têm médias de visualizações por usuário ativo semelhantes, variando de 1,29 a 1,48.
- As páginas da Web em "Outros" têm uma média de 1,37 visualizações por usuário ativo.

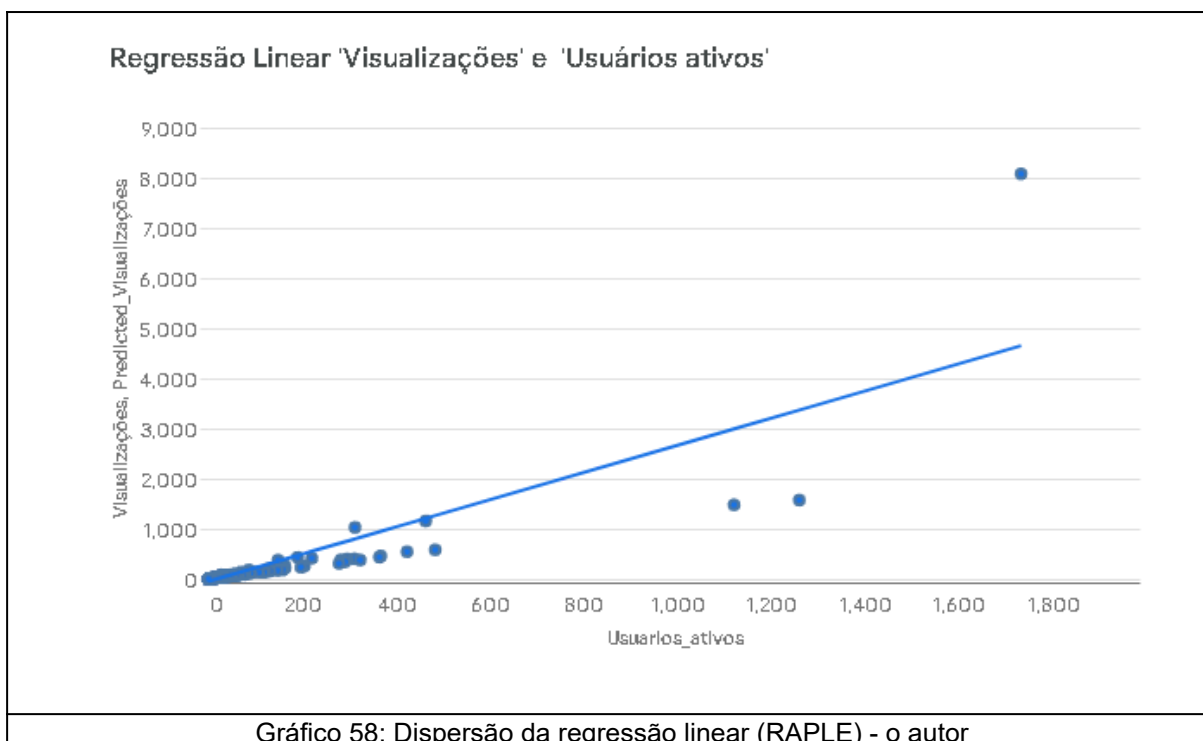


A contagem média de eventos para páginas da Web em diferentes idiomas segue um padrão semelhante ao número de visualizações e usuários ativos. Seguem os gráficos que orientam as inferências acima:

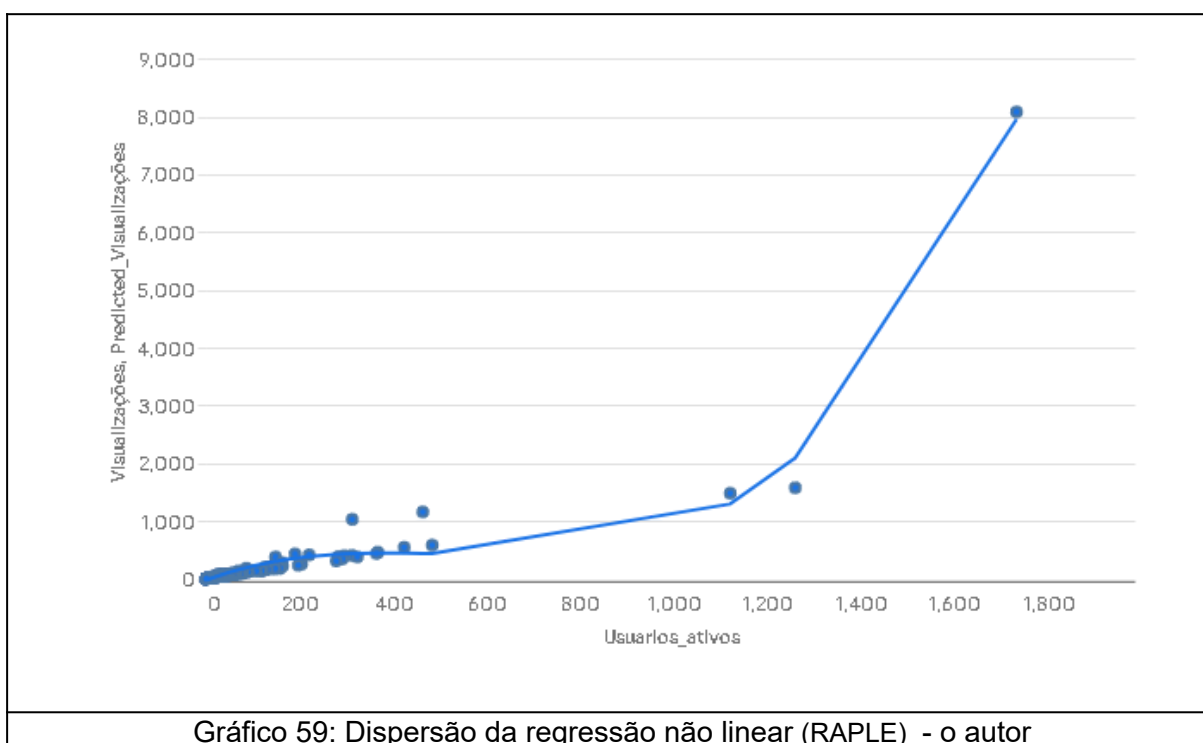
Análise de regressão

A análise de regressão mostra que há uma relação positiva e estatisticamente significativa entre o número de usuários ativos e o número de visualizações. O R-quadrado de 0,722 indica que o modelo explica 72,2% da variância nas visualizações.

O coeficiente para Usuários Ativos é de 2,7133, o que significa que, para cada aumento de um usuário ativo, o número de visualizações aumenta em 2,7133, em média. O valor de p para Usuários Ativos é 0,000, o que é menor que 0,05, indicando que a relação é estatisticamente significativa.



Com o objetivo de avaliar se uma regressão não linear pode oferecer melhores perspectivas, comparamos o desempenho de diversos modelos de regressão não linear com um modelo de regressão linear previamente ajustado e visualizamos o modelo com o maior R-quadrado ajustado.



O modelo de regressão polinomial de grau 3, com um R-quadrado ajustado de 0,978, é o que melhor se ajusta aos dados. Isso significa que o modelo explica 97,8% da variância nas visualizações. Quando comparado, o modelo polinomial de grau 3 teve um desempenho significativamente melhor do que o modelo de regressão linear, que teve um R-quadrado ajustado de 0,721.

É importante destacar que o modelo polinomial de grau 3 apresenta um elevado número de correlações, o que pode indicar a presença de multicolinearidade. Uma alta correlação entre as variáveis independentes presentes no modelo pode afetar a precisão das estimativas dos coeficientes.

Embora o modelo polinomial de grau 3 apresente um ajuste superior aos dados em comparação com o modelo de regressão linear, é crucial reconhecer as potenciais restrições decorrentes da multicolinearidade (FREUND, WILSON, SA, 2006)

Método de coleta de dados	Quando o pesquisador coleta amostras somente no subespaço da região dos regressores definidos.
Restrições no modelo ou população	Escolhas de variáveis ou séries temporais decorrentes da restrição do modelo ou da população na coleta.
Escolha do modelo	As variáveis escolhidas para compor o modelo podem ser linearmente dependentes, causando assim multicolinearidade.
Modelo com excesso de termos	O modelo possui mais variáveis regressoras do que observações
Quadro 12: Fonte primárias de multicolinearidade (MONTGOMERY; PECK; VINING, 2012)	

Para avaliar a colinearidade na regressão, usamos o fator de inflação da variância (VIF) que é uma medida da quantidade de multicolinearidade em um conjunto de múltiplas variáveis de regressão.

Como a Multicolinearidade ocorre quando duas ou mais variáveis preditoras em um modelo de regressão estão altamente correlacionadas, podemos encontrar estimativas de coeficientes instáveis, intervalos de confiança muito amplos ou valores de p inflados. O VIF para uma variável preditora é calculado regredindo essa variável em todas as outras variáveis preditoras no modelo. O R-quadrado dessa regressão é então usado para calcular o VIF da seguinte forma:

$$VIF = 1 / (1 - R\text{-quadrado})$$

Um valor de VIF de 1 indica que não há multicolinearidade. Valores de VIF maiores que 1 indicam multicolinearidade crescente. Um valor de VIF de 5 ou 10 é frequentemente usado como limite para indicar multicolinearidade problemática. No caso do modelo de regressão polinomial de grau 3, os valores de VIF para as variáveis preditoras indicaram haver significativa multicolinearidade no modelo.

Usuários Ativos: 14,81	Usuários Ativos**2: 130,73	Usuários Ativos**3: 74,00
Tabela 7: Valores de Aplicação do fator de inflação da variância (RAPLE) (VIF)		

1. O R-quadrado dessa regressão foi de 0,933.
2. O VIF foi calculado como $1 / (1 - 0,933) = 14,81$.

Os valores de VIF para as outras variáveis preditoras foram calculados de forma semelhante. A análise de regressão demonstrou as vantagens e deficiências dos métodos dos mínimos quadrados no modelo linear e da multicolinearidade não linear, esclarecendo que ambos os métodos utilizam uma fórmula de ajuste, minimizando a soma dos quadrados das diferenças entre os valores observados e a predição do modelo, consequentemente, a variância dos erros quadráticos. Também é importante verificar se as condições do teorema de Gauss-Markov são atendidas para garantir a propriedade de estimadores não tendenciosos ⁵².

Modelo	Equação
Regressão Linear	$y = \beta_0 + \beta_1 x + \epsilon$
Regressão Polinomial	$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_n x^n + \epsilon$
Quadro 13: Equações dos modelos de regressão (RAPLE) - o autor	

A regressão polinomial, embora possa proporcionar um ajuste mais preciso para dados não-lineares, pode ser propensa ao sobreajuste, caso o grau do polinômio seja muito elevado em relação à quantidade de dados disponíveis. A interpretação dos coeficientes de um modelo polinomial pode apresentar complexidade crescente em comparação com um modelo linear (ZEVIANI, RIBEIRO

⁵² <https://mathworld.wolfram.com/LeastSquaresFittingPolynomial.html>

JR, BONAT, 2016). Esse aumento de complexidade é particularmente evidente à medida que o grau do polinômio se eleva⁵³.

Assim como nos modelos lineares, o objetivo é identificar e estabelecer a relação entre variáveis explicativas e resposta. Entretanto, enquanto os modelos lineares definem, em geral, relações empíricas, os modelos não-lineares são, em grande parte das vezes, motivados pelo conhecimento do tipo de relação entre as variáveis (IBIDEM).

A análise dos dados e das regressões forneceu percepções valiosas sobre o comportamento do usuário no site. Várias métricas foram examinadas, incluindo visualizações, usuários ativos, visualizações por usuário ativo e contagens de eventos. Essas percepções podem ser usadas para melhorar a experiência do usuário e o desempenho geral do site em relação a algumas variáveis estudadas.

- As páginas da web em inglês tiveram a maior média de visualizações por usuário ativo, com 1,34.
- As páginas da web em "Outro" tiveram a maior média de visualizações, com 288.
- As páginas da web em francês tiveram a segunda maior média de visualizações, com 104.
- As páginas da web em espanhol tiveram a terceira maior média de visualizações, com 122.
- As páginas da web em alemão tiveram a menor média de visualizações, com 29.
- As páginas da web em "Outro" também tiveram o maior desvio padrão, com 1.295 visualizações, o que indica uma grande variação no número de visualizações para esta categoria.

O número de visualizações está fortemente associado ao número de usuários ativos. Em média, cada aumento de um usuário ativo corresponde a um aumento de 2,7133 no número de visualizações.

O modelo de regressão não linear polinomial de grau 3 foi o que melhor se ajustou aos dados, no entanto, o mesmo tem um alto número de correlações, sugerindo multicolinearidade. O fator de inflação da variância (VIF) foi aplicado para avaliar a colinearidade na regressão. Os valores de VIF para todas as três variáveis preditoras foram altos, indicando multicolinearidade.

Percepções

O engajamento do usuário é fundamental para o sucesso do site ou canal, como demonstrado pela forte correlação entre usuários ativos e visualizações.

⁵³ <https://www.ime.unicamp.br/~cnaber/cursomodelosnaolinearesR.pdf>

Páginas com mais usuários ativos, geralmente apresentam um maior número de visualizações, enfatizando que o envolvimento dos usuários os faz retornar.

O modelo de regressão polinomial de 3º grau indica que a relação entre usuários ativos e visualizações não é estritamente linear. Pode haver um ponto de rendimentos decrescentes, onde aumentar os usuários ativos pode não resultar em um aumento proporcional nas visualizações. Isso pode ser devido a vários fatores, como a capacidade de carga da página da web, o tipo de conteúdo ou o público-alvo.

A complexidade das relações entre as variáveis preditoras é destacada pela presença de multicolinearidade no modelo de regressão. Embora o modelo polinomial de 3º grau apresente um bom ajuste, a multicolinearidade indica que as variáveis preditoras não são independentes. Isso pode afetar a interpretação dos coeficientes de regressão, exigindo cautela na análise e interpretação dos resultados.

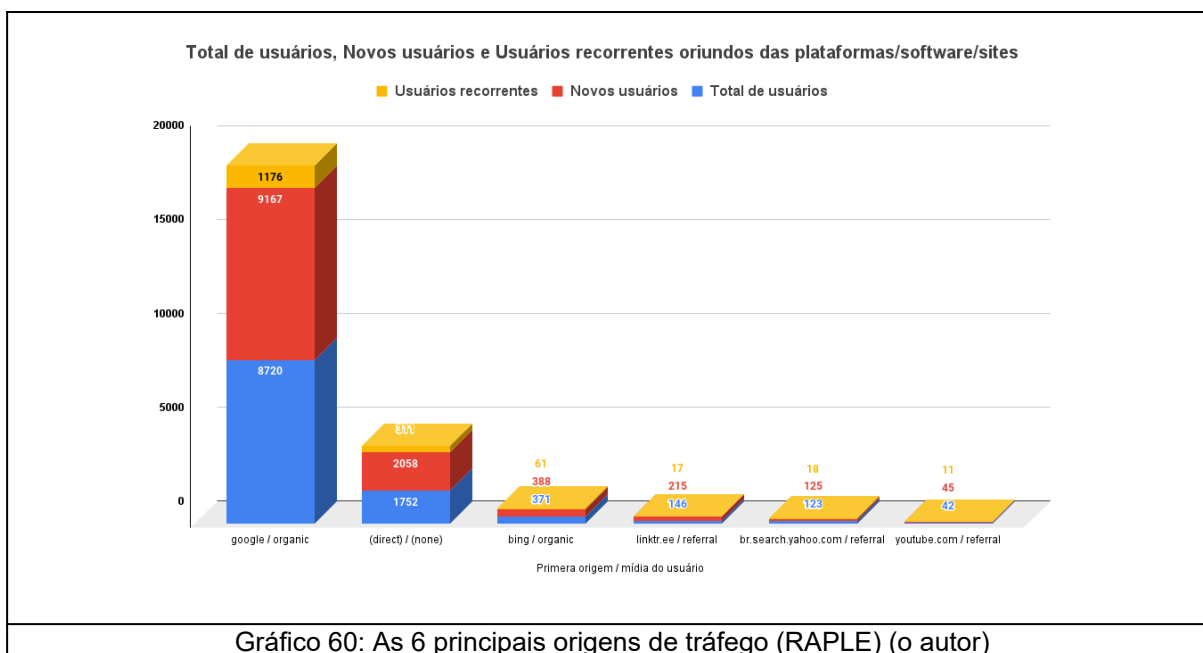
A interpretação pode mostrar que os dados somente se tornam conhecimento se houver um tratamento adequado para transformá-los em informação antes da análise. As correlações são uma importante fonte de conhecimento quando coletadas e analisadas quanti e qualitativamente, como feito na verificação de distribuição de visualizações por língua.

Seguiremos agregando novas variáveis ao modelo e aplicando métodos estatísticos computacionais que possam oferecer novas formas de obter conhecimento através das análises multivariadas.

3.12.5 - Origens do Tráfego do site da RAPLE

Montamos um importante *dataset*⁵⁴ sobre a origem do tráfego via Analytics do site da RAPLE e realizamos uma série de análises sobre as dinâmicas de acesso, obtivemos indicativos de perfil dos usuários e hábitos que podem revelar preferências não previstas no modelo de análise. O gráfico mostra as 6 maiores origens, que concentram 98,63% do total e 98,40% dos usuários recorrentes

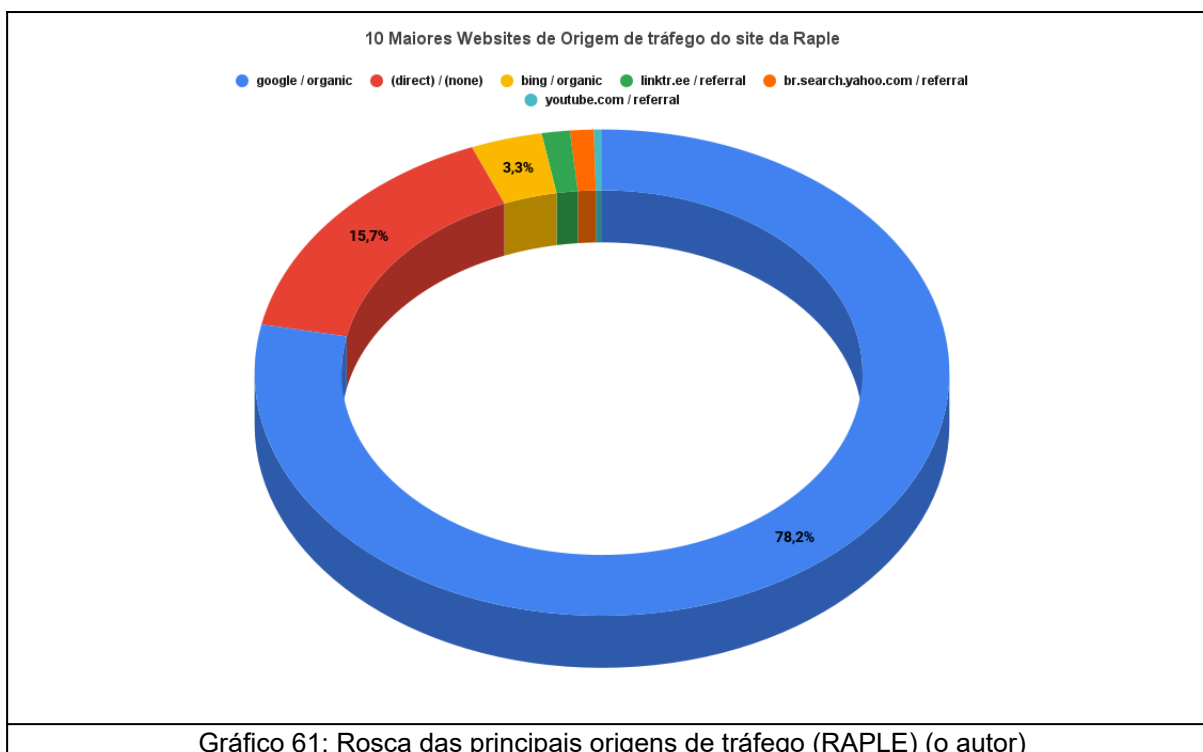
⁵⁴Um dataset constitui um conjunto estruturado de dados organizados e armazenados conjuntamente, viabilizando sua utilização em tarefas de análise ou processamento.



A predominância de sites de busca orgânica, como o Google, Bing e Yahoo, indicam haver boa estratégia de Otimização para Motores de Busca - SEO (Search Engine Optimization) pelo Webmaster do site. A SEO é um conjunto de técnicas e estratégias que visam melhorar a posição de um site nos resultados de busca.

Também é recomendável investigar outras fontes de tráfego, além do Google, para compreender o comportamento dos usuários e suas preferências, pois mesmo que ele represente cerca de 78% do tráfego, os novos recursos de IA nos buscadores tendem a agregar novas origens e pode mudar a predominância esmagadora dele.

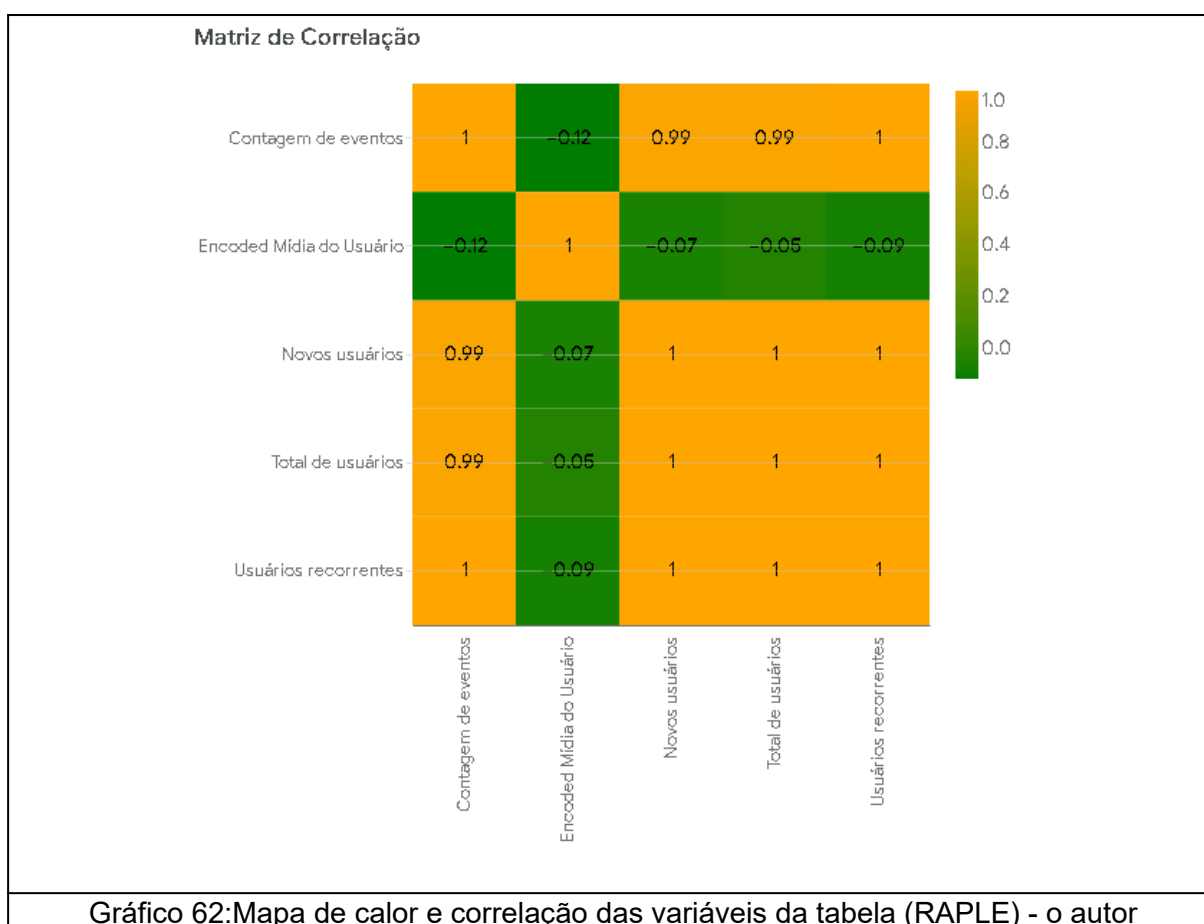
Outro fator importante é que a segunda maior fonte é o tráfego direto, indicando que as pessoas já conhecem e digitam o endereço do site e que a sigla do nome da rede (RAPLE) é reconhecida e bem fixada na mente de mais de 2000 usuários. O Google também lidera na contagem de eventos, seguido por tráfego direto, o que sugere maior engajamento nesses canais, mas a presença de outras fontes mostra diversificação na aquisição de usuários.



Analizamos a relação entre as colunas numéricas com dados quantitativos dos usuários e a coluna Primeira origem / mídia do usuário, usando uma codificação numérica para as origens de mídia e realizando uma análise de correlação. Esta codificação pode simplificar a análise e visualização dos dados, especialmente quando há muitas categorias diferentes na coluna original.

Existem métodos diferentes de codificação, mas a de rótulos (label encoding) com um número inteiro para cada categoria na coluna "Primeira origem / mídia do usuário" foi aplicada para permitir a análise de correlação e regressão. Isso permitiu avaliar a relação entre a origem de mídia e outras variáveis numéricas, como o número de usuários e a contagem de eventos.

A coluna Primeira origem / mídia do usuário, ao ser codificada para Encoded Mídia do Usuário, exibiu uma correlação negativa com as outras colunas. Isso indica que a plataforma de origem possui uma influência menor no tráfego e no engajamento quando comparada a outros fatores, como o tipo de usuário e a contagem de eventos.



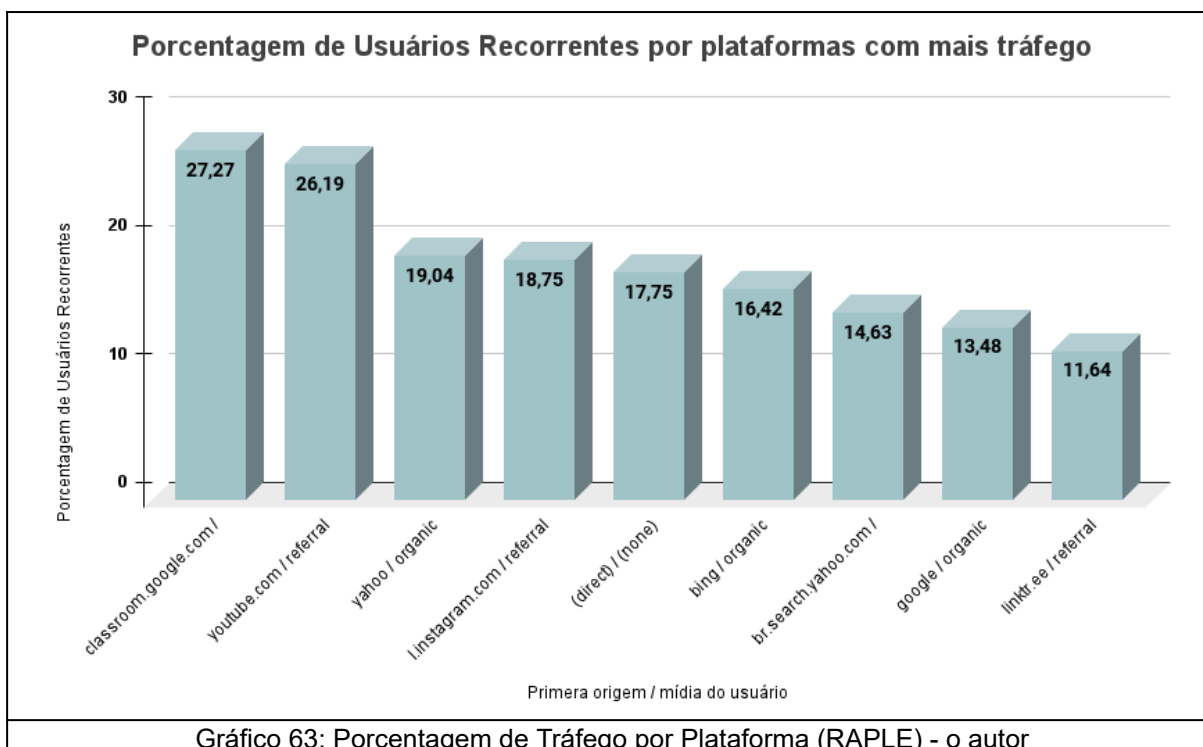
Os dados revelam que o tipo de usuário e a quantidade de eventos são mais significativos para o tráfego e o engajamento do que a plataforma de origem, porém a subjetividade desta afirmação nos motiva a continuar buscando dados, informações e outros fatores que podem influenciar o tráfego e o engajamento.

Buscamos analisar os dados pela proporcionalidade das origens de tráfego e surpreendentemente obtivemos percepções diferenciadas dos mesmos dados.

3.12.6 - Proporcionalidade de Recorrência dos Usuários

Para analisar a proporcionalidade entre “Usuários recorrentes” e “Total de usuários”, calculamos a porcentagem de usuários recorrentes em relação ao total deles, que representa o retorno de usuários ao site.

Calculando a média da “Porcentagem de Usuários Recorrentes” para cada valor de “Primeira origem / média do usuário”, plotamos um gráfico de barras para visualizar a proporção de usuários recorrentes para cada origem de tráfego.



Observando o gráfico com os 9 maiores geradores de tráfego para o site da RAPLE, podemos inferir algumas reflexões:

- **Plataformas Sociais com Altos Índices de Retenção:** Origens como 'classroom.google.com', 'youtube.com / referral', 'google / organic' e 'linkedin.com / referral' apresentam as maiores taxas de retorno de usuários. Esse resultado sugere que essas plataformas direcionam um público mais propenso a retornar ao site, indicando maior nível de engajamento e interesse no conteúdo oferecido.
- **Origens com Baixa Retenção:** Alguns canais de origem, como 'm.facebook.com / referral', 'l.instagram.com / referral' e 'facebook/referral', demonstram baixas taxas de usuários recorrentes. Isso indica que o público proveniente dessas plataformas tende a não retornar ao site e os motivos podem ser descobertos e analisados com os dados..

A retenção de usuários em um site pode variar de acordo com a origem do tráfego. Origens com alta retenção tendem a direcionar usuários que encontram conteúdo relevante e envolvente no site, o que os incentiva a retornar.

Por outro lado, origens com baixa retenção podem trazer usuários que acessam o site por curiosidade ou redirecionamentos, mas não demonstram

interesse em retornar, pois o conteúdo pode não atender às suas expectativas ou necessidades.

Algumas métricas fundamentais também podem gerar inferências valiosas quando articuladas com os resultados totalizados. A tabela mostra as estatísticas descritivas (média, mediana e desvio padrão) para as colunas:

	Total de Usuários	Novos Usuários	Usuários Recorrentes
Média	1123.1	1210.2	160.7
Mediana	82.5	85	14
Desvio Padrão	2721.76	2864.36	369.078

Tabela 8: Análise descritiva sobre a proporção dos tipos de usuários (o autor)

Percepções:

Média

- A média de “Novos usuários” é 1210.2, enquanto a de “Total de usuários” é 1123.1. Isso sugere que há mais novos usuários do que o total deles, o que indica uma perda de 7,19% de usuários ao longo do período de tempo analisado.
- A média de Usuários recorrentes, de 160.7, é significativamente inferior às outras médias, indicando que a maioria dos usuários não acessa o site frequentemente.

Mediana:

- As medianas de “Total de usuários” e “Novos usuários” são 82.5 e 85, respectivamente, bem menores que suas médias. Isso sugere que existem alguns valores muito altos nessas colunas, distorcendo a média.
- A mediana de “Usuários recorrentes” é 14, o que confirma que a maioria dos usuários não retorna ao site com frequência.

Desvio Padrão:

- Os desvios padrão são altos para todas as colunas, especialmente para “Total de usuários” e “Novos usuários”. Isso indica uma grande variação nos dados, com valores dispersos em torno da média.

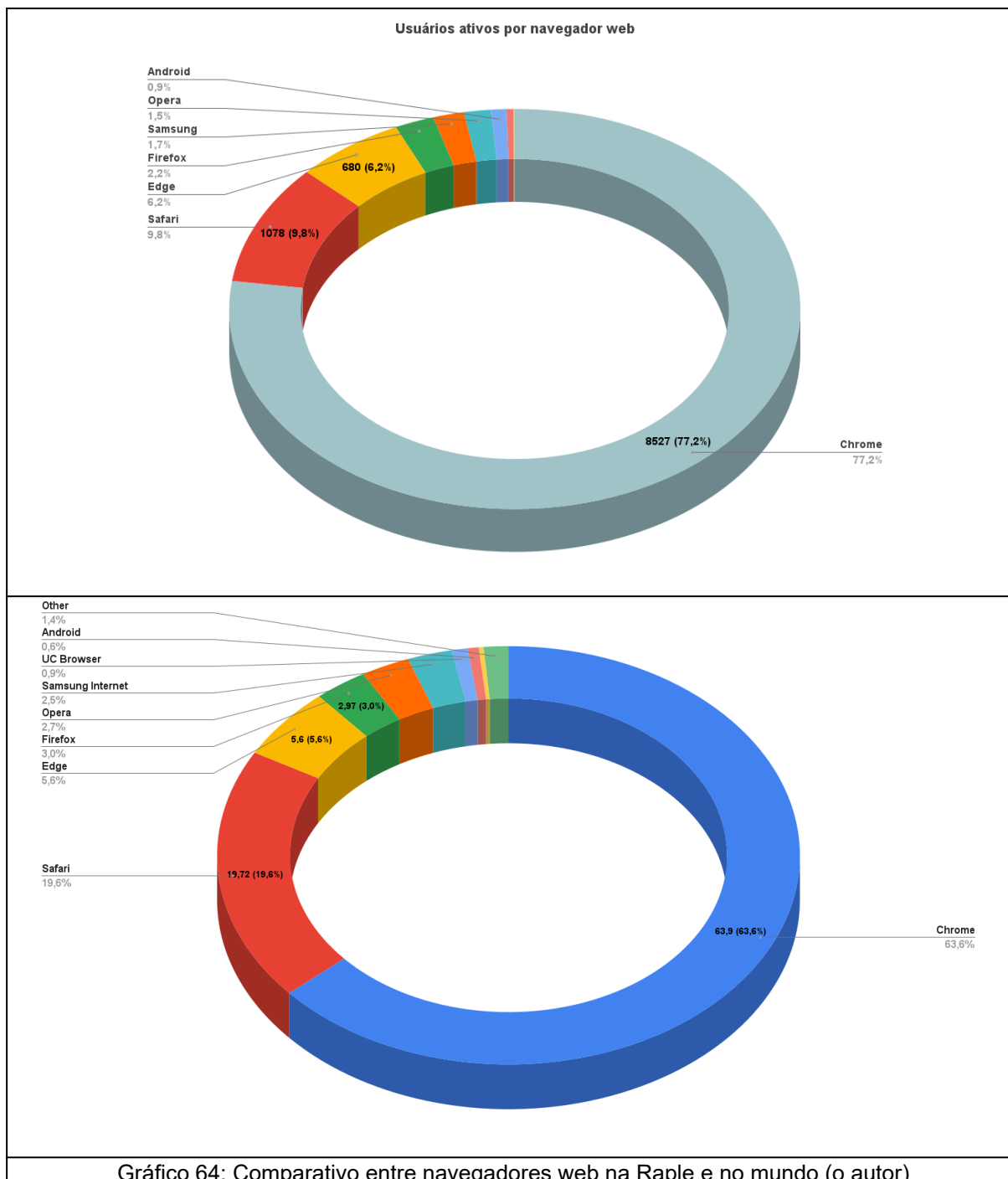
Buscando conhecer e compreender o impacto dessas variáveis, continuamos a agregar novos *data frames* para obter informações e gerar conhecimento sobre o público do canal.

3.12.7 - Usuários ativos por navegador web

Os usuários ativos da Raple utilizam os navegadores web quase nas mesmas proporções que o uso global entre Janeiro de 2023 e Janeiro de 2024⁵⁵, como

⁵⁵ Participação de mercado de navegadores de desktop no mundo de janeiro de 2023 a janeiro de 2024. Dados: StatCounter

demonstrado abaixo em comparação com dados da StatCounter, demonstrando similaridade nas distribuições percentuais, havendo apenas uma alternância entre os navegadores Opera e Samsung, mas com diferenças ínfimas.



Algumas percepções podem ser úteis para entender o cenário atual do mercado de navegadores e as preferências dos usuários da RAPLE:

- **Concentração de mercado:** O mercado de navegadores é altamente concentrado, com o Chrome dominando mais de três quartos do total de usuários.

- **Preferência por navegadores móveis:** A presença de navegadores como Safari (in-app), Opera Mini e Android Webview na lista sugere que muitos usuários estão acessando a RAPLE por dispositivos móveis.
- **Diversidade limitada:** A maioria dos usuários concentra-se em poucos navegadores, indicando pouca diversidade no opções.

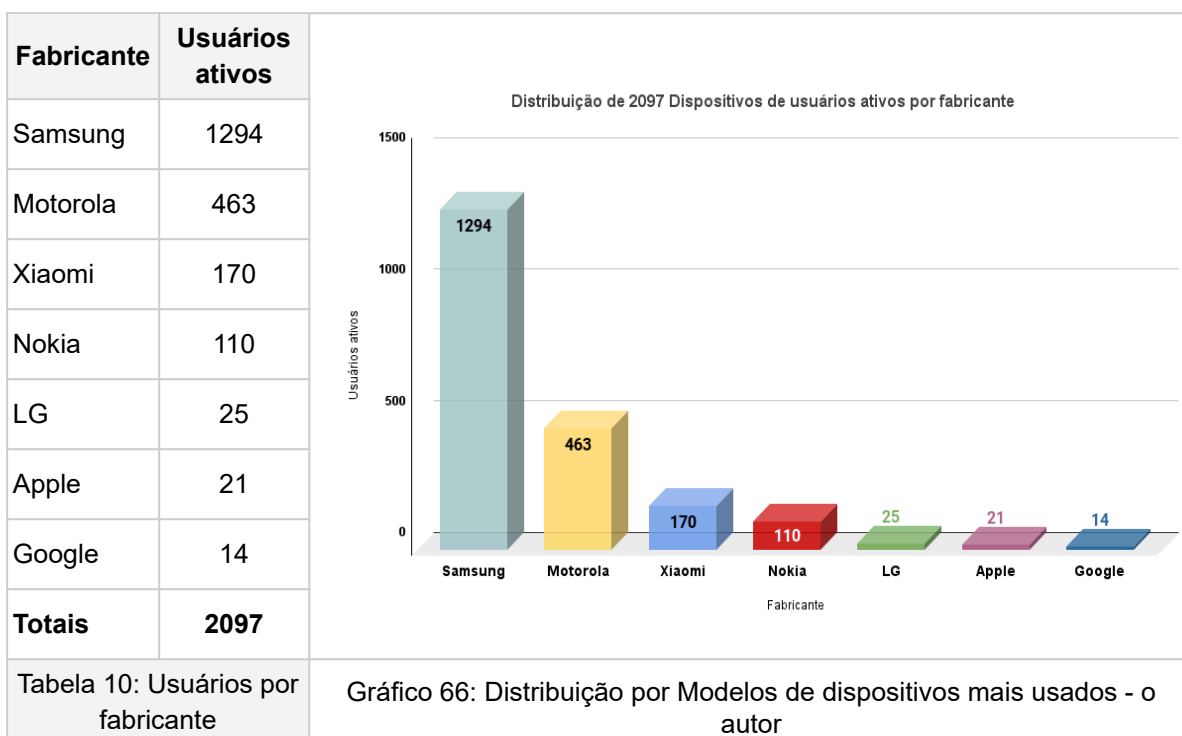
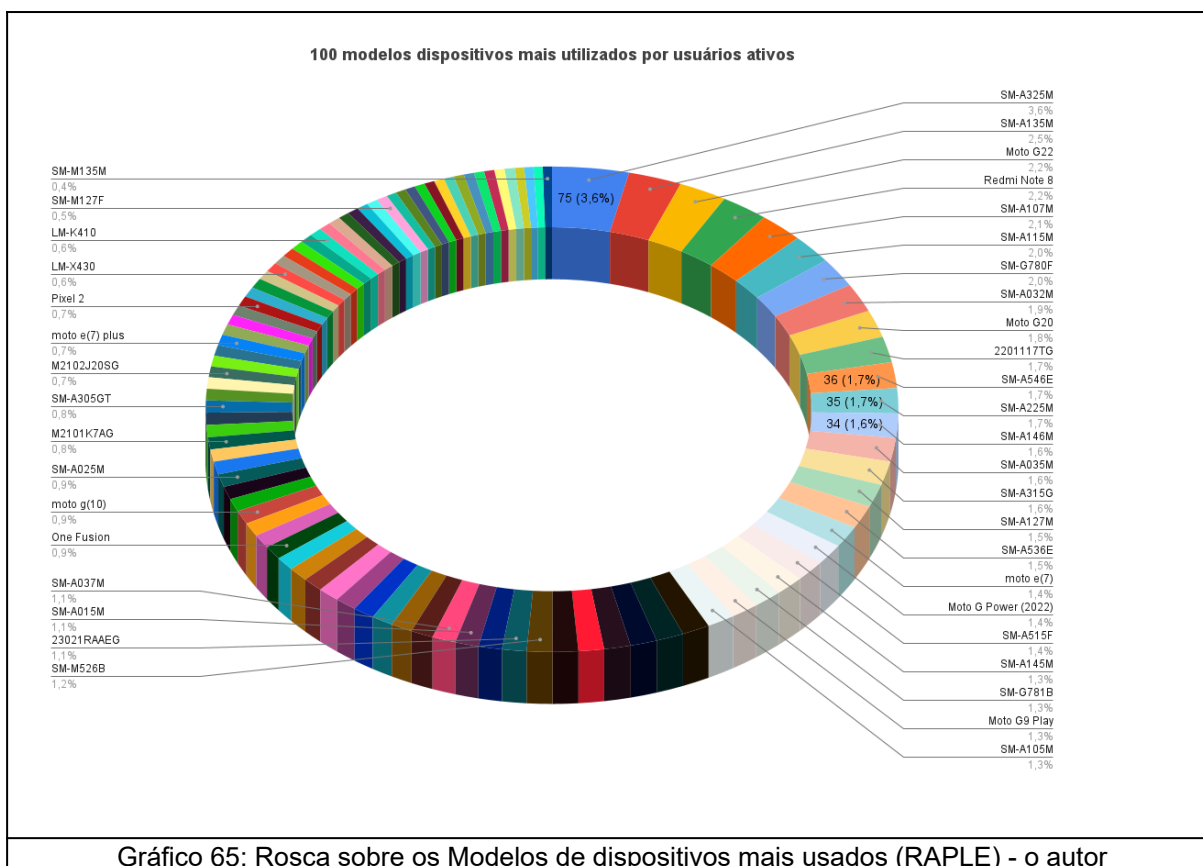
Navegador	Índice RAPLE	Navegador	Índice global
Chrome	77,24	Chrome	63,9
Safari	9,76	Safari	19,72
Edge	6,16	Edge	5,60
Firefox	2,23	Firefox	2,97
Samsung Internet	1,75	Opera	2,72
Opera	1,52	Samsung Internet	2,49
Android Webview	0,87	UC Browser	0,89
Safari (in-app)	0,40	Android	0,61
Opera Mini	0,04	Yandex Browser	0,27
Mozilla Compatible Agent	0,02	Outros	1,38
Android Browser	0,01		
Yandex Browser	0,01		

Tabela 9: Índice do uso de navegadores no site das RAPLE (o autor)

Também avaliamos quais as marcas e modelos de dispositivos dos usuários ativos que mais acessam as páginas da RAPLE em uma amostra de 2097 aparelhos. Há uma clara e inegável predominância de dispositivos da Samsung com 1.294 dispositivos (61,70%), seguidos de Motorola com 463 deles (22,07%), Xiaomi com 170 (8,10%), Nokia com 110 aparelhos (5,24%), LG com 25 aparelhos (1,19%), Apple com 21 deles (1%) e Google com 14 aparelhos (0,66%).

A amostra pode contemplar outros modelos, marcas e fabricantes, porém optamos por trabalhar com uma amostra dos 100 aparelhos com mais usuários ativos e classificamos os mesmos pelos seus fabricantes, baseados no código do modelo que acessou o site da RAPLE, sendo descartados da lista, modelos com menos de 10 acessos ativos ou que não apresentavam o modelo do dispositivo.

No ajuste de dados ele foram suprimidos, embora representem mais de $\frac{2}{3}$ da distribuição de usuários ativos, mas não seria possível analisar os dados adequadamente pelo fato deles fazerem menção somente ao navegador e não ao modelo do aparelho. Dos 10.921 usuários ativos encontrados, somente 3.350 faziam menção ao modelo do aparelho, dos quais foi retirada a amostra de 2097 que originaram a análise e gráfico.



Segue a tabela dos registros excluídos da amostra que podem ser de diversos fabricantes e não possuem nenhuma referência de modelo.

Modelo do dispositivo	Usuários ativos
-----------------------	-----------------

Chrome	4954
iPhone	1078
Edge	669
Safari	310
Firefox	232
ChromeBook	176
Opera	152
Tabela 11: registros excluídos da amostra de distribuição (RAPLE) - o autor	

Há relevância em conhecer estes dados para avaliar a necessidade de fornecer um site responsivo ou mesmo um site Mobile Friendly pela RAPLE para atender as demandas de um público que acessa cada vez mais conteúdos pelo smartphone.

Percepções

Investigar a preferência por diferentes fabricantes em segmentos específicos de mercado (tal como faixa etária, origem e país) pode revelar informações mais detalhadas sobre o comportamento do consumidor. A distribuição de dispositivos entre os fabricantes é desigual, com a Samsung e a Motorola concentrando a maior parte do mercado. A média percentual de dispositivos por fabricante é de 14,29%, o que indica que há uma concentração significativa de dispositivos entre esses dois fabricantes.

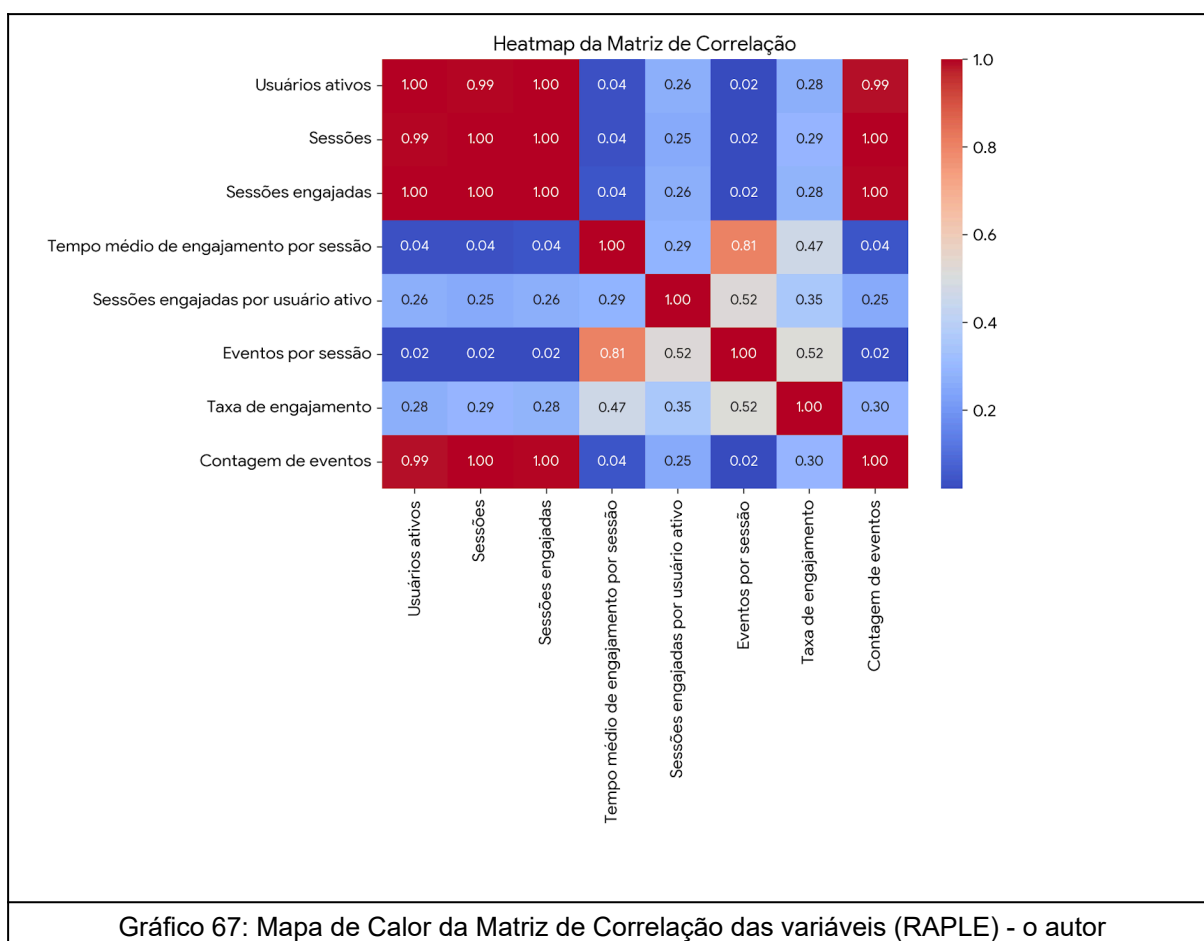
- **Dominância da Samsung:** A Samsung é a fabricante dominante, com 61,71% dos dispositivos, o que corresponde a 1294 usuários. Isso indica uma forte preferência por essa marca do mercado.
- **Motorola em segundo:** A Motorola ocupa o segundo lugar com 22,08% (463 usuários), mostrando uma presença significativa, mas bem menor que a Samsung.
- **Demais fabricantes: Xiaomi, Nokia, LG, Apple e Google** representam fatias menores do mercado, com percentuais que variam de 8,11% (Xiaomi) a 0,67% (Google).
- **Concentração de mercado:** O mercado é concentrado, com Samsung e Motorola juntas detendo mais de 80% dos dispositivos. Isso pode indicar uma menor diversidade de escolha para os consumidores ou uma preferência por marcas estabelecidas.

3.12.8 - Sessões por origem de tráfego

A sessão é um grupo de interações de usuários com seu site que ocorrem em um determinado período. Uma única sessão pode conter várias exibições de página,

eventos, download de conteúdos, redirecionamentos, interação com o site ou outros usuários, além de preenchimento de formulários e enquetes, ou seja, uma infinidade de possibilidades do usuário interagir com o site. Nesta tabela, o número de sessões e o tempo médio de engajamento são variáveis importantes para avaliar a qualidade dos acessos e não apenas a quantidade, ou seja, o tempo que o usuário permanece em cada sessão de seu acesso indica potencial engajamento e recorrência.

Como cada usuário pode abrir várias sessões e elas terminam após 30 minutos de inatividade ou à meia noite, as mesmas podem ocorrer em períodos contínuos ou sazonais, em dias, semanas ou meses diferentes. Também há renovação da sessão quando, dentro dos 30 minutos, o mesmo usuário acessa através de outra origem (link ou plataforma de busca).



A matriz de correlação revela uma forte relação positiva entre as colunas 'Sessões' e 'Sessões engajadas' (0.99), seguidas por 'Contagem de eventos' e 'Sessões engajadas' (0.97). Esses achados sugerem que as sessões e as sessões engajadas estão intimamente relacionadas, e isso não surpreende, já que as

sessões engajadas são um subconjunto das sessões. Além disso, a alta correlação entre 'Contagem de eventos' e 'Sessões engajadas' indica que as sessões engajadas geralmente apresentam uma contagem de eventos mais alta.

A relação entre 'Sessões engajadas' e 'Taxa de engajamento' não é linear. A taxa de engajamento tende a ser maior para um número menor de sessões engajadas, mas não há um padrão claro. Isso sugere que a taxa de engajamento pode ser influenciada por outros fatores além do número de sessões engajadas.

Análise das Métricas:

A tabela abaixo mostra as principais métricas estatísticas descritivas dos seus dados, agrupadas por origem de acesso representando a média, mediana e desvio padrão, respectivamente.

A coluna Origem da campanha manual da sessão foi usada para agrupar os dados e as métricas foram calculadas para as colunas numéricas: “Usuários ativos”, “Sessões”, “Sessões engajadas”, “Tempo médio de engajamento por sessão”, “Sessões engajadas por usuário ativo”, “Taxa de engajamento” e “Contagem de eventos”.

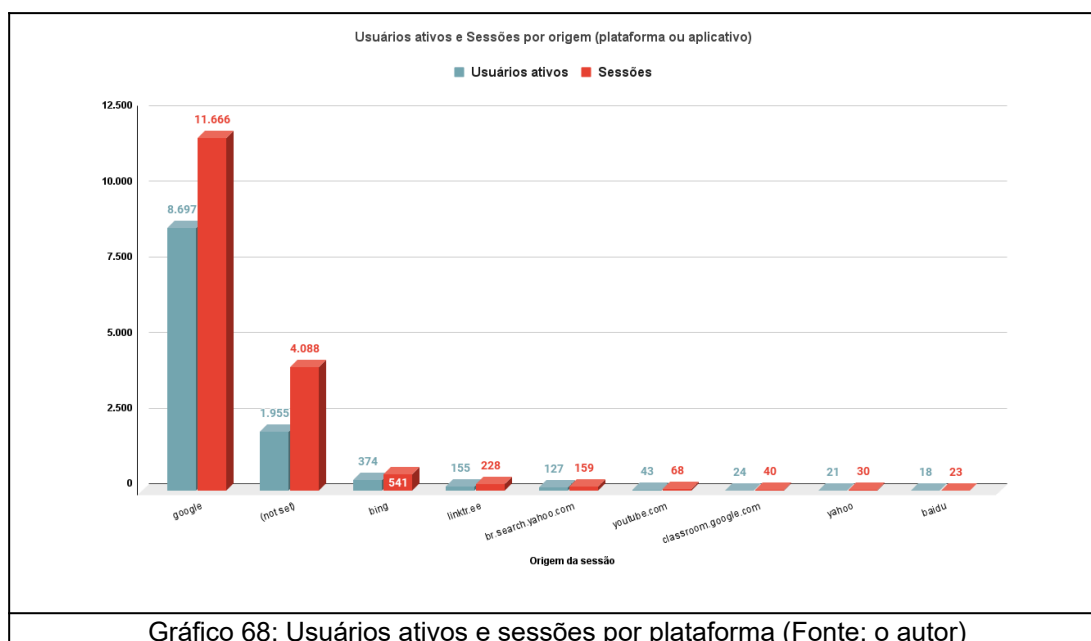
O desvio padrão é uma métrica estatística que quantifica a dispersão dos dados em torno da média. Um desvio padrão elevado indica que os dados estão mais espalhados, enquanto um desvio padrão baixo sugere que os dados estão mais concentrados em torno da média. Os quartis são valores que dividem os dados em quatro partes iguais. O primeiro quartil (Q1) representa o 25º percentil, o segundo quartil (Q2) representa o 50º percentil (ou mediana) e o terceiro quartil (Q3) representa o 75º percentil.

Métrica	Média	Mediana	Desvio Padrão	Mínimo	Máximo
Usuários Ativos	295	3	1416,48	1	8697
Sessões	437	7	1958,45	1	11666
Sessões Engajadas	254	4	1159,48	0	7024
Tempo Médio de Engajamento por Sessão	25	20	17,49	0	80,75
Sessões Engajadas por Usuário Ativo	26	5	34,34	0	87
Taxa de Engajamento	46	63	35,07	0	90

Contagem de Eventos	2.245	38	9701,01	0	56988
Tabela 12: Métricas descritivas das variáveis analisadas (RAPLE) - o autor					

Percepções:

- A origem com o maior número de usuários ativos, sessões, sessões engajadas e contagem de eventos é proveniente do 'google'.
- As origens com as maiores taxas de engajamento são provenientes de 'quora.com' e 't.me', enquanto as origens com as menores taxas são 'Pinterest', 'search.seznam.cz', 'Mamma.com', 'Yandex', 'Rambler', 'Wow.com', 'GMX', 'Seznam', 'Mail.ru' e 'Ecosia'.
- As campanhas com o maior número de sessões engajadas por usuário ativo são provenientes de 'search.yahoo.co.jp' e 'pages.kwai', enquanto a origem "Pinterest" tem a menor média de usuários ativos e em todas as outras métricas.

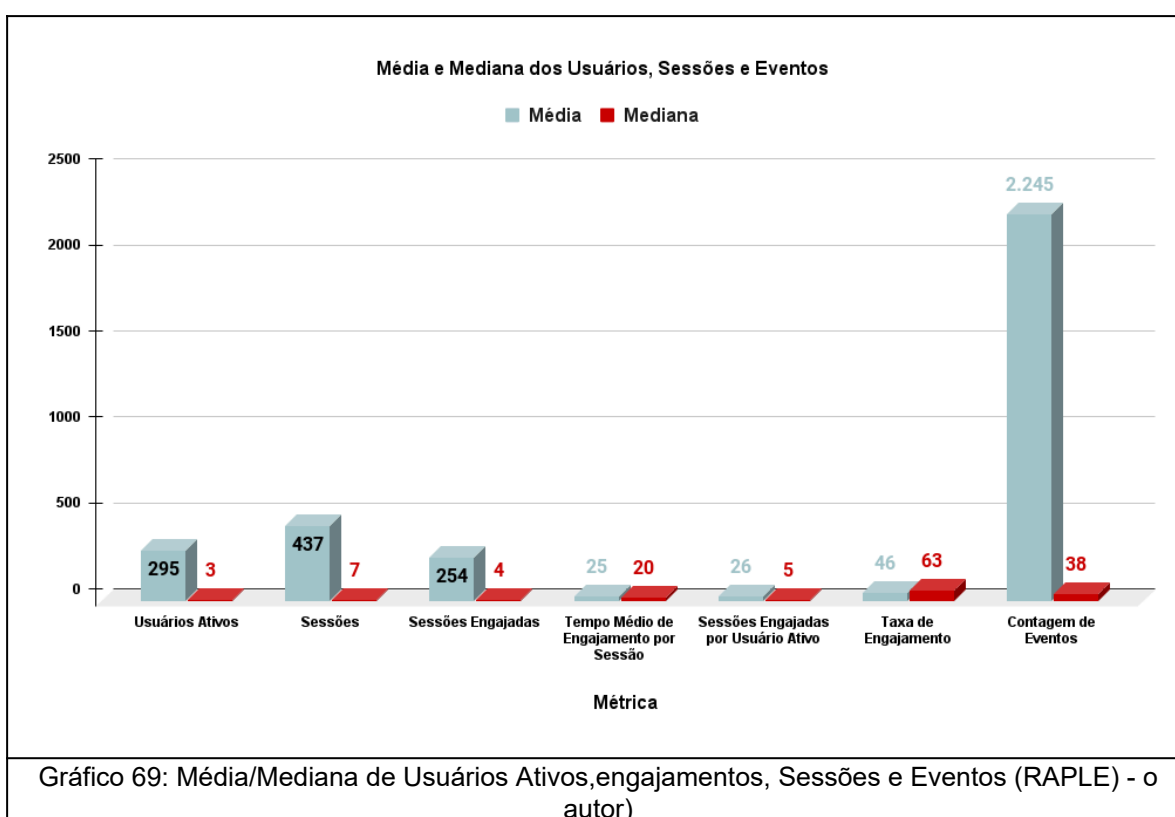


Possibilidades:

- Investigar os fatores que influenciam a taxa de engajamento, além do número de sessões engajadas.
- Analisar as origens das maiores taxas de engajamento e conhecer os dados, visando análises quantitativas que gerem conhecimento sobre suas idiossincrasias e dinâmicas de difusão científica

- Avaliar as origens com as menores taxas de engajamento e obter percepções sobre como dinamizar o acesso aos conteúdos e torná-los mais acessíveis através de buscas e metadados.
- Monitorar o desempenho das origens das campanhas ao longo do tempo para identificar tendências e ajustar as estratégias de duração, frequência e pautas do canal.

O gráfico abaixo ilustra a média e a mediana de diferentes métricas relacionadas a usuários, sessões e eventos, pelo qual podemos inferir algumas informações e correlações:



Usuários Ativos, Sessões e Sessões Engajadas:

- Média: a média de usuários ativos é menor do que a média de sessões. Isso significa que, em média, cada usuário realiza várias sessões.
- Mediana: a mediana, tanto para usuários ativos quanto para sessões, é baixa. Isso sugere que a maioria dos usuários realiza poucas sessões e que a distribuição dos dados é assimétrica, com alguns valores altos influenciando a média.
- Correlação: provavelmente, há uma correlação positiva entre usuários ativos e sessões, e entre sessões e sessões engajadas. Isso porque mais usuários tendem a gerar mais sessões, por sua vez, aumentam a probabilidade de haver mais sessões engajadas.

Tempo Médio de Engajamento por Sessão, Sessões Engajadas por Usuário Ativo e Taxa de Engajamento:

- **Média e Mediana:** As médias e medianas dessas métricas são relativamente próximas, sugerindo uma distribuição mais simétrica dos dados.
- **Correlação:** Pode existir uma correlação negativa entre o tempo médio de engajamento por sessão e o número de sessões engajadas por usuário ativo. Usuários com um maior número de sessões engajadas tendem a ter um tempo médio de engajamento menor em cada sessão. Por outro lado, a taxa de engajamento pode estar positivamente correlacionada com o tempo médio de engajamento e o número de sessões engajadas por usuário ativo.

Contagem de Eventos:

- **Média:** A contagem de eventos apresenta a maior média entre todas as métricas, o que indica um alto volume de eventos registrados.
- **Mediana:** No entanto, a mediana é baixa, sugerindo que a maioria das sessões possui uma quantidade moderada de eventos e que a média é influenciada por alguns valores extremos.
- **Correlação:** Pode haver uma correlação positiva entre a contagem de eventos e as sessões engajadas, pois as sessões com maior engajamento tendem a registrar mais eventos.

Para compreender completamente os dados, é crucial avaliar tanto a média quanto a mediana, conforme demonstrado pelo gráfico e pelas inferências. Embora a média possa ser influenciada por valores extremos, a mediana fornece uma medida mais confiável da tendência central dos dados.

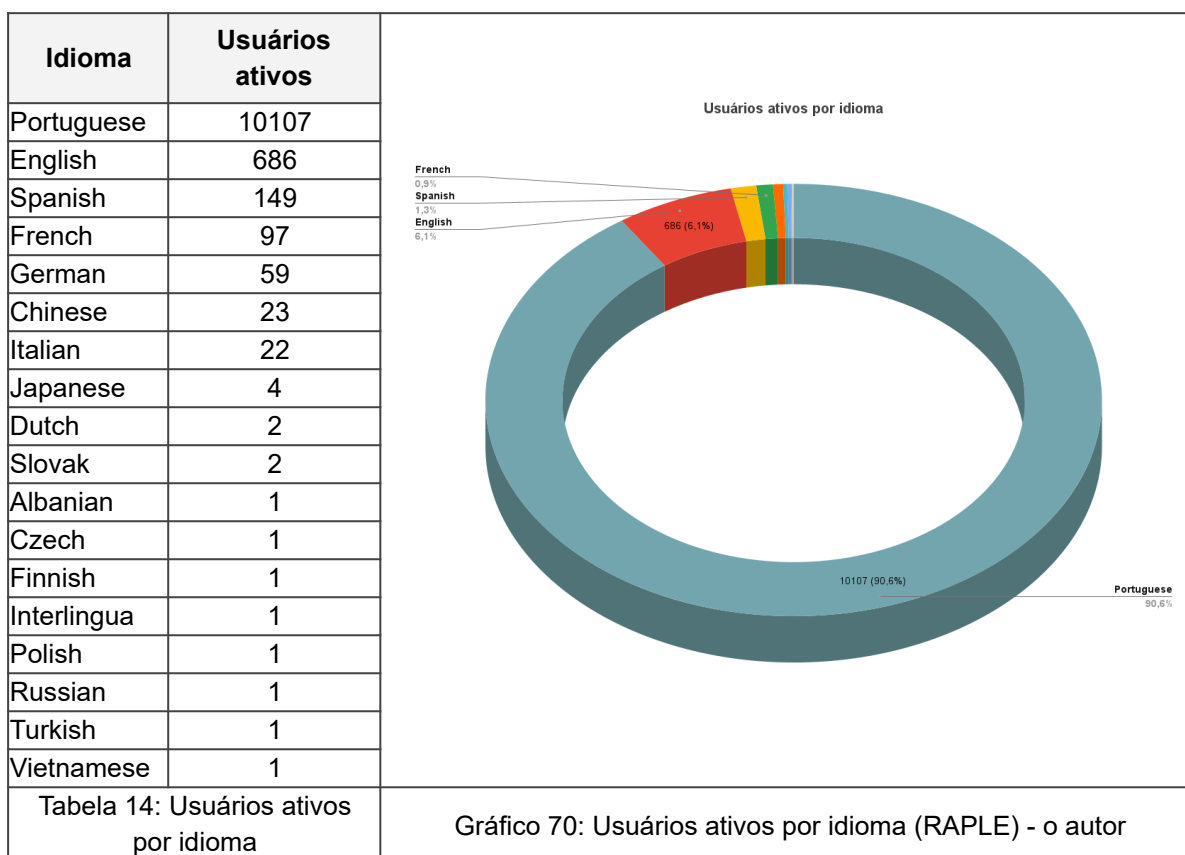
Podemos granular ainda mais os dados e obter percepções visuais mais amplas dos dados através do Mapa de Calor da Tabela analisada, onde podemos notar visualmente as que as métricas de engajamento se diferenciam por serem numericamente diferentes. Elas variam entre medidas de tempo em segundos, quantidade e percentual, onde as tonalidades evidenciam valores discrepantes entre cores muito diferentes.

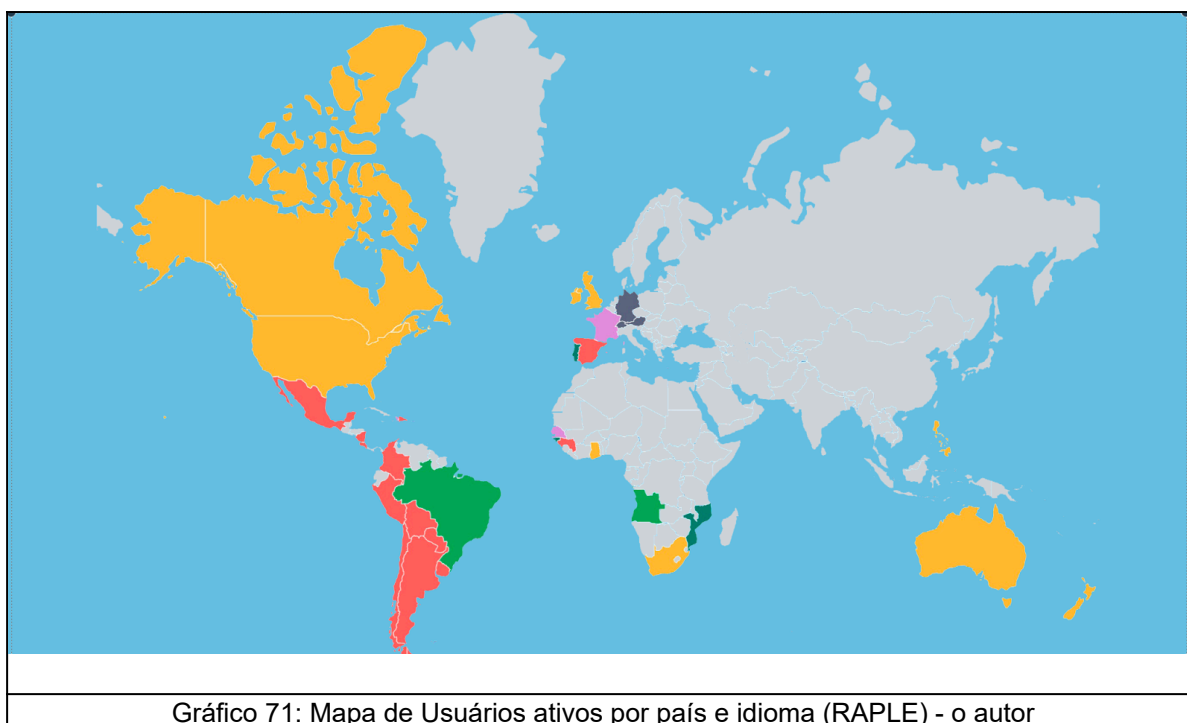
Origem da campanha manual da sessão	Usuários ativos	Sessões	Sessões engajadas	Tempo médio de engajamento por sessão	Sessões engajadas por usuário ativo	Taxa de engajamento	Contagem de eventos
google	8.697	11.666	7.024	43	80	60	56.988
(not set)	1.955	4.088	2.018	36	10	49	22.594
bing	374	541	385	51	10	71	3.188
linktr.ee	155	228	173	81	11	75	1.800

3.12.9 - Métricas de Usuários Ativos por Geolocalização e Idioma

A análise descritiva dos dados revela que o idioma português é o mais utilizado na RAPLE, representando 90,57% dos usuários ativos. O inglês ocupa o segundo lugar com 6,15%, seguido pelo espanhol com 1,34%. Os demais idiomas representam menos de 1% dos usuários ativos cada. O gráfico ilustra visualmente essa distribuição, permitindo uma rápida compreensão da predominância do idioma português em relação aos demais. A grande maioria dos usuários ativos utiliza o idioma português, que é a língua nativa da instituição e de todos os integrantes do projeto, também porque a própria natureza multilíngue do site deriva de sua finalidade de apoiar professores de línguas estrangeiras.

O inglês representando mais de 6% da origem de idiomas pode indicar que professores ou estudantes de português de países não lusófonos estejam usando o site para aprender ou ensinar. Outra hipótese é que usuários de outros países busquem conteúdos específicos das 5 línguas que o site contempla (Alemão, Espanhol, Francês, Inglês e Português para Estrangeiros (PLE)).





A tabela abaixo representa o Mapa de Calor dos usuários ativos por idioma e país, sendo necessária uma simplificação para considerar apenas o Idioma oficial de cada um.

País	Usuários ativos	Idioma oficial	País	Usuários ativos	Idioma oficial
Brazil	9826	Português	Finland	3	Finlandês
Portugal	456	Português	Poland	3	Polonês
Mozambique	118	Português	Sweden	3	Sueco
United States	92	Inglês	Chile	2	Espanhol
France	81	Francês	Colômbia	2	Espanhol
Germany	71	Alemão	Costa Rica	2	Espanhol
Angola	45	Português	Czechia	2	Tcheco
China	45	Chinês	Guinea-Bissau	2	Português
Argentina	40	Espanhol	Macao	2	Português
Switzerland	39	Alemão	New Zealand	2	Inglês
(not set)	36	nan	Nicarágua	2	Espanhol
Canadá	28	Inglês	Uruguay	2	Espanhol
Indonésia	20	Indonésio	Vietnam	2	Vietnamita
Netherlands	17	Holandês	Austrália	1	Inglês
Spain	17	Espanhol	Bulgária	1	Búlgaro
Cape Verde	16	Português	Dominican Republic	1	Espanhol
United Kingdom	16	Inglês	Georgia	1	Georgiano

Luxembourg	13	Luxemburguês	Ghana	1	Inglês
South Korea	13	Coreano	Guinea	1	Espanhol
Belgium	12	Holandês	Haiti	1	Francês
Ireland	12	Inglês	Israel	1	Hebraico
Italy	12	Italiano	Kyrgyzstan	1	Russo
Singapore	11	Inglês	Lebanon	1	Árabe
Hungary	7	Húngaro	Martinique	1	Espanhol
México	7	Espanhol	Philippines	1	Inglês
Japan	6	Japonês	Rússia	1	Russo
Peru	6	Espanhol	Senegal	1	Francês
Áustria	5	Alemão	Slovakia	1	Eslovaco
Bolívia	5	Espanhol	South Africa	1	Inglês
Turkiye	5	Turco			
India	4	Hindi			
Paraguay	4	Espanhol			
São Tomé & Príncipe	4	Português			

Tabela 15: Mapa de Calor dos usuários ativos por idioma e país (RAPLE) - o autor

A análise das variáveis Visualização, Sessões, Engajamento, Tempo Médio de Engajamento e Contagem de Eventos relacionadas às Origens de acesso (Campanha no Analytics) podem fornecer percepções sobre a capilaridade global do site e outras métricas geopolíticas e linguísticas.

Ao analisar os dados, observamos uma correlação positiva fraca entre idioma e usuários ativos, com um coeficiente de correlação de 0,14. Isso sugere que a influência do idioma no número de usuários ativos é limitada e que outros fatores podem ter um impacto mais significativo.

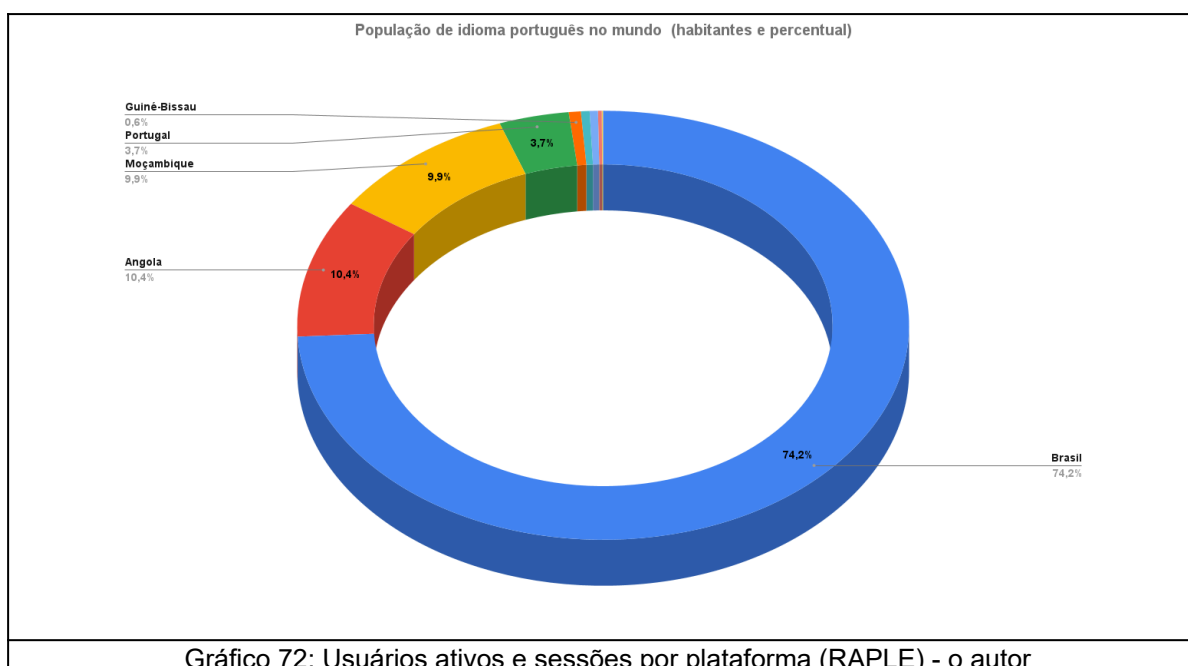
Idiomas

O idioma português tem o maior número de usuários ativos, o que pode ser explicado pelo idioma original do site e proporcionalmente, pela grande população de países como Brasil, Angola, Moçambique e Portugal. Todos os 9 países de idioma português somam 281.160.432 habitantes que têm o português como língua oficial, embora haja exceções de equivalência entre idiomas.

País	Continente	População (habitantes)	Percentual
------	------------	------------------------	------------

Brasil	América	208.494.900	74,31
Angola	África	29.310.273	10,45
Moçambique	África	27.909.798	9,95
Portugal	Europa	10.374.822	3,70
Guiné-Bissau	África	1.792.338	0,64
Guiné-Equatorial	África	1.221.490	0,44
Macao	Ásia	704.149	0,25
Cabo Verde	África	560.999	0,20
São Tomé e Príncipe	África	204.454	0,07

Tabela 16: Países com língua portuguesa como língua oficial (RAPLE) - o autor

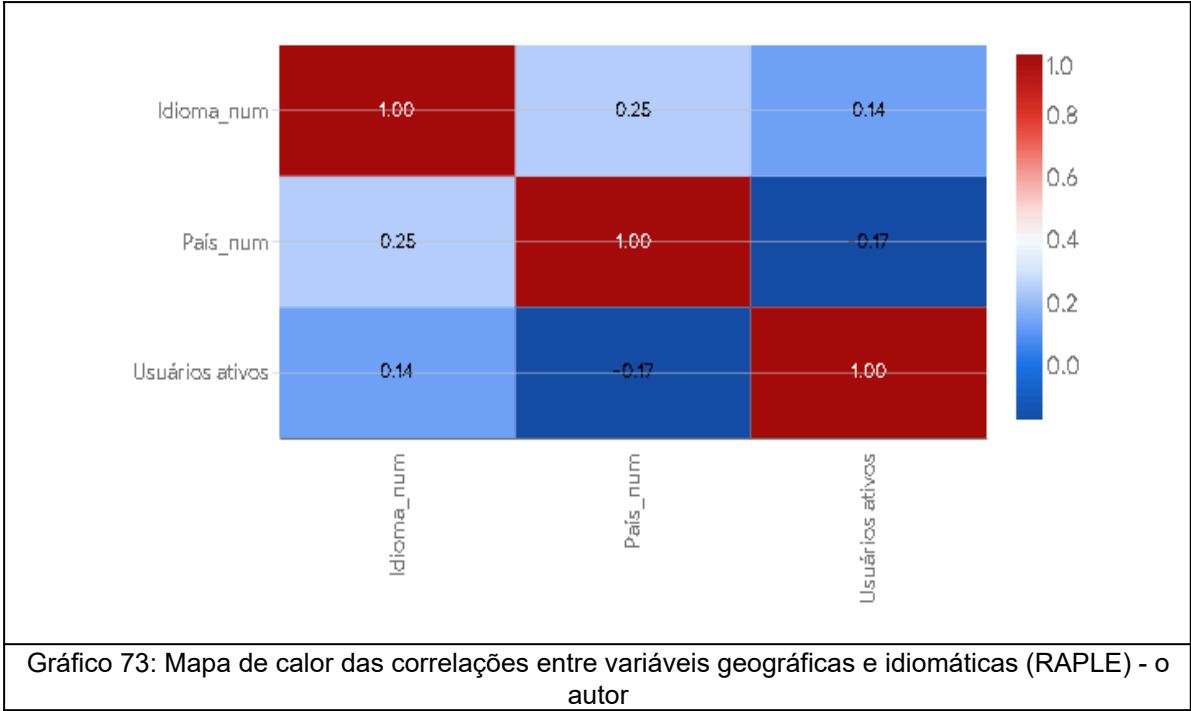


Para possibilitar a análise de correlação, convertemos os campos de texto em numéricos. A análise de correlação requer valores numéricos, e essa conversão permite atribuir um número único a cada valor de texto distinto. Dessa forma, os algoritmos de correlação podem identificar padrões e relações entre as variáveis. O objetivo principal dessa conversão é permitir a análise estatística de dados que não são originalmente numéricos, que permitam:

- **Identificação de Padrões:** Descubra a relação entre diferentes variáveis, como a influência do idioma no número de usuários ativos.
- **Analisar Relações:** Medir a intensidade da correlação entre variáveis, como a conexão entre o país e o idioma.
- **Construção de modelos:** Elaborar modelos preditivos, como a regressão linear, para compreender as relações entre variáveis e antecipar resultados.

A conversão de campos de texto em numéricos é um processo essencial na análise de dados, ao permitir que se obtenha insights profundos que fundamentam decisões mais estratégicas com base nos dados coletados.

Ao converter os campos de texto em numéricos, podemos utilizar técnicas estatísticas e de aprendizado de máquina para identificar padrões, correlações e tendências nos dados. Isso permite compreender melhor os dados e tomar decisões com mais segurança.



O espanhol é falado como língua oficial em mais países do que qualquer outro idioma, possivelmente devido ao vasto número de colonizações espanholas na América Latina e à grande população destes países. O inglês é o segundo idioma em número de países, o que também pode ser atribuído à sua ampla colonização global em países que atualmente exibem alta densidade populacional.

O terceiro idioma com maior capilaridade global também está presente em 3 continentes, principalmente no africano, embora o Brasil corresponda a 74% dos falantes nativos do português no mundo. Somadas, as populações dos quatro mais populosos correspondem a 98,40% da soma global.

Espanhol	Inglês	Português	Francês	Alemão
14 países	10 países	8 países	3 países	3 países
Argentina Espanha México	United States Canadá United Kingdom	Brazil Portugal Angola	France Haiti Senegal	Germany Switzerland Áustria

Peru Bolívia Paraguay Chile Colômbia Costa Rica Nicarágua Uruguay Rep. Dominicana Guinea Martinique	Ireland Singapore New Zealand Austrália Ghana Philippines South Africa	Mozambique Cape Verde S.Tomé & Príncipe Guinea-Bissau Macao		
Tabela 17: Capilaridade dos idiomas dos usuários que acessam (RAPLE) - o autor				

Os três idiomas mais falados correspondem exatamente à metade (32 de 64) dos países e embora o francês e o alemão tenham se difundido pelo mundo, os usuários da RAPLE estão localizados em apenas 3 países de cada um desses idiomas.

Consideramos a viabilidade de uma regressão linear com o número de usuários ativos como variável dependente e os campos país e idioma como variáveis independentes, analisando os dados da tabela. Para possibilitar sua inclusão no modelo de regressão linear, convertemos as variáveis independentes, país e idioma, em valores numéricos.

A análise da matriz de correlação e das primeiras linhas da tabela revelou uma correlação fraca entre as variáveis independentes e a variável dependente, sugerindo que a regressão linear pode não ser a melhor abordagem para modelar os dados. Além disso, o baixo valor de R^2 (0,07) indica que o modelo de regressão linear explica apenas 7% da variabilidade dos dados, confirmando sua inadequação para esse conjunto.

- Relação não linear: a relação entre as variáveis pode não ser linear, dificultando a captura da complexidade da relação.
- Variáveis categóricas: as variáveis independentes (país e idioma) são categóricas e não numéricas, o que pode não representar adequadamente a relação entre elas.
- Influência de outras variáveis: podem existir outras variáveis, não incluídas na análise, que influenciam o número de usuários ativos, como renda, idade e acesso à internet.

Para obter um modelo mais preciso e com maior capacidade preditiva, podemos utilizar alternativas de modelagem, como a regressão não linear ou a

regressão logística. Essas técnicas são mais eficazes para lidar com variáveis categóricas, resultando em modelos mais precisos e com melhor desempenho preditivo.

Regressão não linear

Para construir um modelo de regressão que modelasse a relação entre as variáveis, analisamos os dados da tabela e realizamos uma regressão não linear. Ao utilizarmos as colunas de texto no modelo de regressão, convertemos os valores em numéricos. Em seguida, ajustamos o modelo aos dados usando um modelo de regressão polinomial de grau 2 (HOFFMANN, 2016)

Este modelo apresentou um coeficiente de determinação (R^2) de 0,14, indicando que explica 14% da variabilidade dos dados. O resultado representa um pequeno aumento em relação ao modelo de regressão linear anterior, mas não consegue representar adequadamente a relação entre as variáveis, apesar da melhora observada.

Indicando que a relação pode ser mais complexa do que o modelo pode capturar ou que variáveis não incluídas na análise podem ter influência significativa, construímos modelos de regressão logística para esse objetivo com variáveis categóricas. Apesar da conversão das variáveis categóricas, existem métodos diferentes de representar os dados através de conversões que as aplicam como numeração de caracteres textuais.

Regressão logística

A regressão linear e a regressão logística são métodos de modelagem estatística usados para prever o valor de uma variável de saída com base em uma ou mais variáveis de entrada. Essas variáveis são classificadas em variáveis dependentes (variáveis de saída) e variáveis independentes (variáveis de entrada). Vejamos as equações:

Modelo de Regressão	Equação (HOFFMANN, 2016)
Linear	$y = \beta_0 + \beta_1 x + \epsilon$
Polinomial (não linear)	$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_n x^n + \epsilon$

Logística (não linear)	$y = e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon)} / (1 + e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon)})$
Tabela 18: Equações dos modelos de regressão (RAPLE) - o autor	

As variáveis categóricas são aquelas que assumem valores de uma lista finita de categorias e usam a classificação binária. Um exemplo clássico é o número obtido ao jogar um dado de seis lados. Esse tipo de variável é importante na relação logística, que descreve a relação entre uma variável dependente e uma ou mais variáveis independentes.

Regressão linear	Regressão logística
Um método estatístico para prever um valor de saída com base em um conjunto de valores de entrada.	Um método estatístico para prever a probabilidade de um valor de saída em uma determinada categoria de um conjunto de variáveis categóricas.
Relação linear, representada por uma linha reta.	Relação logística ou relação sigmoidal, representada por uma curva em forma de S.
Equação Linear.	Equação Logarítmica
Aprendizado: Regressão.	Aprendizado: Classificação.
Distribuição: Normal/gaussiana.	Distribuição: Binomial.
Tarefas que exigem uma variável dependente contínua prevista em uma escala.	Tarefas que exigem uma probabilidade prevista de uma variável dependente categórica ocorrer em um conjunto fixo de categorias.
Quadro 14: Diferenças entre modelos de regressão (Baseado em AWS ⁵⁶)	

O modelo de regressão logística resultou em uma acurácia de 0,23, o que indica que o modelo classificou corretamente 23% dos casos. Apesar da baixa acurácia, o modelo de regressão logística é útil para análises descritivas e exploratórias, fornecendo informações sobre a relação entre as variáveis e identificando possíveis tendências.

Aplicamos três métodos alternativos para lidar com variáveis categóricas em modelos de regressão logística:

Modelo	Breve descrição	Acurácia
--------	-----------------	----------

⁵⁶<https://aws.amazon.com/pt/compare/the-difference-between-linear-regression-and-logistic-regression/>

One-Hot Encoding	A codificação one-hot é uma técnica para representar dados categóricos como vetores numéricos, em que cada categoria exclusiva é representada por uma coluna binária com um valor de 1 indicando sua presença e 0 indicando sua ausência.	0,31 (31%)
Dummy Coding	A codificação dummy permite a transformação de variáveis categóricas em um formato adequado para análise. Cada coeficiente associado a uma variável dummy representa a diferença no resultado em relação à categoria que foi omitida	0,31 (31%)
Effect Coding	O effect coding é uma técnica empregada na codificação de variáveis categóricas em modelos de estimação. Em vez de utilizar a codificação fictícia, em que uma categoria é representada por 1 e as demais por 0, o effect coding representa as categorias como -1 e +1.	0,31 (31%)
Quadro 15: Diferenças entre modelos de regressão logística alternativos (HOFFMANN, 2016)		

Estes métodos permitiram que o modelo utilizasse as informações das variáveis categóricas sem a necessidade de convertê-las diretamente em números ordinais, o que pode distorcer a análise. Mesmo assim, atingiram menos de um terço de acerto nas classificações e mais que o triplo da regressão linear, confirmando que a correlação entre as variáveis é fraca e pouco influencia o número de usuários ativos.

No caso da RAPLE, a influência do idioma português nos usuários ativos com predominância em 90,57% deles, está diretamente relacionada à origem dos conteúdos, que embora contenham elementos idiomáticos de outros países, é majoritariamente em língua portuguesa. Essa discrepância cria uma distorção na análise que não pôde ser explicada pelos dados dos outros idiomas nos países que acessam o site, mas fica categoricamente explicado que a quantidade de usuários nos países lusófonos é muito importante, ao mensurar o alcance e prioridades do projeto de extensão, além da internacionalização no âmbito local.

Também é importante analisar os acessos dos materiais de Português para Estrangeiros (PLE), visando verificar a possibilidade do acesso pelos usuários ativos ser estimulado pelo aprendizado do português nos conteúdos de todos os outros idiomas que o projeto atende.

3.12.10 - Análise por Estado, Província ou Região

O recorte geográfico por Estado, Província ou Região apresentou resultados semelhantes, mas demonstra maior capilaridade devido ao aumento da granularidade dos dados, embora o conceito regional seja bem flexível em sua especificidade. No caso do Brasil, as regiões são os estados, mas em alguns registros são províncias ou mesmo cidades ou regiões metropolitanas, que podem receber os nomes de condado, província, região ou mesmo um cantão.



As variáveis 'Usuários ativos', 'Novos usuários', 'Sessões engajadas' e 'Contagem de eventos' possuem correlação forte e positiva, indicando que essas métricas tendem a variar em conjunto, aumentando ou diminuindo simultaneamente. No entanto, a 'Taxa de engajamento' exibe fraca correlação com todas as outras variáveis, sugerindo que seu comportamento seja relativamente independente.

As estatísticas descritivas oferecem uma visão geral das principais características de cada variável. A média de usuários ativos é de 51,92 e um desvio padrão de 255,35 indica uma grande variabilidade nos dados. De forma semelhante, às outras variáveis, também apresentam desvios padrão elevados, o que sugere uma considerável dispersão dos dados em relação à média.

As principais métricas descritivas são:

Métrica	Cálculo
Contagem	Total de observações de valores no conjunto de dados.
Média	Soma dos valores, divididos pela contagem
Desvio Padrão	Dispersão ou variabilidade em relação à média
Mínimo	É o valor mínimo no conjunto de dados

25% ou 1o. quartil	Valor abaixo do qual 25% dos dados se encontram
50%, 2o. ou Mediana	Valor abaixo do qual 50% dos dados se encontram
75% ou 3o. quartil	Valor abaixo do qual 75% dos dados se encontram
Máximo	É o valor máximo no conjunto de dados
Quadro 16: Descrição das fórmulas estatísticas descritivas - o autor	

As métricas de “Contagem de Eventos” é muito discrepante em seu valor máximo, superando mais de 10 vezes em relação aos demais, enquanto a “Taxa de engajamento” é calculada percentualmente e apresenta valores relativamente baixos.

Métrica	Usuários ativos	Sessões engajadas	Contagem de eventos	Taxa de engajamento
Contagem	219	219	219	219
Média	51,92	44,80	39,88	61,60
Desvio Padrão	255,35	303,91	3151,90	56.036
Mínimo	0	0	1	0
25%	1	1	5	40
50%	3	2	14	60
75%	10	6	51.50	100
Máximo	3485	4389	46085	625
Tabela 19: Métricas descritivas de variáveis correlacionadas (RAPLE) - o autor				

As correlações também demonstram alta variabilidade dos dados, mas permitem avaliar o ajuste do modelo aos dados ao comparar os valores das métricas em duplas. Em seguida, fizemos uma série de análise de regressões em modelos lineares e polinomiais que demonstram quais modelos oferecem maior precisão no ajuste aos dados preditivos.

Usuários ativos x Sessões engajadas: Relação altamente linear e positiva com $R^2 = 0,968$, indicando que o modelo se ajusta bem aos dados. Regressão linear de Sessões engajadas em Usuários ativos: Intercepto: -16,003 Coeficiente: 1,171 R^2 : 0,968

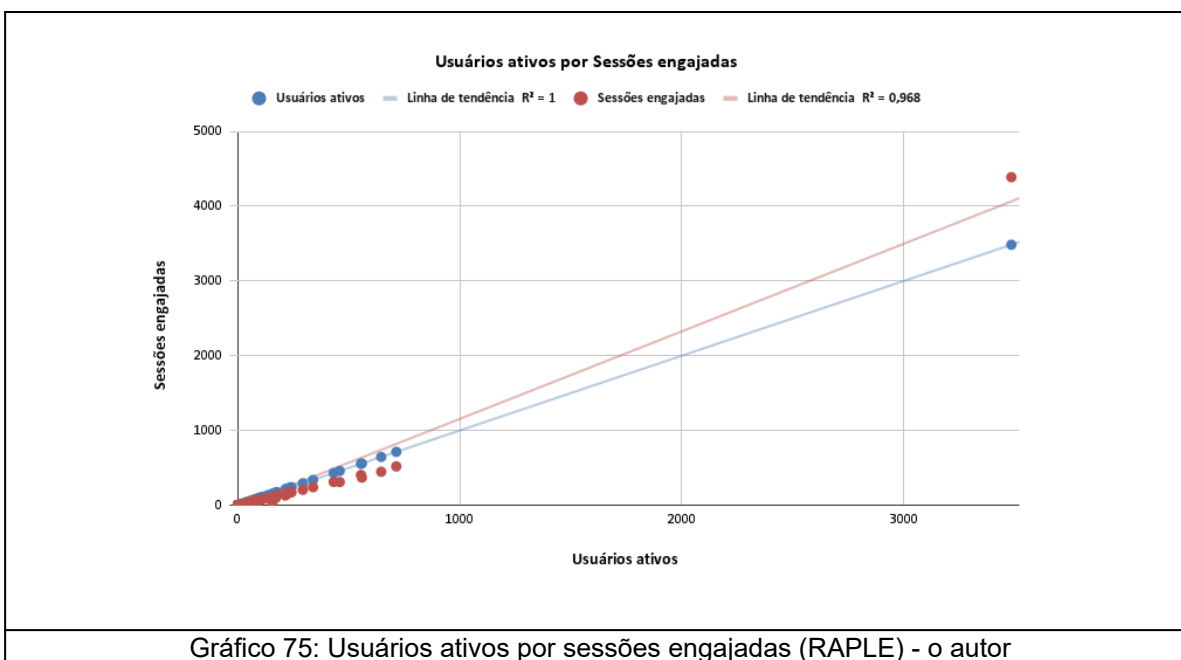


Gráfico 75: Usuários ativos por sessões engajadas (RAPLE) - o autor

Usuários ativos x Contagem de eventos: Relação altamente linear e positiva. O R^2 de 0,942 indica que o modelo se ajusta bem aos dados. Regressão linear de Contagem de eventos em Usuários ativos: Intercepto: -223,145 Coeficiente: 11,979 R^2 : 0,942

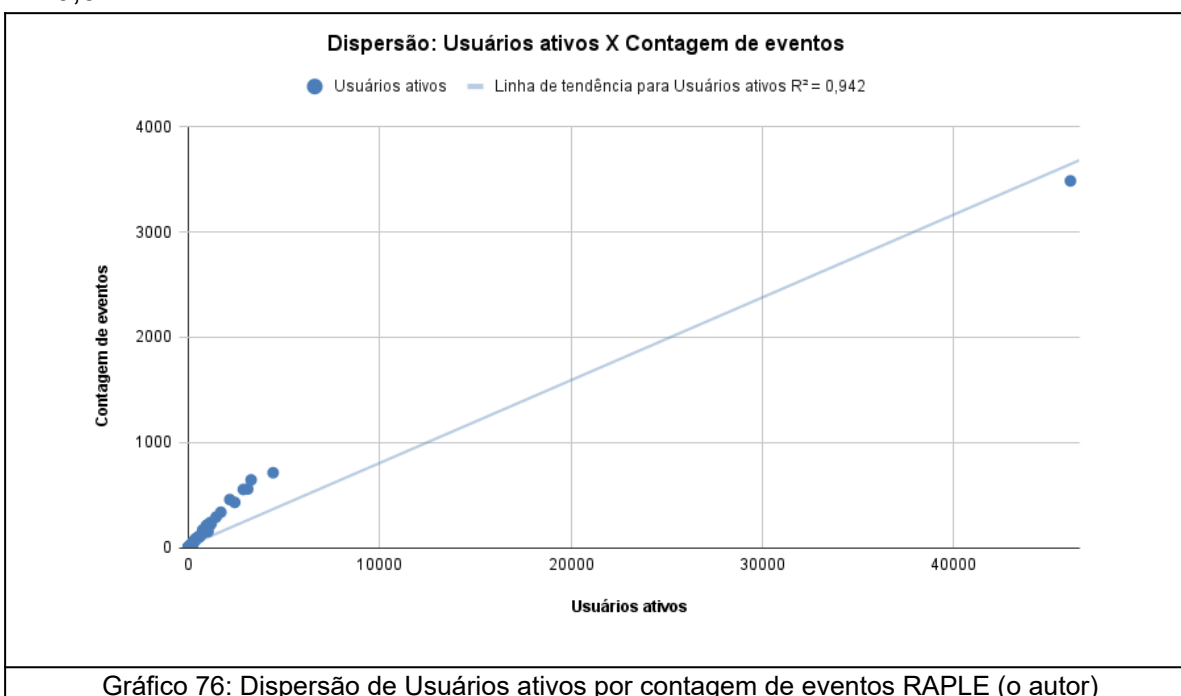


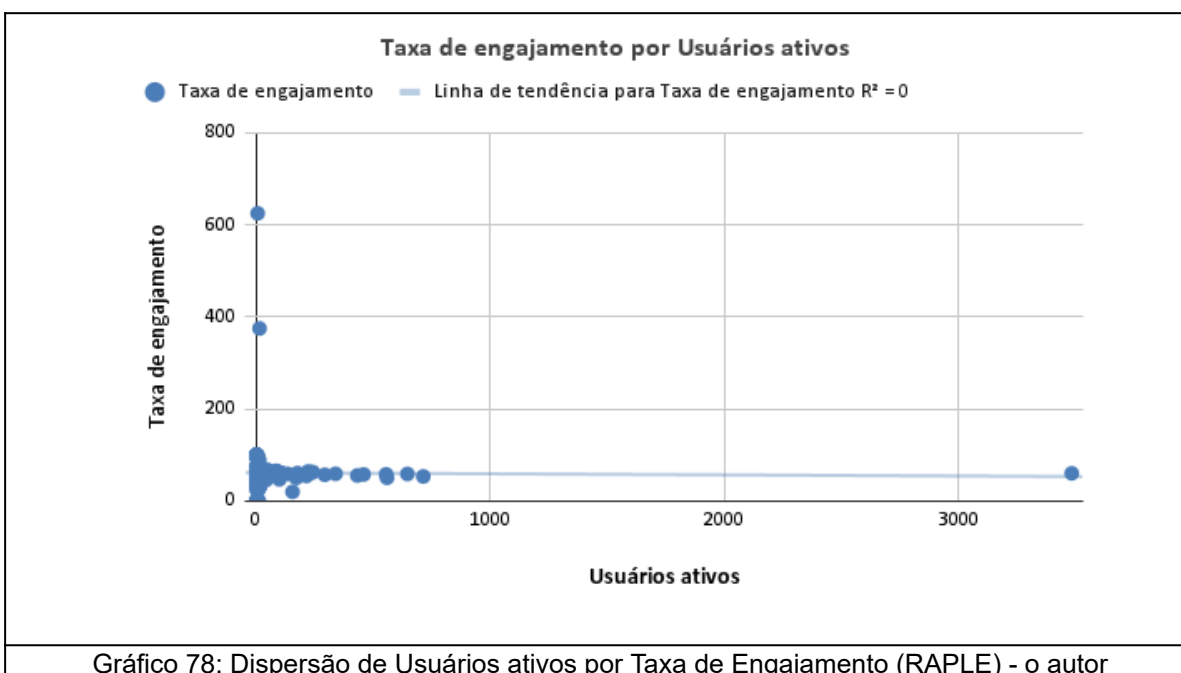
Gráfico 76: Dispersão de Usuários ativos por contagem de eventos RAPLE (o autor)

Taxa de engajamento por Contagem de eventos: Relação linear negativa muito fraca. O coeficiente de determinação (R^2) é de $3e-05$, o que indica que o modelo não se ajusta bem aos dados. A regressão linear da taxa de engajamento em relação à

contagem de eventos tem um intercepto de 61,648 e um coeficiente de $-9,8\text{e-}05$. O R^2 é de $3\text{e-}05$.

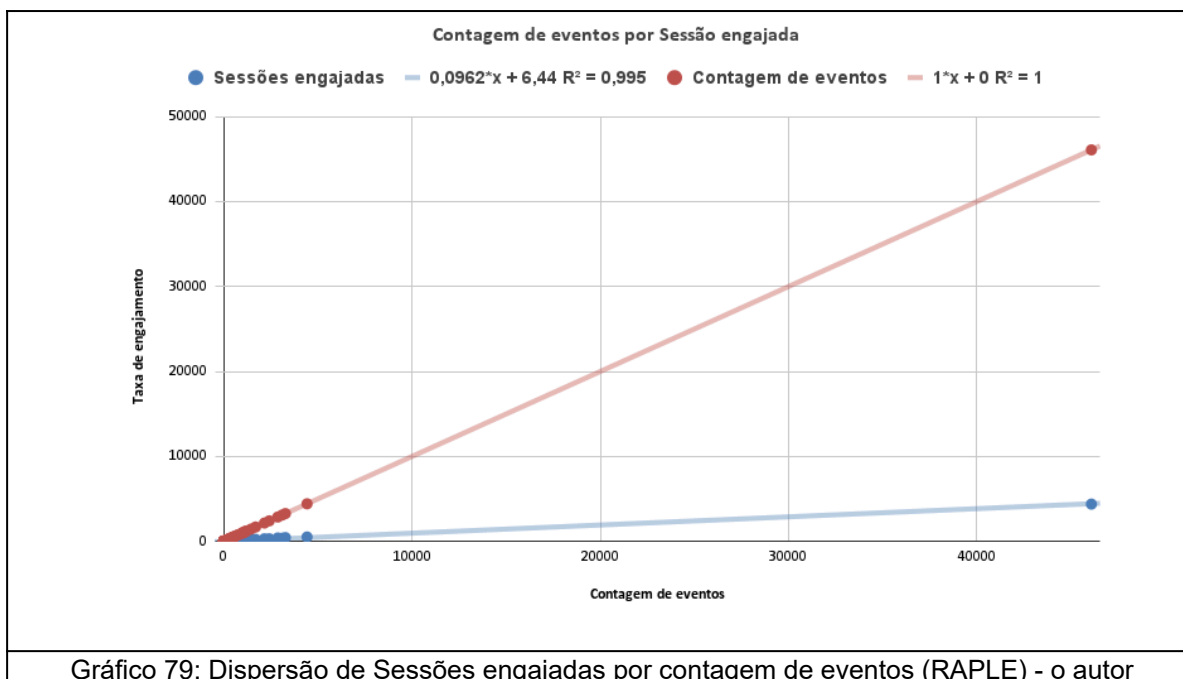


Taxa de engajamento e usuários ativos: Exibem uma relação **não linear**, conforme indicado pelo R^2 de 0,00057. A regressão polinomial da taxa de engajamento indicou um R^2 de 0,00057s, o Intercepto de 62,169, além dos coeficientes $0,00000000\text{e}+00$ -2, $12224334\text{e-}02$ 1, $93735148\text{e-}05$ -3, $85419887\text{e-}09$

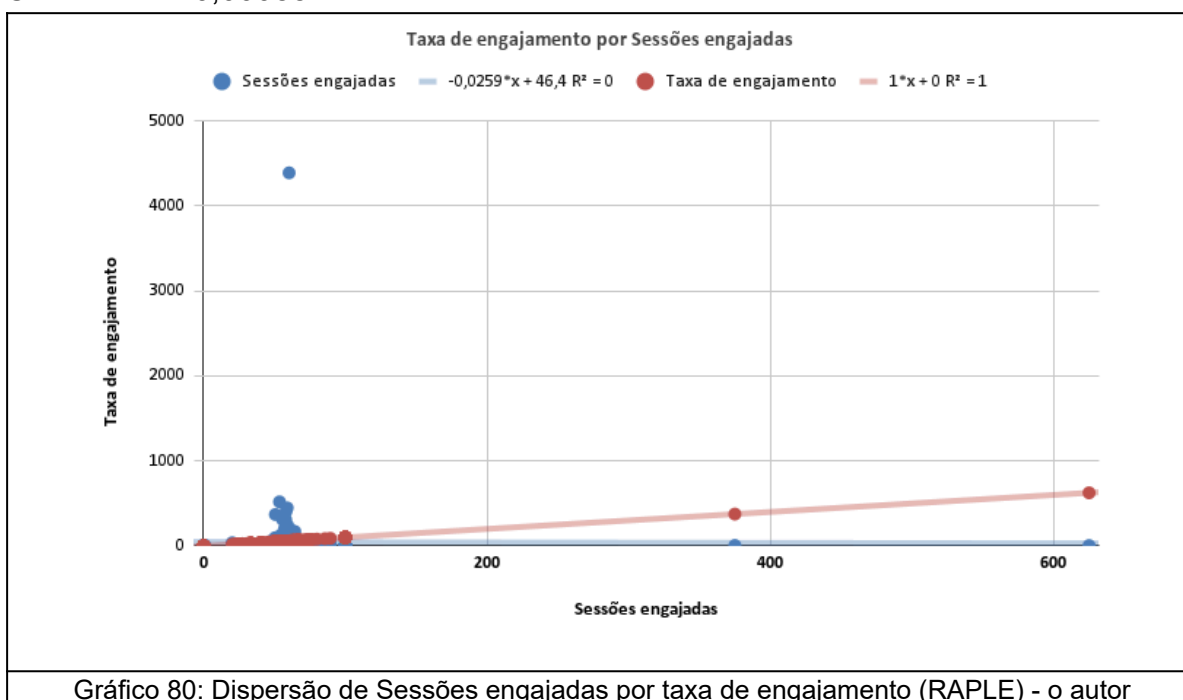


Contagem de eventos por Sessões engajadas: Relação altamente linear e positiva. O coeficiente de determinação (R^2) é de 0,995, indicando que o modelo se ajusta bem

aos dados. A regressão linear da contagem de eventos em relação às sessões engajadas apresenta um intercepto de -64,674 e um coeficiente de 10,346.



Sessões engajadas x Taxa de engajamento: Relação linear negativa muito fraca. O R^2 é $2,3e-05$, indicando que o modelo não se ajusta bem aos dados. Regressão linear de Taxa de engajamento em Sessões engajadas: Intercepto: 61,648 Coeficiente: -0,00088.



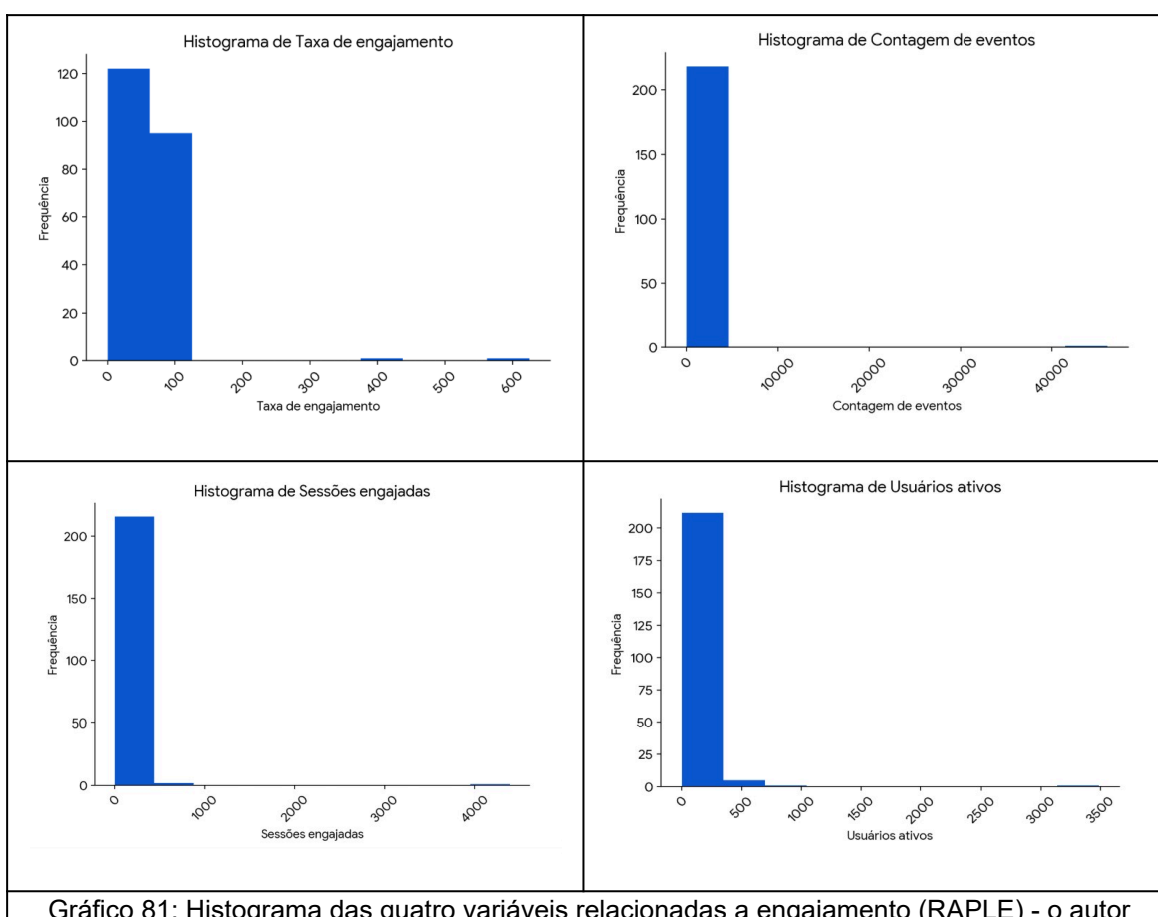
Vamos analisar as correlações entre todos os dados de forma agrupada e através dos histogramas de cada métrica.

Matriz de Correlação

	Usuários ativos	Sessões engajadas	Contagem de eventos	Taxa de engajamento
Usuários ativos	1	0.983997	0.970503	-0.0108096
Sessões engajadas	0.983997	1	0.997533	-0.00476813
Contagem de eventos	0.970503	0.997533	1	-0.00550691
Taxa de engajamento	-0.0108096	-0.00476813	-0.00550691	1
Tabela 20: Matriz de correlação de variáveis de engajamento (RAPLE) - o autor				

Percepções:

- 'Usuários ativos', 'Sessões engajadas' e 'Contagem de eventos' são altamente correlacionados.
- A métrica 'Taxa de engajamento' tem uma correlação negativa com todas as outras
- Os histogramas mostram que os dados são distorcidos para a direita para todas as variáveis.



O gráfico rosca sobre a distribuição por estado demonstra clara predominância de São Paulo (30,6%), seguido por mais dois estados da região

sudeste (Minas Gerais (6,3%) e Rio de Janeiro (5,7%)), três da região Sul (Rio Grande do Sul (4,9%), Paraná (4,9%) e Santa Catarina (3,8%)), que juntamente com outros três estados do Nordeste, (Bahia (4%), Ceará (3,8%) e Pernambuco (2,6%)), contemplam praticamente $\frac{2}{3}$ de todos os usuários ativos da Raple (66%). Os demais 34% estão divididos entre diversas regiões ou estados de diversos países do mundo todo, com destaque para o Distrito de Lisboa em Portugal com 1,5% dos ativos, além dos 1,2% que não definem suas origens geográficas, sendo em maior número que outros estados brasileiros.

A taxa de engajamento acompanha a mesma métrica do número de usuários ativos, com destaque para Maputo (capital de Moçambique) com 61,53%, Roraima com 68,25 %, além de Alagoas (65,97%) e Sergipe (64,5%) que demonstram converter muitas sessões ativas em engajadas e consequentemente, maiores possibilidades de eventos.

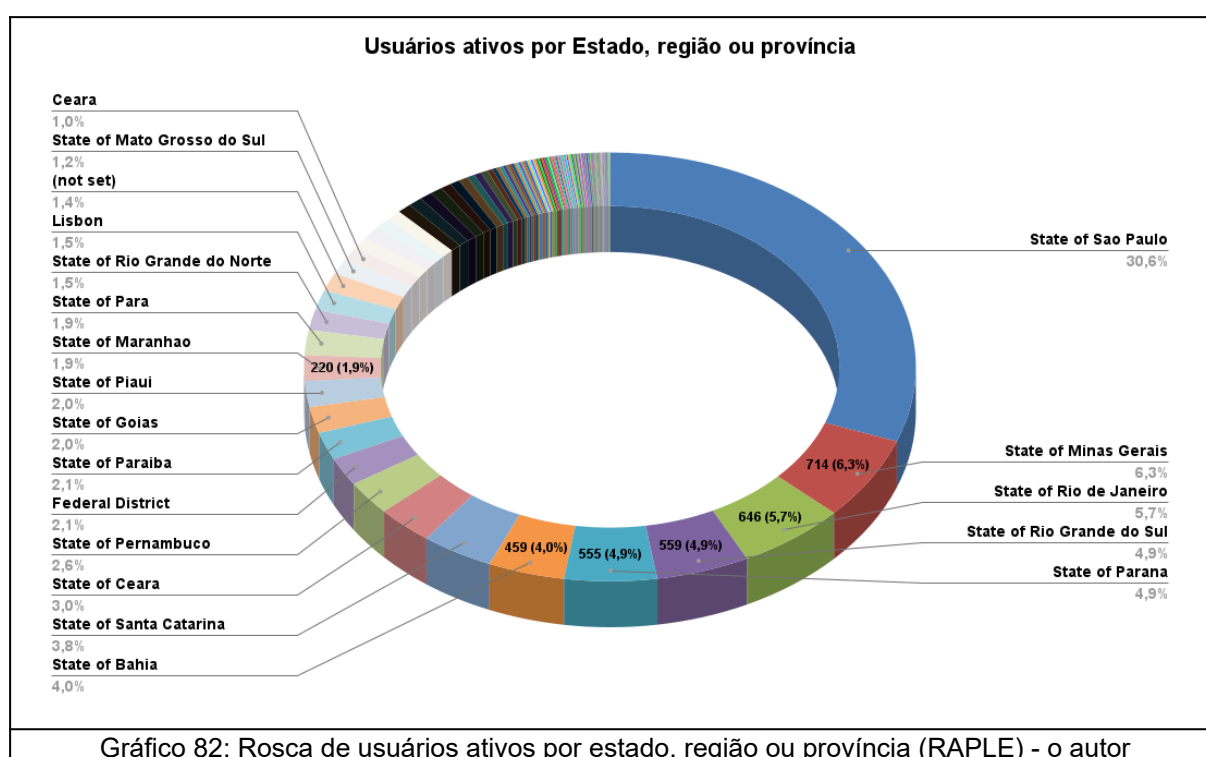


Tabela com métricas sobre estados, regiões, distritos ou províncias

Região	Usuários ativos	Sessões engajadas	Contagem de eventos	Taxa de engajamento
State of São Paulo	3485	4389	46085	60,37
State of Minas Gerais	714	519	4425	53,61

State of Rio de Janeiro	646	448	3276	59,02
State of Rio Grande do Sul	559	370	3099	50,68
State of Paraná	555	405	2870	58,18
State of Bahia	459	310	2162	57,94
State of Santa Catarina	432	311	2420	55,73
State of Ceará	339	241	1694	59,35
State of Pernambuco	293	206	1446	57,22
Federal District	242	173	1156	62,9
State of Paraíba	236	176	1193	62,41
State of Goiás	226	158	1061	61,47
State of Piauí	222	170	1123	64,39
State of Maranhão	220	158	1141	61,96
State of Pará	214	130	944	54,62
State of Rio Grande do Norte	176	139	980	60,96
Lisbon	171	97	747	50,52
(not set)	155	38	1025	20,21
State of Mato Grosso do Sul	135	101	725	59,06
Ceará	115	84	587	60
State of Espírito Santo	106	81	584	60,9
State of Mato Grosso	102	74	551	60,65
State of Amazonas	99	53	479	47,32
State of Sergipe	92	68	443	64,15
Maputo	85	42	360	61,53
State of Alagoas	85	64	408	65,97
State of Amapá	77	56	387	62,22
Porto District	73	50	356	59,52
State of Acre	64	43	291	56,57
State of Tocantins	58	43	298	64,17
Setúbal	57	33	227	54,09
State of Roraima	46	43	246	68,25
State of Rondônia	43	29	193	63,04
Luanda Province	42	22	197	45,83
Ile-de-France	33	23	168	52,27
Braga	23	17	103	68
Aveiro District	22	17	106	70,83
Bavaria	21	15	105	57,69
Quebec	18	13	83	59,09

Nampula Province	17	10	93	52,63
Pays de la Loire	17	7	70	41,17
Coimbra District	16	5	69	31,25
Leiria District	15	12	98	60
Azores	14	6	58	375
England	13	4	76	66,66
North Rhine-Westphalia	13	9	75	56,25
Oregon	13	17	97	89,47
Busan	12	2	38	66,66
Madeira	12	8	52	66,66
Praia	12	9	59	60
Santarém District	12	11	204	73,33
Faro District	11	9	49	0,75
Sofala Province	11	3	41	27,27
Baden-Wurtemberg	10	7	41	63,63
California	10	8	54	72,72
County Dublin	10	4	38	63,63
New York	10	5	79	50
Virginia	10	5	42	54,54
Zurich	10	6	50	50
Buenos Aires	9	6	51	61,53
Buenos Aires Province	9	5	46	41,66
Vaud	9	2	38	66,66
Viseu District	9	7	41	63,63
Guarda District	8	4	33	50
Auvergne-Rhône-Alpes	7	5	43	54,54
East Java	7	12	157	85,71
Hessen	7	4	31	57,14
Brussels	6	5	27	625
Corrientes	6	12	185	63,63
Grand Est	6	3	22	50
Ohio	6	2	30	33,33
Ontario	6	2	27	22,22
Provence-Alpes-Côte d'Azur	6	4	32	57,14
Viana do Castelo District	6	4	29	57,14
Zhejiang	6	0	6	0
Berlin	5	4	26	50

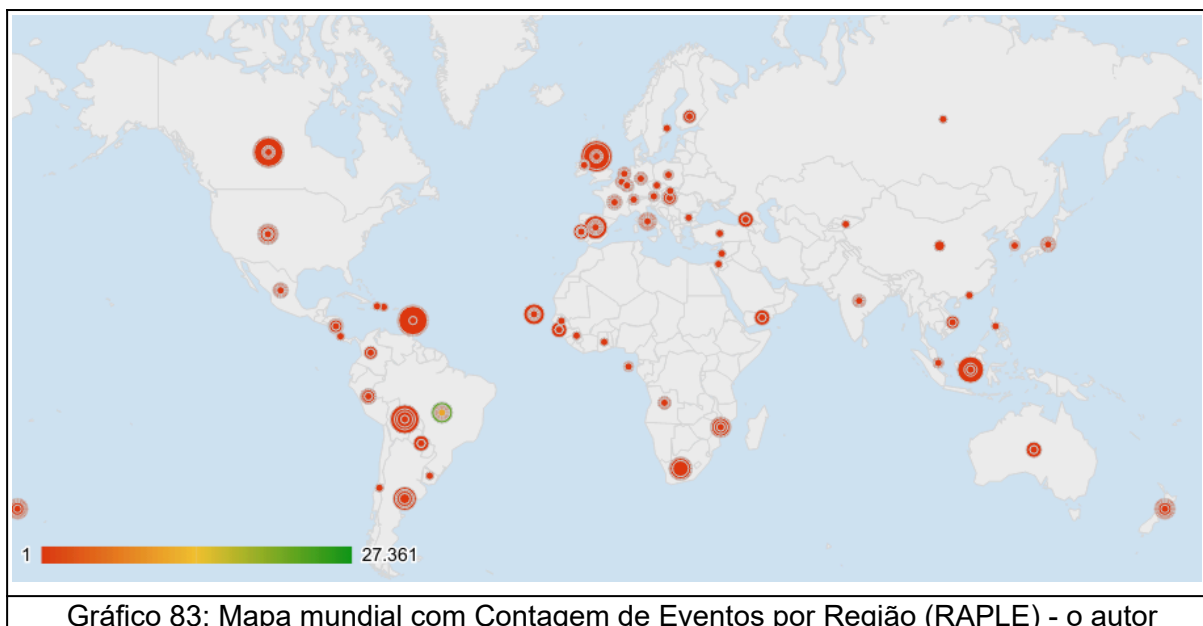
Canton of Bern	5	2	18	40
Community of Madrid	5	3	18	60
Florida	5	3	28	60
Geneva	5	2	20	40
La Paz Department	5	6	82	54,54
Mexico City	5	1	32	76,92
Saxony	5	5	20	100
West Java	5	4	23	66,66
Castile and León	4	6	42	66,66
Flanders	4	2	20	33,33
Georgia	4	4	34	66,66
Hauts-de-France	4	2	14	50
Mendoza Province	4	0	13	0
North Sumatra	4	3	16	50
Saarland	4	3	16	75
Tokyo	4	0	15	0
Vienna	4	1	14	24
Washington	4	2	18	40
Andalusia	3	4	18	100
Beja District	3	3	14	100
Brittany	3	4	23	80
Castelo Branco District	3	2	12	66,66
Chaco Province	3	3	15	100
Fribourg	3	2	14	66,66
Istanbul	3	0	10	0
Lombardy	3	2	18	50
Masovian Voivodeship	3	2	14	66,66
North Holland	3	1	10	33,33
Rhineland-Palatinate	3	1	10	33,33
Shanghai	3	0	8	0
Szabolcs-Szatmar-Bereg	3	2	15	50
Uusimaa	3	3	16	50
Vila Real District	3	2	13	66,66
Wallonia	3	2	11	66,66
Aargau	2	1	7	50
Alto Parana Department	2	2	8	100
Bali	2	4	18	100

Budapest	2	2	16	50
Catalonia	2	2	12	100
Córdoba Province	2	1	8	50
County Cork	2	2	9	100
Evora District	2	1	7	50
Hanoi	2	1	11	33,33
Inhambane Province	2	1	9	50
Iowa	2	0	6	0
Jujuy	2	3	14	75
Kansas	2	0	6	0
Lazio	2	3	31	75
Lima Province	2	2	19	100
Lower Saxony	2	2	7	100
Madrid	2	0	7	0
Managua	2	1	15	50
Maputo Province	2	1	9	33,33
Misiones Province	2	1	14	25
Missouri	2	0	9	0
Norrbotten County	2	0	6	0
Nouvelle-Aquitaine	2	0	8	0
Piedmont	2	3	19	100
Portalegre District	2	1	7	50
South Holland	2	1	7	50
Tamil Nadu	2	0	8	0
Tibet	2	0	2	0
Veneto	2	2	8	100
Aichi	1	1	13	100
Alajuela Province	1	1	4	100
Amambay Department	1	1	5	100
Ankara	1	1	4	100
Antioquia	1	1	6	100
Auckland	1	3	15	100
Banten	1	1	6	100
Beirut Governorate	1	1	4	100
Bekes County	1	0	4	0
Benguela Province	1	1	5	100
Blagoevgrad Province	1	0	3	0

Bogotá	1	0	4	0
Bourgogne-Franche-Comte	1	0	5	0
Braganca District	1	1	4	100
Bratislava Region	1	1	4	100
Bremen	1	0	4	0
British Columbia	1	0	4	0
Cabinda Province	1	1	4	100
Canary Islands	1	0	8	0
Canton of Zug	1	1	3	100
Capital Municipality	1	0	5	0
Castile-La Mancha	1	0	3	0
Center District	1	1	4	100
Central Bohemian Region	1	0	4	0
Centre-Val de Loire	1	1	4	100
Conakry	1	1	4	100
Cordoba	1	2	17	100
Corrientes Province	1	1	12	100
Emilia-Romagna	1	1	7	100
Fort-de-France	1	1	18	100
Gauteng	1	1	11	50
Greater Accra Region	1	1	5	100
Groningen	1	1	7	100
Hamburg	1	1	4	100
Hidalgo	1	1	4	100
Krasnodar Krai	1	0	4	100
Limburg	1	1	10	33,33
Los Lagos	1	1	4	100
Louisiana	1	0	3	0
Madre de Dios	1	1	4	100
Manica Province	1	1	4	100
Massachusetts	1	0	3	0
North Carolina	1	0	3	0
Northern Ireland	1	1	11	100
Ouest Department	1	1	4	100
Overijssel	1	1	4	100
Queensland	1	1	4	100
Region of Murcia	1	1	8	100

Salta Province	1	1	4	100
San Jose Province	1	1	4	100
San Luis Province	1	1	4	100
São Domingos	1	0	5	0
Schaffhausen	1	0	1	0
Schwyz	1	0	4	0
Scotland	1	1	8	50
Shizuoka	1	1	9	100
Solothurn	1	0	4	0
Sonora	1	1	4	100
South Kalimantan	1	1	4	100
South Moravian Region	1	1	4	100
Stockholm County	1	1	4	100
Styria	1	0	7	0
Taranaki	1	1	5	100
Tbilisi	1	1	6	50
Texas	1	1	4	100
Thies Region	1	1	4	100
Thuringia	1	0	3	0
Tianjin	1	0	1	0
Ticino	1	0	3	0
Tuscany	1	0	3	0
Uíge Province	1	0	1	0
Umbria	1	0	4	0
Utrecht	1	1	4	100
Uttar Pradesh	1	2	13	100
Valais	1	1	6	100
Valparaíso	1	1	4	100
Vas County	1	1	4	100
West Bengal	1	0	4	0
Pest County	0	0	2	0
Tabela 21: Tabela com métricas sobre estados, regiões, distritos ou províncias (RAPLE) - o autor				

3.12.11 - Eventos por Região



Inferências

A análise descritiva dos dados revelou uma forte correlação positiva entre "Usuários ativos", "Sessões engajadas" e "Contagem de eventos", mas uma correlação negativa entre "Taxa de engajamento" e as outras três variáveis. Além da matriz de correlação, foram construídos gráficos de dispersão para confirmar os resultados obtidos.

Após realizar análises descritivas e gráficos de dispersão, aplicamos regressões lineares ou não lineares, conforme necessário, para investigar mais profundamente a relação entre as variáveis. Os resultados da regressão demonstraram que a relação entre 'Usuários ativos', 'Sessões engajadas' e 'Contagem de eventos' é bem representada por uma regressão linear.

A relação entre 'Taxa de engajamento' e as outras três variáveis é melhor modelada por uma regressão polinomial e a análise de regressão revela que "Usuários ativos", "Sessões engajadas" e "Contagem de eventos" são preditores significativos uns dos outros. No entanto, "Taxa de engajamento" não é um bom preditor de nenhuma das outras três variáveis.

O modelo de regressão polinomial para "Taxa de engajamento" sugere que pode haver uma relação mais complexa entre essa variável e as outras três, que não podem ser totalmente capturadas por um modelo linear simples.

3.12.12 - Análise de Distribuição por cidades

A capilaridade de alcance da RAPLE no Brasil e no exterior fica mais evidenciada com a análise dos dados por cidade. É notável o alcance e distribuição de usuários ativos no Brasil, mesmo em regiões remotas, como Manaus (81 usuários ativos), Macapá (61 usuários ativos) e Rio Branco (32 usuários ativos).

Os mapas cumprem um importante papel na visualização dinâmica dos dados e síntese da informação para criar conhecimento e variações conceituais dos dados nas análises multivariadas. A projeção de cada uma das colunas das tabelas nos mapas, propiciou um recorte que foi melhor adaptado para ser visualizado por continente em alguns casos, pois as regiões metropolitanas são grandes centros conurbados. É surpreendente que haja professores de línguas estrangeiras em tantos municípios de nosso país continental, onde as fronteiras físicas são superadas pela ubiquidade das redes informacionais, mostrando que o objetivo da difusão, e em certa medida, do letramento científico, estão sendo contemplados nos projetos de extensão aqui estudados e pesquisados.

Há um dado importante que merece ser estudado adequadamente sobre a distribuição de usuários ativos da RAPLE que corresponde ao fato da maioria percentual deles (1445 ou 12,39%) não definir quais as cidades de onde acessam. Como muitos países passaram a regular o uso e privacidade de dados em aplicativos e sites, uma quantidade massiva de usuários (quase 13%) não permite acessar sua localização, conseqüentemente há uma influência considerável no desvio padrão que precisará ser tratado nos dados.

As cidades representam muito mais que simples aglomerações urbanas por mera casualidade, sendo verdadeiras estruturas vivas que crescem a taxas de 15% a cada vez que dobra de tamanho, conforme as equações superlineares de Jeffrey West e Luís Bettencourt (BETTENCOURT, WEST et al., 2010). Elas indicam que as cidades são uma das invenções mais importantes da história humana, possibilitando otimizar nosso potencial econômico e liberar nossa engenhosidade.

"Quando começamos a viver em cidades, fizemos algo que nunca havia acontecido antes na história da vida. Desenvolver insights teóricos sobre elas que possam informar análises quantitativas de sua sustentabilidade a longo prazo em termos da interação entre inovação, apropriação de recursos e consumo e a composição de sua atividade social e econômica." (IBIDEM)

Ao analisar uma infinidade de dados de diferentes fontes e formatos sobre as cidades, West e sua equipe encontraram leis de escala e desenvolveram equações que permitiram percepções bastante precisas. Descobriram que quando cidades dobram de tamanho, exigem 15% menos infraestrutura como resultado da economia de escala, mas também as dinâmicas das relações sociais em rede incrementam outros 15% nos resultados socioeconômicos.

“À medida que as cidades dobram, elas geram 15% mais riqueza e empregos, bem como 15% mais crimes e doenças.” (IBIDEM)

Quando referenciadas pelo tamanho e população de uma cidade, as equações super lineares conseguem prever detalhes como comprimento das ruas e a quantidade de fios elétricos, número de postos de gasolina e de empresas de todos os tipos, faixa de renda, taxa de graduação universitária dos cidadãos, número de casos de AIDS, outras doenças e até mesmo a taxa de homicídios.

3.12.13 - Padrões de Engajamento nas Cidades Analisadas

Os dados revelam padrões de engajamento do usuário em várias cidades. Uma observação importante é a alta taxa de engajamento em grandes centros urbanos, como São Paulo, Rio de Janeiro e Brasília, por exemplo. Elas apresentam taxas de engajamento superiores a 50%, sugerindo que nesses centros urbanos, os usuários são mais propensos a interagir com o conteúdo e utilizar a plataforma por mais tempo.

As taxas de engajamento variam consideravelmente entre as cidades médias, como Araraquara, com taxas acima de 60%, quando comparada com Ribeirão Preto, com engajamento abaixo de 10%. Essa disparidade pode ser influenciada por fatores como acesso à internet, perfil demográfico dos usuários e tipo de conteúdo disponível.

Observamos uma correlação positiva entre o tempo médio de engajamento do usuário e as taxas de engajamento nas cidades analisadas. Cidades com maior tempo médio de engajamento por usuário tendem a apresentar taxas de engajamento mais elevadas.

Um exemplo disso é Araraquara, onde o tempo médio de engajamento de 204 segundos por usuário está associado a uma taxa de engajamento de 63%.

Cidades com muitos usuários ativos tendem a ter taxas de engajamento próximas da média e cidades com poucos usuários apresentam taxas de engajamento mais voláteis.

Analisar padrões de engajamento requer considerar fatores além dos já mencionados, sendo relevante investigar a influência de outras variáveis, tais como dados demográficos de idade e gênero, conteúdo consumido, qualidade e disponibilidade da internet em cada região, além de hábitos de uso e frequência de acesso ao site.

Engajamento por Usuário Ativo		Tempo de Engajamento por Usuário Ativo	
Cidade	Percentual	Cidade	Segundos
Boa Vista	69,49	Araraquara	204,43
Franca	68,52	São Paulo	56,23
Osasco	66,67	Belo Horizonte	51,88
São José do Rio Preto	66,15	Fortaleza	51,61
Goiânia	65,88	Salvador	51,07
Aracaju	65,96	Porto Alegre	52,32
Bauru	64,52	Rio de Janeiro	41,22
Araraquara	63,32	Curitiba	45,06
Brasília	63,57	Campinas	50,37
Salvador	63,49	Brasília	38,68
Tabela 22: Comparativo de engajamentos e tempo das cidades com maior acesso (RAPLE) - o autor			

Ao comparar as diferentes métricas acima que possuem unidades de medida diferenciadas (percentual e segundos), podemos inferir que apenas 3 cidades (Salvador, Araraquara e Brasília) estão presentes em ambas, com maiores engajamentos por percentual de interações e tempo de navegação. Essas cidades possuem métricas indicando haver usuários que acessam com maior frequência e por mais tempo os conteúdos do site.

Também é interessante notar que ao trocarmos a métrica de percentual (Engajamento por usuários ativo) por uma contagem numérica das “Sessões engajadas por usuários ativos” e multiplicarmos pelo “Tempo médio de engajamento”, obtemos um valor que representa o produto da quantidade total de segundos de cada cidade.

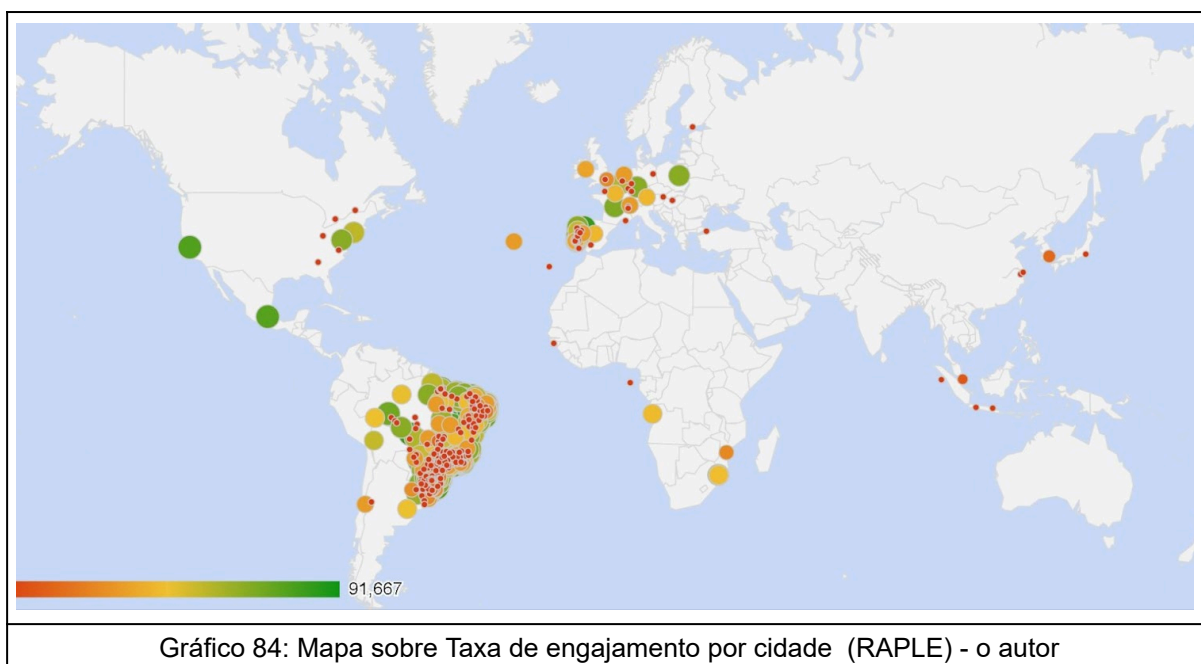
Novamente notamos as mesmas três cidades (Salvador, Araraquara e Brasília) da análise anterior nesta comparação de métricas, mas acrescidas de todas as outras cidades da métrica de tempo de engajamento (segundos). Essa característica pode indicar que a métrica de engajamento por si só, não explica essa classificação de reordenação pelo produto (Sessões x Tempo).

Cidade	Sessões Engajadas por Usuário Ativo	Tempo Médio de Engajamento (segundos)	Produto (Sessões x Tempo)	Minutos Totais
Araraquara	204,42	204,43	41820,82	697,01
São Paulo	81,98	56,23	4609,94	76,83
Campinas	82,1	50,37	4135,46	68,92
Salvador	72,73	51,07	3713,41	61,89
Belo Horizonte	70,7	51,88	3667,82	61,13
Fortaleza	69,44	51,61	3586,32	59,77
Porto Alegre	68,18	52,32	3570,77	59,51
Curitiba	68,71	45,06	3095,87	51,60
Brasília	72,15	38,68	2787,82	46,46
Rio de Janeiro	66,47	41,22	2744,77	45,75

Tabela 23: Métricas comparativas entre as cidades que mais acessam a RAPLE - o autor

Este dataframe fornece insights sobre o engajamento em diferentes cidades, possibilitando uma análise da distribuição geográfica desses indicadores. Esta análise é fundamental para entender como o interesse e a interação do público variam em diferentes localidades. Nota-se que o engajamento se concentra nas grandes cidades, como São Paulo, Rio de Janeiro e Belo Horizonte. Devido à sua maior densidade populacional, essas cidades possuem uma base de usuários potencialmente maior, influenciando as métricas de engajamento.

A Taxa de Engajamento mede a porcentagem de usuários ativos que interagem com a plataforma ou campanha. O Tempo Médio de Engajamento indica a duração média da interação do usuário com o conteúdo. Essas taxas variam significativamente entre diferentes regiões do país, também devido a fatores socioeconômicos, como renda e acesso à internet, além de hábitos de consumo de mídia.



Temos analisado as particularidades da distribuição geográfica dos engajamentos de forma mais granular possível, buscando identificar padrões e tendências específicas de cada localidade pelas métricas complementares de usuários ativos, novos usuários, sessões engajadas e contagem de eventos, fornecendo uma visão mais completa do comportamento do público.

As taxas de engajamento observadas no dataframe variam consideravelmente entre as cidades e a análise mais detalhada é recomendável para compreender os fatores que influenciam o engajamento em cada localidade.

As cidades de Araraquara, Brasília e Salvador destacam-se com taxas de engajamento acima de 60%, sugerindo boa aderência à plataforma e conteúdo do público dessas cidades. Outras cidades com taxas elevadas incluem Maceió (74,42%), Boa Vista (69,49%), Franca (68,52%), Osasco (66,67%), São Gonçalo (61,11%) e Blumenau (54,35%) .

Por outro lado a cidade de Ribeirão Preto, que detém bons índices de desenvolvimento econômico e disponibilidade de acesso à internet de alta velocidade, alcançou baixa taxa de engajamento (6,16%), semelhante a Guarulhos (a segunda cidade mais populosa do estado e 13.^a mais populosa do Brasil), obteve taxas de engajamento significativamente baixas para uma cidade com mais de 1.290.000 habitantes.



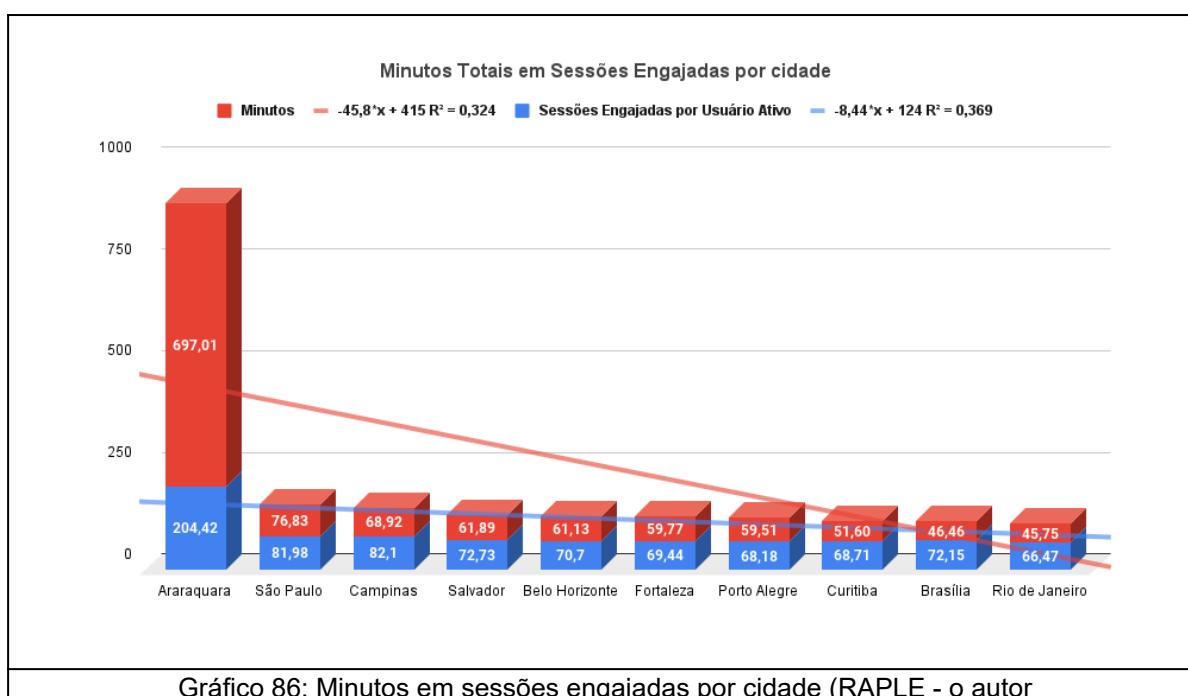
A taxa de engajamento não parece estar diretamente correlacionada ao número de usuários em cidades com grande base de usuários, como São Paulo e Rio de Janeiro, pois apresentam taxas de engajamento moderadas, enquanto cidades menores com menos usuários exibem taxas mais altas.

3.12.14 - Minutos Totais em Sessões Engajadas

Araraquara se destaca com 697,01 minutos, indicando um alto engajamento dos usuários na plataforma, seja em termos de tempo de uso ou número de usuários. As demais cidades apresentam valores bem mais baixos, entre 45 e 80 minutos, com uma tendência de queda à direita no gráfico.

Apesar da proximidade geográfica com Araraquara, as cidades de São Paulo e Campinas apresentam concentrações significativamente menores, sugerindo que a alta concentração observada nela é um fenômeno específico da cidade, e não um reflexo de uma tendência regional.

A linha de tendência (em vermelho) com $R^2 = 0,324$ indica uma correlação negativa moderada entre a posição da cidade no gráfico e o total de minutos, embora essa relação não seja muito forte.



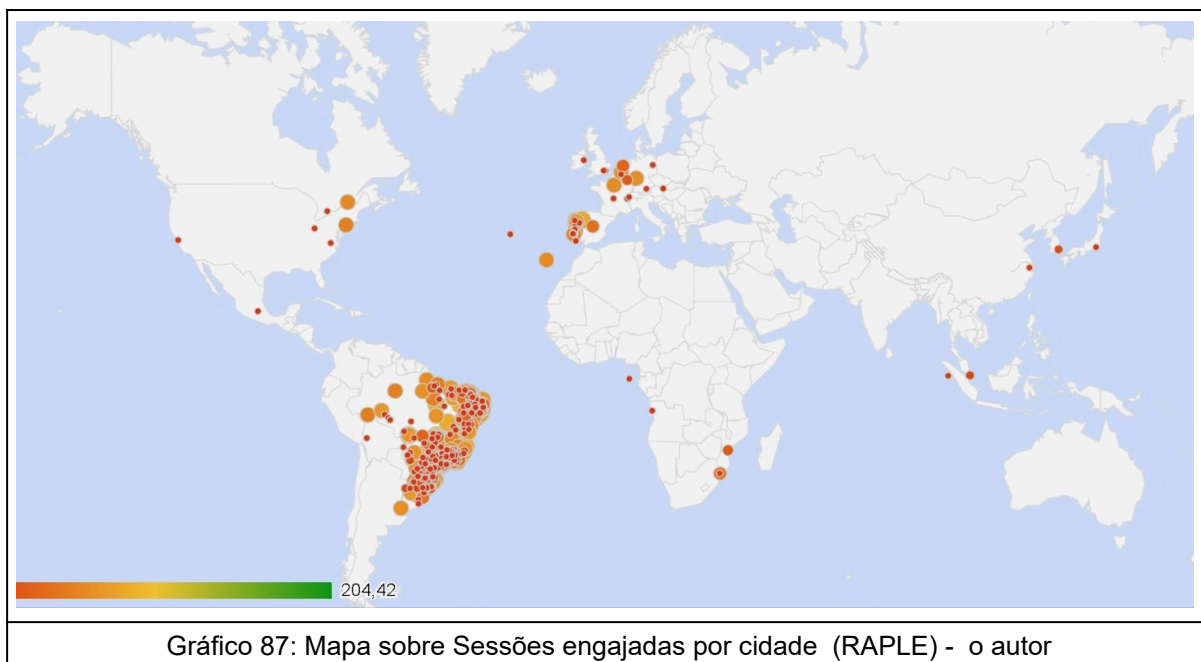
Sessões Engajadas por Usuário Ativo

Araraquara também lidera neste indicador, com cerca de 204 sessões por usuário ativo, o que reforça o alto engajamento na cidade. As demais cidades apresentam valores mais próximos entre si, variando entre 60 e 80 sessões.

Curiosamente, São Paulo tem o menor valor, apesar de ser uma grande metrópole. Isso pode indicar que os usuários em São Paulo, apesar de acessarem a plataforma, o fazem com menos frequência ou por períodos mais curtos. A linha de tendência (em azul) com $R^2 = 0,369$ também indica uma correlação negativa moderada.

Percepções

Araraquara obteve alta participação e engajamento dos usuários por ser a sede do projeto na Unesp. São Paulo apresenta um baixo engajamento em termos de minutos e número de sessões por usuário, concomitante a um expressivo número de usuários em potencial. Investigar as possíveis correlações entre essas variáveis pode explicar algum fator geográfico ou demográfico influenciando o engajamento, além da densidade populacional, e o perfil socioeconômico dos usuários.



3.12.15 - Usuários Ativos

A variável "Usuários Ativos" é um indicador-chave de desempenho (KPI - Key Performance Indicator) para avaliar o alcance e a penetração da plataforma e seus serviços digitais em diferentes cidades. Essa métrica quantifica o número total de usuários que se envolveram com a plataforma em um determinado período de tempo. Ao analisar essa variável em conjunto com outros indicadores, como a taxa de engajamento, é possível estabelecer um entendimento mais abrangente do desempenho da plataforma em cada localidade.

A análise revela uma concentração de usuários ativos em grandes centros urbanos como São Paulo (1445 usuários ativos), Rio de Janeiro (343) e Brasília (237). Essa tendência é esperada, considerando a maior densidade populacional e a maior penetração de internet nesses locais.

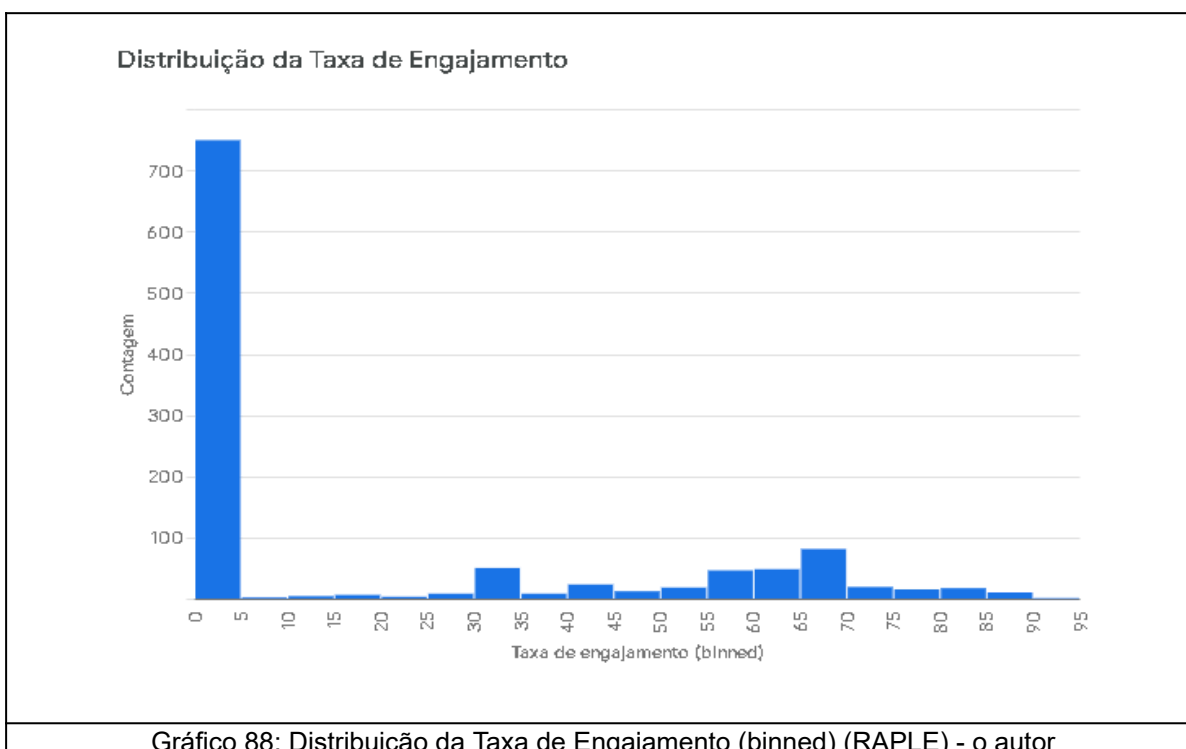
A plataforma também registra usuários ativos em cidades menores, como Araraquara (868) e Ribeirão Preto (130). Essa presença em localidades menores indica um potencial de expansão e crescimento da plataforma.

Como mencionado anteriormente, a correlação entre a taxa de envolvimento e o número de usuários ativos nem sempre é direta. É possível observar taxas de envolvimento moderadas em cidades com um grande número de usuários ativos e cidades menores com altas taxas de engajamento.

Esta análise combinada da variável "Usuários Ativos" com a taxa de engajamento e outros indicadores é crucial para entender a dinâmica de uso da plataforma em cada cidade. Além disso, a variável "Novos Usuários" — que indica o número de usuários que se cadastraram ou acessaram a plataforma pela primeira vez durante o período analisado — pode revelar o potencial de crescimento da plataforma em cada cidade.

A análise dos dados, abrangendo 1121 linhas e 15 colunas numéricas, revelou uma taxa média de engajamento de 19.01, com desvio padrão de 28.12. Os valores de engajamento observados variaram de 0 a 91.67 e o tempo médio de engajamento por usuário ativo foi de 37.83 segundos, com um desvio padrão de 71.89, variando de 0 a 1160 segundos. Existe uma correlação positiva fraca (0,18) entre a "Taxa de envolvimento" e as "Sessões engajadas por usuário ativo".

Correlação positiva muito forte (0,99) é evidente entre as variáveis "Usuários ativos" e "Novos usuários", assim como entre "Sessões engajadas" e "Contagem de eventos" (0,98). Ao plotarmos um histograma para observar a distribuição dos valores na coluna "Taxa de Engajamento", percebemos que a maioria está próxima de zero, indicando que a maioria das interações tem engajamento baixo.



A análise dos dados revela que a taxa de engajamento é relativamente baixa

na maioria das cidades, com a maior concentração de valores situando-se entre 0 e 10, além de um pico expressivo na faixa de 0 a 1. Contudo, a presença de uma cauda longa à direita na distribuição dos dados indica a existência de cidades com taxas de engajamento significativamente acima da média, sugerindo a presença de outliers, ou seja, cidades com níveis de engajamento excepcionalmente elevados.

Correlação com a Taxa de Engajamento

Uma plataforma com alta taxa de engajamento e um grande número de usuários ativos em uma cidade específica pode ser considerada bem-sucedida em atrair e manter usuários. Essa correlação positiva é evidente em cidades como São Paulo e Araraquara, que exibem esse padrão de uso.

As cidades com alto engajamento, mas poucos usuários ativos, podem representar um recorte com grande potencial de crescimento. As cidades com alto número de usuários ativos, mas com baixas taxas de engajamento, indicam possibilidades de melhorar a experiência do usuário para incentivar maior interação com a plataforma.

Correlação com Novos Usuários

Um alto número de novos usuários, aliado a um crescimento no número de usuários ativos, pode indicar que a plataforma está expandindo seu alcance e atraindo novos membros.



Gráfico 89: Mapa sobre Usuários ativos no mundo (RAPLE) - o autor

Há uma clara concentração de usuários ativos na América do Sul e Europa, notadamente concentrada também em alguns países populosos africanos, cujo idioma oficial é o português. Há uma capilaridade no interior do Brasil que reflete o alcance dos conteúdos em quase todo o país, mesmo em regiões remotas e de difícil acesso.

Aproximando a visualização na América Latina, nota-se elevada concentração de usuários em toda a faixa litorânea até quase a metade do interior do Brasil, correspondendo às regiões mais densamente povoadas. Nas áreas de grandes extensões agrárias e florestais nas regiões norte e centro-oeste há uma acentuada diminuição na capilaridade regional, embora apresentem um alcance e engajamento acima de algumas regiões altamente povoadas e de maior desenvolvimento socioeconômico.



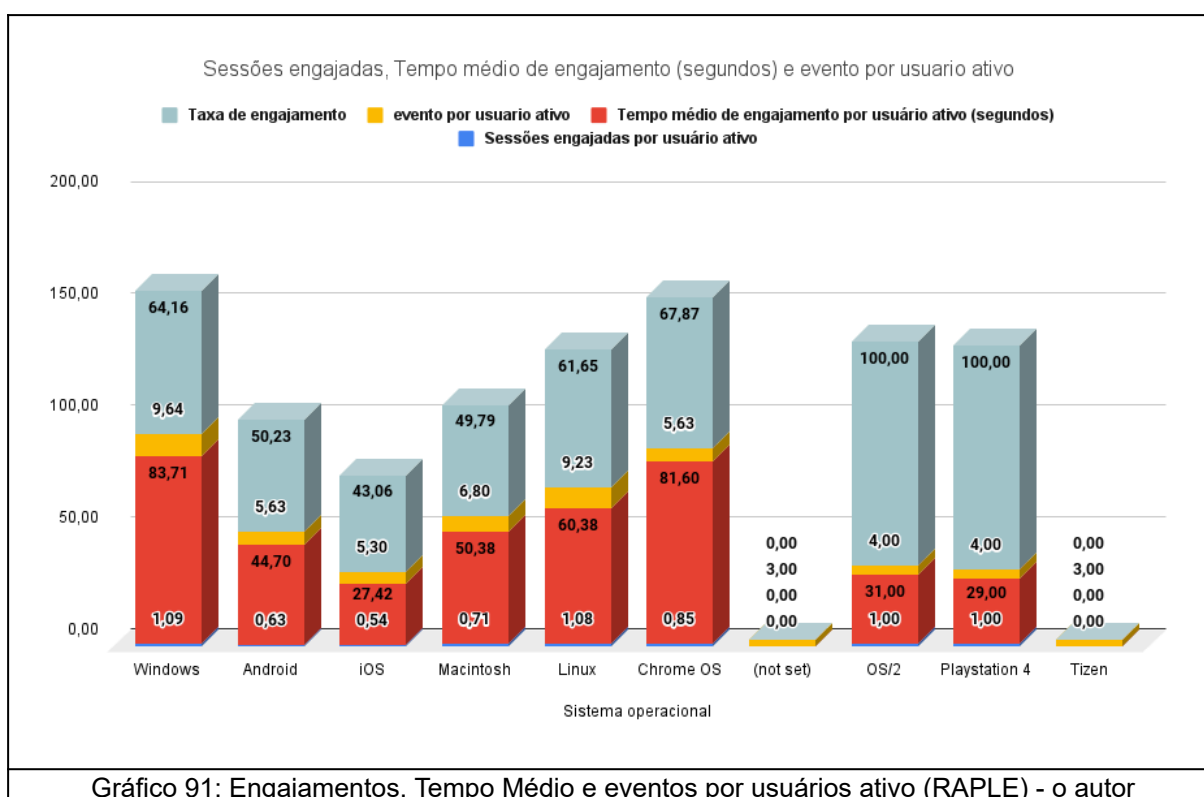
Gráfico 90: Usuários ativos na América Latina (RAPLE) - o autor

3.12.16-Uma Análise da Popularidade e Engajamento por Sistema Operacional

Outra recorte do dataframe fornece informações sobre a popularidade e o engajamento de diferentes sistemas operacionais. O Windows lidera com 5.999 usuários ativos, representando uma porcentagem significativa do total. Essa dominância reflete sua presença em diversos dispositivos, desde computadores

peçoais a servidores, seguido por Android, com 3.388 usuários e pelo iOS com 1.150 usuários.

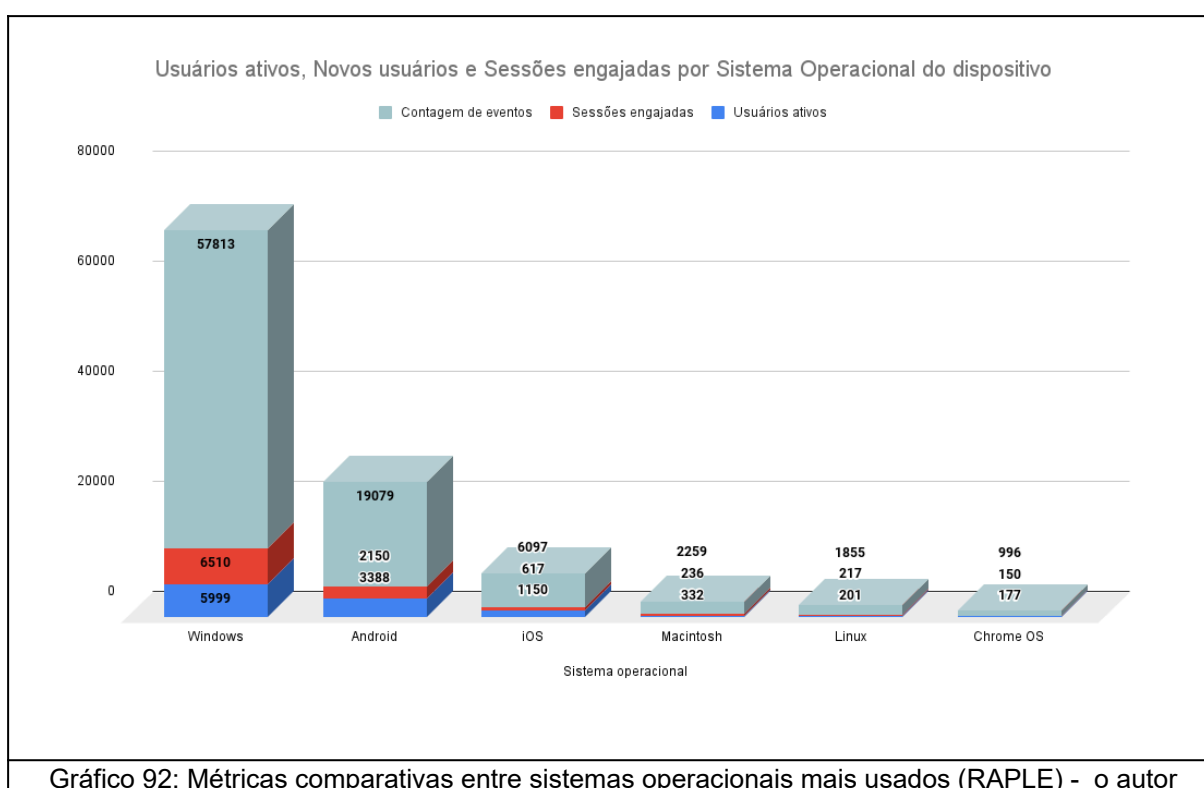
A tabela não apenas mede a popularidade, mas também avalia o engajamento do usuário, oferecendo uma visão detalhada de como os usuários interagem com cada sistema. Apesar de ter apenas um usuário ativo, o sistema OS/2 apresenta uma taxa de engajamento de 100%, com uma sessão engajada. O Playstation 4, também com apenas um usuário, demonstra o mesmo nível de engajamento. Essas altas taxas podem sugerir uma forte satisfação do usuário, mas é importante considerar o tamanho limitado da amostra.



Apesar de sua ampla popularidade, o Windows apresenta uma taxa de engajamento de 64,16%, com uma média de 1,09 sessões engajadas por usuário ativo. Isso indica que uma parcela significativa de usuários não está ativamente envolvida com a plataforma. Em contraste, o Chrome OS, com uma base de usuários menor (177 usuários ativos), demonstra uma taxa de engajamento superior, atingindo 67,87%, com 0,85 sessões engajadas por usuário ativo. Essa taxa de engajamento relativamente alta sugere um melhor nível de satisfação entre os usuários do Chrome OS.

A tabela também apresenta o tempo médio de engajamento por usuário ativo em segundos, o que oferece uma visão adicional sobre a utilização do sistema. O Windows apresenta o maior tempo de engajamento, com 83,71 segundos, seguido de perto pelo Chrome OS, com 81,60 segundos. Linux (60,38 segundos) e OS/2 (31 segundos) também demonstram tempos de engajamento consideráveis.

A análise da tabela revela uma variedade de informações relevantes sobre a popularidade e o engajamento em diferentes sistemas operacionais. Embora o Windows lidere em número de usuários ativos, as métricas de engajamento e tempo médio de uso indicam que outros sistemas operacionais, como Chrome OS e Linux, também apresentam níveis significativos de interação do usuário.



Sistema operacional	Usuário s ativos	Sessões engajadas	Taxa de engajam/o	Sessões engajadas por usuário ativo	Tempo médio engajam/o (segs)	Contagem eventos	Evento por usuário ativo
Windows	5999	6510	64,16	1,09	83,71	57813	9,64
Android	3388	2150	50,23	0,63	44,70	19079	5,63
iOS	1150	617	43,06	0,54	27,42	6097	5,30
Macintosh	332	236	49,79	0,71	50,38	2259	6,80
Linux	201	217	61,65	1,08	60,38	1855	9,23

Chrome OS	177	150	67,87	0,85	81,60	996	5,63
(not set)	2	0	0,00	0,00	0,00	6	3,00
OS/2	1	1	100,00	1,00	31,00	4	4,00
Playstation 4	1	1	100,00	1,00	29,00	4	4,00
Tizen	1	0	0,00	0,00	0,00	3	3,00

Tabela 24: Métricas e mapa de calor por sistema operacional (RAPLE) - o autor

A análise da tabela nos permite traçar um panorama abrangente do cenário dos sistemas operacionais. O Windows, com sua onipresença, lidera em popularidade, mas a taxa de engajamento e os tempos de uso revelam a força de outros sistemas como o Chrome OS e o Linux.

A alta taxa de engajamento em sistemas com base de usuários pequena, como o OS/2 e o Playstation 4, ressalta a importância de considerar diferentes métricas para avaliar a "popularidade" e o sucesso de um sistema operacional.

3.12.17 - Análises de interações por dispositivo

Cada evento representa uma ação específica realizada por um usuário em um site ou aplicativo e fornece informações importantes sobre o comportamento do mesmo, observando a distribuição dos eventos por categoria de dispositivo (desktop, mobile, tablet e smart TV). A análise permite entender como os usuários interagem com a plataforma em diferentes dispositivos, possibilitando a otimização da experiência de cada um deles.

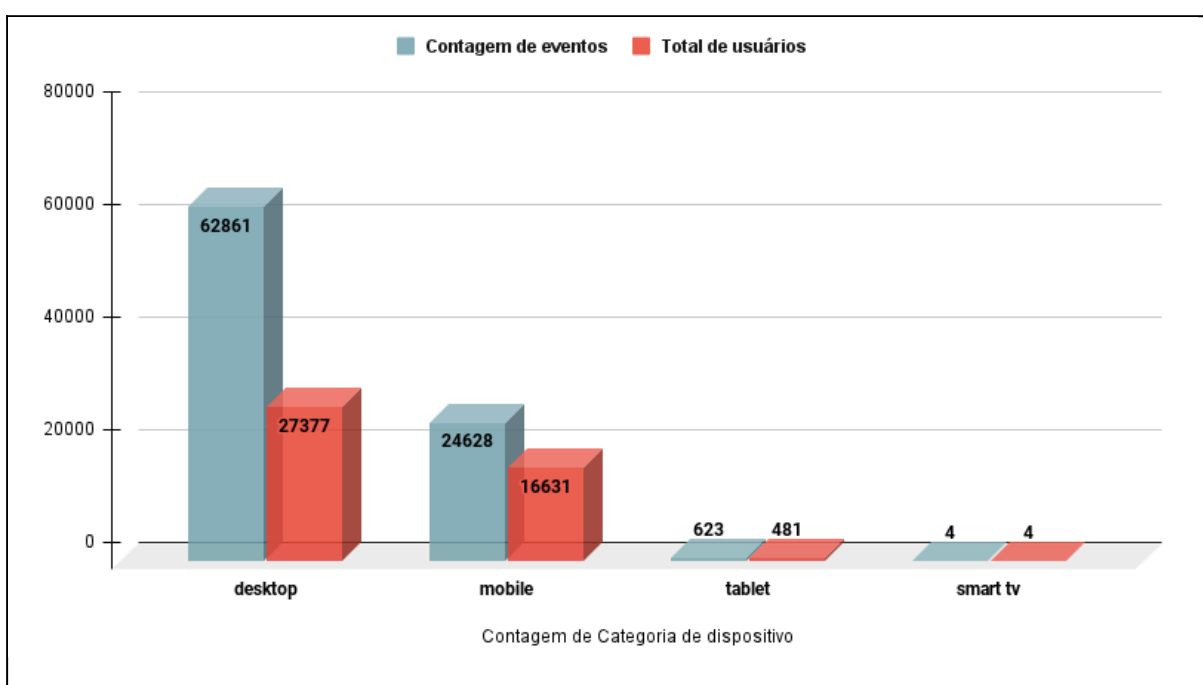


Gráfico 93: Eventos e usuários por dispositivo (RAPLE) -o autor

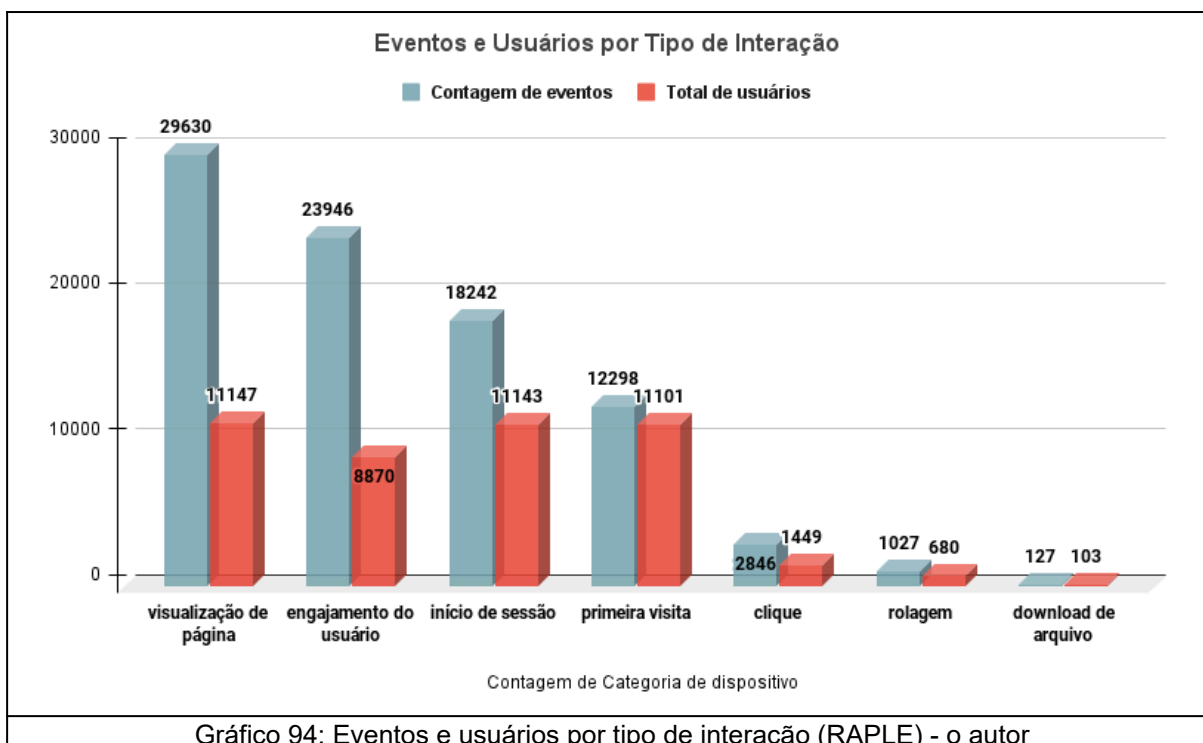
Alguns fatores podem contribuir para explicar as diferenças de uso em diferentes situações e lugares:

Visualizações de Página (page view): Os dispositivos de mesa (desktop e notebooks), com telas maiores, normalmente oferecem visualização e navegação de conteúdo mais fáceis em comparação com os dispositivos móveis. As telas menores dos dispositivos móveis podem restringir a quantidade de informações visíveis, levando a uma navegação mais focada e menos exibições de página.

O uso de desktops geralmente está associado a um ambiente mais propício à navegação, como o trabalho ou a residência. Já o uso de dispositivos móveis costuma ocorrer em situações mais dinâmicas e com maior propensão a interrupções como na rua, no transporte ou em lugares públicos, o que pode resultar em sessões de navegação mais curtas e com menos visualizações de página.

Nos desktops, os usuários tendem a explorar um site ou aplicativo mais detalhadamente, enquanto nos dispositivos móveis, eles podem estar mais concentrados em tarefas específicas, como encontrar informações rapidamente ou buscar sites, mas também assistem vídeos neles.

Embora o número de eventos de “page_view” seja menor nos *mobiles*, é fundamental reconhecer o número significativo de visualizações de página em dispositivos móveis. Isso ressalta a possibilidade de otimizar sites e redes sociais para estes dispositivos, garantindo uma experiência positiva para o usuário em plataformas diferentes.



Visualizações de página (page view): A categoria desktop lidera em número de visualizações de página (21457), indicando que usuários em desktops tendem a navegar mais extensivamente pelo site. Em dispositivos móveis, a contagem (7981) é significativamente menor, sugerindo sessões de navegação mais curtas ou focadas em tarefas específicas.

Nos tablets, a contagem (191) é ainda mais baixa, possivelmente devido à menor popularidade de tablets para navegação web e a presença de visualizações de página em Smart TVs, com apenas 1 evento registrado, aponta para a expansão da plataforma para novos dispositivos.

Engajamento do Usuário (user engagement): O engajamento do usuário é mais alto em desktops (18753), reforçando a ideia de que usuários nesta plataforma se engajam mais profundamente com o conteúdo. Em dispositivos móveis, o engajamento (5073) é menor, o que pode ser atribuído a fatores como maior propensão a distrações em telas menores.

O engajamento em tablets (119) é relativamente baixo, similar à tendência observada nas visualizações de página, enquanto um único evento de engajamento em Smart TV foi registrado.

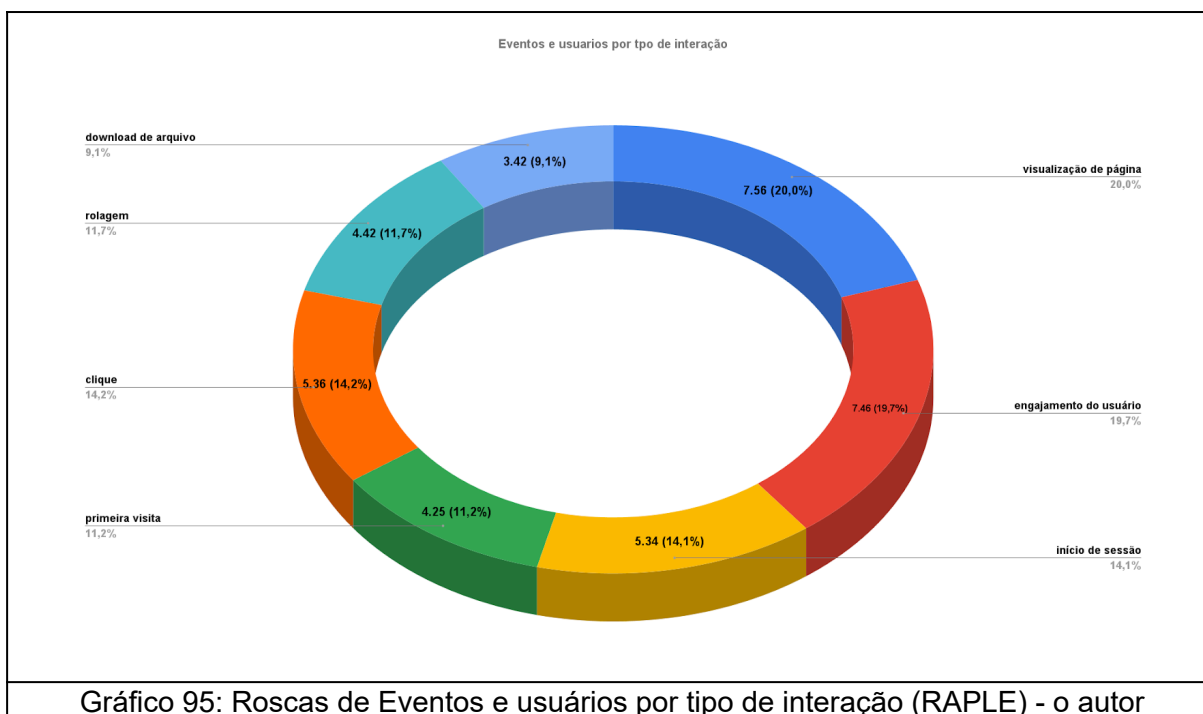
Início de Sessão (session_start): Os dados mostram que os *desktops* são a plataforma preferida para iniciar interações, com 12.200 sessões iniciadas em comparação com 5.894 em dispositivos móveis. Embora o uso de dispositivos móveis seja menor, ele continua significativo, enfatizando a importância de otimizar a plataforma para esses dispositivos.

Tablets registram um número menor de inícios de sessão (147), consistente com a tendência observada em outros eventos. As *Smart TVs*, como nos demais, registram apenas uma sessão iniciada, indicando um uso ainda incipiente.

Primeira Visita (first_visit):

O número de primeiras visitas em *desktops* (7387) é considerável, indicando um fluxo constante de novos usuários que escolhem essa plataforma como ponto de entrada. Os *mobile* também apresentam um número significativo de primeiras visitas (4779), seguido pelos *Tablets*, com uma contagem menor de primeiras visitas (131), o que pode indicar menor atratividade para novos usuários. As *Smart TVs* tiveram apenas uma primeira visita, sugerindo que a plataforma precisa ser melhor analisada em aspectos operacionais da plataforma ou mesmo de metadados no ambiente de aplicativos desses aparelhos.

Eventos Adicionais (clique, scroll, file_download): Os eventos de "clique", "rolagem" e "download de arquivo" seguem, em geral, o padrão observado nos eventos principais, com *desktops* liderando em contagem, seguidos por *mobile* e *tablets*. A presença desses eventos em todas as categorias de dispositivo reforça a importância de monitorar e otimizar a experiência do usuário em cada plataforma.



Percepções:

A análise dos dados sugere que desktops são a plataforma preferida para acessar e interagir com o site, com usuários engajando-se mais profundamente com o conteúdo. Dispositivos móveis, por sua vez, demonstram crescente relevância, com um número expressivo de visualizações de página, engajamento e início de sessões. Tablets e Smart TVs, embora com menor participação, representam novas oportunidades de expansão e exigem atenção em termos de design e usabilidade.

Eventos Mais Frequentes: Os eventos "visualizações de página" (page_view), "engajamento do usuário" (user_engagement) e "início de sessão" (session_start) são os mais frequentes em todas as categorias de dispositivo. Isso indica que usuários estão ativamente acessando a plataforma, interagindo com o conteúdo e realizando ações básicas de navegação.

Eventos Menos Frequentes: Eventos como "download de arquivo" (file_download), "clique" (click) e "scroll" (rolagem) ocorrem com menor frequência, sugerindo que essas ações são menos comuns ou que a interface da plataforma não está otimizada para incentivar tais interações.

Recomendações:

Implantar um framework ou plug-in de desenvolvimento de design responsivo, buscando uma experiência de usuário otimizada em todos os dispositivos. Também é

possível monitorar continuamente os eventos de usuário em cada categoria de dispositivo para identificar novas áreas em potencial, como as Smart TVs considerando as particularidades de interação neste dispositivo.

4 ANÁLISE DE CONTEÚDOS EM REDES COMPLEXAS POR APRENDIZAGEM DE MÁQUINAS RAG-LLM

As redes dos projetos de extensão pesquisados são consideradas complexas por suas características dinâmicas de comunicação, não linearidade, retroalimentação e adaptatividade. Sob a mesma ótica de considerarmos a internet um sistema social auto-organizador, as redes pesquisadas são formadas de pessoas com pouca ou nenhuma supervisão central, que executam tarefas simples, como subir e gerir páginas na rede, criar conteúdos e ligações a outras páginas. Estes sistemas em rede agem de forma sincronizada, informacional e adaptativa, fazendo emergir padrões eficientes de ação, tal como um sistema imunológico, uma rede neural profunda ou um *grid* de energia elétrica .

Nessas Redes de Conhecimento, as conexões (arestas) se dão por relações semânticas, associações, citações em nós que recriam conceitos, ideias e informações (conteúdos). Nelas, a evolução do conhecimento ao longo do tempo, a diversidade de fontes de informação e sua retroalimentação através da difusão e do letramento científico, as tornam resilientes e dinâmicas.

Podemos indicar 8 características delas, inerentes à complexidade, segundo (MITCHELL, 2019; WATTS E STROGATZ, 1998; WILENSKY E RAND, 2019; BARABÁSI, 2009 e BERTELLE et.al, 2009), sendo que nem todas estão presentes simultaneamente em todas as redes de conhecimento.

1. Dinamicidade: As redes de conhecimento não sendo estáticas, estão em constante evolução à medida que novos nós são acrescentados, as ligações são transformadas e o próprio conhecimento muda.

2. Heterogeneidade: São constituídas por diversos elementos, incluindo pessoas com diferentes competências, informações em vários formatos e diferentes ferramentas e tecnologias, podendo gerar ciência e inovação.

3. Interdependência: Os nós da rede de conhecimento não são isolados e influenciam uns aos outros. As ações e decisões de um nó podem afetar os outros, e o conhecimento individual é moldado pela interação com o coletivo.

4. Emergência: Podem gerar resultados não previstos ou planejados individualmente. A interação entre nós incrementa a criação de novas ideias e a capacidade aprimorada de inovação, superando capacidades individuais.

5. Auto-organização: Tendem a se auto-organizar, encontrando padrões e estruturas sem direção centralizada, permitindo que a rede se adapte organicamente às mudanças e aos desafios.

6. Não linearidade: Relações de causa e efeito na rede nem sempre são lineares e

pequenas mudanças em uma parte dela podem impactar significativamente toda ela.

7. Escala: O tamanho das redes de conhecimento pode variar de pequenos grupos a grandes comunidades globais, e a escala da rede afeta sua complexidade, exigindo diferentes abordagens e ferramentas de gerenciamento.

8. Cognição Distribuída: O conhecimento na rede existe tanto nos indivíduos quanto nas conexões entre eles, num tipo de “cognição distribuída”, possibilitando resolver problemas e criar conhecimento mais eficazmente do que os indivíduos sozinhos

A abordagem híbrida nesta tese utilizou recursos computacionais, matemáticos e estatísticos que contemplam algumas dessas características, onde foram aplicadas as melhores fórmulas e práticas possíveis na busca de análises e resultados que reflitam as dinâmicas comunicacionais e informacionais das redes de conhecimento nos dois projetos de extensão analisados.

No capítulo, há uma revisão bibliográfica com articulação de conhecimentos sobre as tecnologias envolvidas na análise de conteúdos e modelos de aprendizagem profunda de máquinas que analisam dinâmicas não lineares de comunicação e troca de saberes. As redes de conhecimento ultrapassam o domínio virtual da internet e se entrelaçam com a rede de colaboradores reais, que também são nós da rede complexa, que hoje é distribuída nos cinco continentes, entre diversas instituições e arranjos sociotécnicos (LATOUR, 2000)

Também realizamos uma breve cronologia das duas vertentes científicas que originaram os atuais avanços em Aprendizado de Máquina e Agentes de Conhecimento. Esses cientistas, alguns ganhadores de Prêmios Nobel⁵⁷, ora se dividiam, ora se agrupavam como adeptos da abordagem simbólica (linearidade, matemática, computação), e os que defendiam uma abordagem conexionista (não linearidade, estatística e redes neurais artificiais).

Nessa cronologia, demonstramos a origem de cada vertente em sua essência conceitual, matemática ou estatística para explicar que a concepção do Perceptron, baseada na visão reducionista e cartesiana com retas e funções lineares, não dava conta de realizar simples operações booleanas combinadas e não servia para lidar com a complexidade do mundo natural.

O Demônio de Maxwell, um dos maiores paradigmas de refutação científica da história, nos remete à última lei geral da física, a 2a. Lei da Termodinâmica de

⁵⁷ <https://jornal.usp.br/radio-usp/nobel-um-premio-cercado-de-prestigio-e-controversias/>

Rudolf Clausius, Lord Kelvin, Constantin Carathéodory e Max Planck (PRIGOGINE, 1996) que ensejou a criação de uma base de conhecimento, chamada mecânica ou física estatística. Ambas as vertentes se beneficiaram da física estatística, cada qual em sua particularidade, criando algoritmos que foram verdadeiros catalisadores tecnológicos e alavancaram a telefonia, a computação, as redes neurais e a biologia.

A Constante de Boltzmann foi desenvolvida no final do século XIX e foi um ponto de virada nas ciências naturais. Boltzmann fez a primeira conexão entre entropia e probabilidade em 1877, mas a relação só foi expressa em uma constante por Max Planck em 1900.

4.1 - Ciências da complexidade

A multiplicidade de subáreas dentro das Ciências Sociais deu origem a um ecossistema de metodologias científicas, que inclui não apenas métodos estabelecidos e amplamente utilizados em Sociologia, Ciência Política e Antropologia, mas também novas abordagens que visam atender à diversidade de conhecimentos necessários para estudar e investigar sistemas sociais complexos.

A complexidade não se refere a coisas complicadas, mas articula múltiplas disciplinas ao oferecer novas abordagens em campos tradicionais. Embora uma única definição seja difícil de ser encontrada, ela pode ser ilustrada examinando-se os fenômenos estudados por cientistas de sistemas complexos, suas disciplinas, objetivos e metodologias. (MITCHELL, 2009).

Em uma abordagem metodológica que caracteriza a sociedade como um sistema complexo, é muito relevante analisarmos se as características do objeto de pesquisa apresentam propriedades inerentes à complexidade. Sistemas financeiros e bancários, redes de computadores, redes sociais ou de cooperação científica, instituições e até mesmo cidades, cooperam em sinergia, e embora haja coordenação de algumas funções entre os órgãos, não possui uma coordenação central (WATTS E STROGATZ, 1998).

A ciência da complexidade, ao contrário da física, não pode usar modelos simples para prever a dinâmica de sistemas complexos como o genoma, o cérebro ou a sociedade, pois eles não podem ser descritos de forma concisa por leis gerais, o que é uma característica fundamental dos sistemas complexos.

As principais características dos sistemas complexos incluem sua construção a partir de componentes simples chamados agentes (WILENSKY E RAND, 2019). Esses agentes interagem de uma forma não linear que desafia o mapeamento simplista ou mecanicista, levando ao princípio central de que o todo é maior do que a soma de suas partes. Esse conceito se originou da noção aristotélica de holismo (ARISTÓTELES, 2002) e foi reafirmado nas Regras do Método Sociológico de Durkheim (DURKHEIM, 2002).

A rede mundial de computadores conectados pelo protocolo IP (Internet protocol) permitiu o surgimento daquilo que chamamos de internet, que pode ser considerada como um sistema social auto-organizador. Nela, pessoas, com pouca ou nenhuma supervisão central, executam tarefas simples: afixar páginas na rede e criar ligações a outras páginas. Alguns desses sistemas agem de forma sincronizada, trocam informações, apresentam emergência de padrões eficientes de ação e se adaptam (MITCHELL, 2009), tal como um sistema imunológico, entre outros.

... cientistas de sistemas complexos descobriram que a internet como um todo, tem muitas propriedades inesperadas em grande escala envolvendo a sua estrutura global, a forma como cresce, como a informação propaga-se sobre as suas ligações, e as relações coevolutivas entre os comportamento dos motores de busca e da estrutura de links da Web, todos os quais conduzem ao que poderia ser chamado comportamento "adaptativo" para o sistema como um todo (MITCHELL, 2009).

O comportamento estruturado não centralizado também é uma dessas propriedades, chamada de emergência, significando que o mesmo não pode ser compreendido apenas pelo comportamento dos agentes (WILENSKY e RAND, 2019), mesmo quando agrupados. Mesmo hierarquizados, alguns sistemas não podem ser explicados pela decomposição de suas partes. O processamento de informação também integra essas características e realimenta a tomada de decisões do sistema, realizando adequações em seu comportamento.

Essas adequações levam a mais uma característica dos sistemas complexos, que são dinâmicos e ajustam seus padrões de forma imprevisível, mas coerentes através do tempo (BERTELLE et.al, 2009). Ao reunir essas condições, pode-se inferir que os sistemas complexos evoluem, se adaptam e aprendem no sentido

darwiniano (DARWIN, 2013), mas também interagem, não só entre seus agentes, mas com o ambiente em que estão situados.

Os sistemas auto-organizados são muito relevantes em diferentes áreas de estudo dos sistemas complexos na biologia, ciências sociais e economia através de modelos de computador que tornam possível visualizar as interações que ocorrem nesse fenômeno natural. A auto-organização pode ser definida, de forma resumida, como a produção de padrões ou comportamentos altamente organizados em sistemas ou grupos, resultante das interações com informações limitadas entre seus membros, sem que haja nenhuma forma de controle central ou liderança de qualquer componente e originada de um comportamento emergente (MITCHELL, 2009).

Sistemas complexos não são estáticos e se adaptam de forma imprevisível ao retroalimentar suas informações em períodos de tempos relativamente curtos (MILLER, PAGE, 2009). Mesmo utilizando modelos computacionais para analisar seus dados, podem consumir recursos e quantidades de memória gigantescos que necessitam algoritmos que viabilizem sua simulação.

As informações do sistema circulam de maneiras peculiares e a imprevisibilidade de suas formas os tornam mais difíceis de avaliar, seja pela aleatoriedade das variáveis em que a mecânica estatística pode ser útil até certo ponto, mas outros aspectos da não-linearidade conectadas a propriedades escalares como leis de potência (MILLER, PAGE, 2009), cadeias causais (BERTELLE et.al, 2009), retroalimentações positivas ou negativas (MITCHELL, 2009) os tornam imprevisíveis, mas possíveis de estudar e melhorar o conhecimento humano acerca de si mesmo e da sociedade.

O estudo de sistemas complexos exige uma abordagem integrada, combinando disciplinas, objetivos e metodologias para garantir análises confiáveis e uma compreensão mais profunda do mundo real. Essa abordagem interdisciplinar pode levar a avanços na previsibilidade e no desenvolvimento científico e tecnológico, desafiando a ideia de uma Teoria Geral da Complexidade. A essência da ciência da complexidade está na integração da pesquisa teórica, do trabalho experimental e da simulação por computador, agora reconhecida como uma terceira

metodologia científica. Podemos descrever quatro disciplinas centrais que, articuladas, oferecem um arcabouço conceitual e científico para esses estudos:

Dinâmica	Estudo das estruturas em constante mudança no comportamento dos sistemas
Informação	Mapeia representações simbólicas e estuda o processo comunicacional entre os agentes e organismos
Computação	Fundamental para entender como as informações são processadas e influem nos comportamentos resultantes da dinâmica
Evolução	Avalia a adaptação e o aprendizado resultante das interações através do tempo
Quadro 17 - Propriedades comuns aos sistemas complexos ((MITCHELL, 2009)	

Ao fazer uma analogia com a computação digital criada pelos humanos, notamos que a computação biológica não é binária e nem estática, e sim ativa e analógica, sendo distribuída pelo tempo e espaço entre os componentes do sistema. No biológico o processamento é descentralizado, local, detalhado e estocástico com retroalimentação positiva ou negativa, com aleatoriedade perpassando todos os sistemas como vantagem (MITCHELL, 2009).

Os cientistas da computação estão se inspirando nos atributos de auto-organização da vida para desenvolver novos sistemas que exibem um comportamento emergente baseado em regras simples, como os algoritmos genéticos. Inspirados na Teoria da Evolução das Espécies de Darwin em redes neurais baseadas no cérebro humano, produzem algoritmos de proteção de computadores em rede inspirados no sistema imunológico e algoritmos de busca inspirados em fungos,

Diferentemente de sistemas simples que podem ser compreendidos pela análise isolada da agregação ou somatória de seus componentes, o conceito de emergência de padrões auto-organizados e de estruturas dissipativas, são características fundamentais da adaptação (PRIGOGINE, 1996).

Ele transcende os fundamentos da física teórica e vem se tornando um marco referencial de diversas disciplinas que vão da biologia à neurociência, da ecologia dos sistemas para a economia, das ciências sociais e urbanas e mais recentemente, das ciências das redes, onde várias propriedades emergentes estão relacionadas a sua estrutura sem escala e dinamicamente interativa (MITCHELL, 2009).

4.2 - Ciência das Redes

Propomos o uso de redes neurais por grafos em algoritmos de aprendizado de máquina para entender melhor as redes complexas, enfatizando que nenhuma metodologia única pode capturar totalmente a complexidade das relações entre os indivíduos. Busca-se a viabilidade técnica de desenvolver algoritmos de aprendizado de máquina que permitam visualizar e registrar as relações entre indivíduos em sua dinâmica comunicacional.

Esta pesquisa foi motivada pela constatação de que na atual sociedade em rede (CASTELLS, 1999), caracterizada pela convergência digital (BLACKMAN, 1998) e midiática (JENKINS, 2003), os indivíduos estão interconectados por várias redes tangíveis e intangíveis. Essas redes, como as de satélites, interoperam com servidores computacionais virtualizados através de redes intercontinentais, trafegando dados e conteúdos de povos de todas as nações se entrelaçando aos fluxos locais das emissoras de rádio e televisão no espectro regional, num contexto que podemos chamar de hipermídia (SANTAELLA, 2007).

Embora o espectro radioelétrico seja regulado pelos estados nacionais, seus contornos imateriais estão em constantes transformações, ocasionadas pela ubiquidade das tecnologias convergentes que transpassam as limitações impostas pelas concessões de faixas de frequência. Uma vez que os conteúdos hipermídia são pervasivos e tornam ubíquas as mediações entre usuários, independente da fronteira territorial, linguística, cultural ou política (SANTAELLA, 2007), também ultrapassam limites mercadológicos que concessões públicas de radiodifusão impõe para manter oligopólios de empresas de mídia.

As redes hipermídia são apenas uma camada das redes complexas que se entrelaçam em clusters (BARABÁSI, 2009) que concentram inúmeros outros conectores da rede que são pessoas (hubs), instituições, empresas ou qualquer outro agrupamento, físico ou virtual, que articula a troca, cíclica ou não, de fluxos de dados e informações que circulam na internet, monetizando inúmeras atividades humanas. Não possuem nenhuma ordem de nó fixo ou ponto de referência e frequentemente são dinâmicas, não lineares e com características multimodais

4.3 - Complexidade, clusters e hubs

A Ciência de Redes é um campo em rápido crescimento dentro da Ciência de Sistemas Complexos devido à prevalência de redes na natureza, na sociedade e na tecnologia, como a Internet, redes elétricas, sistemas bancários e plataformas de mídia social, todas estudadas nesse campo.

O estudo metodológico e a análise de redes originaram-se da teoria matemática dos grafos de Euler, de 1736, e do conceito antropológico de atores sociais por meio de imersão ou socialização secundária (Soares, 1995). A moderna ciência das redes teve início somente após o impacto significativo do artigo de Paul Erdős e Alfréd Rényi, de 1959, sobre o estudo de redes aleatórias na teoria dos grafos. Somente em 1973, o artigo sobre redes sociais “The Strength of Weak Ties” (A força dos laços fracos), de Mark Granovetter, sobre laços fortes e fracos, e o princípio do coeficiente de agrupamento do modelo de rede Small-World (Watts e Strogatz, 1998) atualizaram as teorias até então existentes.

O conceito de Barabási sobre “hubs” foi inovador em 2002, como indivíduos altamente conectados em uma rede, possibilitando o fenômeno dos “graus de separação”. Esses hubs, presentes em diversos sistemas complexos, como economias, células e redes metabólicas, conectam e influenciam vários outros elementos. Eles são uma propriedade fundamental da maioria das redes, surgindo espontaneamente como uma característica da complexidade. A análise das relações de rede, incluindo as conexões e seu peso, revela padrões de comunicação e troca dentro do sistema complexo.

Para melhor compreensão dos conectores e hubs, podemos mapear mais alguns itens da terminologia da área de redes, como o grau de um nó, que corresponde ao número de conexões a um nó ou saindo dele, podendo ser diretos ou indiretos, onde a conexão acontece via outro nó ou hub, existindo ainda os conceitos de “grau-in” indicando os links direcionados ao nó ou “grau-out” que mapeiam saídas dele. (OLIVEIRA, 2019)

O conceito de grau de distribuição da rede mostra quantos nós existem para cada grau e o conceito de pulo (hop), quando um nó chega a outro através de um link, podendo ser otimizado pelo menor caminho e o coeficiente de agregação, que também pode ser relacionado ao conceito de que “os amigos de meus amigos também são meus amigos” (GRANOVETTER, 1973). Ele se refere à razão entre as

ligações que conectam os vizinhos do nó, uns aos outros pelo maior número possível de ligações. Sendo a média dos coeficientes de todos os nós de agregação, um alto coeficiente indica uma rede pequeno mundo. (OLIVEIRA, 2019)

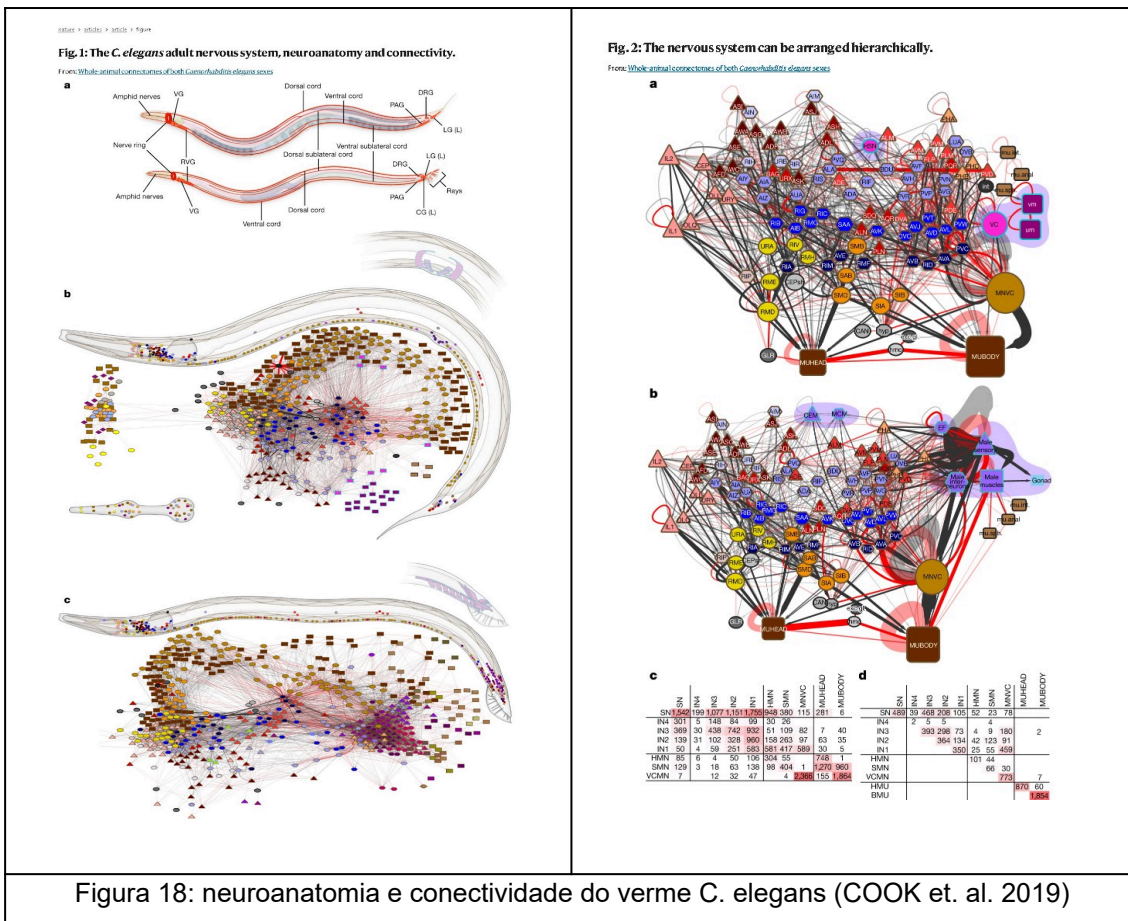
As redes de tráfego aéreo demonstram como as principais cidades atuam como hubs, conectando-se a rotas regionais. As interrupções climáticas nesses centros podem afetar toda a rede. A pesquisa de 1999 de Barabási revelou que a formação da rede segue um padrão baseado na popularidade de nós, e não aleatoriamente. Esse padrão “rico fica mais rico” indicou apenas 19 graus de separação entre duas páginas da Web, mostrando que as redes aleatórias são suscetíveis a interrupções aleatórias, as redes de mundo pequeno são mais resilientes.

4.4 - Redes neurais reais orgânicas (conectomas)

A rede neural de um *Caenorhabditis elegans*, um pequeno verme com 300 neurônios e aproximadamente 7.000 conexões, foi mapeada para exemplificar uma rede com osciladores biológicos. Na representação computacional da conectividade dessas sinapses, uma esfera representa um neurônio, e os links entre elas representam a junção sináptica das conexões. (ACHACOSO et. al, 1989). O tamanho de cada esfera na rede representa o número de suas conexões. As esferas maiores, chamadas de hubs, têm mais conexões, demonstrando que as conexões da rede não são distribuídas igualmente.

As microfotografias eletrônicas do verme publicadas em 1986 de White et. al., 1986) inspiraram o primeiro conectoma que a dupla Yamamoto e Achacoso(1992) publicaram em artigo cujo o subtítulo fazia uma analogia comparativa de que um macaco seria igual a 33 milhões de *Caenorhabditis elegans*, para ilustrar a quantidade aproximada de neurônios do cérebro humano.

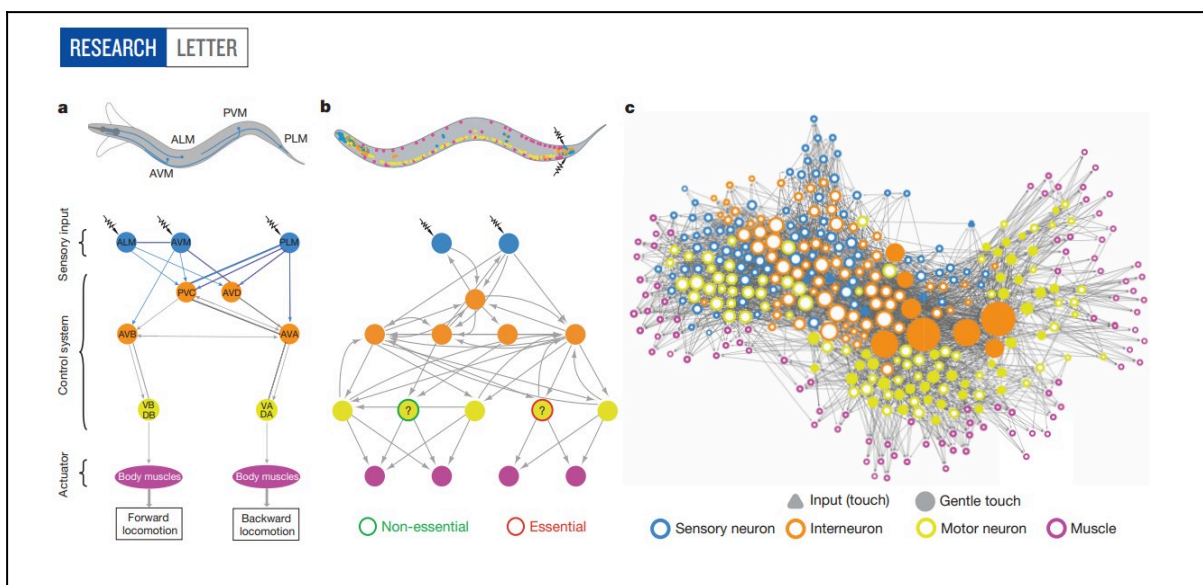
Segundo o artigo, o verme possui sistema nervoso hermafrodita em um complemento total de 302 neurônios, agrupados em 118 classes deles, com morfologias e conectividades semelhantes, em cerca de 5.000 sinapses químicas, 2.000 junções neuromusculares e 600 junções comunicantes (YAMAMOTO E ACHACOSO, 1992).



Afirmam que apesar de avanços teóricos desses estudos, ainda faltava a prova experimental empírica de validação, conseguida pela aplicação de um quadro de controle ao sistema de vias neurais do cérebro e sistema nervoso do *C. elegans*, quando considerado coletivamente (conectoma⁵⁸). O experimento mapeou o envolvimento de cada neurônio nos comportamentos locomotores do verme.

⁵⁸ o sistema de vias neurais em um cérebro ou sistema nervoso, considerado coletivamente.

Figura 19 - Controlando a rede neural de *C. elegans*. (A) Esquema do circuito neural para resposta locomotora ao toque suave em *C. elegans* (adaptado) . (B) Representação gráfica do quadro de controle proposto. Segundo os princípios ilustrados, se a remoção de um neurônio perturbar a controlabilidade dos músculos, a designamos “essencial para a locomoção”; se não, chamamos-lhe “não essencial”. Para fazer esta avaliação, primeiro mapearam os comportamentos locomotores de *C. elegans* reativos numa rede alvo problema de controle, perguntando até que ponto os neurônios sensoriais (azuis) podem controlar os músculos (rosa). Isto nos permite prever antecipadamente o envolvimento desconhecido do PDB na locomoção de *C. elegans*, e diferenças funcionais entre os neurônios individuais dentro da classe neuronal DD. (C) O conectoma usado neste estudo, consiste em 279 neuronios (os 282 neurônios não-faríngeos, excluindo o CANL/R e VC06 que não se conectam com o resto da rede) e 95 músculos. Cada nó é proporcional à soma dos seus graus de entrada e saída. Os nós preenchidos representam os neurônios tradicionalmente designados para os circuitos responsáveis por resposta suave ao toque, sugerindo a complexidade da previsão neuronal apenas a partir do diagrama de conectividade. (YAN et. al., 2017)



Ao propor a utilização dos neurônios motores como alvos para o controle locomotor, conseguiram melhorar as previsões, embora haja fatores que afetam a precisão dos dados do conectoma, como erros humanos de mapeamento e identificação de sinapses e pelo fato de terem reconstruído apenas dois animais.

O estudo também mensurou a probabilidade de determinada classe de neurônios prevista como essencial à locomoção permanecer nessa função, mesmo após a eliminação e adição aleatória de elos fracos e reconexões das ligações (sinapses). Foi observado que o controle dos músculos pelos neurônios motores requerem 12 classes neuronais, permitindo analisar o envolvimento de neurônios individuais em cada classe.

4.5 - O Conexionismo (conectoma) e o Simbolismo (perceptrons)

A neurociência computacional já foi considerada um estudo teórico das conexões cerebrais através de modelos matemáticos e computacionais, sendo atualmente um campo de pesquisa avançada que utiliza equipamentos médicos, como tomógrafos e ressonância magnética. Conectados a bases de dados em redes de computadores de larga capacidade, processam dados para gerar imagens tridimensionais das estruturas e conexões cerebrais em funcionamento, como conectomas usados para compreender a dinâmica da atividade neuronal, por exemplo.

Ao analisar e classificar os dados obtidos, se obtém informações sobre as conexões neuronais em escala micro que permitem gerar imagens de mapeamento cerebral em nível celular e orgânico, obtendo-se mapeamento dinâmico do pensamento ou do aprendizado. É uma ciência cara e sofisticada que gasta muito dinheiro de universidades e centros de pesquisa em todo mundo com o intuito de modelar um arcabouço conceitual teórico e empírico sobre a plasticidade do cérebro e a sua complexidade.

Teoricamente, o campo afirma que as funções cognitivas podem ser compreendidas por meio do estudo da comunicação entre diferentes elementos do cérebro, como regiões cerebrais. As imagens processadas mostram mapas da atividade cerebral modelados em formato gráfico linear do sinal hemodinâmico no cérebro (SALA et. el., 2023). Esta técnica de imagem por ressonância magnética (MRI) corresponde a quase 70% dos estudos de conectividade funcional do cérebro, e vem crescendo exponencialmente nos últimos 30 anos (IBIDEM).

Embora tenha evoluído muito nesse período, a perspectiva ideal para muitos pesquisadores é que apenas a técnica de mapeamento hemodinâmico não contempla toda a complexa interação de sinais bioquímicos e elétricos que emergem da comunicação inter-regional cerebral.

O estudo sugere uma perspectiva multimodal integrada com tomografia por emissão de pósitrons (PET) para medir a atividade neural, neurotransmissão e proteinopatias na cognição normal e patológica (IBIDEM).

A neurociência já admite a não linearidade da conectividade neuronal e geral do cérebro propondo uma abordagem que contemple técnicas eletrofisiológicas e de

neuroimagem que mapeiam a rede multicamada de regiões e estruturas como um conectoma multidimensional.

Embora exista um núcleo estrutural composto por fibras neurais principais, as unidades do conectoma interagem continuamente no espaço e no tempo em nível funcional. O esqueleto estrutural pode ser capturado com dMRI, enquanto um conectoma funcional pode ser capturado com técnicas de ressonância magnética, ferramentas eletrofisiológicas e imagens moleculares (SALA et. el., 2023).

O ênfase neste caso sobre melhorias no conectoma como grafo da atividade neuronal nos remete aos sistemas dinâmicos não lineares, com diversas entidades interdependentes, conectadas e adaptativas, nas quais o conceito de informação é fundamental para a retroalimentação e a auto-organização.

A intenção de contextualizar este estudo de neurociência é salientar a emergência da percepção de pesquisadores da área ao proporem métodos integrados de coleta de dados dos exames por duas ou mais tecnologias disponíveis, cada qual com sua particularidade e vantagens. Essa técnica oportuniza uma coleta mais adequada e compreensiva de dados cerebrais estruturais, funcionais e sistêmicos na interação contínua do espaço-tempo, entregando um conectoma multidimensional atualizado com premissas da complexidade, como informação, ordem e entropia (MITCHELL, 2009).

4.5.1 - Informação, ordem e entropia

Existem relações entre informação e a caracterização de ordem e desordem em sistemas complexos. A entropia é uma forma de caracterizar a desordem em sistemas termodinâmicos em que se define a disponibilidade de energia pelo grau de entropia do mesmo, ou seja, quanto maior a desordem, menor grau de energia se dispõe. A segunda lei da termodinâmica⁵⁹ define que seriam naturalmente desordenados, o que implica na impossibilidade de construir máquinas de movimento perpétuo, pois o tempo possui uma direção, sendo essa, a direção da perda de energia de forma irreversível (PRIGOGINE, 1996).⁶⁰

No século XIX, Maxwell introduziu o paradoxo da 2ª lei da termodinâmica, que foi chamado do “Demônio de Maxwell”, simulando um ser que conseguiria

⁵⁹ Ilya Prigogine foi um cientista que desenvolveu uma extensão da segunda lei da termodinâmica. Ele propôs que o aumento da entropia não significa necessariamente desordem

⁶⁰ <https://www.scielo.br/j/rbef/a/Lp8LbrmfvhLstZnfRNW9zGz/?lang=pt>

separar em dois recipientes, moléculas que se movem rapidamente de outras que se movem mais lentamente. Ao possibilitar que trocassem de lado no recipiente, a entropia geral desse sistema seria diminuída, portanto ela não seria uma lei propriamente, mas uma certeza estatística, pois valeria apenas para uma grande quantidade de moléculas e não a elas individualmente (MITCHELL, 2019).

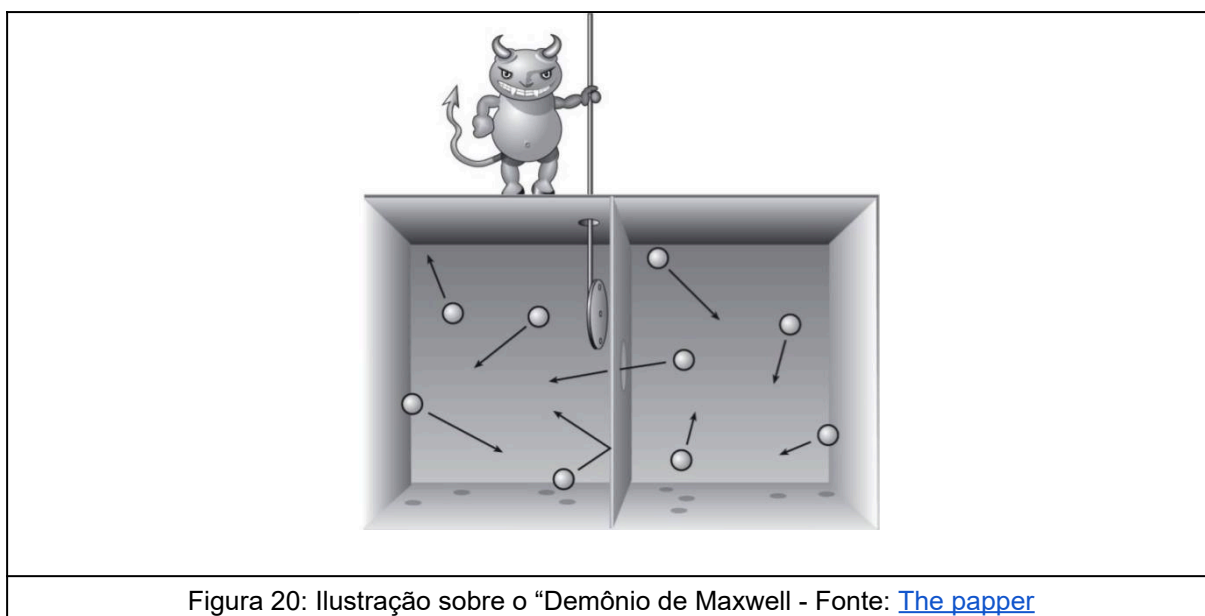


Figura 20: Ilustração sobre o “Demônio de Maxwell” - Fonte: [The papper](#)

Após muitas controvérsias sobre a possibilidade da refutação da 2ª. lei por esse paradoxo, em 1929, Leo Szilard propôs que para se obter a medição da velocidade das moléculas, o demônio precisaria de inteligência para as medições, o que insere no sistema o conceito de informação e que a medição e separação do gás gastam energia, reafirmando a perda de energia e entropia (IBIDEM).

Szilard também propôs o conceito de um bit de informação⁶¹, que seria a medida necessária para a classificação da velocidade das moléculas, definindo-o como sim ou não a qualquer pergunta com duas possíveis respostas. Essa abordagem foi adotada pelo campo da ciência da computação e define a forma de processamento da informação em computadores até hoje (BERTELLE et.al, 2009).

Boltzmann propôs que a entropia seria uma função do número de microestados possíveis de elementos do sistema e que macroestados seriam a combinação de diferentes microestados, sendo que um sistema sempre progride de um macroestado resultante do número máximo de microestados, um número K

⁶¹ Szilard criou o bit (unidade básica de informação) e também a reação nuclear em cadeia, o princípio básico da bomba atômica. *Genius in the Shadows, A Biography of Leo Szilard, the Man Behind the Bomb*, Skyhorse

vezes um logaritmo natural do número W de microestados correspondentes, chamada de constante de Boltzmann.

Já em 1940, Claude Shannon trabalhando em questões relacionadas aos sinais nos fios das redes de telefonia, aplicou algumas ideias da mecânica estatística de Boltzmann, através da qual formulou que, uma fonte ou pessoa, origem de uma mensagem pode enviar a um receptor, diversas fontes de informação.

Em seu artigo "A Mathematical Theory of Communication" de 1948, estabeleceu os limites fundamentais no processamento de sinais e operações de comunicação, além do conceito de compressão de dados⁶².

Constante de Boltzmann		Teoria da Comunicação de Shannon	
Microestado	Descrição detalhada de cada componente de um sistema	Mensagem	Símbolos, números, palavras
Macroestado	Combinação de diferentes microestados	Fonte da mensagem	Conjunto de mensagens possíveis com possibilidades de envio de cada mensagem possível
Entropia	Assume que todos os microestados são igualmente possíveis	Conteúdo de informação	Seja M o número de mensagens possíveis, supõe que todas as mensagens sejam igualmente possíveis
S (macroestado) = $k \log W$	W é o número de microestados correspondente ao macroestado, S é medido em unidades definido por k (Joules por Kelvin)	H (fonte da mensagem)	$\log_2 M$, sendo H medido em bits por mensagem
$1,380649 \times 10^{-23} \text{ m}^2 \text{ kg s}^{-2} \text{ K}^{-1}$		$H(X) = - \sum_{i=1}^m p_i \log_2 p_i$	

Quadro 17: Comparativo básico entre Constante de Boltzmann e Teoria da Comunicação (o autor)

A auto-organização está intrinsecamente relacionada aos sistemas dinâmicos complexos e o conceito de informação é fundamental para a retroalimentação típica desses sistemas. Neles, a informação é transmitida, representada e comunicada

⁶² WEAVER, W. A teoria matemática da comunicação. COHN, Gabriel (Org.). *Comunicação e indústria cultural*. 3. ed. São Paulo: Nacional, 1977. [S.l.: s.n.]

como característica comum a diversos deles e pode oferecer visões iniciais de como funcionam e como podem ser qualificados (MITCHELL, 2009).

4.5.2 - A linearidade do Perceptron

O conectoma multidimensional resultante da ampliação da coleta e análise de dados por duas ou mais técnicas em aparelhos tecnologicamente diferentes, remete ao conceito central do connexionismo, que descreve a mente como resultado de processamentos cognitivos análogos a inúmeros circuitos computacionais simples de entrada, processamento e saída conectadas em redes. O aprendizado ocorre por mudanças nos padrões das conexões em pesos sinápticos resultantes da realimentação contínua e paralela das entradas na rede.

A fixação do conhecimento resulta da distribuição paralela de dados e informações codificadas em diferentes neurônios conectados em rede, pois no connexionismo, essa transformação não ocorre em série como na teoria da informação de Shannon, mas simultaneamente por interconexões distribuídas e processadas paralelamente em locais diversos do cérebro,

Esse fenômeno natural e poderoso está intrinsecamente ligado ao que chamamos de plasticidade cerebral e se relaciona à forma como ele processa, organiza e armazena como conhecimento, as informações resultantes das interações naturais com o mundo exterior.

Nossos sentidos e receptores sensoriais corporais enviam entradas contínuas de dados que, classificados e processados, retroalimentam os circuitos sensoriais e cognitivos dos seres humanos. As redes neurais artificiais foram um importante e concorrido campo de pesquisas que incluía grupos de neurocientistas, cientistas da computação e biólogos que iniciaram historicamente a área após a consolidação dos estudos biológicos das células como unidades constitutivas dos seres vivos em 1840 e dos avançados estudos dos médicos histologistas Santiago Ramón y Cajal e do italiano Camillo Golgi sobre as estruturas neuronais em 1894.

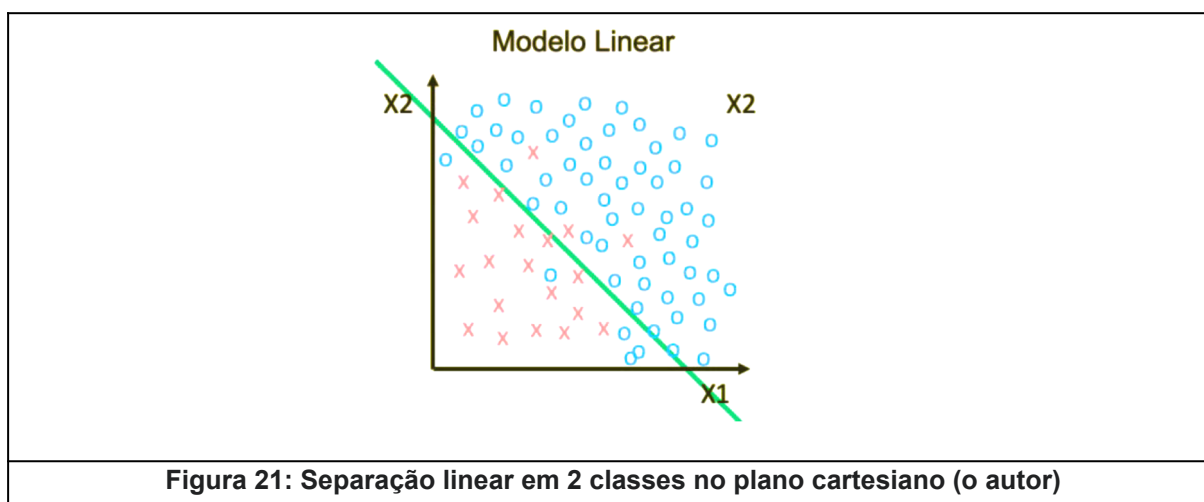
O método de coloração histológica baseado em nitrato de prata do italiano Golgi possibilitou ao espanhol teorizar, após experimentos empíricos, que o sistema nervoso era constituído por bilhões de neurônios que se comunicam entre si através de sinapses (BABINI e MARRANGHELLO, 2007). Esse estudo fundamentou a área

da neurociência que pesquisa não somente a fisiologia do cérebro humano com mais de dez bilhões de neurônios no córtex, mas também a conectividade do mesmo, ao mensurar que ele realize sessenta trilhões de sinapses entre essas conexões.

A concepção estrutural neuronal forneceu um impulso considerável em várias áreas médicas e científicas a partir do início do século XX, mas somente em 1943, em plena 2a. guerra, o psiquiatra e neuroanatomista Warren McCulloch e o matemático Walter Pitts conceberam o primeiro modelo de neurônio artificial baseado na estrutura cerebral humana e na dinâmica conexionista dos neurônios biológicos (BABINI e MARRANGHELLO, 2007); (HAYKIN, 2001).

Nas descrições tradicionais da organização neural, assume-se que uma sinapse é uma conexão simples que pode impor ao neurônio receptivo excitação ou inibição, mas não ambas (IBIDEM).

Após mais de 5 anos foi proposto o primeiro algoritmo de aprendizado através da variação de pesos de entrada dos nós (neurônios) por Hebb em 1949, baseada nos reforços das ligações sinápticas entre nós excitados e em 1958, após 15 anos da concepção do neurônio artificial, Rosenblatt propôs um algoritmo para treinar a rede através de novo modelo neuronal chamado *perceptron* que atua como um classificador que reconhece padrões e divide o espaço de entrada em classes diferentes, de forma linear (BABINI e MARRANGHELLO, 2007).



O algoritmo Least Mean-Square⁶³ (LMS) desenvolvido entre 1959 e 1960 por Widrow e Hoff e nomeado como Adaline (ADaptive LINEar), é fundamentado na

⁶³ Em tradução livre: "Mínimo Quadrado Médio"

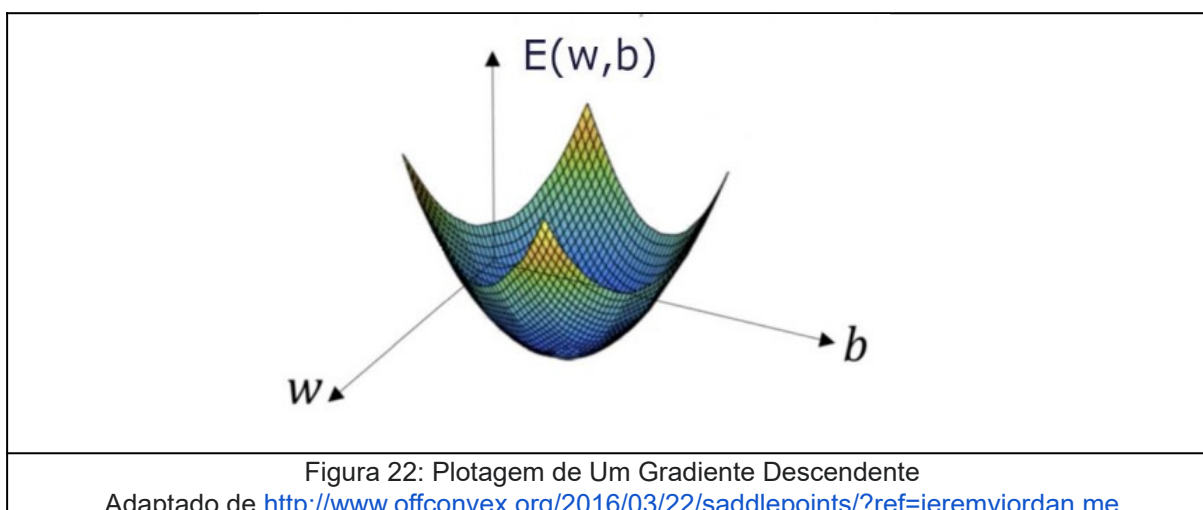
Regra Delta através do método dos mínimos quadrados, buscando superar a incapacidade do Perceptron em classificar dados que não sejam linearmente separáveis. Esse novo procedimento de aprendizagem busca convergir o erro para o mínimo da função quadrática pelo método de gradiente descendente, na qual o ajuste dos pesos são realizados pela soma ponderada das entradas (rede), enquanto no perceptron clássico a saída da função é usada para ajustar os pesos na rede passados na função de ativação (HAYKIN, 2001).

$$\textbf{Adaline: } y = \sum_{j=1}^n x_j w_j + \theta$$

A diferença fundamental entre eles reside no fato de que o algoritmo de aprendizagem minimiza o erro das saídas em relação aos valores desejados como um treinamento e a função de custo a ser minimizada é a soma dos erros quadráticos:

$$E(n) = \frac{1}{2} \sum \left(d_k(n) - y_k(n) \right)^2$$

O método de Gradiente Descendente processa a minimização do erro quadrático $\Delta w_{ki} = -\eta \frac{\partial E}{\partial w_{ki}}$, onde cada peso sináptico i do neurônio k é proporcionalmente atualizado ao negativo da derivada parcial do erro, em relação ao peso $\Delta w_{ki} = -\left(d_k - y_k\right)x_i$



Em 1962 foi proposta uma rede neural em camadas treináveis e múltiplos elementos adaptativos chamada *Madaline* (multiples adalines) na qual um conjunto

de *Adalines* repassam suas respostas para uma única *Adaline* fixa que atua computando as maiorias, sendo que se a metade delas forem +1, sua saída também será igual. Adaline pode ser considerada a primeira rede neural com aplicação prática em um problema real, ao usar os princípios dos filtros adaptativos na eliminação do eco e ruídos em linhas telefônicas analógicas (WIDROW, B; M. A. LEHR, 1990).

Em 1961, Minsky publica um artigo com uma seção que fundamenta o que até hoje denominamos redes neurais e parte da inteligência artificial (IA), chamado “*Steps Toward Artificial Intelligence*”⁶⁴. Outro de seus livros de 1967, *Computation: Finite and Infinite Machines*⁶⁵, teve um impacto bastante positivo na comunidade de pesquisadores ao estender alguns resultados da teoria de McCulloch e Walter Pitts, adaptando-os aos arcabouços científicos das teorias dos autômatos e da computação (HAYKIN, 2001).

O perceptron foi essencial na consolidação das redes neurais como um campo científico promissor que oferecia possibilidades avançadas de resolver problemas computacionais durante a década de 1960. O livro “*Perceptrons: uma introdução à geometria computacional*” lançado em 1969 pelo mesmo Minsky e Papert, trouxe entre inúmeras contribuições aos estudos da área, uma revelação que levou a um período de desinteresse e abandono das promissoras pesquisas com o perceptron de única camada e um redirecionamento de verbas e recursos para outras teorias simbólicas .

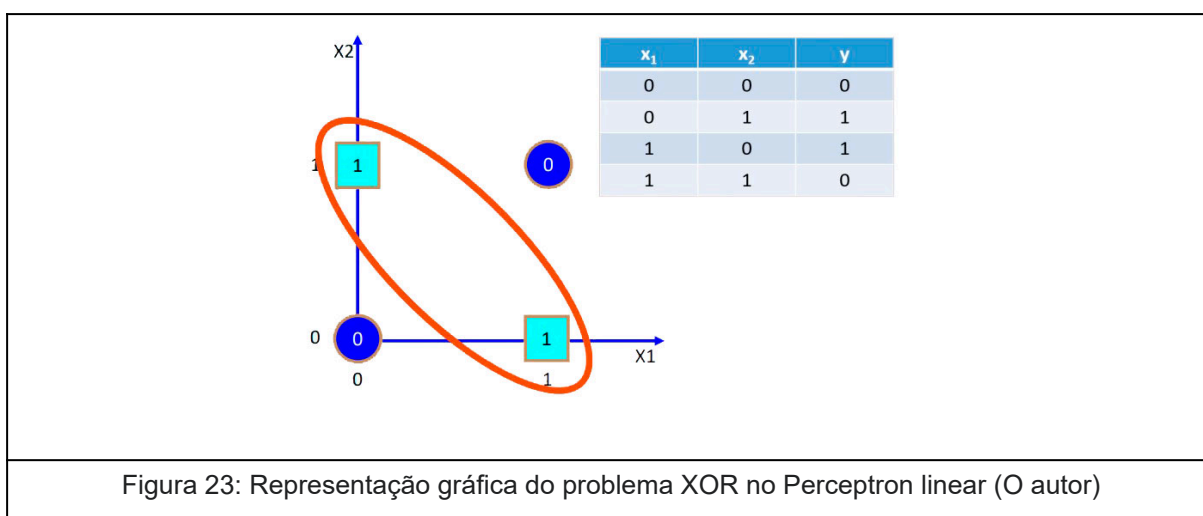
Eles afirmaram que somente separações lineares seriam passíveis de resolução pelo perceptron e que a implementação de outras camadas deles não possibilitaria resolver problemas relativamente simples, como o operador booleano XOR, que corresponde ao traçar uma reta de separação entre dois agrupamentos em um conjunto de dados. O algoritmo de aprendizagem do perceptron (regra delta) não funciona com redes de mais de uma camada.

A constatação contida no livro de Minsky e Papert desencadeou uma série de controvérsias que, embora saudáveis, úteis e necessárias para o desenvolvimento científico, culminou com uma frustração nos meios acadêmicos

⁶⁴ Em tradução livre: “Passos em direção à inteligência artificial”

⁶⁵ Em tradução livre: “Computação: máquinas finitas e infinitas”

proporcional ao ufanismo propagado até mesmo pela mídia, que já explorava comercialmente, em 1969, uma inteligência artificial chamada HAL 9000 como protagonista virtual no filme 2001: Uma odisseia no espaço de Stanley Kubrick. O caráter *hype*⁶⁶ dos enunciados científicos que previam soluções assombrosas para dilemas da humanidade através do perceptron e das redes neurais, deu lugar a um descrédito que fez cessar os financiamentos e doações, que passaram a favorecer os pesquisadores de outros modelos diferentes do conexionista (BABINI e MARRANGHELLO, 2007); (HAYKIN, 2001).



Esse período de aproximadamente dez anos entre 1970 e 1980 foi chamado de “Inverno da Inteligência Artificial” em que muitos pesquisadores das engenharias abandonaram a área, com exceção dos neurocientistas e psicólogos. Existem diversas causas atribuídas ao período de forte queda nos investimentos, tais como a própria natureza interdisciplinar da área que alimentava os conflitos departamentais e consequentes cortes nos orçamentos. A ausência de aplicações práticas das pesquisas e problemas de desempenho nos supercomputadores da época que rodavam quase exclusivamente a linguagem LISP (HAYKIN, 2001) .

Os sistemas especialistas projetados nessas máquinas para tomada de decisões eram muito dispendiosos e lentos, o que colaborou para o descrédito de pesquisas que, inclusive, passaram a usar a denominação “aprendizado de

⁶⁶ Hype é o exagero de algo, ou em marketing uma estratégia para enfatizar alguma coisa, ideia ou um produto. É um assunto que está dando o que falar, é algo que está na moda e que é comentado por todo mundo. Fonte: <https://www.significados.com.br/hype/>

máquina" para se descolar da carga negativa que o connexionismo carregou por ser considerado caro e pouco confiável (CARVALHO et al, 1998).

Somente em 1979 Uttley desenvolveu uma hipótese associada à teoria da informação de Shannon sobre a relação estatística entre os estados flutuantes e a eficiência de ambos os lados de uma sinapse variável do sistema nervoso (HAYKIN, 2001).

Outro importante desenvolvimento ocorrido entre 1973 e 1976 foi o conceito de mapas auto-organizáveis utilizando aprendizagem competitiva ao introduzir a auto-organização no arcabouço das redes neurais, incorporando uma das premissas de abordagem dos sistemas complexos. A inspiração neurofisiológica de mapeamento topológico do cérebro proposto por Malsburg pode ter sido o primeiro a demonstrar a auto-organização em uma simulação computacional (IBIDEM).

4.5.3 - O ressurgimento do perceptron nos anos 1980

A partir de 1980 com o surgimento de computadores e redes mais eficientes, novas proposições sobre aprendizagem competitiva como a de Grossberg, estabeleceu um novo princípio de auto-organização conhecido como teoria da ressonância adaptativa, envolvendo uma camada de reconhecimento de baixo para cima (bottom to up) e uma outra generativa de cima para baixo (top to down) como uma amplificação da atividade neuronal. Esse mesmo princípio, de projeções que avançam e retrocedem, deram suporte a novas abordagens que reabilitaram as pesquisas sobre redes neurais e aproximaram o desenvolvimento do perceptron aos modelos da física estatística (GROSSBERG, 2010) .

Ele também introduziu o uso de sistemas não lineares de equações diferenciais ao teorizar como os elementos cerebrais podem originar funções comportamentais, sendo esse formalismo matemático básico aplicado às pesquisas de redes neurais biológicas, até hoje.

Quadro 18 - Pioneiros das redes neurais que contribuíram com a concepção de armazenamento de informação em redes dinamicamente estáveis	
Cragg e Temperley (1954, 1955)	Observaram que assim como os neurônios podem ser "disparados" (ativados) ou "não disparados" (quiescentes), também os átomos em uma rede têm seus spins apontando "para cima" ou "para baixo"

Quadro 18 - Pioneiros das redes neurais que contribuíram com a concepção de armazenamento de informação em redes dinamicamente estáveis	
Cowan (1967)	Introduziu a característica de disparo "sigmóide" e a condição de disparo suave para um neurônio que 1 era baseada na função logística
Grossberg (1967, 1968)	Apresentou um modelo aditivo de um neurônio, envolvendo equações não-lineares de diferenças/diferenciais e explorou o uso do modelo como uma base para a memória de curto prazo
Amari (1972)	De forma independente, propôs o modelo aditivo de um neurônio para estudar o comportamento dinâmico de elementos semelhantes a neurônios conectados aleatoriamente
Wilson e Cowan (1972)	Derivaram equações diferenciais não-lineares acopladas correspondentes à dinâmica de populações localizadas no, espaço, contendo neurônios tanto excitadores como inibitórios
Little e Shaw (1975)	Descreveram um modelo probabilístico de um neurônio, quer disparando ou não um potencial de ação, e usaram o modelo para desenvolver uma teoria da memória de curto prazo
Anderson, Silverstein, Ritz e Jones (1977)	Propuseram o modelo do estado cerebral em uma caixa (brain-state-in-a-box, BSB), consistindo de uma rede associativa simples acoplada a uma dinâmica não-linear.

Algumas pesquisas e financiamentos iniciaram a ocorrer a partir de 1981 com a proposição de Hopfield sobre aplicar uma função de energia numa nova abordagem de computação executada por redes recorrentes com conexões sinápticas simétricas, demonstrando haver relação entre sistemas físicos e redes recorrentes auto-associativas (BABINI e MARRANGHELLO, 2007).

O Modelo de spins de vidro fundamenta essa concepção de Hopfield, ao considerar o neurônio uma unidade estocástica semelhante a um material magnético na física estatística.

Uma descrição simples do material magnético consiste de um conjunto de spins, chamados de magnetos atômicos, distribuídos em uma latência regular. O spin pode apontar para várias direções, sendo que o número de direções possíveis depende do tipo atômico considerado. Em um caso mais simples possui duas direções. Estas direções são representadas pela variável S_i , onde cada estado (i) pode assumir o valor igual a 1, se o spin está apontado para cima e -1 se apontado para baixo (SILVA, 2003).

As críticas a esta analogia entre os possíveis estados dos neurônios e de spins, fundamentada por física estatística de Cragg e Temperley desde 1955, se referem ao fato de que a superposição linear de estados, algo natural aos spins, não é possível em um mesmo neurônio. Utilizando modelo de Lsing para descrever

claramente as propriedades magnéticas simulando a estrutura de uma substância ferromagnética, funciona como uma simplificação do modelo de transição de fase na hamiltoniana de Heisenberg, pois o modelo do vidro de spins de Sherrington e Kirkpatrick foi inventado somente em 1975 (HAYKIN, 2001) .

O Lsing por sua vez está fundamentado na Teoria do Campo Médio ou Autoconsistente, pois considera que um grande e complexo modelo estocástico pode ser estudado a partir de sua simplificação. Ela teoriza que a interação entre um grande número de componentes individuais possa ser analisada pelo efeito de todos os outros indivíduos em qualquer outro indivíduo, aproximado a um único efeito esperado em interações de curto alcance (KADANO, 2009).

Os muitos mínimos locais característicos dos vidros de spin podem ser usados para representar os padrões apreendidos e sua memória. Ao submeter um padrão semelhante aos já apreendidos, ocorre uma dinâmica de atração aos padrões ensinados indicando a emergência de um atrator de ponto fixo ou caótico, dependendo das condições iniciais.

Antes de expor a dinâmica de ressurgimento da IA e do Perceptron que a teoria de Hopfield desencadeou , cabe esclarecer que a não linearidade e o caos foram os maiores desafios ao retorno da concepção das redes neurais.

4.5.4 - Atratores e caos

Como um exemplo, simulamos um mapa logístico em uma análise de crescimento populacional, onde se apresentou um atrator de ponto fixo que mantinha o crescimento estável em determinadas condições iniciais e comportamento determinístico⁶⁷.

Processos dinâmicos exibem uma mistura de ordem e desordem simultaneamente e alta capacidade de gerar fenômenos emergentes espontaneamente, podendo se estender ciclicamente ao ponto fixo ou ao caos.

Sistemas dinâmicos apresentam crescente complexidade e se auto-organizam, podendo ocorrer a emergência espontânea de padrões. Também são robustos, com transições de fase, podendo gerar eventos extremos, embora a complexidade seja observada mais claramente entre o intervalo de seus oponentes.

⁶⁷ <https://www.complexityexplorer.org/courses/144-introduction-to-complexity/segments/14500>

Outros comportamentos são observados e foram incorporados ao arcabouço linguístico da área, pois constituem sistemas não lineares, com dinâmica ligeiramente afastada do equilíbrio e retroalimentação, que seria a reinserção de dados que o sistema gera e que o auto-regula. A não linearidade remete ao conceito de caos, embora não o defina, pois nem todos os sistemas não lineares são caóticos. Segue abaixo uma simulação de crescimento populacional por função de [mapa logístico desenvolvida em Python](#), na qual o fator inicial influi no determinismo do ponto fixo.

28/04/2022 12:06

Cópia de mapa logístico.ipynb - Colaboratory

```

r = 0
Rval = input("Valor de R - taxa de nascimentos ? ")
R = float(Rval)
Xval = input("Fator inicial ?")
x0 = float(Xval)
x1 = float(Xval)
for a in range(9):
    r = R*(x0-(x0*x0))
    print (r)
    x0 = r

import numpy as np
import matplotlib.pyplot as plt

xn = np.linspace(x1,1,100)
xmap = r*xn*(1-(xn*xn))

plt.figure()
plt.title("Função de mapa logístico")
plt.plot(xn,xmap,"b-", label="r={:.2f}".format(r))
plt.xlabel("$x_n$")
plt.ylabel("$x_{n+1}$")
plt.legend(loc="upper right")
plt.show()

```

Valor de R - taxa de nascimentos ? 2.5
 Fator inicial ? 0.2
 0.4
 0.6
 0.6
 0.6
 0.6
 0.6
 0.6
 0.6
 0.6
 0.6

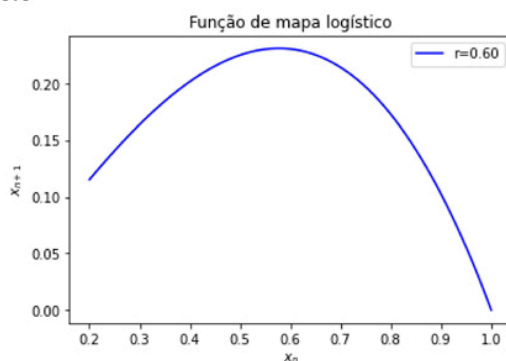


Figura 24 - Simulação de crescimento populacional por função de mapa logístico (o autor)

Como já afirmado anteriormente, pequenas alterações nas condições iniciais podem desencadear um atrator caótico que gera fractais que se iniciam a partir de $r = 3.56995$ começando o comportamento caótico que continua até $r = 4$.

A assinatura do caos neste mapa inicia exatamente aos 3,56995 e os valores da população começam a divergir, emergindo a aleatoriedade que inviabiliza estudos correlacionais e preditivos.

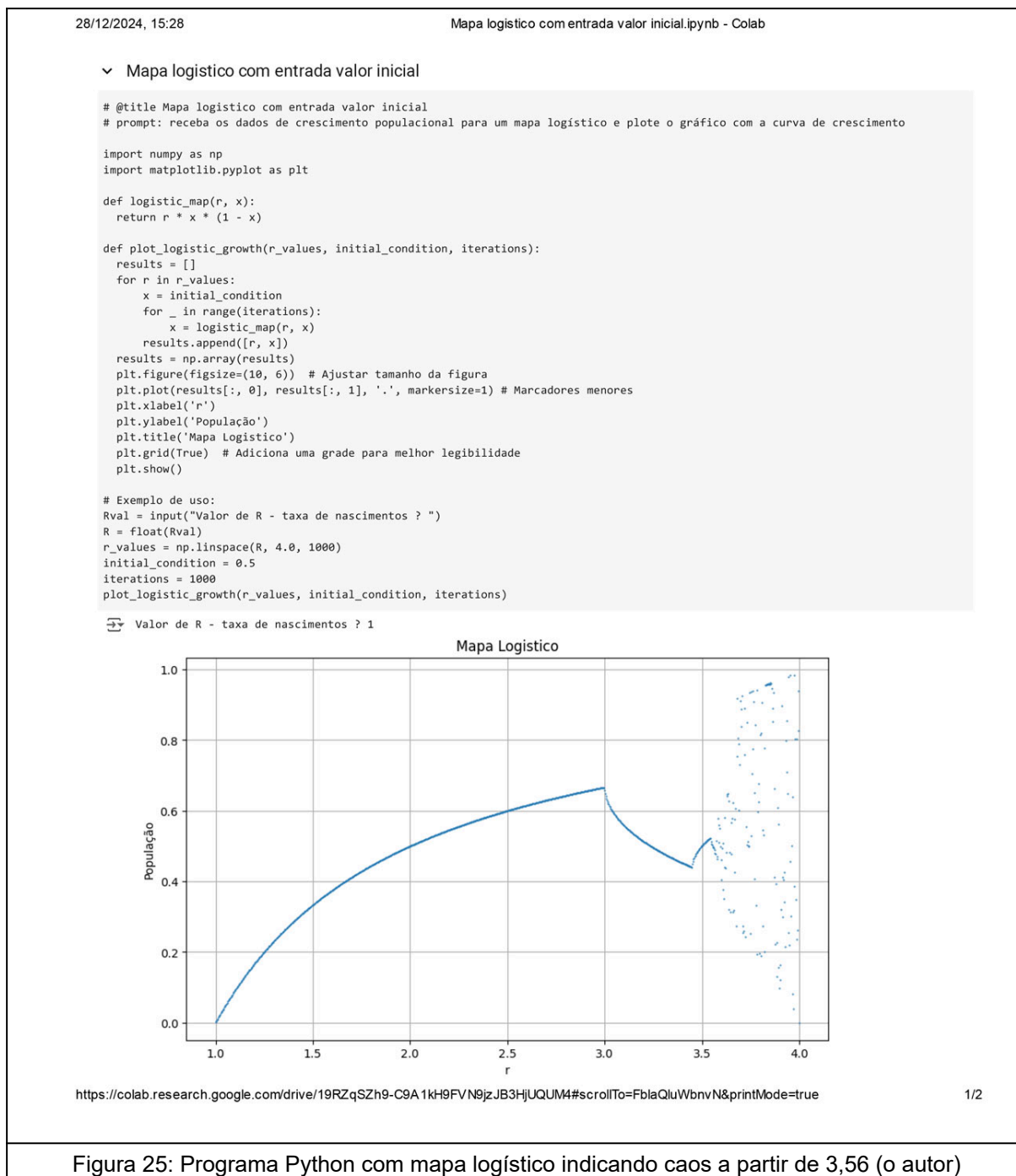


Figura 25: Programa Python com mapa logístico indicando caos a partir de 3,56 (o autor)

A plotagem acima demonstra claramente que lidar com variáveis dinâmicas, como as que representam taxas de crescimento populacionais ou outras métricas de

sistemas dinâmicos, somente podem ser analisadas e resolvidas por modelos não lineares.

4.5.5 - Sistemas Dinâmicos: caos e connexionismo

Em um artigo que delineava limites entre ciência tradicional e da complexidade em 1948, Warren Weaver definiu 3 tipos que se dividem em problemas da simplicidade, da complexidade desordenada e da complexidade organizada. Descrevendo as características de cada um, considerou que os limites da simplicidade teriam entre 2 ou 4 variáveis típicas dos problemas já estudados desde os séculos 17 a 19 pelas ciências físicas (WEAVER, 1948).

Já naquela época (1948), as ciências médicas e biológicas lidavam com dezenas de variáveis inter-relacionadas e suas correlações que descreviam fatos e as analisavam. Naquele mesmo estudo, Warren considerou que a complexidade desordenada exigia abordagens diferentes da ciência da simplicidade, pois lidava com relações dos níveis de conhecimento do macro e microcosmo, que somente com os avanços das teorias da probabilidade e da mecânica estatística, se tornou possível.

Embora a descrição de fenômenos seja plausível, os cálculos exigiam um novo arcabouço. Os processos dinâmicos na complexidade exibem uma mistura de ordem e desordem simultaneamente, tendo alta capacidade de gerar fenômenos emergentes espontaneamente.

Buscando essa nova abordagem, durante 10 anos do “Inverno da IA”, poucos neurocientistas e engenheiros de computação desenvolveram algoritmos de aprendizado de máquina que superaram as limitações e ineficiência dos perceptrons de camada única criticados por Minsky e Papert.

4.5.5.1 - Connexionismo e neurociência para superar o inverno da IA

As concepções sobre sistemas não lineares nortearam o desenvolvimento da teoria connexionista que começou a se diferenciar claramente da abordagem simbolista, que concebia uma mente como sistema cerebral e procurava entender o raciocínio lógico através de símbolos e regras, buscando simular a lógica formal e a linguagem natural (SEJNOWSKI, 2019)

O pioneiro que resgatou o arcabouço teórico de Inteligência Artificial da dormência epistemológica foi John Hopfield, físico, biólogo e neurocientista ganhador do prêmio Nobel de Física 2024, que resolveu o problema de convergência em uma rede não linear para um estado estável chamado “atrator”. Essa abordagem não linear em um sistema dinâmico com eventuais tendências aos caos⁶⁸, possibilitou implementar um modelo de memória de conteúdo endereçável semelhante ao nosso cérebro. Uma memória armazenada pode ser recuperada através de partes dela e a rede que leva seu nome preenche as lacunas dos dados faltantes e recupera toda a informação (IBIDEM), num processo semelhante à nossa memória associativa.

As redes Hopfield foram pioneiras em descrever a função cerebral de forma tão semelhante ao biológico que é reconhecidamente similar ao observado no hipocampo cerebral humano. A neurociência e a física imbricam nesta teoria científica, articulada com a computação e o aprendizado. Ela se tornou um importante solucionador de otimização e foi usada para resolver problemas como do Caixeiro Viajante (PCV) ou do carteiro chinês (PCC), que em última análise, seriam conceitos equivalentes ao Problema das Sete Pontes de Königsberg proposto por Euler em 1736 (MITCHELL, 2009)

Para superar os mesmos problemas de mínimos locais e globais, já vistos na impossibilidade de sobreposição de estados nos vidros de Spin (KADANO, 2009), foi proposta uma nova versão da rede Hopfield, chamada de Máquinas de Boltzmann por Ackley, Hinton e Sejnowski em 1985, buscando um equilíbrio de mínimos locais para gerar um mínimo global, baseados na constante do físico austriaco do século XIX, um dos fundadores da física estatística.

A próxima otimização do modelo foi concebida e proposta pelo psicólogo Donald O. Hebb, que vislumbrou uma eficiência resultante da ativação persistente entre dois neurônios que reforça a sinapse, a qual chamou de plasticidade sináptica. Ela foi descoberta e comprovada anos mais tarde pela neurociência no hipocampo, que é uma área cerebral importante para memórias de longo prazo, desencadeando

⁶⁸ Caos na complexidade significa que, embora seja impossível prever individualmente a evolução de um sistema caótico a partir do seu estado inicial, devido à possibilidade de causas menores desencadearem efeitos significativos, podemos fornecer inferir estatisticamente, como, por exemplo, sistemas meteorológicos, que podem prever uma probabilidade estatística de 20% de chuva num dia.

uma série de experimentos que confirmaram, além da elevação da voltagem neuronal, o receptor especial de NMDA (N-metil-D-aspartato) que ativa a Potenciação de Longo Termo (Long Term Potentiation - LTP) como substrato da memória de longo prazo.

A plasticidade de Hebb nas sinapses é regida por concomitâncias entre as entradas e as saídas, assim como o algoritmo de aprendizado da máquina de Boltzmann (SEJNOWSKI, 2019).

O que realmente surpreendeu toda a comunidade científica foi a constatação de que a máquina de Boltzmann “dormia” para aprender em seu algoritmo de duas fases. Na primeira fase de “despertar”, as unidades de rede são atualizadas até obterem o equilíbrio, seguida da fase de “dormir”, em que as unidades de entrada e saída são fixadas em pares conectados que atualizam, proporcionalmente, a diferença entre as taxas de concomitância nas fases de dormir e despertar.

Essa assombrosa constatação de similaridade entre o cérebro e o modelo, de que a privação do “sono” do algoritmo gerava erros e inconsistências, foi percebida por neurocientistas e psiquiatras que confirmaram que a privação do sono em humanos leva a estados delirantes que são comuns em pacientes de unidades de terapia intensiva (UTIs) com iluminação constante e que não possuem janelas.

Pacientes com esquizofrenia tendem a apresentar distúrbios do sono, o que contribui para seus delírios. Estávamos convencidos de que esse era o caminho certo para entender como o cérebro funciona (SEJNOWSKI, 2019).

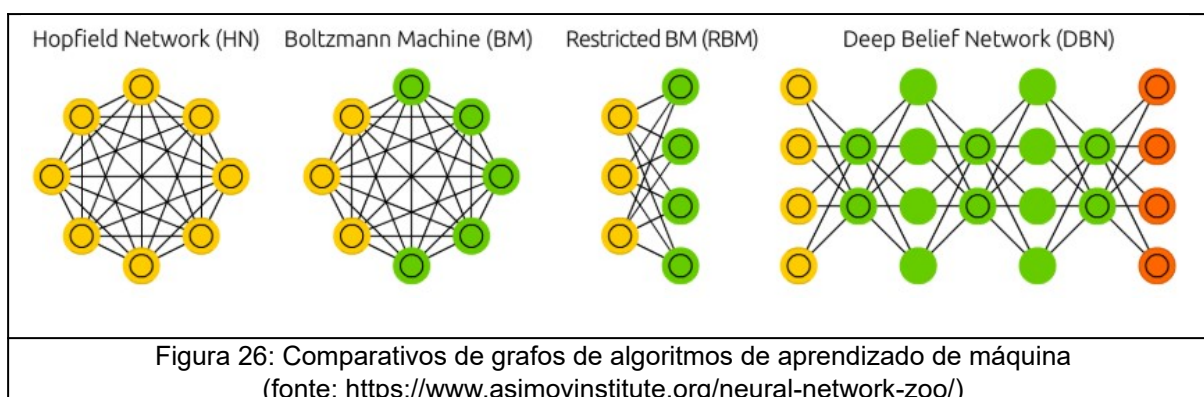
A Máquina de Boltzmann resolveu um problema não linear relativo ao princípio de simetria⁶⁹ que o Perceptron clássico não consegue aprender ou resolver. A Simetria dos Espelhos foi usada para implementar um treinamento da rede que detecta e separa uma imagem simples de seu fundo.

A máquina de Boltzmann é um modelo estocástico gerador de grafos, não supervisionado, probabilístico e baseado em energia, que aproveita a estreita relação entre as distribuições de Boltzmann e a teoria da informação (já demonstrado em quadro deste capítulo na pág. 127). Em outras palavras, esses modelos atribuem uma energia a cada configuração das variáveis que estão sendo

⁶⁹ O corpo humano é bilateralmente simétrico ao longo do seu eixo vertical e a divisão imaginária do corpo em partes iguais, são chamadas de metades especulares. Os seres vivos simétricos podem apresentar simetria bilateral ou radial. https://antigo.uab.ufsc.br/biologia/files/2020/08/Cap_03.pdf

modeladas, intuitivamente, pois o aprendizado nesses modelos significa associar configurações mais prováveis a estados de energia mais baixos.

Na ilustração abaixo, a rede Hopfield é comparada aos modelos baseados na Máquina de Boltzmann, que possuem camadas ocultas de aprendizado, ilustradas em verde e camadas de saída em vermelho.



As máquinas de Boltzmann restritas (RBMs) são uma variação em que os neurônios são restritos para formar um gráfico bipartido. Isso significa que há dois grupos de unidades, comumente chamados de unidades “visíveis” e “ocultas”, e que os nós de cada grupo podem ter uma conexão simétrica entre eles. É importante ressaltar que não há conexão entre os nós do mesmo grupo.

Esta restrição permite algoritmos de treinamento mais eficientes do que os disponíveis para a classe geral de máquinas de Boltzmann, em particular o algoritmo de divergência contrastiva baseado em gradiente.

4.5.6 - Redes Multicamadas

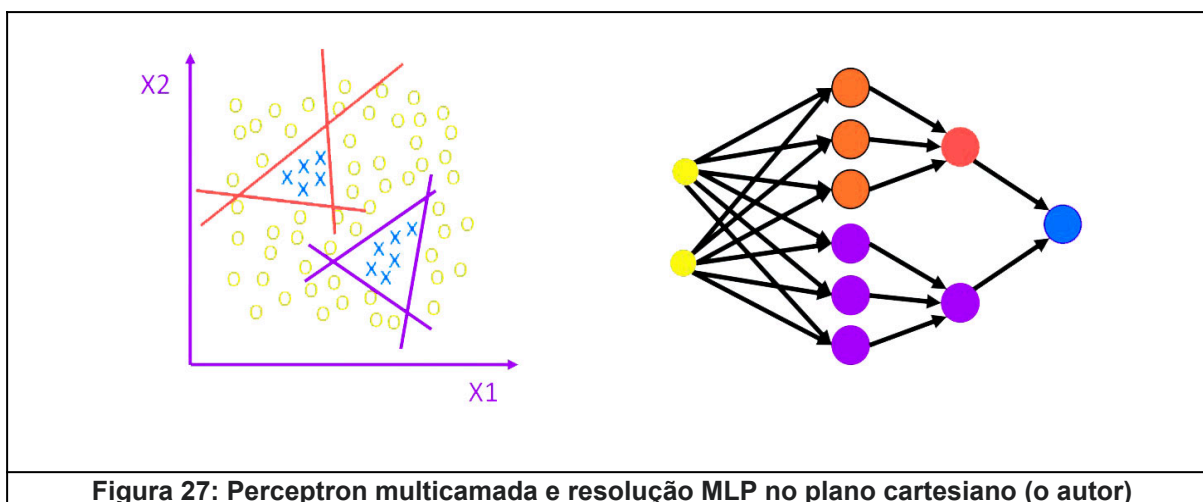
O algoritmo de Retropropagação proposto por Rumelhart, Hinton e Williams em 1986 foram as primeiras redes de múltiplas camadas com resultados eficientes que superaram as inconsistências, resolvendo a não linearidade inerente aos processos dinâmicos a partir da transformação do mesmo em um problema linearmente separável (SILVA, 2003).

As redes multicamadas possibilitaram resoluções da separação não linear como XOR⁷⁰ através da decomposição das tarefas em dois neurônios diferentes,

⁷⁰ Ou exclusivo ou disjunção exclusiva é uma operação lógica entre dois operandos que resulta em um valor lógico verdadeiro se e somente se os dois operandos forem diferentes, ou seja, se um for verdadeiro e o outro for falso.

pois não é possível a separação de duas classes distintas traçando apenas uma reta no eixo cartesiano bidimensional (HAYKIN, 2001).

No exemplo em seguida, são duas camadas, uma com seis neurônios, sendo três para cada operação booleana compondo três retas e uma com dois neurônios para separar cada classe e processar a saída no plano cartesiano. A solução, embora tenha se tornado eficiente em operações booleanas encadeadas, não é possível calcular o gradiente descendente nas camadas ocultas da rede.



A possibilidade de reprocessar os erros da rede em sistemas dinâmicos não lineares para corrigir as saídas é bastante similar ao algoritmo de treinamento com retropropagação em perceptrons multicamadas de redes neurais artificiais. Podem rodar em duas direções, uma das quais os dados propagam da camada de entrada até a camada saída (feed-forward) ou indo da camada de saída até a primeira camada oculta, propagando os erros na direção contrária ao fluxo de dados (feed-backward).

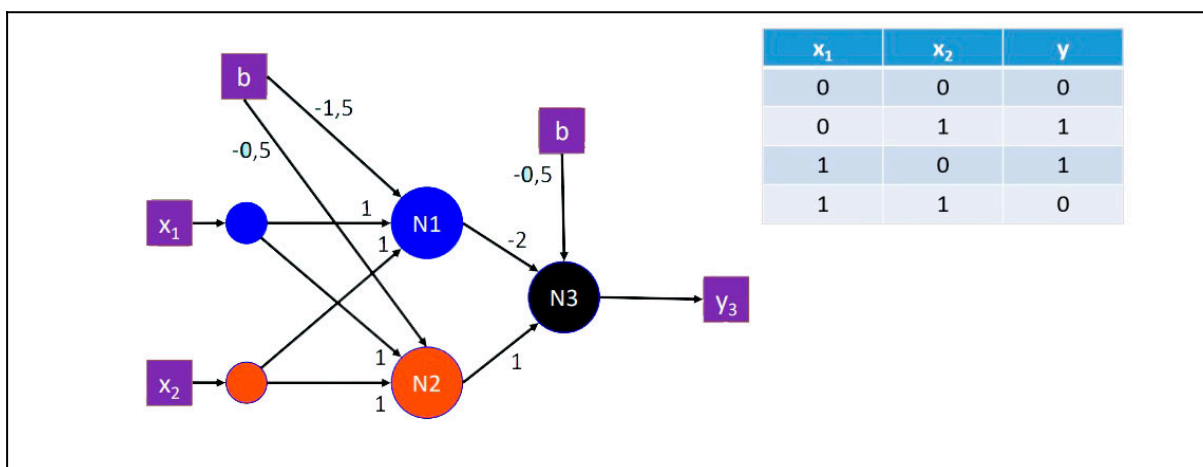


Figura 28: Resolução do problema XOR no Perceptron Multicamada (o autor)
--

Ainda hoje, após décadas de desenvolvimento, as redes de retropropagação se comportam como caixas-pretas em que não há controle ou conhecimento sobre o como os neurônios ocultos alcançam os resultados. Também existem problemas de separação não linear que não podem ser resolvidos apenas com essa abordagem e originaram novos algoritmos mais eficientes. Minsky menciona que o tipo mais comum de nó de rede neural é feito combinando uma parte analógica quase linear que "soma as evidências" com uma parte não linear quase digital que toma uma decisão com base em um limite (MINSKY, 1991)

Outra limitação se refere ao longo tempo de treinamento dessas redes quando executado em processamento serial, embora esteja hoje disponível a aceleração dos pipelines por GPUs ou mesmo processadores neurais que existem desde os anos 1990, como o ETANN (Electronically Trainable Analog Neural Network). Atualmente os processadores neurais estão plenamente disseminados em dispositivos de nossa rotina diária como smartphones, televisores e assistentes virtuais.

A delimitação da quantidade adequada de camadas ocultas para as simulações e representações é uma outra dificuldade encontrada no uso do MLP, ocasionando dimensionalidades superestimadas ou subestimadas durante a fase de treinamento e inviabilizando alguns modelos mais complexos, com enorme volume de dados. Uma das soluções apontadas para superar os problemas dos hiperparâmetros, ainda aplicada atualmente é a utilização de algoritmos genéticos, que fornecem técnicas de busca e otimização paralela, através de seleção natural e reprodução genética, nos moldes do princípio evolutivo da teoria darwiniana.

A terceira guinada teórica na aprendizagem profunda começou em 2006, quando Geoffrey Hinton apresentou a Deep Belief Network (DBN) ou Rede de Crença em Profundidade (HINTON, OSINDERO, TEH, 2006). Isso possibilitou o treinamento de redes com várias camadas, elevando os padrões da inteligência artificial e popularizando o termo "aprendizagem profunda".

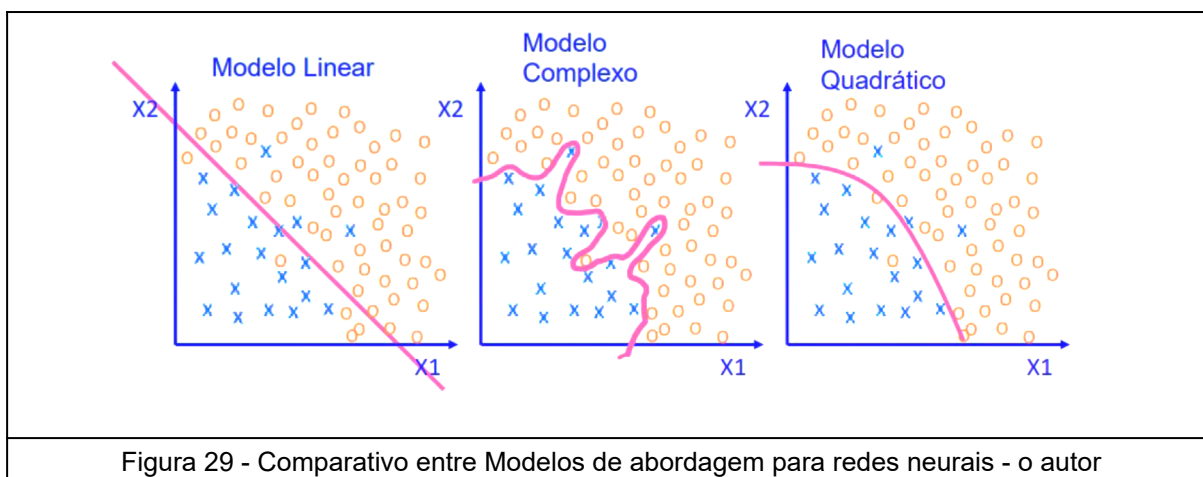


Figura 29 - Comparativo entre Modelos de abordagem para redes neurais - o autor

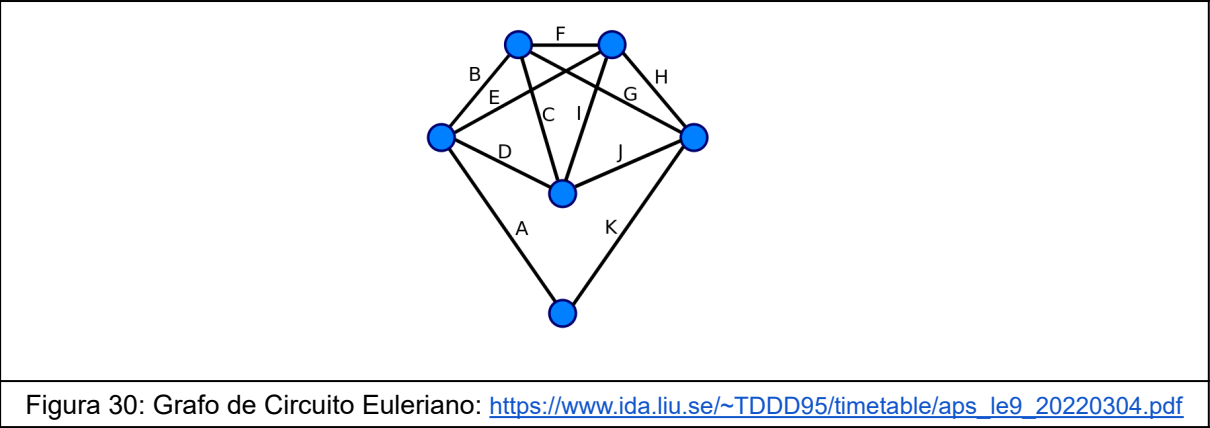
A crescente proeminência dos modelos baseados em energia no aprendizado de máquina se deve, em grande parte, ao sucesso das redes de crenças profundas (DBNs). Isso inclui modelos como máquinas de Boltzmann restritas (RBMs), DBNs e máquinas de Boltzmann profundas (DBMs). Esses modelos, baseados em funções de energia, fornecem uma estrutura sólida para aprendizado não supervisionado e extração de recursos, levando a avanços no reconhecimento de imagens, processamento de linguagem natural e síntese de fala.

O sucesso dos DBNs, com seu pré-treinamento em camadas e a capacidade de capturar distribuições de dados complexas, despertou um interesse e uma exploração mais ampla em modelos baseados em energia para tarefas de aprendizado complexas (CERRI, 2013).

4.6 - A máquina de Boltzmann é um modelo estocástico gerador de grafos

Algoritmos computacionais podem utilizar a Teoria dos Grafos, proposta por Euler no século XVIII, para modelar relações de pares entre objetos como estruturas matemáticas. O problema do caminho mais curto é encontrar um caminho entre dois vértices (ou nós) em um grafo de forma que a soma dos pesos de suas arestas constituintes seja minimizada.

Leonhard Euler desenvolveu a teoria buscando uma solução para otimizar uma rota por todas as Sete pontes de Königsberg, passando apenas uma vez em cada uma delas, deduzindo a relevância do número ímpar de variáveis iniciais na impossibilidade de resolver o problema (FERREIRA e BORGES, 2022).



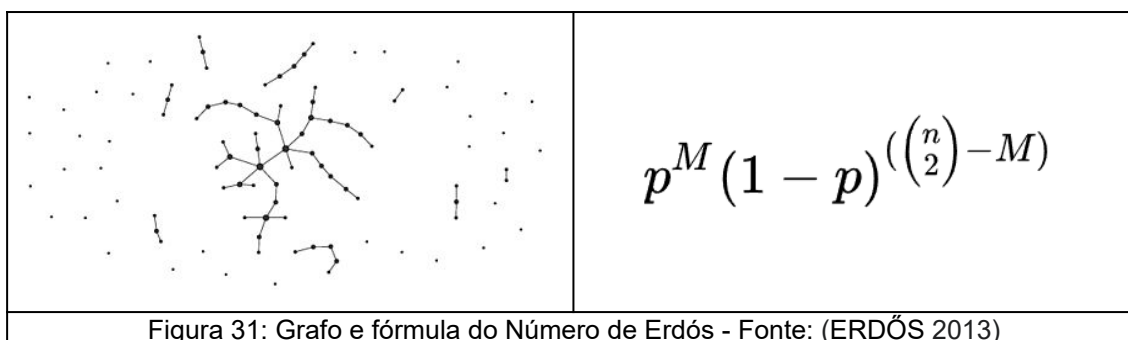
Refutando essa possibilidade, Euler iniciou uma das teorias mais aplicadas em ciência computacional e que deram origem ao estado da arte no aprendizado profundo de máquinas, ao observar que o número de vértices, arestas e faces (V, A e F, respectivamente) satisfazem sempre uma relação aritmética simples e elegante, cuja fórmula é: $V - A + F = 2$. (FERREIRA e BORGES, 2022)

Diversas aplicações tecnológicas foram desenvolvidas com variações e adaptações evolutivas da teoria, sendo aplicada na análise de redes e relações em diversos campos da ciência, desde cadeias alimentares de animais, proteínas de DNA, redes neurais e de computadores.

Uma dessas aplicações foi desenvolvida por Erdős-Rényi que de fato, é uma ligação não ponderada de percolação no grafo completo, examinando um grafo finito ou infinito e eliminando extremidades ou ligações aleatoriamente. Essa abordagem de 1959 introduziu a aleatoriedade ao incluir o limite entre cada par de nós com igual probabilidade, independentemente das extremidades numa rede aberta (ERDŐS 2013).

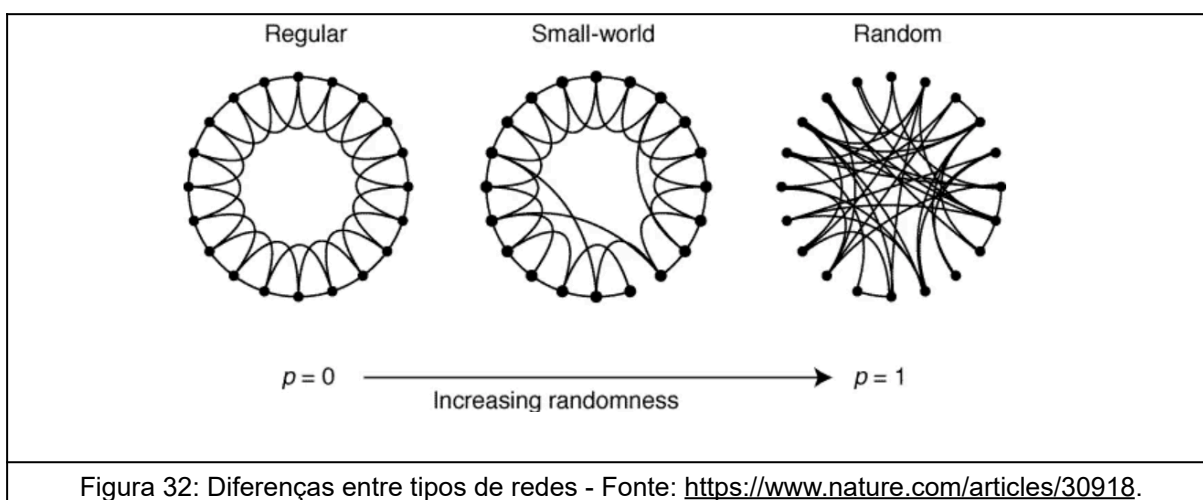
Essa e outras pesquisas deram origem ao Número de Erdős, em homenagem aos milhares de trabalhos acadêmicos desse pesquisador que mais publicou artigos sobre matemática com outros autores, sendo a indicação de parcerias em publicações.

Grafo baseado em Erdős-Rényi	Número de Erdős
------------------------------	-----------------



Nesse sentido em 1998 é publicado o artigo *Collective dynamics of ‘small-world’ networks* de Watts e Strogatz, que em tradução livre seria Dinâmica coletiva de redes “pequeno mundo”, demonstrando que essas redes operam em muitos sistemas auto-organizadores biológicos, tecnológicos e as redes sociais.

Esses modelos de sistemas dinâmicos com exibição de acoplamento com pequenos mundos oferecem maior velocidade de sinal-propagação, melhor potência computacional e sincronismo. Esse artigo de Watts e Strogatz é de grande relevância em diversas áreas científicas e já obteve 153.000 Acessos, 25612 citações, possui índice Altmetric de 410 e no Google Acadêmico atingiu mais de 48.800 citações.



A concepção das redes “pequeno mundo” e sua popularidade na comunidade científica fez reemergir um conceito já existente que expressava a ideia de que o mundo era mesmo relativamente pequeno e que podíamos nos conectar facilmente com qualquer pessoa no mundo através de nossos contatos próximos. A origem dos graus de separação não é científica, mas literária e foi criada em 1929 pelo escritor húngaro Frigyes Karinthy no livro Tudo É Diferente.

Nessa premissa, Barabási propôs três aspectos inerentes às redes, sinalizando que naquela época, a web era a única rede matematicamente comprovada a ter hubs (BARABASI, 2012)

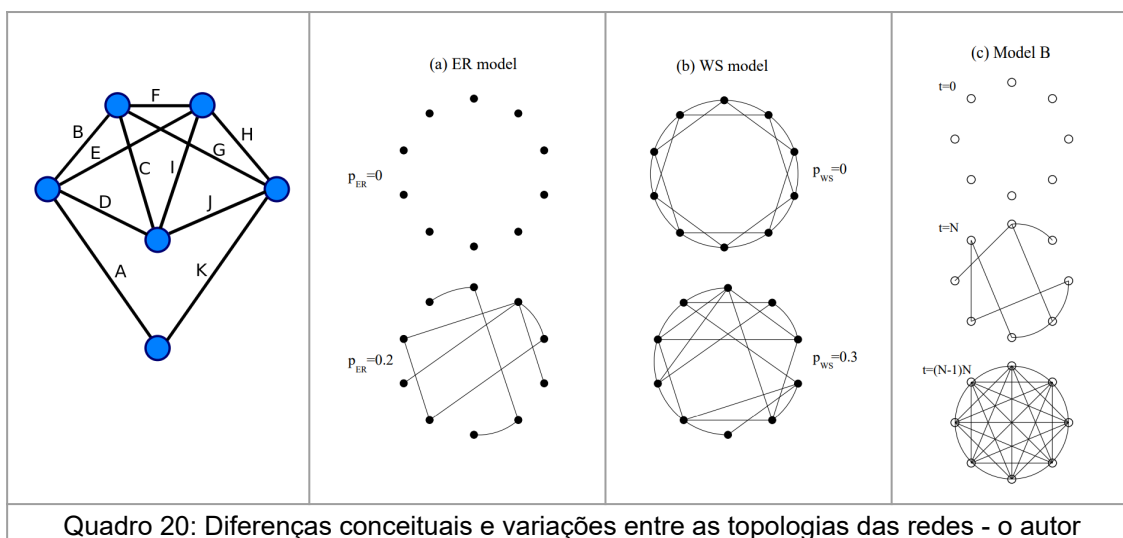
Características	Premissas
Expansão, evolução e adaptação	A rede tem de se expandir, crescer. Tal condição de crescimento é muito importante, assim como a ideia de emergência que a acompanha. Está em constante evolução e adaptação e existe acentuadamente com a rede.
Hubs (conexões preferenciais)	A segunda refere-se a conexões preferenciais, ou seja, novos integrantes vão querer ligar-se a hubs, que participam da rede com mais conexões.
Aptidão Competitiva	A terceira denomina-se aptidão competitiva, que, em termos de rede, implica sua taxa de atração.
Quadro 19: Premissas das redes - The Seventh Link: Rich Get Richer - BARABÁSI	

A proposição de Barabási ocasionou diversas novas abordagens das teorias anteriores e tornou visível a emergência da Ciência das Redes. Podemos então verificar quais são as Principais Revisões (ou novas proposições) da teoria dos Grafos de Euler a Barabási.

A tabela com os gráficos abaixo ilustram visualmente as diferenças conceituais e variações entre as topologias das redes desde Euler (1736 - EU) Erdős-Rényi (1959 - **ER**), passando por Watts-Strogatz (1998 - **WS**) até Barabási (2002 - **BA**) , segue abaixo em grafos comparados.

Euler (1736) EU	Erdős-Rényi (1959) ER	Watts-Strogatz (1998) WS	Barabási (2002) BA
Sete pontes Königsberg	Número de Erdős	Dinâmica pequeno mundo	Dinâmica não aleatória
Grafo Fechado	Rede aberta	Parcialmente aberta	Aberta ou fechada
Prova negativa	Aleatoriedade	Graus de separação e clusters	Hubs e conectores
Redes fechadas e aleatórias são mais vulneráveis a perturbações aleatórias ⁷¹		Redes de pequeno mundo são mais robustas	

⁷¹ <https://Arxiv.Org/Pdf/Cond-Mat/9907068.Pdf>



O conceito de clusterização (que significa uma espécie de agrupamento) e seu coeficiente proposto por Watts-Strogatz reflete o fato de que em redes reais, é muito provável que meus amigos também sejam amigos entre si, como proposto por Granovetter em seu livro de 1973.

Ele considerava haver maiores possibilidades de se conseguir um emprego, experiências, conhecimentos diferentes ou através de pessoas com “laços fracos”, ou seja, menos próximos, do que amigos ou parentes.

O argumento é de que tenham maior relevância na dinâmica de suas redes de contatos por serem em número muito superior às pessoas com quem temos laços fortes. A maioria das redes reais apresenta a coesão de grupos com densidade alta de laços entre diversos membros que se conhecem, em média, sempre com maior probabilidade do que agrupamentos de laços aleatórios.

Ainda segundo Barabási, entendemos a clusterização intuitivamente pelo desejo inato de aderirmos a agrupamentos que nos favoreçam em busca de intimidade, segurança e familiaridade. Ele argumenta que a clusterização, tão visível nas redes sociais não necessariamente virtuais, também se apresenta em inúmeras outras redes, exemplificando com o conectoma do *C. elegans*, já demonstrado anteriormente neste capítulo (BARABÁSI, 2017).

Os diversos exemplos de sistemas biológicos, tecnológicos e sociais que se apresentam clusterizados na bibliografia relatada, podem ser consideradas provas irrefutáveis de que as redes do mundo real não apresentam a aleatoriedade

proposta por Erdős-Rényi, embora possam haver raras exceções ligadas às probabilidades estatisticamente remotas .

A onipresença da clusterização que caracteriza as redes sociais se tornou também uma propriedade genérica das redes complexas, que são objeto deste estudo, representam o fato de que embora sejam muito diferentes em relação às suas formas, dimensões, natureza e escopo, a maioria das redes apresenta princípios de organização em comum, havendo mais semelhanças entre si do que as diferenças que apresentam .

Os impactos na ciência, tecnologia e sociedade ao conceber redes como complexas foram fundamentais no desenvolvimento de novos arcabouços científicos e tecnológicos que colaboram no desenvolvimento de novos fármacos, simulações de física teórica e de partículas.

Também são eficientes na detecção de notícias falsas e desinformação, sistemas preditores e aplicações de internet que já utilizamos diariamente com naturalidade, como sistemas de recomendações de plataformas de streaming, de canais de compra ou nos anúncios monetizados dos buscadores como o Google.

Livres de escala	Não aleatórias
Clusterizadas	Topologicamente variadas
Sem localidade espacial	Sem ordem de nó fixo
Sem ponto de referência	São dinâmicas
São não lineares	Com características multimodais
Quadro 21 - Características das redes complexas (o autor)	

O elemento faltante na tabela anterior é justamente a maior colaboração da Ciência das Redes proposto por Barabási em seu livro *Linked* de 2002. O conceito de *hub* ou “conector” está relacionado com a anexação preferencial como um exemplo de ciclo de feedback positivo, em que variações aleatórias ao se obter maiores números de conexões com outros nós, seja por notoriedade ou pioneirismo, são amplificadas e geram nós com grau elevado que conectam ainda mais nós. Ele utilizou a mesma nomenclatura usada por Merton⁷² em 1968 para explicar o efeito de acumulação de conexões como “efeito Mathew”, ou seja, os ricos ficam mais ricos (BARABÁSI, 2002).

⁷² MERTON, R. K. **The Matthew effect in science. The reward and communication systems of science are considered.** Science (New York, N.Y.), 159(810):56–63, January 1968.

Os conectores são um componente extremamente importante de nossa rede social. Eles criam tendências e modas, fazem contatos importantes, espalham novidades ou ajudam a abrir um restaurante. São a tessitura da sociedade, juntando facilmente diferentes raças, níveis de instrução e diversidade (IBIDEM - tradução livre)

Esses conectores (hubs) que agregam numerosas arestas (links) se apresentam em diversos sistemas complexos que vão desde a economia, biologia, computação, ciências sociais, física, genética, entre outros, sendo um fator intrigante e tema de pesquisas variadas que colabora com a superação completa da cosmovisão randômica.

Em termos simples e resumidos, os modelos de aprendizado de máquina profundo atuais, que permitiram a aceitação dos Agentes de IA disponíveis no mercado global, são baseados em grafos gerados pelas máquinas de Boltzmann, que através de seu modelo estocástico de aprendizado, pode ser representado e emulado em uma rede neural artificial. Para elucidar a necessidade de tão longa cronologia e revisão dos fatos e teorias científicas em torno desses modelos, neste capítulo, é que eles serão os agentes de conhecimento da Análise de Conteúdo desta tese.

4.6.1 - O MODELO BARD NO GEMINI ADVANCED

Um grande modelo de linguagem (LLM) é um recurso de inteligência artificial (IA) treinado com enormes conjuntos de dados em pipelines de aprendizado de máquina por rede neural. O modelo do Google BARD foi desenvolvido como um agente de IA generativa que oferece uma plataforma centralizada para extrair valor e aumentar o LLM com seus próprios dados, usando geração aumentada de recuperação (RAG).

A Geração Aumentada por Recuperação (RAG) aprimora os Modelos de Linguagem Grandes (LLMs) combinando-os com bases de conhecimento externas. Os LLMs são excelentes na geração de conteúdo original, mas podem não ter conhecimento específico do domínio. O RAG resolve esse problema permitindo que os LLMs acessem informações relevantes, o que resulta em respostas mais precisas e contextualmente relevantes

Neles, os pesquisadores inserem palavras-chave ou prompts de comando para retornar uma lista de conjuntos de dados, publicações e outros recursos

relevantes. Desde 2020, o Google aprimorou o MUM (Multitask United Model) para compreensão sintética, assim o Gemini pode compreender pesquisas e gerar respostas humanizadas que consideram vários contextos em uma consulta.

Os recursos multilíngues e multitarefa do modelo MUM permitem que ele compreenda as consultas e extraia insights de texto, imagens e áudio, aumentando a relevância do conteúdo, especialmente quando uma única palavra-chave pode representar vários conceitos.

O BARD, lançado inicialmente em março de 2023 com acesso limitado, concentrava-se em fornecer respostas abrangentes às perguntas. Posteriormente, o Google desenvolveu o Gemini, uma família de LLMs mais poderosa e versátil, projetada para ser multimodal e escalonável. No final de 2023, o Google integrou o Gemini ao Bard, marcando a evolução do modelo e herdando os recursos e as funcionalidades do Gemini.

A otimização de dados passou do foco em uma única palavra-chave para a “otimização multicontextual”, criando conteúdo que abrange uma gama mais ampla de tópicos relacionados e responde a diversas consultas de pesquisa dos usuários. Isso se alinha com a funcionalidade do BERT, e o Gemini tem como objetivo tornar o conteúdo mais focado no usuário, com cada parte fornecendo informações úteis ao contexto fornecido pelo usuário.

Os modelos de linguagem ampla (LLMs) podem ter dificuldades para permanecer relevantes, mesmo com dados de treinamento adequados. O Retrieval Augmented Generation (RAG) resolve isso conectando os LLMs diretamente a fontes de informações atuais, como mídia social e sites de notícias. Isso permite que os LLMs forneçam aos usuários as informações mais atualizadas.

Os modelos de linguagem grandes dependem da entrada do usuário para gerar respostas; o Retrieval Augmented Generation (RAG) introduz um componente de recuperação de informações que usa a entrada do usuário para extrair informações relevantes de uma fonte de dados e as fornece ao LLM para gerar respostas otimizadas⁷³.

Dados externos são novas informações não encontradas no conjunto de dados de treinamento original de um LLM. Eles vêm em vários formatos e de muitas

⁷³ [Página da aws.amazon sobre RAG pipeline](#)

fontes e são convertidos em representações numéricas que os modelos de IA generativos podem entender. Uma pesquisa de relevância converte a consulta do usuário em um vetor, compara-a com bancos de dados de vetores e recupera os documentos mais relevantes.

O modelo RAG adiciona dados recuperados relevantes ao prompt do usuário e usa técnicas de engenharia de prompt para se comunicar de forma eficaz com o LLM. Esse prompt enriquecido permite que grandes modelos de linguagem gerem respostas mais contextualizadas às consultas do usuário.

O Google BERT e o Google BARD são ferramentas poderosas com diferentes funções e aplicações no processamento de linguagem natural. O BERT é uma tecnologia de processamento de linguagem natural (PLN) que ajuda o Google a entender as consultas de pesquisa e a fornecer resultados relevantes, além de ser usado para classificação de textos e tradução de idiomas.

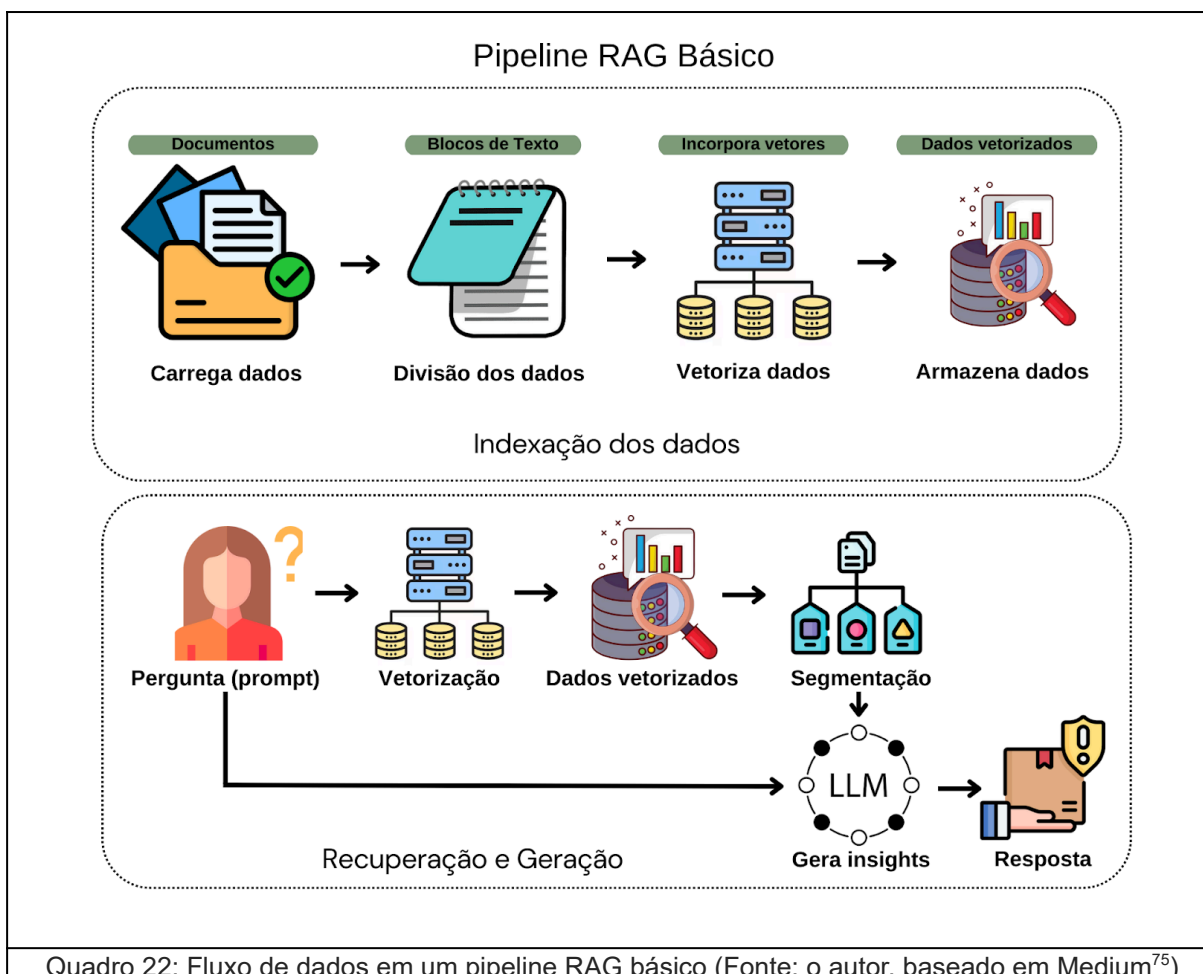
O BARD, por outro lado, é projetado para ter conversas de domínio aberto com usuários e é usado principalmente em chatbots e outros aplicativos de conversação, como bots de atendimento ao cliente e chatbots educacionais. Ambos os modelos têm diferentes dados de treinamento, arquiteturas de modelo e equipes de desenvolvimento, o que enseja diferentes aplicações para diferentes modelos.

A pesquisa semântica aprimora o Retrieval Augmented Generation (RAG), permitindo que pesquisadores integrem vastas fontes de conhecimento externo em seus aplicativos de modelo de linguagem ampla (LLM).

As abordagens tradicionais de pesquisa baseadas em palavras-chave no RAG geralmente produzem resultados limitados, mas as tecnologias de pesquisa semântica automatizam o processo de preparação da base de conhecimento e fornecem trechos de texto semanticamente relevantes, melhorando a qualidade do conteúdo fornecido ao RAG⁷⁴.

Outra vantagem competitiva dos RAGs é que utilizando suas próprias bases de conhecimento ao invés das bases em que os LLMs foram treinados pela internet, a probabilidade de alucinação e delírios é consideravelmente menor.

⁷⁴ <https://aws.amazon.com/pt/bedrock/>



Outra vantagem competitiva dos RAGs é que utilizando suas próprias bases de conhecimento ao invés das bases em que os LLMs foram treinados pela internet, a probabilidade de alucinação e delírios é consideravelmente menor.

4.6.2 - Análises de conteúdos apoiada por LLMs/RAG

As Ciências de Dados Sociais desenvolvem novas formas de geração de conhecimento através da mineração, análise e descrição de percepções sobre os estudos culturais apoiado por métodos estatísticos e probabilísticos quantitativos, que nortearam a análise de conteúdos até o desenvolvimento de ferramentas e plataformas de big data que se tornaram importantes ativos de empresas *big tech*.

As técnicas deste método de pesquisa possibilitam analisar e interpretar conteúdos comunicativos escritos, orais ou audiovisuais, gerando inferências e percepções através de análises sistematizadas e objetivos dos conteúdos, buscando encontrar significados implícitos ao observar uma cultura em sua origem natural.

⁷⁵ <https://medium.com/@drjulija/what-is-retrieval-augmented-generation-rag-938e4f6e03d1>

Os métodos de análise de conteúdo propostos por Laurence Bardin e Bernard Berelson possuem uma boa capilaridade no Brasil, principalmente Bardin, que tem em nosso país seu maior público e citações. Bardin é referenciado como criador do método mais aplicado na academia com uma abordagem de análise de conteúdo categorial, com grande impulso dos brasileiros que o adotaram metodologicamente em muitos estudos, durante décadas.

Um conjunto de técnicas de análise das comunicações visando obter, por procedimentos sistemáticos e objetivos de descrição do conteúdo das mensagens, indicadores (quantitativos ou não) que permitam a inferência de conhecimentos relativos às condições de produção/recepção (variáveis inferidas) destas mensagens (Bardin, 2016, p. 48).

Embora Bardin seja mais rigorosa na categorização como ferramenta para organizar e classificar as informações, ambos analisam os conteúdos de forma objetiva e sistematizada em categorias.

No livro “Análise de Conteúdo”, a autora define a técnica categorial como um conjunto de instrumentos metodológicos que possibilitam abordar diferentes tipos de conteúdo, verbal ou não verbal, pela implementação de procedimentos sistemáticos e objetivos. Sua abordagem permite uma avaliação exaustiva e mais ampla dos conteúdos, incluindo aspectos manifestos e latentes, a derivação de significados e a compreensão de fenômenos no mundo real.

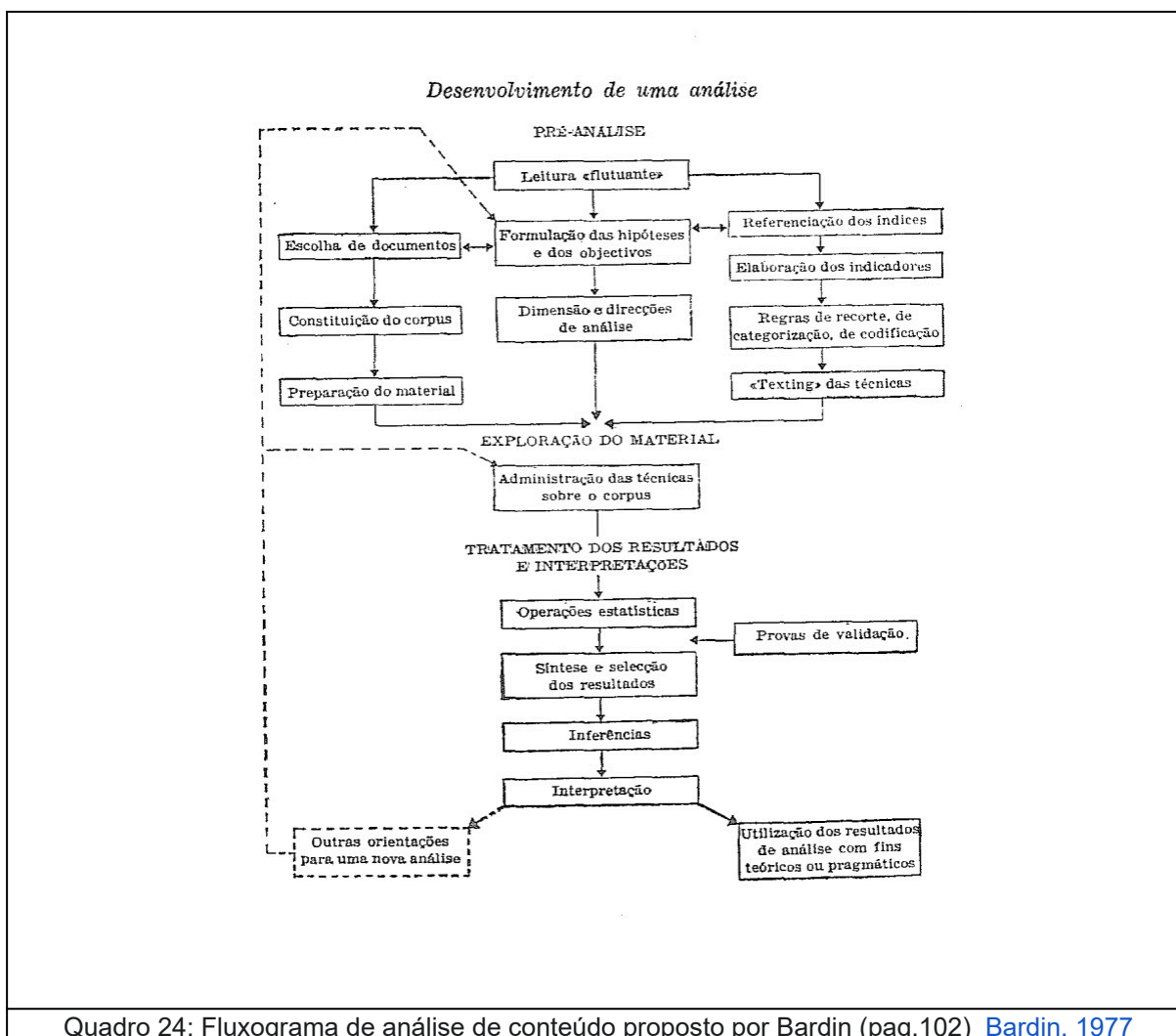
Também aceita que diferentes tipos de materiais sejam adaptados aos objetivos da pesquisa, embora dificulte a interpretação por incidência de subjetividades e vieses, pois segundo ela, é difícil analisar dados, pois exige, ao mesmo tempo, rigor e delicadeza na interpretação (BARDIN, 2016).

A análise categorial de conteúdos concebe a codificação completa dos conteúdos conversacionais em suas respectivas categorias, definindo sua especificidade.

A base da análise de conteúdo categorial, se divide em três fases:

Organização prévia	Etapa em que o pesquisador se familiariza com o material, suas hipóteses e problemas e define objetivos sobre a análise. Dimensionando o material nessa fase, Bardin classifica-a em etapas: a leitura flutuante , para iniciar o contato com o material; as escolhas dos que serão analisados; hipóteses e objetivos ; indexação de referências e indicadores
Exploração do material	No próprio material, as unidades de códigos, como palavras, frases e parágrafos são segmentadas em categorias através de critérios pré-definidos.

Tratamento de resultados, inferência e interpretação	Resultados apresentados a partir da análise são agrupados, interpretados e associados ao referencial teórico o que possibilita produzir inferências e conclusões .
Quadro 23: Três requisitos base da análise de conteúdo categorial (o autor)	



Quadro 24: Fluxograma de análise de conteúdo proposto por Bardin (pag.102) [Bardin, 1977](#)

4.6.3 - Comparativo Berelson x Bardin

O método de Berelson funciona como contraponto metodológico e conceitual a Bardin, embora sejam complementares e usados em pesquisas com enfoque híbrido que busca legitimidade e rigor científico na coleta e análise de dados quantitativos que possam sustentar a análise qualitativa.

Analisando comparativamente estes dois autores, podemos perceber as diferenças e vantagens de cada método em diferentes aplicações. Uma diferença fundamental reside na abordagem quantitativa de Berelson que combina rigor

estatístico com a descrição detalhada dos métodos qualitativos. Bardin enfatiza a interpretação mais intuitiva dos significados, buscando compreender fenômenos implícitos nas inter-relações entre atores e elementos.

Essa ênfase no conteúdo latente proposto por Bardin contrapõe a preferência de Berelson pela explicitude daquilo que está presente, embora apenas priorizasse uma ou outra abordagem em determinados casos. Enquanto Bardin desenvolveu métodos com abordagens mais qualitativas, no sentido de buscar compreender os fenômenos objetivamente, Berelson priorizou a quantificação e a análise estatística³, concentrando-se na análise lexical do vocabulário e em análises documentais ou empíricas.

Ele teve importante contribuição no desenvolvimento de técnicas de análise de conteúdo durante a Segunda Guerra Mundial para decifrar e contextualizar a propaganda nazista. Sendo um dos pioneiros do método, identificava padrões e tendências nas palavras, temas ou pessoas através da contagem de frequência de termos específicos no conteúdo analisado. Descreveu 17 diferentes finalidades do método, comparando diversas fontes para avaliar a eficácia da comunicação, que é um dos objetivos deste estudo para analisar se os conteúdos de 2 projetos de extensão realizam dinâmicas de difusão e letramento científico como cultura científica.

Como unidade elementar da análise de conteúdo, através do código pode-se resumir, filtrar ou agregar dados objetivamente, que por sua vez são agrupados em categorias que sintetizam os aspectos que serão estudados.

Como já explicado, o código é a unidade elementar da AC. Ele irá resumir, filtrar ou condensar dados de acordo com os objetivos e com os interesses da pesquisa. Grupos de códigos, por sua vez, são agrupados em categorias, ou seja, unidades analíticas que materializam as questões a serem verificadas.

Uma análise de conteúdo será produtiva na medida em que suas categorias forem claramente formuladas e bem adaptadas ao problema de pesquisa e ao conteúdo sendo estudado. BERELSON (1952)

Berelson propôs classificar pela tripla polaridade (positivo, negativo ou neutro) como nas atuais técnicas de análise de sentimentos da netnografia, o que indica um reducionismo característico em alguns de seus trabalhos (BARDIN, 2016). Por outro

lado propôs as características das categorias que considerava indispensáveis para analisar conteúdos:

- **Homogêneas:** As categorias e seus códigos devem ser homogêneas entre si e apresentar distinção absoluta (homogeneidade intra-grupo e heterogeneidade entre-grupos)
- **Exaustivas:** Os conteúdos precisam ser codificados em consonância com os objetivos da pesquisa, embora sejam aceitas generalizações residuais agrupadas em Outros para codificações inviáveis ou nulas..
- **Exclusivas:** As categorias e seus códigos precisam ser mutuamente exclusivos, não sendo possível qualquer redundância
- **Objetivas:** diferentes codificadores apresentam resultados iguais
- **Pertinentes ou Adequadas:** adaptadas ao objetivo e ao conteúdo

Estas definições foram confirmadas na época e posteriormente por uma parte considerável da literatura acadêmica especializada, que tornaram absolutamente indispensáveis três delas: exclusivas, exaustivas e homogêneas, nesta ordem de importância, tanto para códigos como para categorias (SAMPAIO e LYCARIÃO, 2021).

Se observadas essas três condições e as premissas estiverem de acordo com os interesses da pesquisa, a tipologia alcançará também a pergunta da mesma, suas hipóteses e os testes de validação desta tese podem ser encadeados para obter-se percepções multivariadas. O referencial de codificação e categorização obtidos são transformados no que o método chama de Livro de Códigos, do inglês codebooks, que descrevem ambos detalhadamente, bem como as regras aplicadas e seguidas explicitamente (IBIDEM).

Para exemplificar a estrutura fundamental do livro (LdC), podemos citar seis elementos principais que auxiliam criar definições e instruções simples, fáceis de seguir e livre de ambiguidades:

1. Categorias e seus códigos
2. Descrição resumida
3. Definição completa
4. Regras de aplicação ou não dos códigos
5. Exemplos

Após classificar e numerar códigos e categorias, é elaborado um formulário de codificação (FdC), que pode ser uma planilha ou aplicativo que contemple todas as categorias em seus possíveis códigos (IBIDEM). Os dados do formulário também pode ser preenchido com variáveis oriundas de dataframes estatísticos e relacionais⁷⁶ de redes sociais e plataformas de streaming.

Os tipos de coleta possíveis para análise de conteúdo são descritos extensivamente e explicadas com diferentes casos explicados em 9 diferentes modalidades, divididas entre técnicas probabilísticas randômicas e não randômicas (não probabilísticas) em (SAMPAIO e LYCARIÃO, 2021), a partir de uma síntese dos manuais de Neuendorf (2002) e Krippendorff (2004):

1. Técnicas probabilísticas (randômicas):
 - a. Amostragem aleatória simples
 - b. Amostragem sistemática
 - c. Amostragem por agrupamento
 - d. Amostragem estratificada
 - e. Amostragem em múltiplos estágios
 - f. Amostragem em bola de neve
2. Técnicas não probabilísticas (não aleatórias)
 - a. Amostragem por conveniência
 - b. Amostra por propósito ou relevância
 - c. Amostra por cotas
 - d. Censo

O mesmo manual detalha a importância do tamanho ou dimensão da amostra, sendo um dos desafios recorrentes em qualquer tipo de coleta, pois a definição precisa atribuir aos resultados obtidos pela análise da amostra, as mesmas características de todos os registros que não fora efetivamente analisados, por terem ficado fora da amostragem.

A atribuição da representatividade de uma amostra populacional, em termos mais simples, implica a amostragem de uma população com alguma probabilidade (ou seja, naturalmente, de forma aleatória). No entanto, Krippendorff observa que essa determinação probabilística se baseia em suposições da teoria estatística que frequentemente não são totalmente legítimas no caso da análise de conteúdo.

Portanto, para amostras de análise de conteúdo minimamente

⁷⁶ Um banco de dados relacional é um conjunto de informações que organiza dados em relações predefinidas em uma ou mais tabelas (ou "relações") de colunas e linhas, facilitando a visualização e a compreensão. fonte: <https://cloud.google.com/learn/what-is-a-relational-database?hl=pt-BR>

representativas, mesmo quando se usam métodos de seleção probabilísticos apropriados e muitas tentativas deliberadas, é improvável que haja representatividade, como explica Krippendorff (2004, pp. 110-113)

O Manual da Análise de Conteúdos (SAMPAIO e LYCARIÃO, 2021) apresentou uma espécie de roteiro de condução de análise de conteúdo categorial, baseado na sugestão de Riffe, Lacy, Fico (2014) da divisão em três grandes etapas, cujo detalhamento segue:

ETAPAS DA ANÁLISE DE CONTEÚDO	
Conceituação	
1. Identificar o problema (revisão de literatura)	
2. Questões de pesquisa e hipóteses	
Desenho	
3. Selecionar a(s) unidade(s) e subunidade(s) de análise	
4. Criar e definir categorias	
a. elaboração do livro de códigos	
b. elaborar a planilha de codificação	
5. Amostragem	
6. Pré-teste das categorias e das regras de codificação	
a. treinamento	
b. revisão do livro de códigos	
c. teste de confiabilidade-piloto	
7. Treinamento final e teste de confiabilidade das categorias	
8. Codificação	
9. Testes de confiabilidade intermediário e final	
Análise	
10. Tabulação e aplicação de procedimentos estatísticos	
11. Interpretar e reportar os resultados	
12. Validação e replicabilidade	

Quadro 25: Detalhamento das etapas de AC em 3 etapas⁷⁷ (Fonte: (SAMPAIO e LYCARIÃO, 2021))

A análise de conteúdo originou de um caráter meramente quantitativo, criado como necessidade de análise de propaganda durante a Segunda Guerra Mundial, evoluindo para um método que introduziu as premissas de interpretação e significado (BERELSON, 1952), (HOLSTI, 1969). Posteriormente, influenciada por Neuendorf (2002), Krippendorff (2004), Bardin (2016) e outros⁷⁸, adotou uma

⁷⁷ Na especificação das subetapas, os autores fizeram um amálgama entre as sugestões de outros autores (Downe-Wamboldt, 1992; Weber, 1990; Neuendorf, 2002). pág. 49

https://repositorio.enap.gov.br/bitstream/1/6542/1/Analise_de_conteudo_categorial_final.pdf

⁷⁸https://repositorio.enap.gov.br/bitstream/1/6542/1/Analise_de_conteudo_categorial_final.pdf

interpretação essencialmente qualitativa que valoriza as inferências fundamentadas em análises híbridas (estatísticas e analíticas).

Concluindo o denso e explicativo Manual de Análise de Conteúdo (cf. Sampaio et al, 2021), as revisões atuais da literatura mostram que as pesquisas brasileiras relatam superficialmente os procedimentos de análise de conteúdo e a maioria delas ainda se baseia excessivamente no manual de Bardin. Não sendo exclusividade do Brasil, revistas internacionais demonstraram a percepção clara de que a maioria dos artigos não aplicaram métodos de amostragem adequados e nem foram transparentes sobre os critérios de classificação na construção da amostra, não usando censo e nem amostra probabilística (SAMPAIO e LYCARIÃO, 2021).

Para exemplificar de forma mais visual e dinâmica, sintetizamos um resumo das características chave dos dois autores que encerra a análise comparativa.

Característica	Bardin	Berelson
Ênfase	Conteúdo manifesto e latente	Conteúdo manifesto, mas reconhecendo a existência do conteúdo latente
Abordagem	Qualitativa e quantitativa	Quantitativa, com ênfase na análise estatística
Objetivo	Compreensão profunda dos significados	Descrição objetiva e identificação de padrões
Fases	Pré-análise, exploração do material, tratamento dos resultados	Codificação, contagem, análise estatística
Quadro 26: Comparativo entre Bardin e Berelson (o autor)		

Atualmente auxiliada por softwares de Processamento de Linguagem Natural (PLN), de Aprendizado de Máquina (AM) e Inteligência Artificial Generativa (LLMs e RAGs) a prática estatística informacional na Ciência de Dados Sociais, possibilita um alto volume de análise de dados em grande escala e de diferentes fontes simultaneamente, oferecendo um novo arcabouço científico e metodológico.

4.6.4 - Os recursos do software de análise qualitativa MAXQDA

Esta pesquisa desenvolveu e implementou um modelo de chatbot integrado à base de conhecimento armazenada vetorialmente em uma aplicação no AWS Bedrock (LLMs/GraphRAGs). O desenvolvimento obteve sucesso ao alimentar a

base de conhecimento com os conteúdos da tese de forma estruturada por menus e treinar o modelo com os textos do site.

Ficou caracterizado que o modelo desenvolvido não atendia a necessidade de coletar e manipular conteúdos audiovisuais, chats e comentários, o que levou a opção de assinar uma versão acadêmica paga do software e plataforma MAXQDA⁷⁹. A opção pelo software de análise qualitativa com inúmeras opções de coleta, tratamento e exposição de dados foi fundamental para tratar devidamente os conteúdos audiovisuais.

O MAXQDA é um software acadêmico para pesquisa qualitativa e de métodos mistos e pode analisar vários dados não estruturados, incluindo entrevistas, grupos de foco, arquivos de áudio/vídeo/imagem e dados do Twitter ou Youtube. Ele também funciona como um minerador avançado com recursos de IA generativa, classificador com inúmeras ferramentas de organização e consistência, além de gerador de gráficos quantitativos, tabelas, nuvens de palavras e transcrições de áudios e vídeos dos canais de streaming e dos comentários para análises de sentimento e do discurso.

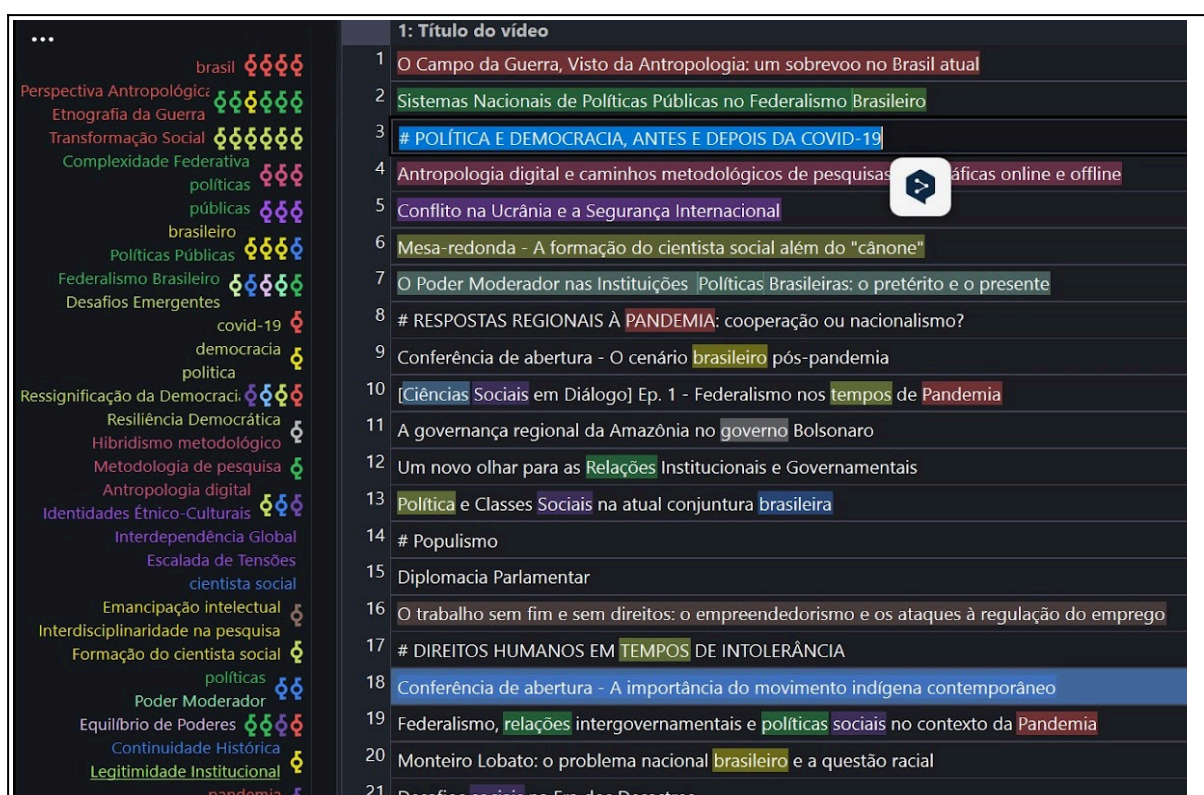
Através dos recursos do Gemini Advanced (versão paga também) e do MAXQDA 2024, foram realizadas as análises qualitativas da tese em um modelo híbrido com as análises descritivas. A articulação de ambas forneceram percepções e insights sobre as dinâmicas de cultura científica em dois projetos de extensão que buscam divulgar e difundir conhecimento e levar o letramento científico para a comunidade interna e externa da instituição de ensino superior.

Os recursos dos dois softwares se complementam e resultaram em análises comparativas entre os resumos, análises de sentimento e do discurso sobre os mesmos conteúdos, oferecendo uma visão crítica e síncrona dos resultados de diferentes pipelines de aprendizado de máquina (LLMs/RAG). O MAXQDA possibilitou a aplicação do método de análise, como anteriormente proposto neste capítulo (págs. 158/164), mas com adaptações que respondem às novas formas de coletar, minerar, classificar e vetorizar dados e que afetam diretamente a forma como eles são processados e analisados.

⁷⁹ <https://www.maxqda.com/pt>

Todas as etapas da análise de conteúdo proposto no detalhamento em 3 etapas⁸⁰ e na especificação das subetapas de outros autores (Downe-Wamboldt, 1992; Weber, 1990; Neuendorf, 2002 APUD SAMPAIO e LYCARIÃO, 2021) não são mais necessárias, pois os conteúdos coletados já são limpos e até mesmo classificados e vetorizados pelos recursos do MAXQDA.

As ferramentas de codificação, categorização, geração do livro de códigos, nuvem de palavras e tabela de frequência delas e dos códigos atende, tanto as etapas propostas por Bardin, que são Pré-análise, Exploração do Material e Tratamento dos Resultados, bem como as definidas por Berelson como Codificação, Contagem e Análise estatística.



Quadro 27: Interface visual do MAXQDA exibindo codificação dos títulos audiovisuais (o autor)

A codificação dos conteúdos audiovisuais é automatizada após a transcrição pelos recursos do MAXQDA que permite gerar nuvem e tabela de frequência de palavras, que alimenta o livro de códigos. Por sua vez, os títulos de cada uma das transmissões gera até três codificações em diferentes categorias que delimitam o assunto abordado pelo mesmo assunto com abordagens diferentes.

⁸⁰ Na especificação das subetapas, os autores fizeram um amálgama entre as sugestões de outros autores (Downe-Wamboldt, 1992; Weber, 1990; Neuendorf, 2002). pág. 49

https://repositorio.enap.gov.br/bitstream/1/6542/1/Analise_de_conteudo_categorial_final.pdf

Uma amostra de 99 codificações entre todos os títulos das transmissões foi impressa para exemplificar o livro de códigos, aqui neste capítulo e que estará integralmente disponível no site da tese (tecnociencia.net) com dezenas de páginas.

A amostra inicia com uma lista seguida das três codificações propostas para cada título e seus códigos, oferecendo maior flexibilidade e efetividade aos buscadores e eficiência nas buscas por IA generativa e Aprendizado de Máquina (AM). Esta codificação e categorização possibilita que os conteúdos audiovisuais sejam localizados e seus recursos em forma textual possam ser aplicados em diferentes finalidade de pesquisa, ensino e extensão.

Abaixo, um exemplo da amostra de 99 codificações do Livro de Códigos dos títulos dos conteúdos audiovisuais do Projeto Ciências Sociais em Diálogo e suas respectivas sínteses. Ao final há uma tabela com as palavras recorrentes

Livro de códigos Projeto Ciências Sociais em diálogo.mx24 06/01/2025	Lista de Códigos		
	1 Reconfiguração de prioridade	34 Poder Moderador	68 Compartilhamento de Informações
	2 Desafios estruturais	35 Legado literário e cultural	69 Programa de Rádio/TV
	3 Transformações sociais	36 Resistência e transformação	70 Engajamento cívico
	4 Identidade regional	37 Memória coletiva e identidade	71 Participação
	5 Interdependência global	38 Perspectiva histórico-social	72 cientista social
	6 Pandemia	39 Memória e identidade	73 Bolsonaro
	7 Emancipação intelectual	40 Vulnerabilidade	74 Portugal
	8 Interdisciplinaridade na pesquisa	41 Socioambiental	75 linguagem
	9 Formação do cientista social	42 Computacionais	76 trabalho
	10 Identidades Étnico-Culturais	43 Afastamento dos militares	77 pública
	11 Interdependência Global	44 Liderança de Lula	78 política
	12 Escalada de Tensões	45 Papel dos EUA	79 Brasil
	13 Hibridismo metodológico	46 Golpe de Estado	80 Importação do YouTube
	14 Metodologia de pesquisa	47 Presença militar no poder	81 mesa
	15 Antropologia digital	48 Destruição do Estado brasileiro	82 governo
	16 Ressignificação da Democracia	49 Terceira via militar	83 estados
	17 Resiliência Democrática	50 Guerra híbrida	84 eleitoral
	18 Desafios Emergentes	51 Guerra híbrida	85 democracia
	19 Complexidade Federativa	52 Projeto de poder dos generais	86 brasileiro
	20 Federalismo Brasileiro	53 Desafios da carreira científica	87 tempos
	21 Políticas Públicas	54 Trajetória de vida	88 públicas
	22 Transformação Social	55 Ciência	89 crise
	23 Etnografia da Guerra	56 Discriminação institucional	90 brasileira
	24 Perspectiva Antropológica	57 Problema público	91 relações
	25 Equilíbrio de Poder	58 Intolerância religiosa	92 eleições
	26 Descentralização Decisória	59 Empoderamento comunitário	93 covid-19
	27 Políticas Públicas	60 Importância	94 ciências
	28 Violência Estrutural	62 Conexão com a Realidade	95 sociais
	29 Contexto Nacional	63 Política Educacional	96 políticas
	30 Antropologia	64 Transição para o Mercado de Trabalho	97 política
	31 Continuidade Histórica	65 Reforma do Ensino Médio	98 Brasil
	32 Legitimidade Institucional	66 apresentação	99 pandemia
	33 Equilíbrio de Poderes	67 Construção de Narrativa	

1 Reconfiguração de prioridades Mudanças nos valores e na agenda pública pós-pandemia. 2 Desafios estruturais Problemas de longa data que se agravaram durante a pandemia. 3 Transformações sociais Mudanças na dinâmica social brasileira após a pandemia. 4 Identidade regional Sentimento de pertencimento e coesão entre regiões. 5 Interdependência global Relações e conexões entre regiões em um contexto globalizado. 6 Pandemia Desafios e impactos causados pela crise sanitária global. 7 Emancipação intelectual Busca por uma formação acadêmica mais abrangente e transformadora. 14 Metodologia de pesquisa Abordagens e técnicas utilizadas na investigação. 15 Antropologia digital Estudo da cultura e sociedade no ambiente digital. 16 Ressignificação da Democracia Interpretação da ressignificação do conceito e da prática da democracia.	8 Interdisciplinaridade na pesquisa Integração de diferentes áreas do saber na formação do cientista social. 9 Formação do cientista social Discussão sobre a educação e o desenvolvimento do pesquisador social. 10 Identidades Étnico-Culturais Papel das diferenças étnicas, linguísticas e culturais como fontes de divisão e conflito. 11 Interdependência Global Conexões e interdependências que tornam os conflitos locais em questões internacionais. 12 Escalada de Tensões Aumento gradual das hostilidades e do risco de uma maior confrontação. 13 Hibridismo metodológico Integração de métodos de pesquisa online e offline. 17 Resiliência Democrática Avaliação da resiliência e da capacidade de recuperação da democracia. 18 Desafios Emergentes Identificação dos desafios políticos e democráticos decorrentes da pandemia.	19 Complexidade Federativa Desafios de implementação de políticas em contexto federativo. 20 Federalismo Brasileiro Modelo de organização política e administrativa do país. Estudo antropológico das práticas e significados da guerra. 24 Perspectiva Antropológica Abordagem da antropologia para compreender a realidade social. 25 Equilíbrio de Poder Tensões e negociações entre União, estados e municípios. 26 Descentralização Decisória Distribuição de poder e autonomia entre níveis federativos. 27 Políticas Públicas Conjunto de ações governamentais para atender demandas sociais. 28 Violência Estrutural Dinâmicas de poder e opressão enraizadas no país. 29 Contexto Nacional Análise da realidade brasileira em sua totalidade. 30 Antropologia Perspectiva acadêmica para compreender a realidade social.	21 Políticas Públicas Conjunto de ações governamentais para atender demandas sociais. 22 Transformação Social Análise das possibilidades de mudança e reconstrução da realidade. 23 Etnografia da Guerra 31 Continuidade Histórica Influência do passado nas estruturas políticas atuais. 32 Legitimidade Institucional Percepção e aceitação do Poder Moderador. 33 Equilíbrio de Poderes Papel do Poder Moderador na balança de poderes. 34 Poder Moderador Conceito histórico e político no Brasil. 35 Legado literário e cultural Importância da obra dos autores para a compreensão da cultura brasileira. 36 Resistência e transformação Potencial de mudança e superação das estruturas de poder. 37 Memória coletiva e identidade nacional Investigação da formação da identidade brasileira.																												
38 Perspectiva histórico-sociológica Análise do contexto social e histórico da sociedade rural brasileira. 39 Memória e identidade Exploração das raízes e da formação da identidade brasileira. 40 Vulnerabilidade 41 Socioambiental 42 Computacionais 43 Afastamento dos militares da política Discussão sobre a necessidade de proibir a participação de militares na política. 44 Liderança de Lula Análise do papel e da posição de Lula nesse cenário de disputa de poder. 45 Papel dos EUA Investigação sobre a possível participação e influência dos Estados Unidos nesse contexto. 46 Golpe de Estado Discussão sobre a possibilidade de um autogolpe militar e suas implicações. 47 Presença militar no poder Discussão sobre a influência dos militares no cenário político brasileiro.	48 Destruição do Estado brasileiro Visão de que há um projeto para dismantlar as instituições do país. 49 Terceira via militar Possibilidade de os militares se apresentarem como uma alternativa política. 50 Guerra híbrida Referência a uma estratégia de guerra não convencional. 51 Guerra híbrida Análise do conceito de guerra híbrida e sua aplicação no contexto brasileiro. 52 Projeto de poder dos generais Preocupação com o suposto plano de tomada de poder pelos militares. 53 Desafios da carreira científica Reflete sobre os obstáculos enfrentados por mulheres na ciência. 54 Trajetória de vida Análise a evolução e as experiências de Carlota. 55 Ciência Retrata a natureza e o processo da ciência. 56 Discriminação institucional Práticas e políticas que excluem e marginalizam minorias religiosas.	57 Problema público Questão social relevante que requer atenção e ação pública. 58 Intolerância religiosa Conflitos e discriminação envolvendo crenças e práticas religiosas. 59 Empoderamento comunitário Fortalecimento da autonomia e do sentido de pertencimento dos povos indígenas. 60 Importância Relevância e significado do movimento indígena para a sociedade. 62 Conexão com a Realidade Alinhamento do ensino médio com as demandas do mundo real. 63 Política Educacional Diretrizes e estratégias governamentais para o ensino médio.	64 Transição para o Mercado de Trabalho Facilitação da entrada dos estudantes no mundo do trabalho. 65 Reforma do Ensino Médio Mudanças estruturais no currículo do ensino médio. 67 Construção de Narrativa Construção de uma narrativa sobre o programa. 68 Compartilhamento de Informações Transmissão e troca de informações. 69 Programa de Rádio/TV Referência a um programa de rádio ou TV. 70 Engajamento cívico Participação em assuntos públicos ou comunitários. 71 Participação Envolvimento em uma atividade ou iniciativa.																												
<table> <tr> <td>72 cientista social</td><td>73 Bolsonaro</td><td>74 Portugal</td><td>75 linguagem</td></tr> <tr> <td>76 trabalho</td><td>77 pública</td><td>78 política</td><td>79 brasil</td></tr> <tr> <td>80 YouTube</td><td>81 mesa</td><td>82 governo</td><td>83 estados</td></tr> <tr> <td>84 eleitoral</td><td>85 democracia</td><td>86 brasileiro</td><td>87 tempos</td></tr> <tr> <td>88 públicas</td><td>89 crise</td><td>90 brasileira</td><td>91 relações</td></tr> <tr> <td>92 eleições</td><td>93 covid-19</td><td>94 ciências</td><td>95 sociais</td></tr> <tr> <td>96 políticas</td><td>97 política</td><td>98 brasil</td><td>99 pandemia</td></tr> </table>				72 cientista social	73 Bolsonaro	74 Portugal	75 linguagem	76 trabalho	77 pública	78 política	79 brasil	80 YouTube	81 mesa	82 governo	83 estados	84 eleitoral	85 democracia	86 brasileiro	87 tempos	88 públicas	89 crise	90 brasileira	91 relações	92 eleições	93 covid-19	94 ciências	95 sociais	96 políticas	97 política	98 brasil	99 pandemia
72 cientista social	73 Bolsonaro	74 Portugal	75 linguagem																												
76 trabalho	77 pública	78 política	79 brasil																												
80 YouTube	81 mesa	82 governo	83 estados																												
84 eleitoral	85 democracia	86 brasileiro	87 tempos																												
88 públicas	89 crise	90 brasileira	91 relações																												
92 eleições	93 covid-19	94 ciências	95 sociais																												
96 políticas	97 política	98 brasil	99 pandemia																												
Quadro 28: Livro de Códigos do Canal Ciências Sociais em Diálogo (o autor)																															

Através da codificação dos títulos dos conteúdos, que por si só já carregam informações sobre o escopo e finalidade do debate, foi possível realizar análises das

A nuvem de palavras a seguir resultou da frequência de códigos de todos os títulos das transmissões e apresenta uma distribuição diagonal das palavras em 45 graus, além das horizontais e verticais, diferenciando estas nuvens cujos dados se originaram de agregações de conteúdos.



A tabela de frequência ranqueia e classifica cada palavra quanto ao seu tamanho e frequência de ocorrências, gerando os dados aplicados na nuvem de palavras, que organiza as mais frequentes pelo seu tamanho e centralidade. A nuvem de palavras capturou aproximadamente 4 anos de transmissões de extensos debates de mais de 1 hora e meia, em média, muitas vezes influenciada pela sazonalidade dos assuntos ou emergências inesperadas como a pandemia e seus desdobramentos.

Por inferência relativa aos títulos com maior número de visualizações e tempo total de exibição, classificamos os temas mais relacionados no conjunto de conteúdos do canal. De acordo com os vídeos mais vistos e tempo total assistido, os temas mais relevantes indicam importantes pautas da política e sociedade brasileira, tal como a crise da COVID-19, crise política, governo Bolsonaro, a questão do racismo, o federalismo e as políticas públicas.

Outro tema de grande interesse é a debate sobre o papel das Ciências Sociais e da universidade na sociedade contemporânea, com vídeos como "Mesa - O lugar do cientista social na sociedade digital" , "Antropologia digital e caminhos metodológicos de pesquisas etnográficas online e offline", e "Conferência de Abertura: as transformações do século XXI" .

Além disso, questões relacionadas a movimentos sociais, como o movimento indígena, economia e trabalho e a reforma trabalhista, também foram abordadas em vídeos com boa repercussão com atenção específica à pandemia, à crise política, às questões raciais, ao federalismo, às políticas públicas e aos atuais papéis das universidades e das ciências sociais.

Quadro 29: Resumo de IA (LLM/RAG) do MAXQDA (Bedrock Claude)

O conjunto apresenta uma série de palestras e debates sobre temas relevantes nas Ciências Sociais, abordando questões políticas, econômicas, sociais e culturais no Brasil e no mundo. Alguns dos tópicos abordados incluem o federalismo brasileiro, a pandemia da COVID-19 e seus impactos, a crise política no país, as relações internacionais, as desigualdades sociais, a questão ambiental, a educação, a participação social, os movimentos sociais, a atuação dos militares, a extrema direita, o racismo, a violência, a mídia e as redes sociais, entre outros.

As discussões envolvem pesquisadores, professores e especialistas de diversas áreas, como Sociologia, Antropologia, Ciência Política, Relações Internacionais e Economia, que apresentam análises e reflexões sobre esses temas, buscando compreender os

fenômenos contemporâneos e suas implicações.

Há uma ênfase na interdisciplinaridade e na abordagem crítica, com o objetivo de contribuir para o debate público e a formulação de políticas públicas. Ao longo das palestras e mesas-redondas, são discutidos os desafios enfrentados pela sociedade brasileira, as transformações em curso e as perspectivas futuras, sempre considerando a complexidade dos fenômenos e a necessidade de uma análise aprofundada.

O canal "Ciências Sociais em Diálogo" se propõe a ser um espaço de divulgação científica e de fomento ao debate sobre temas relevantes para a compreensão da realidade social, política e cultural do país.

O gráfico de tendência de palavras é calculado a partir das diferenças entre as frequências observadas e esperadas, divididas pelas frequências esperadas, sendo o teste Chi Quadrado de Pearson, muito usado para comparar duas variáveis categóricas e verificar se há correlações entre elas.

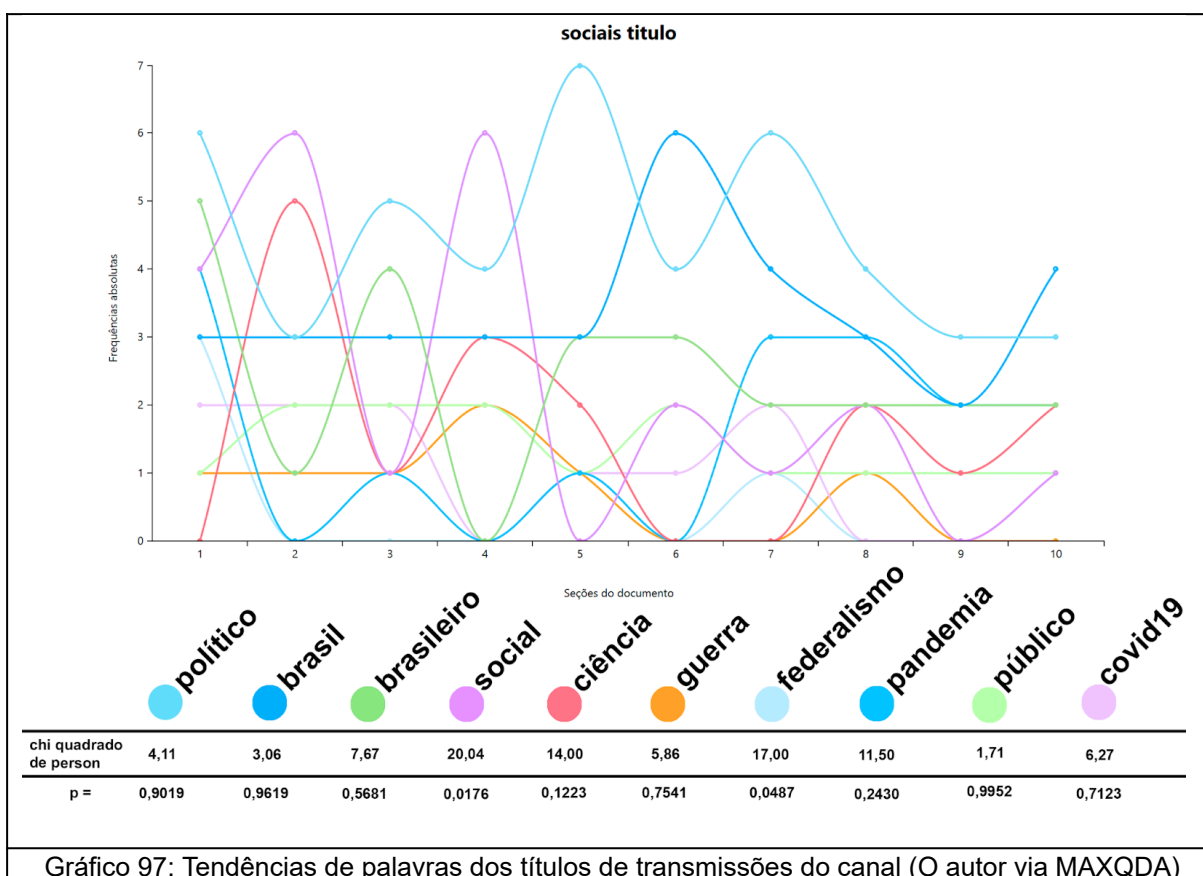


Gráfico 97: Tendências de palavras dos títulos de transmissões do canal (O autor via MAXQDA)

Não foram localizados estudos ou pesquisas acadêmicas que citaram ou utilizaram os mesmos métodos de Processamento de Linguagem Natural e pipelines

de Aprendizagem de Máquina para as análises multivariadas empregadas nesta tese.

Não sendo possível afirmar sobre o uso efetivo dos mesmos recursos em outras pesquisas ou artigos publicados, podemos declarar que não foram encontrados usos desses métodos na forma como foram aplicados nos softwares MAXQDA 24 e nem nas IAs generativas Gemini Advanced, Notebook LM ou Botpress.

4.7 - Análises de conteúdos auxiliado por Aprendizado de Máquina RAG/LLM

Os vídeos transmitidos e armazenados no Youtube, seus comentários e chats são os conteúdos analisados qualitativamente, iniciando pela transcrição do mesmo em arquivo de texto PDF pela ferramenta Maxqda e disponibilizado juntamente com acesso ao audiovisual incorporado no site. Qualquer pessoa pode ler a transcrição e acessar o trecho do conteúdo no contexto da transmissão.



Figura 33 - Tela do site com a transcrição do vídeo analisado e incorporado - Fonte: o autor

Após a transcrição, são gerados a tabela de frequência e a nuvem de palavras usadas na live, além do gráfico de tendência de palavras gerados pela análise no Maxqda. A nuvem permite uma visualização dinâmica da predominância dos assuntos pela classificação das 1631 palavras desta live ⁸¹.

A transcrição, a nuvem e a tabela de frequência de palavras são submetidos a um treinamento de aprendizado de máquina em framework Python, além de uma janela

⁸¹ <https://tecnociencia.net.br/piero-leirner/>



Transcrição: Análise e resumo IA
Gemini Advanced



Transcrição: Análise e resumo IA
MAXQDA

Quais os principais assuntos do texto ?

O texto é uma transcrição de uma palestra do antropólogo Piero Leirner sobre o papel dos militares na política brasileira. Os principais assuntos abordados são:

- * A pesquisa de Leirner sobre os militares, iniciada nos anos 90, e a percepção de que eles se veem como protagonistas na condução do país.
- * A mudança de doutrina militar após o fim da Guerra Fria e a busca por um novo protagonismo político, mesmo em tempos democráticos.
- * A influência dos militares no governo Bolsonaro, incluindo a disseminação da ideia de "guerra híbrida" e o uso de táticas de guerra psicológica.
- * O projeto de um "Estado permanente" controlado pelos militares, com foco no controle da informação e na segurança.
- * A relação entre militares e empresas, a formação de um complexo administrativo-militar e a influência de acordos internacionais na política brasileira.
- * O papel de Bolsonaro como um "projeto" dos militares, possivelmente descartável quando não for mais útil.

A palestra também inclui perguntas da plateia sobre como tirar os militares do poder, a relação entre militares e grupos religiosos, e o futuro do projeto militar após a eventual saída de Bolsonaro.

<https://g.co/gemini/share/c8fb5c2c85da>

Resumo detalhado do texto

O texto "O Campo da Guerra... o Brasil atual" aborda a relação dos militares com o poder no Brasil contemporâneo, destacando sua influência em diferentes áreas da sociedade e do governo. Inicialmente, a narrativa apresenta a experiência do autor, que começou a pesquisar militares na década de 90, sob orientação da Professora Maria Lúcia Montes, e sua imersão em estudos antropológicos nesse contexto 9-14 53-56.

O autor descreve sua pesquisa de campo na região da Amazônia, onde inicialmente buscava estudar o projeto Calha Norte, mas acabou se envolvendo em uma etnografia sobre a atuação militar na região 72-97. Ele discute a ideia de "guerra híbrida" e como os militares brasileiros estão pensando estratégias de defesa e segurança, inclusive em relação a possíveis ameaças políticas e sociais 426-442 552-556.

Além disso, o texto aborda a relação dos militares com outros setores, como empresários, políticos e igrejas, e como essas interações podem influenciar as decisões estratégicas do país 100-108 938-940. Há uma análise sobre a possível continuidade do projeto militar no Brasil, mesmo em cenários de mudança política, e como os militares buscam manter sua influência e protagonismo no cenário nacional 109-112 586-588.

O autor também discute a presença dos militares em projetos de segurança e defesa, como a possível parceria do Brasil com a OTAN, e como essas ações podem impactar as relações internacionais do país 669-671 815-817. Por fim, o texto levanta questões sobre o controle das informações e o papel dos militares na gestão do Estado, apontando para um possível projeto de Brasil pensado por eles, com foco em segurança e defesa, e como essa visão pode se refletir nas políticas e estratégias nacionais 948-950 953-958.

Pode fazer um resumo detalhado ?

Qual o papel dos militares segundo Leirner ?

E sobre a guerra ?

Qual o assunto do texto ?

Figura 35 - Tela do site com análises de IA sobre a transcrição - Fonte: o autor

Um link com acesso ao contexto do chat no Gemini Advanced é disponibilizado para que qualquer usuário com o mesmo nível de conta Google possa prosseguir com novas perguntas ao modelo, cujo recurso não existe no Maxqda, que mostra as linhas onde estão as palavras que orientam sua análise. Após processar os dados da transcrição e filtragem, são geradas tabela, nuvem de palavras e relatório curto de análise de sentimentos nos comentários do vídeo no canal (on demand).⁸²



⁸² <https://tecnociencia.net.br/piero-leirner/>

Figura 36 - Tela do site com nuvem de palavras e análises de IA sobre comentários - Fonte: o autor

Após processar os dados da transcrição, são gerados resumos, análises de sentimentos e de discurso dos diálogos nos comentários conteúdo (On demand).



Figura 37 - Tela do site com Análise de sentimentos dos comentários - Fonte: o autor

Este modelo não restringe as análises de sentimento ao reducionismo da tripla polaridade (positivo, negativo ou neutro), mas possibilita a análise semântica ou mesmo linguística simultaneamente aos estudos estatísticos transversais. Também foram analisados os emoticons presentes nos diálogos, coletando e analisando importantes recursos de comunicação já incorporados à realidade linguística e dinâmicas conversacionais online.

Os “Comentários” ocorrem após a transmissão e disponibilização pelo Youtube da versão “Por demanda”, nos quais surgem novos questionamentos e manifestações que podem se desdobrar em um diálogo entre os espectadores, ou mesmo entre eles e os palestrantes. É possível inferir que os comentários sejam estratégicos no relacionamento com os inscritos no canal, afinal, após a live, é o único espaço conversacional disponível, tendo um caráter mais informal e reflexivo que o chat, que acontece em tempo real.

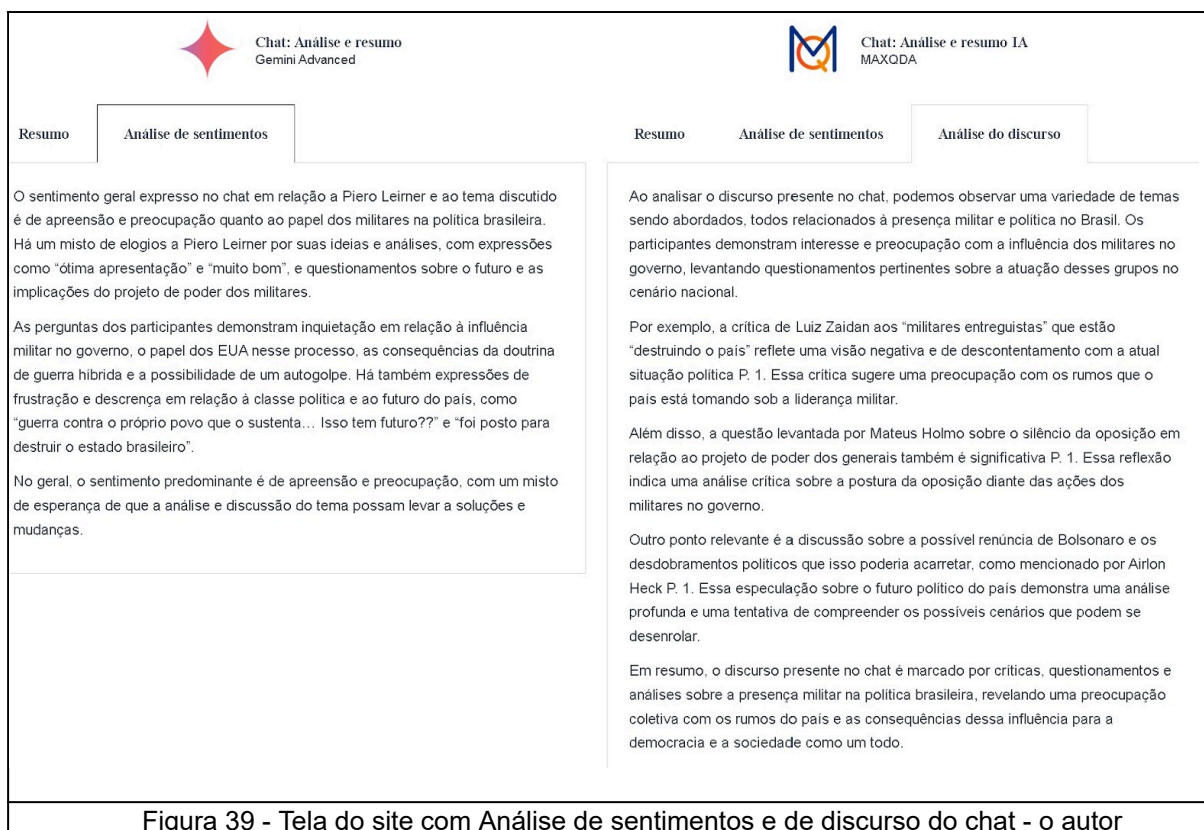


Figura 39 - Tela do site com Análise de sentimentos e de discurso do chat - o autor

As análises de sentimento de ambos os modelos de IA generativa (Gemini e MaxQDA) permitem contextualizar sua própria base de conhecimento, ao agregar sites, pdfs, planilhas, tabelas, infográficos, manuais, documentos e qualquer conjunto de dados, estruturados ou não.

Este recurso possibilita que seu chatbot ou agente de análise não precise acessar conteúdos externos da internet, caso isso seja relevante, como é o caso desta tese, ou seja, somente dados e informações da base de conhecimento que você mesmo cria, atualiza e realimenta, treinando o modelo continuamente.

Se o canal dispuser de suporte adequado, além dos dados implícitos nas conversas, pode-se estabelecer diálogos com os espectadores e se relacionar com seus inscritos, estimular novas inscrições e oferecer recursos que aumentem o engajamento. O acesso a novos conteúdos do canal e redes sociais pode gerar oportunidades de aumentar o consumo e compartilhamento de mais conteúdos por meio dessas plataformas.

4.8 - ANÁLISES DE IA DA LIVE DE PIERO LEIRNER

Seguem as perguntas e resumos sobre a Live de Piero Leirner, realizadas no Gemini Advanced e MaxQDA e disponíveis no site da tese, sempre sinalizados com um ícone de robô



4.8.1 - Quais os principais assuntos da live ?

 Gemini Advanced	 MaxQDA AI
O texto é uma transcrição de uma palestra do antropólogo Piero Leirner sobre o papel dos militares na política brasileira. Os principais assuntos abordados são:	O texto aborda principalmente a relação entre os militares e o poder no Brasil contemporâneo, com destaque para a influência e ações dos militares em diferentes esferas da sociedade e do governo.
A pesquisa de Leirner sobre os militares, iniciada nos anos 90, e a percepção de que eles se veem como protagonistas na condução do país.A mudança de doutrina militar após o fim da Guerra Fria e a busca por um novo protagonismo político, mesmo em tempos democráticos.	Ele discute a presença dos militares em projetos políticos e estratégicos, como a chamada “guerra híbrida”, e como essa influência se manifesta em diversos setores, incluindo a defesa, a segurança e até mesmo a política partidária.
A influência dos militares no governo Bolsonaro, incluindo a disseminação da ideia de “guerra híbrida” e o uso de táticas de guerra psicológica.O projeto de um “Estado permanente” controlado pelos militares, com foco no controle da informação e na segurança	Além disso, o texto também aborda a relação dos militares com outros atores, como igrejas e grupos religiosos, e como essas interações podem influenciar o cenário político e social do país.
A relação entre militares e empresas, a formação de um complexo administrativo-militar e a influência de acordos internacionais na política brasileira. O papel de Bolsonaro como um “projeto” dos militares, possivelmente descartável quando não for mais útil.	

Quadro 30: Respostas comparadas de RAGs (Base de Conhecimento própria + LLMs) - o autor

No caso das análises feitas pelo Gemini Advanced, é possível continuar a interagir e realizar perguntas ao modelo treinado com a transcrição da live e ajustar o pipeline de aprendizado de máquina para obter resultados e percepções dos conteúdos e dados de interação e redes sociais.

Quanto mais documentos sobre o assunto houver na base de dados, mais precisão do modelo de linguagem na entrega e quanto melhor a elaboração das perguntas e proposições (prompts)

No caso da live de Piero Leirner estão disponíveis no site da tese no link <https://g.co/gemini/share/b35829cf34db>

Pode fazer um resumo detalhado ? Gemini Advanced

 Gemini Advanced	 MaxQDA AI
Pesquisa de Leirner: Começou nos anos 90, com foco inicial na Amazônia. Mudou para entender a visão dos militares sobre seu papel no Brasil. Leirner foi impedido de pesquisar dentro das instituições militares após seu mestrado. Em 2016, retomou a pesquisa, focando na “guerra híbrida” e na ascensão de Bolsonaro.	O texto “O Campo da Guerra...o Brasil atual” aborda a relação dos militares com o poder no Brasil contemporâneo, destacando sua influência em diferentes áreas da sociedade e do governo. Inicialmente, a narrativa apresenta a experiência do autor, que começou a pesquisar militares na década de 90, sob orientação da Professora Maria Lúcia Montes, e sua imersão em estudos antropológicos nesse contexto 9-14 53-56 .
Visão dos Militares: Se veem como vanguarda e protagonistas na condução do país. Acreditam que o Brasil não tem líderes competentes fora das Forças Armadas. Confusão entre o que é o Brasil e o que são as Forças Armadas. Revisam a história para justificar seu protagonismo, como a ênfase na Batalha de Guararapes (1648).	O autor descreve sua pesquisa de campo na região da Amazônia, onde inicialmente buscava estudar o projeto Calha Norte, mas acabou se envolvendo em uma etnografia sobre a atuação militar na região 72-97 . Ele discute a ideia de “guerra híbrida” e como os militares brasileiros estão pensando estratégias de defesa e segurança, inclusive em relação a possíveis ameaças políticas e sociais 426-442 552-556 .
Doutrina da Guerra Híbrida: Inspirada em autores como Clifford Geertz, visa dissolver a fronteira entre política e guerra. Criação de um estado de conflito permanente, onde o controle da informação é chave. Bolsonaro é visto como um “agente operacional terceirizado” para implementar essa doutrina. As contradições de Bolsonaro são parte da estratégia de guerra psicológica.	Além disso, o texto aborda a relação dos militares com outros setores, como empresários, políticos e igrejas, e como essas interações podem influenciar as decisões estratégicas do país 100-108 938-940 . Há uma análise sobre a possível continuidade do projeto militar no Brasil, mesmo em cenários de mudança política, e como os militares buscam manter sua influência e protagonismo no cenário nacional 109-112 586-588 .
Ascensão de Bolsonaro: Resultado de um projeto das Forças Armadas, não o contrário. Bolsonaro é um “para-raios” e “fusível” para os militares. O objetivo é estabelecer um “Estado permanente” nos moldes de um Pentágono brasileiro. O controle da informação é central	O autor também discute a presença dos militares em projetos de segurança e defesa, como a possível parceria do Brasil com a OTAN, e como essas ações podem impactar as relações internacionais do país 669-671 815-817 .

nesse projeto, com todos os ministérios reportando aos militares.	
Militares e o Capital: Reconfiguração do Estado e suas relações com o capital. Militares ocupam cargos chave no Estado e em empresas privadas. Acordos com empresas americanas e israelenses. Lava Jato vista como oportunidade para os militares assumirem o controle de setores estratégicos.	Por fim, o texto levanta questões sobre o controle das informações e o papel dos militares na gestão do Estado, apontando para um possível projeto de Brasil pensado por eles, com foco em segurança e defesa, e como essa visão pode se refletir nas políticas e estratégias nacionais 948-950 953-958 .
O Futuro: Bolsonaro é descartável, mas o projeto militar é duradouro. Dificuldade em remover os militares do poder devido à estrutura criada. O projeto de Brasil dos militares visa um Estado com forte controle social e alinhado aos EUA. A “queda” de Bolsonaro não significa o fim do projeto militar. A participação dos EUA é fundamental nesse processo, com acordos e cooperação em diversas áreas.	
Conclusão: O Brasil vive um momento de reconfiguração profunda do Estado e da sociedade. Os militares desempenham um papel central nesse processo, com um projeto de longo prazo. A “guerra híbrida” é a doutrina que guia suas ações, com foco no controle da informação e na criação de um estado permanente. A participação dos EUA é crucial para a consolidação desse projeto.	
Quadro 31: Resumo detalhado comparado de RAGs (Base de Conhecimento própria + LLMs) - o autor	

Qual o papel dos militares segundo Leirner ?

Gemini Advanced
Piero Leirner argumenta que os militares se veem como a vanguarda do Brasil, acreditando que são os únicos capazes de liderar o país em direção a um projeto de nação. Essa visão se intensificou com o fim da Guerra Fria, quando os militares brasileiros questionaram a aliança com os Estados Unidos e passaram a defender um novo protagonismo político. Leirner também destaca a dificuldade dos militares em se verem como parte da sociedade, tendendo a acreditar que a sociedade é um produto deles. Essa percepção distorcida, juntamente com a crença em seu próprio papel heroico na história do Brasil, reforça a convicção dos militares de que são indispensáveis para a condução do país.
Quadro 32: Pergunta sobre papel dos militares via RAGs + LLMs - o autor via Gemini 1.5

E sobre a guerra ?

Gemini Advanced 
A guerra é um tema central na entrevista com Piero Leirner, abordada sob a perspectiva da doutrina militar da “guerra híbrida”.

Guerra Híbrida: Leirner explica que a guerra híbrida é uma doutrina militar que visa dissolver a fronteira entre política e guerra, criando um estado de conflito permanente. Nela o controle da informação é uma arma chave, utilizada para manipular e influenciar a sociedade.
Bolsonaro e a Guerra Híbrida: O antropólogo argumenta que a ascensão de Jair Bolsonaro à presidência foi um projeto das Forças Armadas brasileiras, que utilizaram a doutrina da guerra híbrida para alcançá-lo. As contradições e ações polêmicas de Bolsonaro são vistas como parte de uma estratégia de guerra psicológica para manter a sociedade em constante estado de tensão e manipulação.
Militares e o Controle da Informação: O controle da informação é central no projeto de poder dos militares. Leirner destaca que todas as informações dos ministérios e outros poderes são direcionadas aos militares, que as utilizam para consolidar seu domínio sobre o Estado.
O Projeto de Brasil dos Militares: O projeto de Brasil idealizado pelos militares, segundo Leirner, é um estado com forte controle social, alinhado aos interesses dos Estados Unidos e com um controle centralizado da informação. A guerra híbrida é a ferramenta para alcançar esse objetivo, mantendo a sociedade em constante estado de alerta e manipulação.
Quadro 33: Pergunta sobre a guerra via RAGs + LLMs - o autor via Gemini

4.8.2 - Resumo dos comentários

Gemini Advanced 
O documento contém uma série de comentários que elogiam o trabalho de Piero Leirner, um teórico e observador do cenário político. Os comentários parabenizam Piero por sua participação em uma transmissão e elogiam sua entrevista e trabalho. Há sugestões de temas para futuras discussões, como a participação da Profa. Angélica Lovatto sobre teorias burguesas e pensamento pós-moderno. Também  elogios à visão de Piero sobre o militarismo e reconhecimento de sua importância para a compreensão da nova realidade brasileira. Em geral, os comentários destacam a excelência e a importância do trabalho de Piero Leirner.
Quadro 34: Resumo dos Comentários via RAGs + LLMs - o autor via Gemini

Análise de sentimentos nos comentários

Gemini Advanced
Os comentários sobre Piero Leirner são extremamente positivos, indicando admiração e apreço por seu trabalho. As palavras e frases utilizadas expressam emoções como entusiasmo, gratidão e reconhecimento pela contribuição de Piero. A linguagem utilizada é elogiosa e respeitosa, evidenciando a consideração que os comentaristas têm por ele. Além disso, a sugestão de temas para futuras discussões demonstra o interesse contínuo em seu trabalho e a expectativa por mais conteúdo relevante. Em geral, o sentimento predominante é de admiração e reconhecimento pelo trabalho de Piero Leirner.
Quadro 35: Pergunta sobre a guerra via RAGs + LLMs - o autor via Gemini

4.8.3 - Análise do chat durante a live - Gemini Advanced

Resumo do chat

 Gemini Advanced	 MaxQDA AI
O chat contém uma discussão sobre a participação dos militares na política brasileira, com foco nas ideias do antropólogo Piero Leirner. Os participantes levantam questões sobre como remover os militares do poder, o silêncio da oposição em relação ao projeto de poder dos generais, a demissão de José Viegas Filho, a participação dos EUA na dinâmica militar, o impacto da pandemia de Covid-19 no projeto de guerra híbrida, as ações militares após a queda de Bolsonaro, A relação entre o projeto de poder dos militares e igrejas/grupos religiosos, o futuro do projeto de guerra contra o próprio povo, a renúncia de Bolsonaro e a possibilidade de um autogolpe, a destruição do Estado brasileiro, a polarização esquerda x direita e a terceira via militar, o papel de Lula nesse contexto, a participação de militares na política e a influência de um “inimigo” externo.	O texto consiste em uma série de perguntas e comentários feitos por diferentes pessoas a Piero Leirner, aparentemente em um ambiente virtual. As perguntas abordam temas como a presença militar no Brasil, a influência dos Estados Unidos nessa dinâmica, a relação entre o projeto de poder dos militares e grupos religiosos, a possibilidade de renúncia de Bolsonaro e a polarização política no país. Além disso, há questionamentos sobre a atuação dos militares na política e a influência de figuras como Lula nesse contexto. Também há comentários sobre a Covid-19 e seu impacto no projeto de guerra híbrida, bem como a preocupação com o futuro do país diante da atual situação política. Ao longo das interações, há uma preocupação constante com a presença e influência dos militares no poder e suas possíveis consequências para o Brasil.
Quadro 36: Chats comparados em RAGs (Base de Conhecimento própria + LLMs) - o autor	

Análise de sentimentos do chat

 Gemini Advanced	 MaxQDA AI
O sentimento geral expresso no chat em relação a Piero Leirner e ao tema discutido é de apreensão e preocupação quanto ao papel dos militares na política brasileira. Há um misto de elogios a Piero Leirner por suas ideias e análises, com expressões como “ótima apresentação” e “muito bom”, e questionamentos sobre o futuro e as implicações do projeto de poder dos militares. As perguntas dos participantes demonstram inquietação em relação à influência militar no governo, o papel dos EUA nesse processo, as consequências da doutrina de guerra híbrida e a possibilidade de um autogolpe. Há também expressões de frustração e descrença em relação à classe política e ao futuro do país, como “guerra contra o próprio povo que o sustenta... Isso tem futuro??” e “foi posto para destruir o estado brasileiro”.	Ao analisar os sentimentos presentes no chat, podemos identificar uma mistura de preocupação, críticas, perplexidade e até certo ceticismo em relação à situação política e militar do Brasil. Por exemplo, o participante LZ expressa uma forte crítica ao mencionar os “militares entreguistas” que estão “destruindo o país” P.1. Essa afirmação reflete uma visão negativa e indignada em relação à atuação dos militares no poder. Da mesma forma, MH levanta a questão do silêncio da oposição em relação ao projeto de poder dos generais, o que pode indicar uma sensação de perplexidade ou descontentamento com a postura política vigente P.1. Já o participante RCN demonstra preocupação ao mencionar a “guerra contra o próprio povo” e questionar se isso tem futuro P.1. Essa reflexão sugere uma visão pessimista em relação aos rumos do país.

No geral, o sentimento predominante é de apreensão e preocupação, com um misto de esperança de que a análise e discussão do tema possam levar a soluções e mudanças.	Além disso, há quem levante a possibilidade de renúncia de Bolsonaro e discuta sobre a polarização política que isso poderia gerar, como apontado por AH P. 1. Essa discussão indica uma certa incerteza e especulação sobre os desdobramentos políticos futuros. Em resumo, os sentimentos presentes no chat refletem uma atmosfera de preocupação, críticas e incertezas em relação ao cenário político e militar do Brasil, evidenciando a complexidade e a tensão presentes nas discussões.
Quadro 37: Análise de sentimentos comparadas do chat via RAGs + LLMs - o autor	

Análise do discurso do chat

 Gemini Advanced	 MaxQDA AI
A entrevista com Piero Leirner apresenta um discurso crítico e contundente sobre a atuação dos militares na política brasileira. Para aprofundar a análise, podemos observar alguns elementos chave:	Podemos observar uma variedade de temas sendo abordados, todos relacionados à presença militar e política no Brasil. Os participantes demonstram interesse e preocupação com a influência dos militares no governo, levantando questionamentos pertinentes sobre a atuação desses grupos no cenário nacional.
Construção do discurso: Leirner constroi seu discurso com base em sua pesquisa acadêmica e experiência profissional, utilizando uma linguagem clara e direta, com termos técnicos específicos da área militar e política. Ele apresenta argumentos e evidências para sustentar suas afirmações, como a doutrina da guerra híbrida e a participação dos militares em cargos-chave no governo e em empresas.	Por exemplo, a crítica de LZ aos “militares entreguistas” que estão “destruindo o país” reflete uma visão negativa e de descontentamento com a atual situação política. Essa crítica sugere uma preocupação com os rumos que o país está tomando sob a liderança militar.
Estratégias discursivas: O antropólogo usa diversas estratégias discursivas para fortalecer seu argumento, como: • Generalização: utiliza termos como "os militares" para se referir a um grupo homogêneo com objetivos e interesses comuns. • Nominalização: transforma ações em substantivos abstratos, como "controle da informação" e "ascensão de Bolsonaro", para enfatizar o caráter estratégico e planejado dessas ações. • Metáforas: compara Bolsonaro a um "para-raios" e "fusível" para os militares, ilustrando a sua função no projeto militar. • Adjetivação: utiliza adjetivos como "contraditório" e "polêmico" para	Além disso, a questão levantada por MH sobre o silêncio da oposição em relação ao projeto de poder dos generais também é significativa. Essa reflexão indica uma análise crítica sobre a postura da oposição diante das ações dos militares no governo.

descrever Bolsonaro, reforçando a ideia de que ele é parte de uma estratégia de manipulação.	
Ideologia e poder: O discurso de Leirner revela a ideologia por trás da atuação dos militares, que se veem como os únicos capazes de liderar o país e garantir a ordem e o progresso. Ele denuncia o projeto de poder dos militares, que buscam controlar o Estado e a sociedade através da guerra híbrida e da manipulação da informação.	Outro ponto relevante é a discussão sobre a possível renúncia de Bolsonaro e os desdobramentos políticos que isso poderia acarretar, como mencionado por AH. Essa especulação sobre o futuro político do país demonstra uma análise profunda e uma tentativa de compreender os possíveis cenários que podem se desenrolar.
Implicações do discurso: O discurso de Leirner tem implicações importantes para a compreensão da política brasileira atual. Ele alerta para os perigos da crescente influência dos militares na política e para a necessidade de resistir a esse projeto autoritário.	Em resumo, o discurso presente no chat é marcado por críticas, questionamentos e análises sobre a presença militar na política brasileira, revelando uma preocupação coletiva com os rumos do país e as consequências dessa influência para a democracia e a sociedade como um todo.
Contexto sociopolítico: A entrevista foi concedida em um momento de grande polarização política no Brasil, após a queda de Bolsonaro. O discurso de Leirner contribui para o debate sobre o papel dos militares na democracia e para a construção de um futuro mais justo e igualitário.	
Em suma, a análise do discurso de Piero Leirner revela um panorama preocupante sobre a atuação dos militares na política brasileira. Ele denuncia um projeto de poder autoritário que visa controlar o Estado e a sociedade através da guerra híbrida e da manipulação da informação. O discurso do antropólogo é um chamado à resistência e à construção de um futuro democrático para o Brasil.	
Quadro 38: Análise do discurso comparadas do chat via RAGs + LLMs - Gemini/Claude	

4.9 - ANÁLISES DE IA DA LIVE DE FERNANDO ABRUCIO

A transmissão da palestra de Fernando Abrucio, até a coleta de dados em outubro de 2024, obteve 4617 de visualizações, das quais apenas 46 originadas da transmissão ao vivo(live) e 4571 na modalidade por demanda (on demand). Esses números demonstraram a viabilidade de classificar as “Visualizações” em duas vertentes que refletem dinâmicas diferentes e não convém comparar um período de menos de 2 horas durante a live a anos de visualizações por demanda, mesmo porque algumas pessoas assistiram na live e on demand.

O importante é observar e estudar essas idiossincrasias pela métrica adequada, pois se aplicarmos uma função inadequada e demasiadamente focada na quantidade de visualizações, sem considerar o tempo de exibição e demais variáveis dependentes, estaremos recortando o objeto por retas cartesianas que não contemplam a complexidade e densidade das relações que são geradas e resultam das lives. Em relação às análises do conteúdo, podemos propor inferências através de percepções de ambas modalidades de coleta e análise de dados quantitativos e qualitativos.

4.9.1 - Transcrição da live

Quais os principais assuntos do texto ? Gemini Advanced

 Gemini Advanced	 MaxQDA AI
O texto é uma transcrição de uma palestra do Professor Fernando Abrucio sobre sistemas nacionais de políticas públicas no federalismo brasileiro. Os principais assuntos abordados são:	Origem e Evolução dos Sistemas Nacionais de Políticas Públicas: O texto explora a origem e a evolução dos sistemas nacionais de políticas públicas, destacando a importância da coordenação entre os diferentes níveis de governo para lidar com desafios sociais complexos.
A origem e a importância dos sistemas nacionais de políticas públicas, especialmente em países federativos como o Brasil.	Federalismo Brasileiro: Há uma análise detalhada sobre o federalismo brasileiro, destacando a autonomia dos estados e municípios, bem como a necessidade de coordenação e cooperação entre os entes federativos para garantir a efetividade das políticas públicas.
A relação entre o desenvolvimento do estado de bem-estar social no Brasil e a criação de sistemas nacionais de políticas públicas.	Desigualdades Territoriais: O texto ressalta a desigualdade territorial no Brasil e a importância de sistemas nacionais de políticas públicas para lidar com essas disparidades, especialmente em áreas como saúde e educação.
O histórico do federalismo no Brasil, desde o período imperial até a atualidade, e seus impactos na implementação de políticas públicas.	Governança e Participação Federativa: É discutida a importância da governança e da participação federativa na elaboração e implementação de políticas públicas eficazes, enfatizando a necessidade de construir agendas nacionais que considerem a diversidade do país.
As características e desafios dos sistemas nacionais de políticas públicas no Brasil, incluindo o papel da União, dos estados e dos municípios. A importância da coordenação entre os diferentes níveis de governo e da participação social para o sucesso das políticas públicas.	Coordenação de Esforços e Ações: O texto destaca a necessidade de coordenar esforços e ações entre os diferentes entes federativos para enfrentar desafios como a desigualdade, ressaltando a importância da cooperação e da articulação entre os diversos níveis de governo.

Exemplos de sistemas nacionais de políticas públicas no Brasil, como o SUS (Sistema Único de Saúde), o SUAS (Sistema Único de Assistência Social) e o SNIRH (Sistema Nacional de Gerenciamento de Recursos Hídricos).	A importância da coordenação e ação conjunta entre os diferentes níveis de governo (federal, estadual e municipal) para o sucesso das políticas públicas 1 4 7 19-20 23 47; a necessidade de fortalecer as capacidades estatais locais para garantir a efetividade das políticas públicas 63 66 69 78 85 87;
Os desafios futuros para a consolidação e o aprimoramento dos sistemas nacionais de políticas públicas no Brasil, como a necessidade de maior coordenação intersetorial e de participação social.	a descentralização e municipalização das políticas públicas pós-88 e os desafios enfrentados pelos municípios na implementação dessas políticas 53 55-57 67 70-71 74 77 87; a importância dos sistemas nacionais de políticas públicas para garantir a igualdade e eficácia das ações governamentais [50, 52, 79, 80]
Quadro 39: Respostas comparadas sobre assuntos do texto de Abrucio- RAGsLLMs Gemini/Claude	

Resumo detalhado do texto - Gemini Advanced

 Gemini Advanced	 MaxQDA AI
A transcrição da live com o Professor Fernando Abrucio aborda a temática dos sistemas nacionais de políticas públicas no federalismo brasileiro, com foco na sua importância para a construção de um estado de bem-estar social e uma sociedade mais igualitária. O professor inicia contextualizando a origem do debate sobre sistemas nacionais de políticas públicas, que surge no pós-guerra com a expansão do Welfare State e a necessidade de articular políticas públicas em países com diferentes níveis de autonomia governamental. Ele destaca a relevância do tema para o Brasil, um país federativo com grande desigualdade social e territorial. Abrucio traça um panorama histórico do federalismo brasileiro, desde o período imperial, marcado pela centralização do poder, até a redemocratização e a Constituição de 1988, que descentralizou as políticas públicas e conferiu maior autonomia aos municípios. Essa mudança, apesar de positiva, gerou desafios de governança, pois muitos municípios não possuíam capacidade estatal para implementar as políticas. O professor argumenta que os sistemas nacionais de políticas públicas são cruciais para a coordenação federativa, o combate às desigualdades territoriais, a coordenação de esforços em ações comuns e a criação de	O texto aborda a importância do Sistema Nacional de Políticas Públicas no contexto do federalismo brasileiro. Inicia destacando a origem e evolução dos sistemas nacionais de políticas públicas, ressaltando a necessidade de coordenação entre os diferentes níveis de governo para lidar com desafios sociais complexos. Destaca a relevância do debate sobre desigualdades territoriais e a necessidade de construir agendas nacionais que considerem a diversidade do país. Além disso, enfatiza a importância da governança e da participação federativa na elaboração e implementação de políticas públicas eficazes. O texto também aborda a complexidade do federalismo brasileiro, destacando a autonomia dos estados e municípios, bem como a necessidade de coordenação e cooperação entre os entes federativos para garantir a efetividade das políticas públicas. O texto aborda a importância da coordenação e articulação entre os diferentes níveis de governo (União, estados e municípios) para o sucesso das políticas públicas no Brasil. O autor destaca a necessidade de ações conjuntas e esforços coordenados para lidar com questões complexas que exigem cooperação entre os entes federativos. Além disso, são discutidos temas como descentralização, capacidades estatais locais,

capacidades estatais locais. Ele cita exemplos como o SUS, o SUAS e o SNIRH, destacando seus sucessos e desafios. Na parte final, Abrucio discute os desafios futuros para os sistemas nacionais de políticas públicas no Brasil, como a necessidade de criação de sistemas em áreas ainda não contempladas, o combate às tendências centralizadoras da União, o fortalecimento do papel dos estados na coordenação federativa, a superação da fragmentação municipal e a promoção de maior participação social. Ele conclui enfatizando que o fortalecimento desses sistemas é essencial para a efetividade das políticas públicas e a garantia do direito à igualdade no Brasil. A live também inclui um debate com a participação da Professora Teresa, que levanta questões sobre o impacto da pandemia na educação e o protagonismo dos municípios na implementação de políticas públicas. Abrucio responde, destacando a importância da coordenação federativa e da criação de capacidades estatais locais para o sucesso das políticas públicas, especialmente em momentos de crise. https://gemini.google.com/app/fbc84c8837d43123	federalismo, sistemas nacionais de políticas públicas e a relação entre autonomia local e coordenação nacional. O autor ressalta que a falta de sistemas nacionais bem estruturados, como na área da educação, pode resultar em descoordenação e impactar negativamente a efetividade das políticas. Também destaca a importância dos municípios, que, apesar de terem ganhado protagonismo, enfrentam desafios de capacidade e dependência de recursos da União. Sobre a descentralização fiscal, o autor aponta que, apesar de existir uma descentralização significativa, ainda há desafios na gestão financeira dos municípios, especialmente os menores, que dependem fortemente de repasses federais. A reforma tributária e administrativa são citadas como possíveis soluções para equacionar essas questões e fortalecer a autonomia local. Por fim, o texto ressalta a importância da coordenação federativa, a necessidade de combater desigualdades regionais e a crítica ao modelo de mais Brasil – Brasília, que tende a centralizar decisões e recursos, prejudicando a efetividade das políticas públicas. A falta de diálogo e cooperação entre os entes federativos é apontada como um obstáculo para a melhoria da governança e da qualidade das políticas públicas no Brasil.
Quadro 40: Respostas comparadas sobre resumo do texto de Abrucio RAGs + LLMs Gemini/Claude	

O que o texto fala sobre o estado de bem estar social ?

Gemini Advanced  O federalismo brasileiro é um tema central na discussão sobre sistemas nacionais de políticas públicas. O texto aborda sua trajetória histórica, desde o período imperial, marcado pela centralização do poder, até a Constituição de 1988, que descentralizou as políticas públicas e conferiu maior autonomia aos municípios. Essa mudança, apesar de positiva, trouxe desafios de governança, pois muitos municípios não possuíam capacidade estatal para implementar as políticas. O professor Fernando Abrucio destaca que o federalismo brasileiro é caracterizado por: • Uma federação tripartite: União, estados e municípios são entes federativos, com o município tendo um papel de destaque na implementação de políticas públicas. • Grande heterogeneidade entre os municípios: Há uma enorme desigualdade em termos de recursos e capacidades estatais, o que dificulta a implementação de políticas públicas de forma homogênea em todo o país. • Relação complexa entre os entes federativos: A Constituição de 1988 estabelece competências comuns e concorrentes entre a União, os estados e os municípios, o que exige mecanismos de coordenação para evitar conflitos e garantir a efetividade das políticas.
--

- Importância da coordenação federativa: A criação de sistemas nacionais de políticas públicas é fundamental para articular as ações dos diferentes entes federativos, combater as desigualdades territoriais e garantir o acesso universal aos direitos sociais.

O texto também menciona os desafios enfrentados pelo federalismo brasileiro, como as tendências centralizadoras da União e a dificuldade de coordenação entre os diferentes níveis de governo. O professor Abrucio defende que o fortalecimento dos sistemas nacionais de políticas públicas e a criação de mecanismos de cooperação entre os entes federativos são essenciais para superar esses desafios e garantir a efetividade das políticas públicas no Brasil.

Quadro 41: Resposta sobre o estado de bem estar social de RAGs + LLMs - Gemini Advanced

4.9.2 - Resumo detalhado dos comentários

 Gemini Advanced	 MaxQDA AI
O texto apresenta uma série de comentários e interações em uma plataforma online, possivelmente YouTube, relacionados a uma aula ou palestra ministrada por um professor. Os comentários são majoritariamente positivos e elogiosos, com os espectadores expressando admiração pela inteligência do professor, pela clareza das explicações e pela relevância do conteúdo apresentado. Alguns comentários específicos mencionam o federalismo e a falta de ações do Governo Central, indicando que esse foi um dos temas abordados na aula. Outros comentários agradecem ao professor pelos esclarecimentos e pela oportunidade de aprendizado. Além dos elogios, há um comentário que sugere uma avaliação do governo atual em relação às políticas públicas daqui a dois anos, e outro que parabeniza o professor pelo trabalho e agradece pela ajuda aos concurseiros do CNU (Concurso Nacional Unificado). Em resumo, o texto oferece um panorama das reações positivas do público à aula ou palestra, destacando a qualidade do conteúdo e do professor, e mencionando alguns dos temas abordados, como federalismo e políticas públicas.	Para fazer um resumo detalhado do texto fornecido, é importante destacar que se trata de uma série de comentários positivos e elogios a um professor, provavelmente após uma aula ou palestra sobre federalismo e ações do Governo Central. Os comentários são bastante elogiosos e demonstram a satisfação dos alunos com a didática e clareza do professor em abordar o tema. Os comentários variam em termos de tempo, com alguns sendo feitos há 1 mês, 2 meses e até 4 meses. Alguns alunos expressam gratidão, como "Obrigado professor" 45 47, "Gratidão prof." 56, e "Obrigado pelos esclarecimentos e oportunidade" 63 65. Além disso, há menções de que as explicações são claras e diretas, sem enrolação, como em "explica as coisas com poucas palavras, muita teoria, sem enrolação" 8-9. Os elogios são variados, desde "Muito boa aula!" 13-14, "Excelente!!!!" 22 24, "Muito bom!" 39 41, até "Isso foi muito interessante..." 19. Alguns alunos destacam a inteligência do professor, como em "Nossa vc é muito inteligente" 7 9, e outros mencionam a importância das informações compartilhadas, como em "Vale a pena daqui a dois anos avaliar o governo atual no que tange as políticas públicas de uma forma geral" 51 53. Em resumo, o texto reflete uma atmosfera de apreciação e reconhecimento pelos ensinamentos do professor, destacando sua clareza, inteligência e relevância das informações transmitidas durante a aula ou palestra sobre federalismo e ações do Governo Central.

Quadro 42: Resumos detalhados dos comentários on demand Abrucio - RAGs/LLMs - Gemini/Claude

Análise de sentimentos dos comentários

 Gemini Advanced	 MaxQDA AI
Os comentários transmitem um sentimento extremamente positivo em relação à aula ou palestra do professor. As palavras e frases utilizadas expressam admiração, gratidão e entusiasmo em relação ao conteúdo e à didática do professor. Exemplos de expressões que demonstram sentimentos positivos: • “Muito boas as colocações” • “Parabéns prof” • “Nossa vc e muito inteligente” • “explica as coisas com poucas palavras” • “muita teoria ,sem enrolação” • “Muito boa aula!” • “Isso foi muito interessante...” • “Excelente!!!!” • “amei lhe ouvir professor!!” • “Obrigada professor” • “Muito bom!” • “Obrigado professor” • “Gratidão prof” • “Parabéns pelo trabalho e mt obrigado” • “Muito bom professor ! Obrigado pelos esclarecimentos e oportunidade” Não há comentários com sentimentos negativos. A única observação neutra é um comentário que sugere uma avaliação do governo em relação às políticas públicas daqui a dois anos. No entanto, essa observação não expressa um sentimento negativo em relação à aula ou ao professor.	Ao analisar o sentimento expresso nos comentários do texto fornecido, percebe-se predominantemente um sentimento positivo e de apreciação por parte dos alunos em relação ao professor e ao conteúdo abordado. Os elogios como “Muito boa aula!” 13-14, “Excelente!!!!” 22 24, “Muito bom!” 39 41, e “Obrigado professor” 45 47 refletem uma atmosfera de satisfação e reconhecimento. Além disso, expressões de gratidão como “Obrigado pelos esclarecimentos e oportunidade” 63 65, “Gratidão prof.” 56, e “Obrigada professor” 33 35 demonstram um sentimento de apreço e respeito pelo trabalho do professor. A menção de que as explicações são claras e diretas, sem enrolação, como em “explica as coisas com poucas palavras, muita teoria, sem enrolação” 7 9, também indica uma valorização da didática do professor. Em suma, a análise de sentimento revela que os alunos expressam principalmente emoções positivas, como satisfação, gratidão, respeito e apreço em relação ao professor e ao conteúdo apresentado durante a aula ou palestra sobre federalismo e ações do Governo Central.
Quadro 43: Comparativo análises de sentimento dos comentários on demand Abrucio - Gemini/Claude	

4.9.3 - Resumo do chat durante a live

 Gemini Advanced	 MaxQDA AI
O chat se inicia com saudações entre os participantes. LB inicia o debate perguntando	O texto consiste em uma conversa entre várias pessoas, incluindo FA, VB, LB, JO, KM, TR e

sobre a dificuldade dos municípios em gerir suas políticas locais devido à dependência da União, mesmo com a descentralização promovida pela Constituição. Ele questiona como solucionar essa questão e se a reforma tributária em discussão poderia ajudar. MM responde, mencionando a descentralização administrativa determinada pela Constituição Federal, mas aponta que o governo federal mantém controle sobre a distribuição de recursos. Ele expressa a ideia de que um projeto de “mais Brasil e menos Brasília” não se efetiva e lamenta essa situação. LB agradece a oportunidade de ouvir o professor e aos organizadores. MM, por sua vez, deseja sucesso a todos.	MM. A discussão gira em torno da descentralização das políticas públicas e da questão fiscal, com LB levantando a dificuldade enfrentada pelos municípios, especialmente os pequenos, em gerir suas políticas locais devido à dependência da União. Ele questiona de que forma seria possível equacionar essa questão e se a reforma tributária em discussão no congresso poderia ajudar. MM comenta sobre a determinação de descentralização administrativa pela Constituição Federal, mas ressalta que o governo federal mantém controle na distribuição de recursos públicos para as políticas públicas nos demais entes subnacionais. A conversa termina com agradecimentos e desejos de sucesso.
Quadro 44: Comparativo resumo do chat na live de Abrucio - RAGs + LLMs - Gemini/Claude	

Análise de sentimento do chat durante a live

 Gemini Advanced	 MaxQDA AI
O chat demonstra um sentimento de frustração e insatisfação com a gestão das políticas públicas no Brasil. LB expressa a dificuldade dos municípios em gerir suas políticas locais devido à dependência da União, mesmo com a descentralização promovida pela Constituição. MM reforça essa frustração ao mencionar que o governo federal mantém o controle sobre a distribuição de recursos, impedindo um projeto de “mais Brasil e menos Brasília”. Apesar disso, há um sentimento de esperança na reforma tributária em discussão, que pode trazer soluções para a questão. No geral, o sentimento predominante é de crítica construtiva e busca por soluções para a gestão das políticas públicas no país.	Podemos observar que a maioria das mensagens transmite um clima positivo e de interação amigável entre os participantes. Inicialmente, vemos saudações cordiais como “Bom dia pessoal” e “Olá!! Quem quiser fazer perguntas é só postar aqui”, o que indica uma atmosfera acolhedora e receptiva. Além disso, há mensagens de agradecimento e elogios, como “análises lúcidas” e “Muito bom te ouvir, professor”. Obrigado aos organizadores”, que demonstram apreço e reconhecimento pelas contribuições e pela organização do evento. No entanto, também há menções a questões mais sérias e críticas, como a dificuldade dos municípios em gerir suas políticas locais devido à dependência da União 6-8. Essas mensagens refletem uma preocupação legítima com questões políticas e fiscais que impactam diretamente a gestão pública em nível local. Em suma, a análise de sentimento do texto revela uma combinação de positividade, cortesia e engajamento, juntamente com reflexões críticas e questionamentos pertinentes sobre temas relevantes.
Quadro 45: Comparativo análise de sentimento do chat na live de Abrucio - Gemini/Claude	

Análise de discurso do chat durante a live

 Gemini Advanced	 MaxQDA AI
O texto, transcrição de uma palestra do Professor Fernando Abrucio, apresenta um discurso estruturado e informativo sobre a complexa relação entre o federalismo brasileiro e a implementação de políticas públicas. O discurso é direcionado a um público acadêmico ou com interesse na área, evidenciado pelo uso de linguagem técnica e referências a autores e obras relevantes. Elementos do Discurso: • Construção da Credibilidade: Abrucio estabelece sua autoridade no tema logo no início, mencionando suas qualificações e experiência, como doutorado em Ciência Política e pesquisas na área. • Uso de Dados e Evidências: O discurso é permeado por dados históricos e estatísticos, além de referências a estudos e pesquisas, o que confere maior credibilidade e embasamento às suas afirmações. • Tom Crítico e Analítico: Abrucio não se limita a descrever o funcionamento do sistema, mas tece críticas sobre as dificuldades e desafios na implementação de políticas públicas no Brasil, como a desigualdade, a falta de coordenação e a fragilidade das capacidades estatais. • Linguagem Clara e Objetiva: Apesar da complexidade do tema, o professor utiliza linguagem clara e objetiva, com frases curtas e bem estruturadas, facilitando a compreensão do conteúdo. • Recursos Retóricos: Abrucio utiliza recursos retóricos como analogias e metáforas para ilustrar suas ideias e torná-las mais acessíveis ao público. • Interação com a Audiência: O discurso inclui momentos de interação com a audiência, respondendo a perguntas e comentários, o que demonstra abertura ao diálogo e receptividade às diferentes perspectivas. Estratégias Discursivas: • Apresentação Histórica: Abrucio traça um panorama histórico do federalismo brasileiro para contextualizar a situação atual e evidenciar as raízes dos problemas contemporâneos. • Comparação com Outros Países: A comparação com outros países, como os	Ao analisar o discurso presente no texto fornecido, podemos identificar diferentes elementos que contribuem para a construção de significados e interações entre os participantes. Em primeiro lugar, observamos uma sequência de saudações cordiais e educadas, como “Bom dia pessoal” e “Olá!! Quem quiser fazer perguntas é só postar aqui”, que estabelecem um tom amigável e acolhedor no início da conversa. Além disso, há a presença de questionamentos e reflexões mais aprofundadas, como o questionamento feito por LBS sobre a descentralização das políticas públicas e a dependência dos municípios em relação à União 6-8. Essas intervenções demonstram um interesse em questões políticas e fiscais, bem como a busca por soluções e entendimento sobre esses temas. Por outro lado, também podemos notar expressões de gratidão e reconhecimento, como “Muito bom te ouvir, professor. Obrigado aos organizadores” 13. Essas manifestações de apreço contribuem para fortalecer os laços de colaboração e respeito mútuo entre os participantes. Em resumo, a análise do discurso no texto revela uma combinação de elementos que vão desde a cordialidade nas interações iniciais, passando por questionamentos críticos e reflexões mais aprofundadas, até manifestações de gratidão e reconhecimento. Essa diversidade de discursos contribui para enriquecer a troca de ideias e experiências no contexto da conversa apresentada.

Estados Unidos e países europeus, serve para destacar as particularidades do federalismo brasileiro e a complexidade dos desafios enfrentados. • Ênfase na Desigualdade: A desigualdade social e territorial é um tema recorrente no discurso, sendo apontada como um dos principais obstáculos para a implementação de políticas públicas eficazes. • Defesa dos Sistemas Nacionais: Abrucio argumenta que os sistemas nacionais de políticas públicas são essenciais para superar os desafios do federalismo brasileiro e garantir o acesso universal aos direitos sociais. Conclusão: O discurso do Professor Fernando Abrucio é uma análise consistente e crítica sobre o federalismo brasileiro e os sistemas nacionais de políticas públicas. Sua fala, construída com base em dados, evidências e linguagem clara, contribui para o debate sobre a construção de um estado de bem-estar social mais justo e igualitário no Brasil.	
Quadro 46: Comparativo análise de discurso do chat na live de Abrucio - Gemini/Claude	

5 - As demandas de difusão e letramento científico

São investigados dois projetos e instâncias que demandam serviços de comunicação, difusão e letramento científico multiplataforma através de sites multimídia e canais de streaming. Essas demandas aumentaram consideravelmente durante a pandemia e permanecem em crescimento devido à virtualização de eventos, aulas, congressos e cursos de extensão como complemento às atividades presenciais que retornaram em meados de 2022.

Atualmente a comunidade universitária da FCLAr demanda 21 canais oficiais de streaming no Youtube que agregam mais de 15.000 inscritos e atendem as demandas de Projetos de Extensão Universitária, de Projetos de Educação Tutorial (PETs), de Grupos de Estudo e Pesquisa, de Centros Acadêmicos e Atlética, além de alguns eventos anuais regulares que realizam transmissões cíclicas durante o ano todo.

As transmissões por streaming na FCLAr iniciaram em 2009 na intranet da faculdade, após aquisições de equipamentos que permitiam a difusão por programas (*players*) diretamente na rede, pois os servidores de streaming não eram online e os serviços ainda não estavam disponíveis no Brasil. Somente em 2011, quando a Rede Nacional de Ensino e Pesquisa (RNP) disponibiliza o Serviço de Transmissão de Vídeo ao Vivo⁸³ com servidores que atuavam como refletores de transmissão distribuídos por todo o Brasil através da Rede ANSP⁸⁴ (atual Rednesp), a faculdade passou a transmitir na internet através de players que suportavam o codec VC1⁸⁵ da Microsoft, um dos primeiros disponíveis e gratuitos a oferecer uma qualidade adequada aos eventos acadêmicos.

A FCLAr foi pioneira na difusão científica por streaming em todo o Brasil, atendendo demandas dos eventos de maior relevância, nos quais a comunidade compartilha os conhecimentos, divulga seus resultados e colabora no letramento científico da sociedade, de forma aberta e gratuita.

Somente em 2013 a FCLAr inicia seu canal sob demanda, após o serviço ser disponibilizado pela Rede Nacional de Ensino e Pesquisa, tornando possível o acesso posterior às transmissões em uma plataforma pública confiável, gratuita e de

⁸³ https://www.rnp.br/arquivos/documents/guiadousuariotransmissaosalv_1.pdf

⁸⁴ <https://www.ansp.br/>

⁸⁵ <https://learn.microsoft.com/pt-br/windows-hardware/drivers/display/vc-1>

finalidade científica, havendo um crescimento acentuado dos acessos a partir de 2018, está relacionado aos novos tratamento dos metadados dos conteúdos e publicações incorporadas (embedded) no site da faculdade, que incrementou a difusão científica por aquele canal.

Em 2021 a RNP atualizou os serviços e lançou a plataforma Eduplay, ampliando ainda mais o alcance das transmissões da FCLAr e considerando que a especificidade dos conteúdos do canal na RNP não serem amplamente buscados pelos cidadãos de forma geral, o alcance da difusão e do letramento científico cumprem algumas das finalidades de democratizar o acesso aos conteúdos, comunicar a produção acadêmica da faculdade e, em certo sentido, adaptar a decodificação do discurso científico à linguagem usual de comunidades tão diferenciadas de nosso país continental.

As estatísticas do canal demonstram o amplo alcance de conteúdos do canal, tanto no Brasil como no exterior, atendendo mais uma demanda estratégica da universidade, ao comunicar e difundir seus conteúdos internacionalmente, indicando acessos a partir de países dos cinco continentes.

5.1 - Estratégias metodológicas e resultados

As técnicas e métodos quali e quantitativos oriundos das metodologias estatísticas e de Análise de Conteúdo, forneceram muitas percepções e inferências sobre as dinâmicas de difusão e letramento científico. As variáveis analisadas descritivamente foram capazes de fornecer diferentes perfis de usuário em cada projeto e em cada tipo de plataforma (site ou streaming).

Os aspectos socioeconômicos e geopolíticos, como origem, quantidade, tempo de acesso, engajamento, idade e gênero foram relacionados aos dados tecnológicos como origem das sessões, tipo de dispositivo, sistemas operacionais e resolução de telas. As inferências possibilitaram as percepções das análises descritivas que seguem agrupadas por tipos de variáveis compartilhadas nos dois projetos estudados.

No caso do projeto Ciências Sociais em Diálogo , a coleta e filtragem dos dados foram realizados no [Youtube](#) através do Analytics do canal e os dados

coletados do projeto da Rede de Apoio aos Usuários de Língua Estrangeira vieram do Google Analytics do site da RAPLE (raple.fclar.unesp.br) e do seu canal de [Youtube](#).

Ambas as origens (site e canais) tiveram dados coletados desde o início dos projetos ou o mais anteriormente possível. O canal Ciências Sociais teve dados coletados desde o início das atividades em abril de 2020 até setembro de 2024, perfazendo 1621 dias, ou 231 semanas, ou 54 meses que totalizam 4 anos e 4 meses de coleta.

O projeto da Rede de Apoio aos Professores de Língua Estrangeira (RAPLE) originou duas coletas de dados com datas de início e intervalos diferentes e que não contemplam todos os anos de atuação do projeto. Os dados coletados do site da RAPLE se iniciam em abril de 2022 até outubro de 2024, totalizando 943 dias, ou 134 semanas, ou 31 meses, perfazendo 2 anos e 6 meses de coleta. O canal do projeto no Youtube foi criado em março de 2021 e foram coletados dados das dinâmicas de transmissão, do chat e dos comentários até outubro de 2024, totalizando 1322 dias, ou 188 semanas ou 44 meses, perfazendo 3 anos e 7 meses de coleta de dados.

Mídia Digital	Período	Dias	Semanas	Meses	Anos
Youtube Ciências Sociais em Diálogo	04/2020 a 09/2024	1621	231	54	4,4
Site da RAPLE na Internet (Analytics)	04/2022 a 10/2024	943	134	31	2,6
Youtube da RAPLE	03/2021 a 10/2024	1322	188	44	3,7

Tabela 25: Recorte temporal da coleta de dados nos projetos de extensão analisados - o autor

As coletas pelo Google Analytics do site da RAPLE iniciaram em 2022, mas anteriormente houve uma avaria em um servidor físico durante o primeiro ano da pandemia, o que ocasionou um longo período de meses para recompor o site nos moldes anteriores. Ao reconstruir o site em novo servidor virtualizado, houve perdas de dados de acesso e impossibilidade de recuperação das métricas anteriores.

5.2 - Variáveis e métricas analisadas

Foram realizadas análises de diversas variáveis dos dados online, YouTube, e mídias sociais, comparadas e correlacionadas para entender melhor o comportamento dos usuários e o desempenho dos conteúdos. Elas geraram as inferências, percepções e relatórios que orientam os coordenadores do projeto em

tomadas de decisões e prestação de contas, além de comprovarem resultados que ajudam na obtenção de bolsas de estudo para os estagiários e voluntários do projeto.

5.2.1 - RAPLE - Rede de Apoio aos Usuários de Língua Estrangeira

Resumidamente, a análise descritiva do uso do site e das redes sociais da Rede de Apoio a Professores de Línguas Estrangeiras (RAPLE), correlacionou métricas de variáveis como Usuários Ativos e Contagem de Eventos para entender o engajamento e identificar padrões de uso e distribuição por página através do Idioma e Dispositivo.

Encontramos correlações entre Visualizações e Usuários Ativos e entre Sessões, Sessões Engajadas e Tempo médio de engajamento por sessão. A taxa de engajamento é analisada em relação ao número de Usuários Ativos e Sessões Engajadas e as origens do tráfego do site e do canal indicaram predominância de links diretos e redirecionamentos do youtube, com pouca engajamento vindo de links das outras redes sociais (facebook e instagram).

Os tipos de dispositivos para acesso aos conteúdos indicam a predominância de computadores desktop, smartphones (celulares), tablets e Smart TV (tvs conectadas), indicando que, embora a maioria dos acessos seja por smartphone, o tempo de visualização é proporcionalmente maior nos desktops e nas TVs.

Ao analisar visualizações e usuários ativos por idioma, notamos que as páginas em inglês têm a maior média de visualizações com 1,34 por usuário ativo e a influência do idioma português nos usuários ativos é alto, com predominância em 90,57% deles, estando diretamente relacionada à origem dos conteúdos, que embora contenham elementos idiomáticos de outros países, é majoritariamente em língua portuguesa.

Examinamos diversas correlações de variáveis indexadas por país, estado, província ou cidade e encontramos uma notável capilaridade em um grande número de países em todos os continentes. É notável o alcance e distribuição de usuários ativos no Brasil e as projeções estatísticas nos mapas oferecem uma visualização dinâmica de dados articuladas a uma síntese das informações e variações conceituais de dados que criam conhecimento.

5.2.2 - Canal Ciências Sociais em Diálogo

Coletamos dados do canal do Youtube do projeto e o recorte temporal da coleta, contemplou os 4,4 anos do mesmo e confirma a resiliência do canal em manter o número de inscritos em crescimento linear contínuo. Do total de 2981 inscritos, o canal obteve e mantém uma taxa de retenção média de 89,51% em relação à todas inscrições obtidas, tendo perdido, em média, 10,49% dos inscritos ao longo do período.

Cada usuário inscrito visualizou 4 horas em média e houve 3.142 impressões para cada um dos conteúdos, sendo que os usuários não inscritos superam os inscritos em 28,56% com 16.538 usuários a mais. Os gráficos permitiram observar picos pontuais e não constantes das visualizações e dos espectadores recorrentes, com número de visualizações abaixo de 200 e o número de espectadores recorrentes abaixo de 50.

As visualizações dos gráficos de conteúdos por semana foram analisados e comparados aos dados do gráfico de espectadores recorrentes por semana, inferindo-se haver uma correlação entre ambas, indicando que as lives mais assistidas atraem mais visualizações.

Pelo fato das duas tabelas possuírem a mesma estrutura de dados, foram indexadas e vetorizadas pela coluna chave "Data" . Uma análise de correlação com modelo de regressão linear para entender a relação entre as visualizações dos dois grupos, identificou um coeficiente de Correlação de Pearson = 0.526, que indica uma correlação moderada e positiva entre as duas variáveis, com crescimento explicado em 52,60% da variação.

A variável "Tempo de Exibição" foi analisada por um modelo de regressão linear para correlacionar as variáveis "Origem do tráfego" e "Status da inscrição" à variável "Visualizações". Aplicamos o método ANOVA bidirecional (LARS & WOLD, 1989), usado quando há uma variável quantitativa e duas variáveis nominais, investigando o efeito simultâneo das duas variáveis independentes na variável dependente.

O teste F para a não normalidade (análise de variações)⁸⁶, analisou as diferenças entre as médias das "Visualizações" para cada categoria de "Origem de

⁸⁶ Embora haja literatura com resultados polêmicos, há evidências a favor e contra sua robustez (BLANCA MENA, 2017).

Tráfego" e "Tempo de Exibição". As análises ANOVA indicaram diferenças significativas nas médias de "Visualizações" entre as diferentes categorias.

Mesmo que a distribuição da variável dependente não seja normal, a análise descritiva e a regressão linear, articuladas, sugerem que as variáveis "Origem do tráfego" e "Status da Inscrição" exercem uma influência significativa em "Visualizações". Ambos os níveis de status de inscrição são preditores significativos de visualizações, com valores de p inferiores a 0,05. No gráfico "Visualizações por Status de Inscrição", o valor de R ao quadrado de 0,311 indica que o modelo explica 31,1% da variância nas visualizações

Para granularizar ainda mais a análise, obtivemos dados sobre quais as plataformas e sites originam as visualizações, a duração média delas e o percentual médio do tempo de exibição em horas. Os 10 principais sites e/ou plataformas agregam 75% da origem de tráfego, totalizando 10.375 visualizações, que também correspondem às maiores médias de Tempo de Exibição.

Compreendemos melhor as métricas entre a idade do espectador e as visualizações através de técnicas de correlação e regressão, buscando determinar e medir estatisticamente a relação entre a idade do espectador e o tempo gasto assistindo aos conteúdos.

Os usuários que mais acessam estão na faixa entre 18 e 44 anos (78,95%) assim como o maior tempo de exibição acumulado (78,95%), porém o percentual médio de visualizações é maior pelos usuários entre 45 e mais de 65 anos. Os mais jovens geram um maior volume de visualizações, enquanto os de maior idade demonstram maior engajamento com o conteúdo assistido.

A baixa representatividade do público acima de 45 anos pode ter relação com dificuldades de acesso ou falta de familiaridade com a plataforma, além da probabilidade dos conteúdos não serem atrativos para essa faixa etária. Essas percepções são úteis para otimizar o conteúdo direcionado a públicos específicos de diferentes idades.

A análise por gênero considerou um cenário não binário com a possibilidade que o Youtube oferece de escolha da opção Outros, porém ninguém a assinalou. A distribuição percentual de visualizações dos conteúdos online de acordo com a faixa etária e o gênero dos espectadores, revelou a predominância do público feminino na

faixa de 25 a 35 anos. Entre 18 e 24, ambos os gêneros alcançam porcentagem aproximadas e acima de 45 anos há uma leve predominância do gênero masculino, sendo que nas demais faixas etárias, a proporção entre os gêneros é relativamente equilibrada.

Os termos mais buscados no Youtube foram analisados pela soma de visualizações de cada título pela origem da busca. Esta análise permitiu classificar os principais assuntos ordenados pelo total e percentual de visualizações. O termo mais buscado foi Ciências sociais em diálogo (40,64%), que representa todo o canal, mas excluindo-o, “Federalismo e descentralização de políticas públicas no Brasil” obteve 22,9% das visualizações em três registros na tabela com títulos ligeiramente diferentes, seguido por “Piero Leirner” (18,9% das visualizações) e “Organização e funcionamento dos sistemas de programas nacionais” (11,13% das mesmas), com dois títulos ligeiramente diferentes.

Este recorte ficou restrito aos termos de busca no Youtube, mas também há análises das origens de tráfego interno e externo nas páginas 170-174 que indicam predominância da origem externa. O modelo de regressão para “Origem do Tráfego” explica 49,5% da variância em “Visualizações” e o modelo para “Status da Inscrição” explica 31,1% da variância em visualizações, indicando as correlações em ambas.

As métricas de visualizações nas Playlists prejudicaram as análises devido ao fato da plataforma não separar as visualizações diretas e oriundas do playlists, que na verdade são listas dos conteúdos visualizados. No caso dos inscritos, que também são contabilizados duplamente, geraram discrepâncias que foram assinaladas e desprezadas na preparação.

5.3 - Análises Comparativas e Regressões

Ao comparar métricas e correlacionar variáveis, geramos análises comparativas que identificaram forte correlação entre visualizações, usuários ativos e contagem de eventos. Para melhor analisar essas e outras relações, aplicamos modelos de regressão linear e não linear para prever a contagem de eventos com base em visualizações e usuários ativos.

5.3.1 - RAPLE - Rede de Apoio aos Usuários de Língua Estrangeira

A análise das variáveis Visualizações x Usuários Ativos, com índice 0.85 indica forte correlação positiva, sugerindo que o aumento das visualizações da página está fortemente associado ao aumento do número de usuários ativos.

Já comparando Visualizações X Contagem de Eventos, notamos claramente uma relação quase linear com 0.998 de forte correlação positiva, assim como quando comparamos Usuários Ativos x Contagem de Eventos, que sugere uma forte correlação positiva entre as variáveis (0.87) .

A análise descritiva dos dados revelou uma forte correlação positiva entre "Usuários ativos", "Sessões engajadas" e "Contagem de eventos", mas uma correlação negativa entre "Taxa de engajamento" e as outras três variáveis. Além da matriz de correlação, foram construídos gráficos de dispersão para confirmar os resultados obtidos.

Aplicamos modelos de regressão linear e não linear para prever a contagem de eventos com base em visualizações e usuários ativos. As análises apresentaram relações lineares significativas entre as variáveis estudadas. Visualizações de páginas mostraram alta correlação, tanto com o número de usuários ativos quanto com a contagem de eventos. Além disso, o número de usuários ativos também apresentou forte correlação com a contagem de eventos.

O modelo de regressão logística resultou em uma acurácia de 0,23, o que indica que o modelo classificou corretamente 23% dos casos. Apesar da baixa acurácia, o modelo de regressão logística foi útil para aplicarmos três métodos alternativos para lidar com variáveis categóricas em modelos de regressão logística, que apresentaram resultados 8% maiores.

Método	Especificidade	Acurácia
One-Hot Encoding	Representa dados categóricos como vetores numéricos, com uma coluna binária (1) presença e (0 ausência)	0,31 (31%)
Dummy Coding	Transforma variáveis categóricas associando uma variável dummy que é a diferença em relação à categoria omitida	0,31 (31%)
Effect Coding	Emprega modelos de estimação representa as categorias como -1 e +1	0,31 (31%)
Tabela 26: Métodos alternativos de regressão logística e acurácia - o autor		

5.3.2 - Canal Ciências Sociais em Diálogo

As regressões realizadas também buscaram vincular o Título do Vídeo aos dados quantitativos das Visualizações ao vivo e Sob demanda, separando as duas modalidades de oferta de conteúdo. As análises mostraram forte predominância das Visualizações sob demanda, mas as obtidas na modalidade "Ao vivo" são mais consistentes e com maior "Duração de visualização", inferindo que os conteúdos permanecem ativos e agregando até 3 vezes mais visualizações, porém com uma redução considerável nos minutos visualizados.

Obtivemos um número positivo do coeficiente de correlação de 0,24, sugerindo uma correlação fraca. Isso indica que a cada aumento de uma unidade em "Visualizações ao Vivo" corresponde a um aumento de 2,01 unidades em "Visualizações sob demanda". Um R^2 de 0,058 explica apenas 5,8% da variabilidade sob demanda.

5.4 - Análise de Séries Temporais

Realizamos análises de séries temporais para identificar padrões, tendências e sazonalidades em visualizações e compartilhamentos, separando componentes de tendência e sazonalidade e testando as hipóteses para comparar as médias de diferentes grupos nos dados.

5.4.1 - Canal Ciências Sociais em Diálogo

O modelo analisa dados estacionários para verificar se a média, sua variância e autocorrelação são constantes durante a série temporal, diferenciando os dados. No caso destes dados, o modelo ARIMA foi capaz de prever com precisão os valores futuros das variáveis "Visualizações" e "Compartilhamentos". Isso sugere que o modelo é um bom ajuste para os dados e que pode ser usado para fazer previsões confiáveis sobre o futuro.

A análise das previsões revelou uma tendência ascendente ao longo do tempo, com picos nos meses de verão, indicando uma sazonalidade nos dados. A decomposição das séries temporais demonstrou que a tendência é o componente mais significativo. Embora a sazonalidade também seja relevante, o componente residual é relativamente pequeno, então a média do teste T indicou uma diferença significativa entre as médias dessas duas variáveis.

5.5 - Análises Descritivas

Empregamos estatísticas descritivas, como média, mediana e desvio padrão para resumir os dados, facilitar a classificação, encontrar inconsistências e discrepâncias que possam afetar as análises.

5.5.1 - RAPLE - Rede de Apoio aos Usuários de Língua Estrangeira

Estatisticamente, o histograma de visualizações mostrou uma contagem de visualizações relativamente baixa por título de página, com poucos conteúdos apresentando uma contagem significativamente alta. Essa distribuição sugere que as visualizações sejam influenciadas por alguns poucos títulos de página com alto tráfego ou engajamento, enquanto a maioria dos títulos de página tem um impacto menor na contagem geral.

Este agrupamento desproporcional e enviesado pode indicar *outliers*, que são dados que se desviam significativamente do padrão geral, não sendo necessariamente relacionado a diferenças expressivas ou positivas. Uma baixa contagem inesperada de eventos em certas páginas pode indicar a existência de gargalos ou desafios de usabilidade que dificultam o acesso dos usuários.

Geramos estatísticas descritivas como média, mediana, desvio padrão, mínimo, máximo e quartis para cada variável numérica e plotamos histogramas para visualizar suas distribuições. Três histogramas também exibem uma assimetria positiva que sugere um desempenho modesto da maioria dos títulos em termos de visualizações, usuários ativos e contagem de eventos, enquanto um pequeno número de títulos de página apresenta um desempenho excepcionalmente alto.

Essa similaridade na distribuição das três métricas pode ser explicada pelo fato de que as mesmas estão relacionadas ao engajamento do usuário com o conteúdo do site, no entanto, é importante observar que a correlação não implica causalidade.

Um título de página com muitas visualizações não significa necessariamente um grande número de usuários ativos e uma alta contagem de eventos, pois outros fatores, como o tipo de conteúdo, o design da página e as estratégias de divulgação, também influenciam essas métricas. A análise comparativa dos histogramas permitiu identificar os destaques em cada métrica, tanto para um desempenho abaixo do esperado como para um número superestimado que não reflete a amostra, e sim

exceções que podem ser tratados como outliers para preservar a qualidade da amostragem e da análise.

Os usuários inscritos são como indicadores de engajamento e a comparação entre as duas modalidades são estratégicas, porém não explicam e não são preditoras das visualizações. As métricas de visualizações variam consideravelmente ao longo do tempo, mas com pequenos períodos de alternância e variabilidade que diminuiriam nas séries temporais.

5.5.2 - Canal Ciências Sociais em Diálogo

As métricas sobre os dados de “Espectadores Recorrentes” se iniciaram em agosto de 2020, seguindo com valores totais proporcionalmente equivalentes até abril de 2023, quando a correlação diminui consideravelmente até os dias atuais. Essa visível alteração na variável pode ser estudada para obter percepções sobre as fontes de recorrência e seu impacto no número de seguidores, mas no geral, os dados mostram picos de visualizações pontuais, mas não constantes.

Uma grande quantidade de usuários recorrentes pode ser um indicativo de boa amplitude de público que assiste de forma continuada e usual, mas há picos de acesso. No gráfico de 'Relação entre Visualizações e Espectadores Recorrentes' é possível observar uma correlação positiva entre as duas variáveis, ou seja, quanto maior o número de visualizações, maior o número de espectadores recorrentes.

O gráfico de 'Distribuição de Inscritos, Inscrições Obtidas e Inscrições Perdidas por Origem da Inscrição' mostra que a principal origem de inscrições é a 'Página de exibição do YouTube', seguida por 'Seu canal do YouTube' e 'Pesquisa do YouTube'. As inscrições obtidas e perdidas seguem a mesma proporção, sendo que a principal origem de inscrições, 'Página de exibição do YouTube', é também a que mais perde e que mais ganha inscritos.

Os gráficos permitem inferir que as visualizações e os espectadores recorrentes têm picos pontuais e não constantes, com número médio de visualizações menor que 200 e o número de espectadores recorrentes abaixo de 50 (25%), ou seja um em cada 4 usuários é um espectador recorrente..

5.6 - Análises Qualitativas e de Conteúdo

Sobre os dados qualitativos de ambos os projetos foram realizadas análises de conteúdo, do discurso e de sentimento dos comentários e transcrições de vídeos, com auxílio de LLMs/RAG (Grandes Modelos de Linguagem referenciadas por Bases de Conhecimento).

Vídeos transmitidos e armazenados no Youtube, seus comentários e chats são transcritos e gravados em PDF na ferramenta Maxqda e disponibilizado juntamente com acesso ao audiovisual incorporado no site, onde qualquer pessoa pode ler a transcrição e acessar o vídeo ao mesmo tempo, além das análises geradas.

Após transcrito, através do texto são gerados a tabela de frequência e a nuvem de palavras usadas na live, além do gráfico de tendência de palavras gerados pela análise no Maxqda. A nuvem permite uma visualização dinâmica da predominância dos assuntos pela classificação das palavras da live que também são submetidos ao treinamento de aprendizado de máquina em framework Python no Notebook Colab, no Notebook LM e no Bedrock do MAXQDA.

Nas RAGs criamos as bases de conhecimento com os conteúdos e interações, que foram vetorizados e armazenados em servidores de busca elástica (Elastic Search) no Amazon OpenSearch Service.

Os LLMs do AI Assist do Maxqda e o Gemini Advanced do Google forneceram análises de sentimento e discurso coerentes e claros, através de comandos específicos e sequenciais no chat. Entregaram resumos com boa redação e sínteses pertinentes, além de gerar inferências sequer notadas via NLTK/Python. Um link com acesso ao contexto do chat no Gemini Advanced foi disponibilizado em cada texto de IA generativa, para que qualquer usuário com o mesmo nível de conta Google possa prosseguir com novas perguntas ao modelo, cujo recurso não existe no Maxqda, mas por ser um software .

Nosso modelo não restringe as análises de sentimento ao reducionismo da tripla polaridade (positivo, negativo ou neutro), mas possibilita a análise semântica ou mesmo linguística simultaneamente aos estudos estatísticos transversais. Também foram analisados os emoticons presentes nos diálogos, coletando e analisando importantes recursos de comunicação já incorporados à realidade linguística e dinâmicas conversacionais online.

Também foram analisados os “Comentários” nos conteúdos sob demanda (*on demand*), nos quais surgem novos questionamentos e manifestações que podem se desdobrar em um diálogo entre os espectadores que se tornam estratégicos no relacionamento com os inscritos no canal. Após a live, é o único espaço conversacional disponível, tendo um caráter mais informal e reflexivo que o chat, que acontece em tempo real.

As interações no chat podem ser úteis para analisar o conteúdo e avaliar a atmosfera da transmissão, mas como os moderadores nem sempre estão disponíveis, a tarefa geralmente fica a cargo da equipe de transmissão, incluindo os próprios mediadores.

As falas dos recursos “Chat” e “Comentários” apoiam as narrativas audiovisuais dos conteúdos, conferindo uma legitimidade única de manifestações que podem ser analisadas em seu contexto, no qual o lugar de fala é descorporificado e desterritorializado, mas ao mesmo tempo politizado e ideológico.

Ambos os modelos de IA generativa (Gemini e MaxQDA) permitem contextualizar sua própria base de conhecimento, ao agregar sites, pdfs, planilhas, tabelas, infográficos, manuais, documentos e qualquer conjunto de dados, estruturados ou não. Este recurso possibilita que seu chatbot ou agente de análise não precise acessar conteúdos externos da internet, caso isso seja relevante, como é o caso desta tese, ou seja, somente dados e informações da base de conhecimento que você mesmo cria, atualiza e realimenta, treinando o modelo continuamente.

No site da tese (tecnociencia.net.br) foi disponibilizado um robô de conversação (chatbot) treinado com os conteúdos da pesquisa via RAGs/LLMs que responde perguntas aos que visitam o site. Ele oferece um interessante recurso interativo de aprendizado que tira dúvidas sobre diversos tipos de dados, gráficos, tabelas, textos, fórmulas e imagens que constituem a tese.

Esse recurso possibilitou a aplicação prática de várias teorias citadas e estudadas nesta tese, através do treinamento do robô Tesebot, que aprende e aprimora suas respostas através das perguntas dos leitores da tese. É possível inferir que haja uma retroalimentação nos robôs, criadas pelos humanos para superar vieses e delírios que os modelos sem memória permanente apresentam.

Quando se “conversa” com o robô treinado em dados restritos, você pode corrigi-lo imediatamente após uma pergunta enviesada, delirante ou alucinada, ensinando o modelo via diálogo ou agregando conteúdos e variáveis que podem estabelecer um limiar mínimo de coerência sobre todos os dados, estruturados ou não, que você apresentou a ele. É um funcionamento similar ao cérebro humano mas com limitações importantes, entende a linguagem, dialoga com o conhecimento em sua memória, que é a base de conhecimento no qual é treinado e responde perguntas que podem ser reapresentadas ao robô em forma de retroalimentação.

6 - CONCLUSÃO DA TESE

Após cinco anos de pesquisa e coleta de dados, dos quais 6 meses finais foram de análises multivariadas dos dados estatísticos de acesso, conteúdos e interações dos dois projetos selecionados para as amostragens, **a conclusão, baseada nos resultados, inferências e percepções, é de plena eficácia das dinâmicas de difusão, divulgação e letramento científico, conceituado como cultura científica**⁸⁷ (VOGT, 2018).

Articulamos métodos de pesquisa e ferramentas computacionais para analisar conteúdos comunicacionais e audiovisuais de forma consistente e documentada, sempre buscando manter as premissas e técnicas dos métodos plenamente estabelecidos pelas Ciências Sociais e pela Ciência de Dados.

A palavra articular foi considerada e selecionada para fazer clara referência a ponderação de dois arcabouços que se complementam, sem que haja predomínio ou pretensão de superioridade de nenhum método. São complementares e não excludentes ou antagônicos, o que significa que podem compartilhar variáveis, percepções e inferências sem se contraporem epistemologicamente.

Os dois canais do Youtube e o site da RAPLE que forneceram os dados da análise, são ativos informacionais de grande valia e cumprem, categoricamente, as finalidades previstas nas atividades de extensão da Universidade Estadual Paulista (UNESP). Também atendem as normas de atuação que permitem, inclusive, que sejam tecnicamente gerenciáveis por alunos de graduação que recebem bolsas de estudos de agências de fomento e de editais da universidade.

As modalidades de difusão e letramento científico, oferecidas por ambos os projetos, atenderam as premissas básicas da cultura científica, não somente através dos acervos audiovisuais e de objetos de aprendizagem, mas também pelas redes sociais, sites de notícias da universidade e pela imprensa.

As dinâmicas dos laboratórios das ciências humanas podem ser observadas no anfiteatro, na biblioteca e atualmente, também nas redes complexas de cultura científica, que se utilizam das mídias digitais e redes sociais para difundir a ciência e a pesquisa desenvolvida, comunicar seus resultados e oferecer seus conteúdos de forma gratuita e acessível aos cidadãos, em forma de difusão e letramento científico.

87

https://www.espiritualidades.com.br/Artigos/V_autores/VOGT_Carlos_tit_Espiral_da_cultura_cientifica-A.pdf

Como dito por Latour através de seu método de etnografia da ciência, os cientistas seriam uma das “tribos da modernidade” e a dinâmica dos laboratórios mostrava que havia práticas coletivas, códigos e uma linguagem específica entre eles. Nos laboratórios das ciências humanas, essas dinâmicas são, em sua grande maioria, conversacionais e reflexivas, e como ocorrem no ambiente virtualizado das plataformas de conteúdo e mídias sociais, sua análise foi realizada com ferramentas que permitiram obter um recorte quali e quantitativo de suas idiossincrasias.

É possível inferir que os dados mostraram uma comunidade que se formou em torno desses projetos de extensão, que atravessa fronteiras e alcança dezenas de países em mais de duas centenas de cidades ou regiões no mundo todo, em diferentes idiomas e culturas, com pessoas que existem também fora do virtual.

O isolamento social ocasionado pela pandemia de Covid-19 acelerou a convergência das redes e a flexibilização da presença física, favorecendo a criação de canais de interação na internet para garantir a continuidade dos eventos acadêmicos, mas em sua forma virtualizada.

A pesquisa que resultou nesta tese, ofereceu um caminho metodológico que articulou os já consolidados métodos das Ciências Sociais e da Ciência de Dados em uma abordagem híbrida (qualitativa e quantitativa), permitindo análises com resultados surpreendentes sobre as métricas e dinâmicas de difusão e letramento científico.

Análises multivariadas não são inovadoras e nem são consideradas um método, mas se tornaram uma realidade latente e de inegável valor aos pesquisadores que optam por articular ambas as modalidades complementarmente.

O que se verifica ao fim da tese, é que sua corroboração é inevitável e carregada de certezas estatísticas e inferenciais, que refletem e legitimam a complexidade das pesquisas científicas na era da ciência pós-acadêmica, da tecnociência, do big-data, do aprendizado de máquina e da inteligência artificial.

Articulando as ferramentas do fazer científico, da análise sociológica e da mineração de dados multivariados (estatísticos e conteúdos), a alternativa possível foi acertada em compreender todas as variáveis e instâncias envolvidas, sem perder o sentido de observar a cultura científica presente nos projetos de extensão pesquisados.

No mesmo sentido, há uma certa ousadia metodológica na forma como a pesquisa foi conduzida pela metodologia da pesquisa-ação, sem a qual não haveria a possibilidade de corrigir rotas em tempo hábil. Foi o caso da abrupta e não justificada decisão da *big tech* Meta, de interromper definitivamente todas as formas de acesso autorizado aos dados do facebook e instagram em 2024.

Esse duro golpe na comunidade científica internacional ocasionou ajustes de rotas metodológicas impensáveis para muitos candidatos ao doutorado, como o autor desta tese. Cerca de 30% da coleta de dados desta pesquisa precisou ser abandonada e ignorada por não haver mais como desenvolver as análises sem acesso aos dados atualizados, afinal a Meta havia parado de fornecer dados já em 2023, na época da qualificação desta tese.

Exatamente nos dias finais de escrita desta tese, a Meta anunciou que não fará mais a mediação profissional dos conteúdos em suas redes sociais, o que significa que, assim como o Twitter, deixarão que apenas os usuários digam o que é ou não desinformação, conteúdo violento e negacionismos científicos ou climáticos.

Isso talvez explique o movimento anterior de cortar acesso dos cientistas aos dados, pois segundo a imprensa mundial, as *big techs* estadunidenses estão se alinhando a presidentes de extrema direita para contrapor a legislação de países que criam leis contra a desinformação e a manipulação algorítmica, geralmente referenciadas pelos dados das pesquisas científicas. O negacionismo, a agnotologia e a desinformação remuneradas se tornaram, inegavelmente, ideológicos e monetizados pelas *big techs*, como afirmado no texto inicial desta tese.

O caminho metodológico da tese se iniciou com revisões bibliográficas sobre cultura algorítmica que abordaram a flexibilização da presença física, gig economy, a desinformação remunerada e ideológica e os algoritmos de anúncios na internet que pagam mais e melhor aos sites de desinformação (GDI) ⁸⁸. Em contraponto, justificou-se a necessidade das instituições de ensino e pesquisa científica enfrentarem e combaterem a desinformação remunerada e o negacionismo científico através de projetos de extensão universitária que auxiliam o letramento científico e a troca de conhecimentos e saberes.

⁸⁸ <https://www.disinformationindex.org/research>

A pandemia acelerou essa nova dinâmica de virtualizar os eventos acadêmicos para atender as demandas de difusão científica e debates culturais, criando novos acervos que, literalmente, se tornaram canais de conteúdos audiovisuais, que ao contrário do que pensa a maioria, são mais visualizados por demanda (*on demand*) do que ao vivo em transmissões (streaming).

O objetivo principal foi avaliar a eficácia de difusão e letramento científico do site da Rede de Apoio aos Professores de Língua Estrangeira (RAPLE) com seus planos de aulas e objetos de ensino aprendizagem e dos conteúdos audiovisuais do canal de Youtube do Ciências Sociais em Diálogo e da RAPLE.

A diversidade de conteúdos e mídias digitais presentes em ambos os projetos, levou-nos a adotar a abordagem híbrida aqui proposta, mesmo cientes dos percursos metodológicos ainda não delineados e que ainda não haviam sido relatados na literatura acadêmica contemporânea.

Justamente por essa possibilidade de inovar alguns métodos consolidados com adaptações cuidadosas, surgiu a demanda de identificar ferramentas de análise quanti e qualitativas para nos ajudar a coletar e “raspar” dados na internet, transcrever parte deles, codificar, categorizar, minerar e realizar as análises multivariadas.

Não avaliamos apenas aspectos operacionais, mas se atendem aos protocolos das Análises Descritivas e de Conteúdo, que requerem alguns recursos que somente foram encontrados no software MAXQDA. Suas ferramentas integradas de IA foram um diferencial que valeu cada centavo de Euro pago por 1 ano de assinatura do mesmo.

As exigências metodológicas foram adaptadas para a realidade objetiva da mineração de dados baixados nas plataformas e suas limitações impostas, como por exemplo, o fato do Youtube não permitir todo e qualquer combinação de dados disponíveis, apenas nos recortes autorizados.

Até o último ano da pesquisa, o autor buscou conciliar, epistemologicamente na tese, a etnografia virtual com a análise de conteúdos, porém o rigor metodológico nos mostrou não serem complementares, com finalidades semelhantes, mas com procedimentos, resultados e relatórios muito diferentes. A análise etnográfica foi então abandonada e o foco nos conteúdos motivou as análises de sentimento, de discurso e

de interações, obtendo-se uma rica interpretação da cultura científica nas atividades de difusão e letramento dos projetos de extensão.

Enfim, estudamos os discursos, sentimentos, engajamentos, linguagem, estatísticas, idiossincrasias e propomos uma série de inferências e percepções sobre essas “tribos” das ciências humanas que buscam clara e inegavelmente, difundir conhecimento científico e oferecer formas lúdicas de letramento para estudantes, pesquisadores, professores e o público em geral.

A conclusão final é de que essas dinâmicas foram eficazes em difundir a cultura científica através dos projetos de extensão “Ciências Sociais em Diálogo” e “Rede de Professores de Língua Estrangeira (RAPLE)” da Faculdade de Ciências e Letras da Unesp de Araraquara, onde o autor atua no suporte acadêmico a estes projetos e vários outros.

6.1 - Considerações Finais

Um resultado incontestável desta tese é o método híbrido proposto e aplicado, que resultou numa forma eficiente de análise multivariada de dados e conteúdos provenientes de redes e plataformas informacionais. As ferramentas foram estrategicamente articuladas para fornecer os melhores resultados, mesmo considerando suas particularidades e limitações metodológicas e práticas.

Um exemplo é o fato de que houve coletas manuais dos chats e automatizadas das transcrições e comentários das lives, sendo que ambas complementam as duas modalidades (streaming e on demand) de visualizações e interações nos projetos, durante a transmissão e após ela, quando os conteúdos ficam anos sendo acessados sob demanda.

As métricas de cada modalidade foram avaliadas por diferentes variáveis que ora separavam, ora uniam as mesmas e forneceram uma percepção de que há uma demanda diferente de divulgação para cada uma delas. O acervo sob demanda precisa ser continuamente divulgado através de técnicas de edição compactadas, resumos e compilações que atraiam novos usuários e mantenham a recorrência de visualizações após a live.

Outro resultado secundário é que a publicação de chats, comentários e transcrições dos conteúdos audiovisuais publicados no site, os tornaram vivos e

ativos aos buscadores semânticos (Google, Mozilla, Safari) e assistentes de inteligência artificial generativa (Gemini, GPT, Perplexity, etc). Essa perspectiva recria um espaço de interação com os conteúdos que anteriormente só poderiam ser acessados se o usuário assistisse o audiovisual por inteiro.

Os acervos audiovisuais passam a ser pesquisáveis por palavras e assuntos, trazendo os conteúdos ao alcance de qualquer pessoa, inclusive as com deficiência auditiva ou visual. Essa nova abordagem pode conferir acessibilidade aos conteúdos audiovisuais de uma forma inédita, pois é possível aos deficientes refinar suas buscas por frases e trechos específicos, sem necessidade de ouvir linearmente as narrativas e debates de cada audiovisual.

O robô Tesebot desenvolvido com ferramentas de Geração Aumentada por Recuperação para Grandes Modelos de Linguagem (RAGs/LLMs), foi incorporado à tese como prova empírica da aplicação de pipelines de Aprendizado de Máquina, sem a necessidade de retreinamento do modelo de linguagem. Essa recente abordagem melhora substancialmente os resultados dos chamados “modelos generativos” de inteligência artificial ao restringir as buscas do modelo de linguagem (LLMs) às bases de conhecimento criadas pelo usuário.

Converter os textos, tabelas, gráficos e *dataframes* (Base de Conhecimento) em bancos de dados de vetores ou grafos (servidores de Busca Elástica⁸⁹), torna-os acessíveis aos grandes modelos de linguagem (LLMs). A busca elástica é uma ferramenta para buscas de código aberto (*opensource*), que tem capacidade para tratar de grandes quantidades de dados em tempo real

Um Agente de Conhecimento (chatbot - na tese o robô TeseBot) articula as três instâncias de armazenagem (1 - Base de Conhecimento), vetorização e/ou grafo-representação (2 - Servidor de busca distribuído) e acesso aos modelos de linguagem (Assistentes de IA generativa já treinados em petabytes de dados (LLMs) e responde perguntas sobre a base aumentada de conhecimento (RAG).

Esse recurso reduz consideravelmente as alucinações e delírios dos assistentes generativos (IAs semi recorrentes como Gemini Advanced, Perplexity

⁸⁹ Desenvolvido por Shay Bannon em 2010, o servidor de buscas distribuído é baseado em [Apache Lucene](#) e desenvolvido em java, utilizando uma interface comum — JSON sobre HTTP.

Pro, GPT, Claude) ao processarem somente dados, estruturados ou não, fornecidos pelo usuário ou instituição.

Quando se realiza uma pesquisa somente em LLMs que são treinados com bancos de dados genéricos e dados da internet, além das alucinações, corre-se o risco de receber respostas carregadas de preconceitos, desinformação e vieses presentes nas notícias falsas e agnotologia ideológica ou remunerada (fake news e negacionismo científico).

Neste ponto da tese, o argumento acima dialoga com a introdução no capítulo um, que discorre sobre porque as *big techs* não moderam a desinformação, justamente para proteger o componente mais importante de seu modelo de negócios atual, os agentes de Inteligência Artificial Generativa (IA).

Nossos argumentos sobre vieses algoritmos no capítulo dois desta tese, dialogam também aqui com as possíveis respostas enviesadas que os agentes de IAs semi recorrentes (que respondem de forma genérica baseadas apenas nos conteúdos públicos e na internet) fornecem corriqueiramente no uso diário dessas tecnologias, que também foram treinadas com dados enviesados e algoritmos tendenciosos.

Sequer percebemos a intensidade de uso recorrente de aplicações com recursos de IA generativa em nossos smartphones e dispositivos, como assistentes virtuais (Alexa, Siri), relógios com análises corporais de bioimpedância, veículos autônomos e robôs de atendimento (chatbots). Todos foram e serão (re)treinados com grandes massas de dados que podem refletir vieses sociopolíticos, racistas, de gênero e etnia.

Também podem conter desinformação e negacionismo científico, que realimenta bolhas ideológicas e conservadoras, manipuladas por startups contratadas para a produção deliberada e intencional de desinformação e difamação da ciência, conhecida como agnotologia (PROCTOR e SCHIEBINGER, 2008)

Destacamos por último, a abordagem híbrida quali e quantitativa, que promoveu um debate epistemológico sobre aspectos interdisciplinares que envolvem duas ciências, seus métodos, técnicas e arcabouços. As Ciências Sociais e a Ciência de Dados se tornam complementares, não excludentes e se articulam naturalmente, pois em ambas, a estatística fundamenta o método.

Essa é um motivo para uma imbricação de ambas à ciência da complexidade, que compartilham desde a constante de Boltzmann e a mecânica estatística de gases no séc. XIX, até os atuais algoritmos de aprendizado de Máquinas Restritas de Boltzmann (BRMs), que possibilitaram a IA Generativa. Esta tese conecta duas vertentes científicas que se entrelaçam em alguns períodos históricos para viabilizar a continuidade de pesquisas científicas indispensáveis ao desenvolvimento tecnológico em que nos situamos, mas em sua mais fundamental forma, a ciência básica.

O livro seminal de Durkheim sobre o suicídio⁹⁰ foi um dos pioneiros na aplicação de métodos quantitativos para estudar fenômenos sociais, através de estatísticas de diferentes países e também por dados coletados por ele no interior da França sobre os suicídios e suas circunstâncias. Fica latente o fato de que as duas ciências fazem uso da estatística para corroborar e apoiar inferências sobre moléculas de gases ou fatos sociais desde o século XIX.

Fizemos uma compilação de bibliografias sobre a origem dos algoritmos lineares ou não lineares desde a constante de Boltzmann-Gibbs-Planck⁹¹ que fundamentam essas duas correntes de pensamento que ora convergem, ora divergem e realimentam a pesquisa com seus erros e acertos. Estas duas linhas não seriam apenas de pensamento, mas também de métodos e técnicas que atravessaram mais de 148 anos, desde que o físico austríaco propôs a constante física que relaciona temperatura e energia de moléculas (1877), quase ao mesmo tempo que Durkheim aplicava o mesmo princípio quantitativo da estatística para explicar fenômenos sociais (1897).

A física estatística relacionada ao problema da refutação da 2a.Lei da Termodinâmica sobre entropia gerou dois conceitos, que estão entre os mais importantes da ciência desde aquela época até os dias atuais. O conceito de bit de informação por Szilard e o conceito de mínimos locais e mínimos globais das Máquinas de Boltzmann, foram fundamentados na mesma mecânica estatística adaptada por Heisenberg para formular a teoria da dispersão estatística na física

⁹⁰ "O Suicídio" é uma obra fundamental da sociologia por sua metodologia, inovadora à época, de análise rigorosa de dados estatísticos que tornou a sociologia uma ciência empírica.

⁹¹ Planck introduziu a constante que leva seu nome, a relação entre a frequência e energia de fótons e a equação de Boltzmann-Planck (por vezes chamada apenas de equação de Boltzmann).^[3]

quântica. As RBMs (Máquinas de Boltzmann Restritas) foram um dos principais algoritmos de aprendizagem de redes neurais e servem de base para modelos mais recentes com retropropagação ou adversariais.

Esses Grandes Modelos de Linguagem (LLMs), como ChatGPT, Gemini, Deep Seek, Claude etc) entregam performances que já são usadas em vários segmentos socioeconômicos globalmente e são uma realidade irreversível. No final de janeiro de 2025, o nome de um desses modelos chineses foi o assunto mais comentado no Google Trends em nível global durante dias.

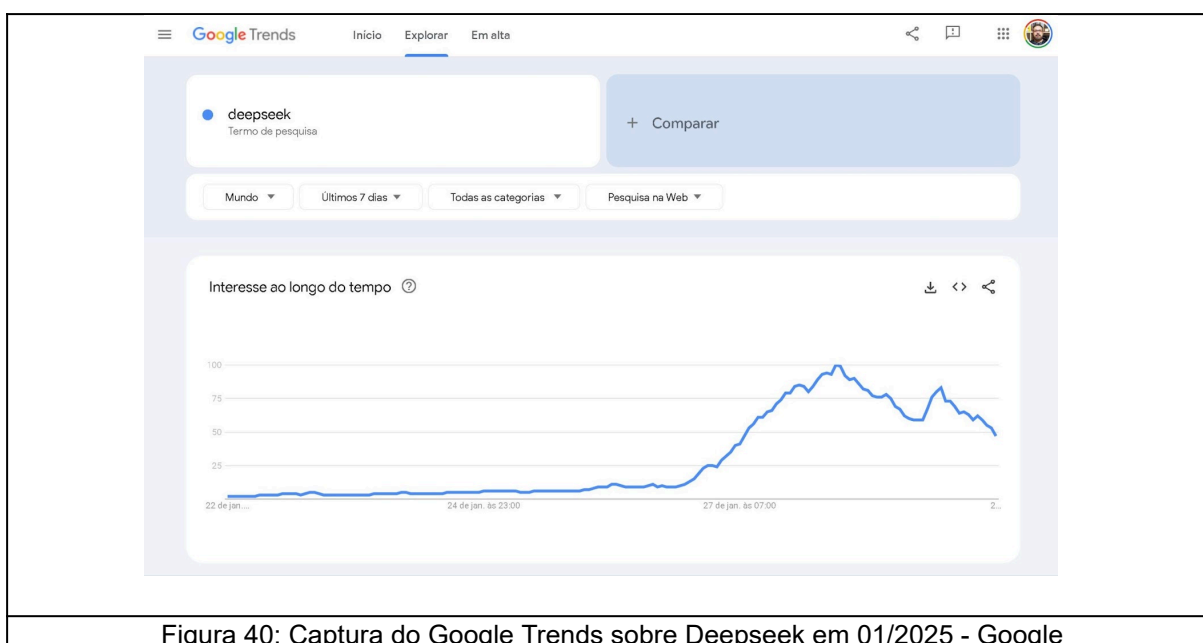


Figura 40: Captura do Google Trends sobre Deepseek em 01/2025 - Google

O Deep Seek abalou até mesmo o mercado de ações de empresas de tecnologia quando anunciou um modelo mais competitivo com custos de treinamento bastante reduzidos, chegando a custar uma pequena fração dos gastos das big techs norte americanas.

Outra vantagem do modelo é sua disponibilidade de código aberto (Open Source) que por suas características colaborativas e descentralizadas, pode acelerar o desenvolvimento de um framework robusto de desenvolvimento de IA a custos moderados, inclusive para países como o Brasil.

A disruptura como vantagem competitiva, novamente evidencia a disparidade de velocidade entre a evolução cultural e evolução genética é um indicativo desta nova realidade, contrapondo a forma como ela é vista na teoria do processo evolucionário de dupla herança (BOYD, RICHERSON, 1985).

Esta tese, defendida no âmbito de um programa de Pós-graduação que pesquisa aspectos sociopolíticos da ciência e tecnologia, ofereceu uma análise extensa sobre uma cronologia do desenvolvimento científico de ciência básica que culminou com os algoritmos de Aprendizado de Máquina Profundo.

Os próximos passos no pós-doc devem se concentrar na consolidação do modelo do TeseBot e aperfeiçoamento de seu treinamento como Agente de Conhecimento que consegue analisar os conteúdos e responder até mesmo questões metodológicas e epistemológicas.

7 - Referências Bibliográficas

[ACHACOSO](#), Theodore B.; FERNANDEZ, Victor; NGUYEN, Duc C.; YAMAMOTO, William S. **Computer Representation of the Synaptic Connectivity of Caenorhabditis Elegans**. (1989) Proceedings of the Annual Symposium on Computer Application in Medical Care: 330–334. [ISSN 0195-4210](#). [PMC 2245716](#)

ACKLAND, R. (2013). **Web social science: Concepts, data and tools for social scientists in the digital age**. SAGE Publications Ltd, <https://doi.org/10.4135/9781446270011>

ALONSO, Cipriano Barrio, **La apropiación social de la ciencia: nuevas formas**, Revista CTS, nº10, vol.4, Enero de 2008 (pág. 213-225)
https://minciencias.gov.co/sites/default/files/ckeditor_files/apropiacion-nuevasformas.pdf

AMARAL, Fernando. **Introdução à Ciência de Dados: Mineração de Dados e Big Data** - Alta Books; 1ª edição (2018)
<https://pt.scribd.com/document/750708084/Introducao-a-Ciencia-de-Dados-Fernando-Amaral>

ARAÚJO, E. A.; MELO, A. V. C. **Capital informacional e construção do poder simbólico: uma proposta epistemológica a partir de Pierre Bourdieu**. Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação, 8., 2007, Salvador - Disponível em: <http://hdl.handle.net/20.500.11959/brapci/171478>.

ARISTÓTELES, **Metafísica vols. I, II, III**, 2ª edição. São Paulo. Edições Loyola. 2002.
https://moodle.ufsc.br/pluginfile.php/1332285/mod_resource/content/1/Aristoteles-Metafisica-Edicoes%20Loyola%20%282002%29.pdf

ARRUDA, Adriana G.; SANTOS, Juliana Cristina Santicioli; ZAMBON, Sueli Aparecida e SOUSA, Cidoval Moraes. **Comunicação pública da ciência e apropriação social da ciência e tecnologia: apontamentos para uma leitura CTS**. REVISTA CEREUS, v. 9, n. 3, p. 66-80, 2017. - <https://www.ojs.unirg.edu.br/index.php/1/article/view/1881>

BABINI, Maurizio; MARRANGHELLO, Norian . **Introdução às Redes Neurais Artificiais** - Cultura Acadêmica - 2007

BALTAR, R.; SIQUEIRA BALTAR, C. **As ciências sociais na era do Zettabyte**. Mediações - Revista de Ciências Sociais, Londrina, v. 18, n. 1, p. 11–19, 2013. DOI: 10.5433/2176-6665.2013v18n1p11. <https://ojs.uel.br/revistas/uel/index.php/mediacoes/article/view/16805>

BARDIN, L. **Análise de conteúdo**. Edição revista/ampliada. São Paulo: Edições 70, 2016. https://edisciplinas.usp.br/pluginfile.php/7684991/mod_resource/content/1/BARDIN_L_1977_Analise_e_de_conteudo_Lisboa_edicoes_70_225.20191102-5693-11evk0e-with-cover-page-v2.pdf

BARRETO JUNIOR, Irineu Francisco, PELLIZZARI, Bruno Henrique Miniuchi; **Bolhas Sociais e seus efeitos na Sociedade da Informação: ditadura do algoritmo e entropia na Internet**. Revista de Direito, Governança e Novas Tecnologias 2019-12-20 Journal article. DOI: [10.26668/IndexLawJournals/2526-0049/2019.v5i2.5856](https://doi.org/10.26668/IndexLawJournals/2526-0049/2019.v5i2.5856)

BAUMAN, Zygmunt. **Vigilância líquida**. Rio de Janeiro: Zahar, 2014

BERELSON, B. **Content analysis in communication research**. Nova York: Univer. Press, 1952. - <https://archive.org/details/contentanalysisi0000bern>

BERTARELLI, Admir A Junior. **Conteúdo 1: Introdução à Estatística Multivariada** - UFJF https://www2.ufjf.br/lates/files/2016/12/Conte%3%bado-1-%e2%80%93N_Introdu%3%a7%3%a3o-%c3%a0-Estat%3%adstica-Multivariada-e-an%3%a1lise-dos-dados.pdf

BERTELLE, Cyrille; DUCHAMP, Gerard H.E.; KADRI-DAHMANI, Hakima. **Complex Systems and Self-organization Modelling**. ISBN: 978-3-540-88072-1; DOI 10.1007/978-3-540-88073-8 - Springer-Verlag Berlin Heidelberg 2009

BEZERRA, A. C. **Vigilância e cultura algorítmica no novo regime de mediação da informação**. Perspectivas em Ciência da Informação, v. 22, n. 4, p. 68-81, 2017. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/38848>.

BLACKMAN, Colin R. **Convergence between telecommunications and other media: How should regulation adapt?**. Telecommunications policy, v. 22, n. 3, p. 163-170, 1998. ScienceDirect <http://www.sciencedirect.com/science/article/pii/S0308596198000032>

BOYD, Robert; RICHERSON, Peter J. **Culture and the Evolutionary Process**. Chicago: University of Chicago Press, 1985, Politics and the Life Sciences. Cambridge University Press, 5(1), pp. 149–149. doi: [10.1017/S0730938400001763](https://doi.org/10.1017/S0730938400001763).

BUNGE, Mario. **Tecnología, ciencia y filosofía**. In: **Filosofía de la tecnología y otros ensayos**. Lima: Fondo Editorial de la Universidad Inca Garcilaso de la Veja, 2012, p. 47-75. https://nanopdf.com/download/filosofia-de-la-tecnologia-y-otros-ensayos-mario-bunge_pdf

CAIRES, Luiza; **Nos países desenvolvidos, o dinheiro que financia a ciência na universidade é público**; Texto jornalístico acessado em 09/06/2019 em <https://jornal.usp.br/ciencias/nos-paises-desenvolvidos-o-dinheiro-que-financia-a-ciencia-e-publico/>
_____, **Avaliação da Pós-Graduação Stricto Sensu, Produção Intelectual**, 2016,

<https://dadosabertos.capes.gov.br/organization/a994a478-5aaf-4e09-8372-134a0af7c0e8?grupos=detalhes-da-producao-intelectual&tags=Produção+Intelectual>

CARIBÉ, Rita de Cássia do Vale; **Comunicação científica: reflexões sobre o conceito**, Informação & Sociedade; João Pessoa. Vol. 25, Ed. 3, (2015)

CARVALHO, André Carlos Ponce de Leon Ferreira e BRAGA, Antônio de Pádua e LUDERMIR, Teresa Bernarda. **Fundamentos de redes neurais artificiais**. Rio de Janeiro: DCC/IM, COPPE/Sistemas, NCE-UFRJ. 1998

CASSINO, João F, SOUZA Joyce, Silveira, Sérgio Amadeu da, **Colonialismo De Dados: Como Opera A Trincheira Algorítmica Na Guerra Neoliberal** - 1ªED.(2021)

CASTELLS, Manuel. **Sociedade em Rede** - Vol I - São Paulo: Paz e Terra, 1999 - https://edisciplinas.usp.br/pluginfile.php/392268/mod_resource/content/1/ASociedadeEmRedesVol.I.pdf

CERRI, Ricardo. **Redes neurais e algoritmos genéticos para problemas de classificação hierárquica multirrótulo**. 2013. Tese (Doutorado em Ciências de Computação e Matemática Computacional) - Universidade de São Paulo, São Carlos, 2013. doi:10.11606/T.55.2013.tde-24032014-163900.

CETIC 2018, **TIC cultura 2018 - pesquisa sobre o uso das tecnologias de informação e comunicação nos equipamentos culturais brasileiros** Núcleo de Informação e Coordenação do Ponto BR. -- São Paulo : Comitê Gestor da Internet no Brasil, 2019. 3700 Kb ; PDF

CGEE 2019, **Percepção Pública da C&T no Brasil – 2019. Resumo Executivo**. Brasília: Centro de Gestão e Estudos Estratégicos supervisionado pelo Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC), 2019. https://www.cgee.org.br/documents/10195/734063/CGEE_resumoexecutivo_Percepcao_pub_CT.pdf

CGEE 2023, **Percepção pública da C&T no Brasil - 2023. Resumo Executivo**. Centro de Gestão e Estudos Estratégicos, 2024. 30 p. ISBN - 978-65-5775-080-3 - https://www.cgee.org.br/documents/10195/4686075/CGEE_OCTI_Resumo_Executivo-Perc_Pub_CT_Br_2023.pdf

CHIZZOTTI, Antonio. **Pesquisa em ciências humanas e sociais**. Cortez editora, 2018. Acessado em 19/10/2018 em <https://scholar.google.com.br/citations?user=Y96oF2oAAAAJ&hl=pt-BR&oi=sra>

COLLINS, Harry; EVANS, Robert. **Repensando a Expertise**. Fabrefactum - 2010 - https://www.researchgate.net/publication/333671358_Repensando_a_Expertise

CONTE, Rosaria et al. **Manifesto da ciência social computacional**. Mediações - Revista de Ciências Sociais, Londrina, v. 18, n. 1, p. 20–54, 2013. DOI: 10.5433/2176-6665. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/mediacoes/article/view/16806>. Acesso em: 22 dez. 2024.

COOK, S.J., JARRELL, T.A., BRITTIN, C.A. et al. **Whole-animal connectomes of both *Caenorhabditis elegans* sexes.** Nature 571, 63–71 (2019).
<https://doi.org/10.1038/s41586-019-1352-7>

D'ANCONA, Matthew. **Pós-Verdade: A Nova Guerra Contra os Fatos em Tempos de Fake News.** Faro Editorial; 1ª edição - ISBN-10 : 8595810176

DAGNINO, Renato. **Tecnociência solidária: um manual estratégico** /. Marília : Lutas Anticapital, 2019. 161 p. ISBN 978-85-53104-36-9.
https://www.mpba.mp.br/sites/default/files/biblioteca/direitos-humanos/programas_sociais/dagnino_renato_tecnociencia_solidaria_-_um_manual_estrategico_marilia_2019.pdf

DARWIN, C. **A Origem das Espécies.** Hemus – Livraria Editora Ltda, São Paulo, SP. 16. 2013. Disponível em <http://ecologia.ib.usp.br/ffa/arquivos/abril/darwin1.pdf>

DASGUPTA, Sanjoy; PAPADIMITRIOU, Christos H. VAZIRANI, Umesh Vir Kumar. **Algorithms.** New York: McGraw-Hill Higher Education, 2008.
<http://algorithmics.lsi.upc.edu/docs/Dasgupta-Papadimitriou-Vazirani.pdf>

DURKHEIM, ÉMILE, **As Regras do método sociológico** - Companhia Editora Nacional (IB ; ISBN-10, 8504002268 ; ISBN-13, 9788504002263

EDELMANN, A., WOLFF, T., MONTAGNE, D., & BAIL, C. A. (2020). **Computational Social Science and Sociology.** Annual review of sociology, 46(1), 61–81.
<https://doi.org/10.1146/annurev-soc-121919-054621>

EMPOLI, Giuliano da. **Os engenheiros do caos: como as fake news, as teorias da conspiração e os algoritmos estão sendo utilizados para disseminar ódio, medo e influenciar eleições.** 3. ed.São Paulo: Vestígio, 2019.

ERDŐS, László. KNOWLES, Antti. YAU, Horng-Tzer. YIN, Jun. **Spectral statistics of Erdős–Rényi graphs I: Local semicircle law.** Ann. Probab. 41 (3B) 2279 - 2375, May 2013. <https://doi.org/10.1214/11-AOP734>

FERREIRA, A. F.; BORGES, L. T. M. **As Pontes de Königsberg.** C.Q.D. - Revista Eletrônica Paulista de Matemática, Bauru, v. 5, 2022. Disponível em: <https://sistemas.fc.unesp.br/ojs/index.php/revistacqd/article/view/62>.

FERREIRA, J. **Adaptação, disrupção e regulação em dispositivos midiáticos.** MATRIZES, v. 10, n. 2, p. 135-153, 2016. DOI: 10.11606/issn.1982-8160.v10i2p135-153.
<https://www.revistas.usp.br/matrizes/article/view/120016>. Acesso em: 16 mar. 2023.

FREITAS, Christiana Soares de, **O Capital Tecnológico-Informacional.** Estudos de Sociologia. Revista do Programa de Pós-graduação em Sociologia da UFPE. v. 8. n. 1.2. p. 163-183. <https://periodicos.ufpe.br/revistas/revsocio/article/download/235444/28433>

FREUND, Rudolf J.; WILSON, William J.; SA, Ping. **Regression analysis.** Elsevier, 2006.

GILLESPIE, Tarleton; BOCZKOWSKI, Pablo e FOOT, Kirsten. **The Relevance of Algorithms. in Media Technologies: Essays on Communication, Materiality, and Society**, ed. Cambridge, MA: MIT Press. 2013. Disponível em: https://www.microsoft.com/en-us/research/wp-content/uploads/2014/01/Gillespie_2014_The-Relevance-of-Algorithms.pdf

GRANOVETTER, M. S. ,The Strength of Weak Ties. American Journal of Sociology, V. 78, issue (may, 1973), 1360-1380 - <https://snap.stanford.edu/class/cs224w-readings/granovetter73weakties.pdf>

GROSSBERG, Stephen. **Towards building a neural networks community**. Neural Networks 23 (2010) 1135–1138. <https://sites.bu.edu/steveg/files/2016/06/GrossbergNNeditorial2010.pdf>

HAYASHI, MCPI., SOUSA, CM., and ROTHBERG, D., orgs. **Apropriação social da ciência e da tecnologia: contribuições para uma agenda** [online]. Campina Grande: EDUEPB, 2011. 323 p. ISBN 978-85-7879-187-2. Available from SciELO Books .

HAYKIN, Simon. **Redes Neurais: Princípio e Prática**. 2a.ed. Porto Alegre: Bookman, 2001 <https://pdfcoffee.com/redes-neurais-principios-e-pratica-2-ed-red-ocr-pdf-free.html>

HOLSTI, O. **Content analysis for the social science and humanities**. [s.l.]: Addison-Wesley, 1969.

IANNI, Octávio. **As Ciências Sociais na era da Globalização**. Rev. bras. Ci. Soc. 13 (37) • Jun 1998 • <https://doi.org/10.1590/S0102-69091998000200002>

IANNI, Octávio. **Globalização: novo paradigma das ciências sociais**. Estud. av. 8 (21) • Ago 1994 • <https://doi.org/10.1590/S0103-40141994000200009>

IBGE – Instituto Brasileiro De Geografia E Estatística . **Censo Brasileiro 2023**. Rio de Janeiro: IBGE, 2023.

_____,. **PNAD Contínua**. Estudos e Pesquisas • Informação Demográfica e Socioeconômica • n.41 - Rio de Janeiro: IBGE, 2018/2023.

_____, **Desigualdades Sociais por Cor ou Raça no Brasil**. Estudos e Pesquisas • Informação Demográfica e Socioeconômica • ISBN 978-85-240-4513-4n.41 - 2019 https://biblioteca.ibge.gov.br/visualizacao/livros/liv101681_informativo.pdf

IBLC 1, **Letramento científico: um indicador para o Brasil**, Instituto Brasileiro de Letramento Científico. São Paulo (SP): IBLC, 2017 <http://iblc.org.br/wp-content/uploads/2018/01/3-publicacao-ilc.pdf>

IBLC 2, **Indicadores de Letramento Científico: Relatório Técnico**, Instituto Brasileiro de Letramento Científico. São Paulo (SP): IBLC, 2017 <http://iblc.org.br/wp-content/uploads/2018/01/2-relatorio-tecnico-ilc.pdf>

GÓES, Geraldo Sandoval, MARTINS, Felipe dos Santos, FIRMINO, Antony Teixeira e RANGE, Leonardo Alves. **A proteção social dos trabalhadores da Gig Economy de Transporte do Brasil - Carta De Conjuntura - IPEA 2023**

HINTON, G. E., OSINDERO, S., & TEH, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, 18(7), 1527–1554. <https://doi.org/10.1162/neco.2006.18.7.1527>

HOFFMANN, Rodolfo, **Análise estatística de relações lineares e não lineares** [recurso eletrônico] / ESALQ - USP - Piracicaba: 2016. 246 p. : il. ISBN: 978-85-921057-1-6 - <https://www.livrosabertos.abcd.usp.br/portaldelivrosUSP/catalog/view/74/65/314>

IPSOS, **Global Trustworthiness Index 2022 - Who does the world trust ?;** <https://www.ipsos.com/en/global-trustworthiness-index-2022>

ISO, **Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability**, ISO, Tech. Rep. ISO 3534-1:2006, 2006. [Online]. Available: <https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standards/04/01/40145.html>

ISO, **Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability**, ISO, Tech. Rep. ISO 3534-1:2006, 2006. [Online].

JENKINS, Henry. **Cultura da convergência**. 2. ed. São Paulo: Aleph, 2009 <https://cienciaslinguagem.eca.usp.br/L3JenkinsConvergencia.pdf>

JUST, N. & LATZER, M. **Governance by Algorithms: Reality Construction by Algorithmic Selection on the Internet**. Media, Culture & Society 2016. https://mediachange.ch/media/pdf/publications/Just_Latzer2016_Governance_by_Algorithms_Reality_Construction.pdf

KADANO, Leo P. . **More is the Same; Phase Transitions and Mean Field Theories**. The James Franck Institute The University of Chicago email: leop@UChicago.edu September 14, 2009. <https://ui.adsabs.harvard.edu/abs/2009JSP...137..777K/abstract>

KELLNER, Douglas, **A Cultura da Mídia: Identidade Política Entre o Moderno e o Pós-moderno**, Editora Edusc:2001 - https://ufabcpoliticacultural.wordpress.com/wp-content/uploads/2015/08/kellner_a-cultura-da-mc3addia_2001.pdf

KERR A; BARRY M and KELLEHER J. D, **Expectations of artificial intelligence and the performativity of ethics: Implications for communication governance**, *Big Data & Society*, vol. 7, no. 1, p. 205395172091593, Jan. 2020. [Online]. Available: <http://journals.sagepub.com/doi/10.1177/2053951720915939>

KNORR-CETINA, Karin, and Andrew Pickering. **Science as practice and culture**. The Couch, the Cathedral, and the Laboratory: On the Relationship between Experiment and Laboratory in Science (1992): 113-38. Disponível em <https://d-nb.info/1103026992/34>

KNORR-CETINA, K. D. **A Comunicação na Ciência**. 1a. publicação in *A Ciência Tal Qual se Faz*. Lisboa. GIL, F. (Org.): 1999. - [Acessado: em https://kops.uni-konstanz.de/server/api/core/bitstreams/2c75d8f0-9b8e-452f-bcb7-a7a29f1cf643/content](https://kops.uni-konstanz.de/server/api/core/bitstreams/2c75d8f0-9b8e-452f-bcb7-a7a29f1cf643/content)

KOSLOWSKI, Adilson. **É o conceito de tecnociência confuso?** Revista Philosophos, v. 20, n. 1, 2015. - <https://revistas.ufg.br/philosophos/article/view/36115>

KRIPPENDORFF, Klaus. **Content analysis: An introduction to its methodology**. Los Angeles: SAGE, 2019. - https://www.daneshnamehicsa.ir/userfiles/files/1/9-%20Content%20Analysis_%20An%20Introduction%20to%20Its%20Methodology.pdf

LARS, Stig; SVANTE, Wold, **Analysis of variance (ANOVA)**, Chemometrics and Intelligent Laboratory Systems, Volume 6, Issue 4, 1989, Pages 259-272, ISSN 0169-7439 (<https://www.sciencedirect.com/science/article/pii/0169743989800954>)

LATOUR, Bruno. **Ciência em ação: como seguir cientistas e engenheiros sociedade afora**. Editora Unesp, 2000.

_____, Bruno. **A esperança de Pandora**. SciELO-Editora UNESP, 2017.

_____, **Reagregando o Social: uma introdução à teoria do Ator-Rede**. Edufba, 2012.

_____, **Jamais fomos modernos**. Editora 34, 1994.

_____, **Políticas da natureza: como associar as ciências à democracia**. Editora Unesp, 2019.

_____, **Onde Aterrar**. Bazar do Tempo, 2017

_____, Bruno e WOLGAR, Steve. **A vida de laboratório: a produção dos fatos científicos**. Rio de Janeiro: Relume Dumará. 1991

https://edisciplinas.usp.br/pluginfile.php/4882193/mod_resource/content/2/latourewoolgar_1997_a-vida-de-laboratorio-e28093-a-producao-dos-fatos-cientificos_book.pdf

_____, **Network theory: networks, societies, spheres: Reflections of an actor-network theorist**. International journal of communication, v. 5, p. 15, 2011. <https://ijoc.org/index.php/ijoc/article/view/1094/558>

LATOUR, B., Jensen, P., Venturini, T., Grauwin, S. and Boullier, D. (2012) 'The whole is always smaller than its parts' – a digital test of Gabriel Tarde's monads. British Journal of Sociology, 63 (4), 590 – 615

LAZER, David; PENTLAND, Alex; ADAMIC, Lada; Aral, Sinan; BARABÁSI, Albert-László; BREWER, Devon; CHRISTAKIS, Nicholas; CONTRACTOR, Noshir; FOWLER, James, [...], and ALSTYNE, Marshall Van. **Computational Social Science**. Science 323,721-723 (2009). DOI: [10.1126/science.1167742](https://doi.org/10.1126/science.1167742) - <https://www.science.org/doi/10.1126/science.1167742>

LÉVY, Pierre **Cibercultura**. São Paulo: Editora 34, 1999

_____, **O que é virtual?** Trad. Paulo Neves. São Paulo, Ed. 34, 1996.

_____, **A inteligência coletiva: por uma antropologia do ciberespaço**, 3. ed. São Paulo: Loyola, 2000.

LESKOVEC Jure, CS224W: **Machine Learning with Graphs**; Stanford University - 2021 <http://cs224w.stanford.edu>

LYOTARD, Jean-François, **O pós moderno explicado às crianças** (Le postmoderne expliqué aux enfants), Tradução Pedro Henrique Ciucci da Silva, 1986 http://revistapandorabrasil.com/revista_pandora/traducoes/lyotard.pdf

_____, **O Pós-Moderno**, 3a. edição, 1988, José Olympio Editora

LUHMANN, N. **A realidade dos meios de comunicação**. São Paulo: Paulus, 2005.

LUPTON, Deborah - Digital Sociology (2015), [S.l.], Routledge.
[https://edisciplinas.usp.br/pluginfile.php/4408475/mod_resource/content/1/Lupton-Digital%20Sociology\(2014\).pdf](https://edisciplinas.usp.br/pluginfile.php/4408475/mod_resource/content/1/Lupton-Digital%20Sociology(2014).pdf)

MAYRING, P. **Qualitative content analysis: Theoretical foundation, basic procedures and software solution**. Klagenfurt: [s.n.], 2014.

MCFARLAND, DA; LEWIS K , GOLDBERG, A . **Sociology in the era of big data: The ascent of forensic social science** - The American Sociologist, 2015 - Springer -
<https://www.gsb.stanford.edu/sites/default/files/publication-pdf/amsoc.pdf>

MERTON, R. K. (1973), **The sociology of science**. Chicago, The University of Chicago Press. Disponível
https://sciencepolicy.colorado.edu/students/envs_5110/merton_sociology_science.pdf

_____, **The Normative Structure of Science**, in: Merton, Robert K., [The Sociology of Science: Theoretical and Empirical Investigations](#), ISBN 978-0-226-52091-9, Chicago: University of Chicago Press, pp. 267-278, [1942],

_____, **Sociologia, Teoria e Estrutura**. São Paulo, Mestre Jou, 1970

_____, **The Matthew effect in science. The reward and communication systems of science are considered**. Science (New York, N.Y.), 159(810):56–63, January 1968.

MILLER, J. H.; PAGE, S. E. **Complex adaptive systems: an introduction to computational models of social life**. New Jersey: Princeton University Press, 2009.
https://www.researchgate.net/publication/275389038_Complex_Adaptive_Systems_An_Introduction_to_Computational_Models_of_Social_Life

MINSKY, M. (1991). **Logical versus analogical or symbolic versus connection or neat versus scruffy**. AI Magazine, 12(2), 34–51
<https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/894>

MITCHELL, Melanie. **Complexity: A Guided Tour**. [S.l.]: Oxford University Press, USA, 2009. Disponível em: <<http://www.librarything.com/isbn/0195124413>>

MONTEIRO, Marko Synésio Alves. **Reconsiderando a etnografia da ciência e da tecnologia: tecnociência na prática**. Revista Brasileira de Ciências Sociais, 2012.
<https://doi.org/10.1590/S0102-69092012000200009>

MONTGOMERY, Douglas C. , PECK, Elizabeth A, VINING, G. Geoffrey - **Introduction to linear regression analysis** – 5a. edição - ISBN 978-0-470-54281-1
https://www.kwcsangli.in/uploads/3--Introduction_to_Linear_Regression_Analysis_5th_ed_Douglas_C_Montgomery_Elizabeth_A_Peck_and_G_.pdf

MOROZOV, E. **Big Tech: a ascensão dos dados e a morte da política**. São Paulo: Ubu, 2018. 189 p. ISBN 978 85 7126 012 2

NEUENDORF, K. **The content analysis guidebook**. Londres: Sage, 2002 - <https://www.daneshnamehicsa.ir/userfiles/files/1/9-%20The%20Content%20Analysis%20Guidebook.pdf>

Schwartz, R. , Vassilev, A. , Greene, K. , Perine, L. , Burt, A. and Hall, P. (2022), **Towards a Standard for Identifying and Managing Bias in Artificial Intelligence**, Special Publication (NIST SP), National Institute of Standards and Technology, Gaithersburg, MD, [online], <https://doi.org/10.6028/NIST.SP.1270>,

NIST - National Institute of Standards and Technology. **Face Recognition Technology Evaluation**

NONAKA, Ikujiro. TAKEUCHI, Hirotaka **Gestão do Conhecimento** eBook Kindle - 2008

NORDMANN, Alfred. **Science in the context of technology**. Boston: Springer, 2011, p. 467-482. <https://philpapers.org/rec/CARSIT-3>

OLIVEIRA, Rute Melo de. **Ligações preferenciais em redes complexas: modelo de desafinidade** / Rute Melo de Oliveira. - Natal, 2019

O'NEIL, Cathy. **Algoritmos de Destruição em Massa**. Editora Rua do Sabão, 2020

ORESQUES, Naomi & CONWAY, Erik M. **Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues From Tobacco Smoke to Global Warming**. Bloomsbury Press. (2010)

PAGE, Larry e BRIN, Sergey, **The PageRank Citation Ranking: Bringing Order to the Web** January 29, 1998, <http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf>

PARK, Patrick S; BLUMENSTOCK, Joshua E. ; MACY, Michael W. **The strength of long-range ties in population-scale social networks**. SCIENCE, 2018, Vol 362, Issue 6421, pp. 1410-1413, DOI: 10.1126/science.aau9735 https://www.jblumenstock.com/files/papers/jblumenstock_2018_science.pdf

PARISER, Eli. **O filtro invisível: o que a internet está escondendo de você**. Rio de Janeiro: Zahar, 2012. https://ciberativismoeguerra.files.wordpress.com/2017/02/filtro-invisivel_eli-pariser.pdf -

PAVESI, P. P.; VALENTIM, J. **Ciências Sociais Computacionais: um novo paradigma para as Ciências Sociais?**. Simbiótica. Revista Eletrônica, [S. l.], v. 8, n. 4, p. 1–16, 2021. DOI: 10.47456/simbitica.v8i4.37344. <https://periodicos.ufes.br/simbiotica/article/view/37344>

PENTLAND A. **Social Physics: How Social Networks Can Make Us Smarter**. New York, Penguin Books 2015 - https://toc.library.ethz.ch/objects/pdf_ead50/4/E50_010066771TB-Inhalt_004891462.pdf

PORTO, CM.

_____, **Difusão e cultura científica: alguns recortes** [online]. Salvador: EDUFBA, 2009. 230 p. ISBN 978-85-2320-912-4. Available from SciELO Books .

_____, **Produção e difusão de ciência na cibercultura: narrativas em múltiplos olhares** / Cristiane Porto, Kaio Eduardo Oliveira, Flávia Rosa (organizadores). Editus, 2018. <http://www.labi.ufscar.br/wp-content/uploads/2019/01/Anexo-9.pdf>

PRENSKY, M. **Nativos digitais, imigrantes digitais**. 2001. Disponível em: <https://www.marcprensky.com/writing/Prensky%20-%20Digital%20Natives,%20D>

PRIGOGINE, Ilya, **O Fim das Certezas - Tempo, Caos e Leis da Natureza** (Editora UNESP, São Paulo, 1996).

PRZEYBILOVICZ, Erico. **Governando iniciativas de cidade inteligente: compreendendo os arranjos de governança sócio-técnica e o uso de tecnologias da informação nos casos de Curitiba e São Paulo** / Erico Przebylłowicz - 2019. <https://bibliotecadigital.fgv.br/dspace/handle/10438/27258>

PROCTOR, R. Schiebinger, L - *Agnology: The Making and Unmaking of Ignorance*. 2008. <https://philpapers.org/archive/PROATM.pdf>

REIS, Verusca Moss Simões dos e Videira, Antonio Augusto Passos, **John Ziman e a ciência pós-acadêmica: consensibilidade, consensualidade e confiabilidade**. Scientiae Studia <<https://doi.org/10.1590/S1678-31662013000300007>>. ISSN 2316-8994. 2013

SALA, Arianna, LIZARRAGA, Aldana, CAMINITI, Silvia Paola, CALHOUN, Vince D., EICKHOFF, Simon B. , HABECK, Christian, JAMADAR, Sharna D., PERANI, PEREIRA, Joana B., VERONESE, Mattia , YAKUSHEV, Igor, **Brain connectomics: time for a molecular imaging perspective?**, Trends in Cognitive Sciences, Volume 27, Issue 4, 2023. (<https://www.sciencedirect.com/science/article/pii/S136466132200300X>)

SAMPAIO, Rafael Cardoso; LYCARIÃO, Diógenes - **Análise de conteúdo categorial: manual de aplicação** -- Brasília: Enap, 2021. 155 p - ISBN: 978-65-87791-18-0

SANTAELLA, Lúcia, **Gêneros discursivos híbridos na era da hipermídia** / Hybrid Discursive Genres in the Hypermedia Era, Bakhtiniana, São Paulo, 9 (2): 206-216, Ago./Dez. 2014. <http://www.scielo.br/pdf/bak/v9n2/a13v9n2.pdf>

SANTOS, Davilene Souza; SOUZA, Adriana Serafim de; ROSA, Flávia Goulart Mota Garcia. **A comunicação científica das práticas interdisciplinares em processos de aprendizagem: uma análise entre 2010-2020**. Informação em Pauta, Fortaleza, v. 5, n. 2, p.56-70, jul./dez. 2020. DOI: <https://doi.org/10.36517/2525-3468.ip.v5i2.2020.44072.56-70>.

SANTOS 2, J. C. F. dos. **Da sociedade da informação e do conhecimento à era dos dados**: perspectivas interdisciplinares contemporâneas das áreas jurídicas e da ciência da informação. Revista do Instituto de Direito Constitucional e Cidadania, [S. l.], v. 7, n. 2, p. e062, 2023. DOI: [10.48159/revistadoidcc.v7n2.e062](https://doi.org/10.48159/revistadoidcc.v7n2.e062).

SEJNOWSKI, Terry, **A revolução do Aprendizado Profundo**. Alta Books; 1ª edição (2020) ISBN 8550807885

SHANNON, C.E. (1948) **A Mathematical Theory of Communication**. Bell System Technical Journal, 27, 379-423.-

<https://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf>

SILVA, Joao et al. **From symbolic to sub-symbolic information in question classification**. Artificial Intelligence Review, v. 35, n. 2, p. 137-154, 2011. <https://link.springer.com/content/pdf/10.1007/s10462-010-9188-4.pdf>

SILVA, Juliana Mendes N. **Redes Neurais Artificiais: Rede Hopfield e Redes Estocásticas**. Orientador: Lúcia Drummond e Roseli Suzi Wedemann, v. 40, 2003.

SISMONDO, Sergio, **An Introduction to Science and Technology Studies**, 1993, https://www.academia.edu/40817422/An_Introduction_to_Science_and_Technology_Studies_Second_Edition

SOARES, Kleber Toldo. **Ciência de Dados Sociais: uma visão sociológica das Humanidades Digitais**. Simbiótica. Revista Eletrônica, [S. l.], v. 8, n. 4, p. 46–70, 2021. DOI: [10.47456/simbitica.v8i4.37346](https://doi.org/10.47456/simbitica.v8i4.37346).

SOUSA, Cidoval Moraes de, **Comunicação Midiática, Ensino de Ciências e Sustentabilidade**, Revista Ciências Humanas - Educação e Desenvolvimento Humano, v. 10, n. 1, edição 18, p. 46 - 55, 2017. <https://www.rchunitau.com.br/index.php/rch/article/view/348/213>

MARTINS, Osvaldo Stella. **Determinação do potencial de sequestro de carbono na recuperação de matas ciliares na região de São Carlos- SP**. 2004. Tese em https://cetesb.sp.gov.br/wp-content/uploads/sites/36/2014/04/martins_determinacao_potencial_2005.pdf

VARSHNEY LR, CHEN BL, PANIAGUA E, HALL DH, CHKLOVSKII DB. **Structural properties of the Caenorhabditis elegans neuronal network**. PLoS Comput Biol. 2011 Feb 3;7(2):e1001066. doi: [10.1371/journal.pcbi.1001066](https://doi.org/10.1371/journal.pcbi.1001066). PMID: 21304930; PMCID: PMC3033362.

VESSURI, Hebe. **Ciencia, tecnología y desarrollo: una experiencia de apropiación social del conocimiento**. INCI, Caracas , v. 27, n. 2, p. 88-92, feb. 2002 . Disponible <http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S0378-18442002000200010&lng=es&nrm=iso>

VOGT C., GOMES, MUNIZ, **ComCiência e divulgação científica** / Carlos Vogt, Marina Gomes, Ricardo Muniz (Organizadores). – Campinas, SP: BCCL/ UNICAMP, 2018.

VOGT, C.; MORALES, A.P. **Espiral da cultura científica” em O discurso dos indicadores de C&T e de percepção de C&T**. Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura, 2016

VOGT, Carlos. **Cultura Científica: Desafios**. São Paulo: Editora da Universidade de São Paulo, 2006.

WEAVER, W. Science And Complexity. American Scientist, 36(4), 536–544. - (1948)
<http://www.jstor.org/stable/27826254>

WEAVER, W. **A teoria matemática da comunicação**. COHN, Gabriel (Org.). *Comunicação e indústria cultural*. 3. ed. São Paulo: Nacional, 1977. [S.l.: s.n.]

WHITE JG, SOUTHGATE E, THOMSON JN, BRENNER S., **The structure of the nervous system of the nematode *Caenorhabditis elegans***. 1986 Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences. 314 (1165): 1–340.
 Bibcode:1986RSPTB.314....1W. PMID 22462104. doi:10.1098/rstb.1986.0056

WIDROW, B; M. A. LEHR M. A., **30 years of adaptive neural networks: perceptron, Madaline, and backpropagation** in *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1415-1442, Sept. 1990, doi: 10.1109/5.58323. - <https://ieeexplore.ieee.org/document/58323>

YAMAMOTO, W. S., & ACHACOSO, T. B., **Scaling up the nervous system of *Caenorhabditis elegans*: is one ape equal to 33 million worms?**. (1992) Computers and biomedical research, an international journal, 25(3), 279–291.
[https://doi.org/10.1016/0010-4809\(92\)90043-a](https://doi.org/10.1016/0010-4809(92)90043-a)

YAN, G., VÉRTES, P., TOWLSON, E., BARABÁSI Albert-László, et al. **Network control principles predict neuron function in the *Caenorhabditis elegans* connectome**. Nature 550, 519–523 (2017). <https://doi.org/10.1038/nature24056>

ZEVIANI, Walmes Marques; RIBEIRO JÚNIOR, Paulo Justiniano; BONAT, Wagner Hugo - Curso: **Modelos de regressão não linear** - Laboratório de Estatística e Geoinformação Departamento de Estatística Universidade Federal do Paraná, 22 à 26 de julho de 2013 -
<https://www.ime.unicamp.br/~cnaber/cursomodelosnaolinearesR.pdf>

ZIMAN, John M. **An introduction to science studies: The philosophical and social aspects of science and technology**. Cambridge University Press, 1987.

_____, J. **Conhecimento público: a dimensão social da ciência** Tradução R. R. Junqueira. São Paulo: Edusp, 1979 [1968]

_____, J. (2000). **New modes of knowledge production. In *Real Science: What it Is and What it Means*** (pp. 56-82). Cambridge: Cambridge University Press.
 doi:10.1017/CBO9780511541391.005

ZUIN, V. G.; ZUIN, A. Álvaro S. **A indústria cultural algorítmica na era da Internet das Coisas**. Educação e Filosofia, Uberlândia, v. 32, n. 66, p. 1131–1156, 2018. DOI: 10.14393/REVEDFIL.issn.0102-6801.v32n66a2018-07.
<https://seer.ufu.br/index.php/EducacaoFilosofia/article/view/41478>

Rede de Apoio aos Professores de Língua Estrangeira

Análises de Conteúdos

Análises e Resumos por modelos de IA: Gemini Advanced e MAXQDA

Transcrição da live de Ana Maria Barcelos: Emoções de Professores de Línguas



Transcrição Online do vídeo
MAXQDA



Arquivo PDF com transcrição da live

Clique para abrir arquivo em nova aba do navegador



Transcrição: Nuvem de Palavras e Tabela de Frequência
MAXQDA



Mostrar 10 registros

Pesquisar:

Palavra	Comprimento da palavra	Frequência	%	Ranking
emoção	6	115	3.42	1
ser	3	92	2.74	2
gente	5	79	2.35	3
professorar	11	66	1.97	4
ai	2	58	1.73	5
estar	5	52	1.55	6
aqui	4	44	1.31	7
achar	5	37	1.10	8
trabalhar	9	32	0.95	9
língua	6	29	0.86	10

Editar

A mostrar 1 a 10 de 277 registros



Transcrição: Análise e resumo IA
Gemini Advanced



Quais os principais assuntos da transcrição ?

A transcrição do live discute principalmente as emoções no ensino de línguas. A professora Ana Maria, especialista em Linguística Aplicada, aborda a importância das emoções na construção da identidade profissional de professores e alunos, e como elas influenciam suas decisões e ações dentro e fora da sala de aula.

A apresentação destaca a necessidade de reconhecer e discutir as emoções no contexto da formação de professores, defendendo uma abordagem pedagógica que promova um clima de segurança e confiança para que professores e alunos possam expressar e refletir sobre suas emoções.

Além disso, a transcrição oferece um panorama histórico sobre o estudo das emoções na área de Linguística Aplicada, mencionando trabalhos e pesquisas relevantes no Brasil e no exterior. Também são discutidos conceitos-chave como o “trabalho emocional”, que destaca as demandas emocionais da profissão docente, e a relação entre emoções, crenças, motivações e identidades.

Em suma, a transcrição gira em torno da importância das emoções no ensino e aprendizagem de línguas, enfatizando a necessidade de uma abordagem mais humana e compreensiva nesse processo.



Pode fazer um resumo detalhado ?



Análise do discurso



Transcrição: Análise e resumo IA
MAXQDA



Resumo detalhado da transcrição

O texto apresenta uma discussão sobre as emoções de professores de línguas estrangeiras. Ele começa com uma breve introdução sobre o projeto “Rede de Apoio a Professores de Línguas Estrangeiras” e seus objetivos de criar uma conexão entre professores. 1-20

Em seguida, a professora Ana Maria é apresentada e sua extensa experiência na área de linguística aplicada, com ênfase em crenças, emoções de professores e alunos, e pedagogia amorosa. 61-90

O foco principal do texto é explorar o que se sabe sobre as emoções de professores de línguas. É destacado que as emoções moldam as identidades profissionais dos professores, influenciam suas decisões e a qualidade da interação com os alunos. 159-171 Elas têm raízes na combinação de sentimento e percepção, e o ensino é visto como uma atividade emocional. 183-205

São apresentadas algumas emoções comuns entre professores, como alegria, entusiasmo, satisfação, mas também tristeza, raiva, ansiedade. 218-230 É enfatizado que o equilíbrio entre sentir, pensar e agir é essencial para o ensino efetivo. 231-244

O texto também traz um breve histórico da pesquisa sobre emoções na linguística aplicada, desde os anos 70 até os dias atuais. 360-440 São mencionados alguns grupos de pesquisa brasileiros que trabalham com o tema das emoções de professores. 446-465

Por fim, são apresentados alguns estudos realizados no programa de pós-graduação da Universidade Federal de Viçosa sobre emoções de professores de línguas, suas metodologias e principais achados. 469-550



Quais os principais assuntos abordados ?

Não houve comentários ao video no Youtube

Análise do chat no Youtube



Tabela de frequência e nuvem de palavras do chat
MAXQDA



Mostrar registros

Pesquisar:

CHAT	INICIAIS	EMOTICON
Estou cada dia mais fã da RAPLE	IC	
Ana Maria é Show!!	IC	
verdade	EV	
siim	EV	
Eu a turma da Let 615 rsrs	IC	
que lindo	EV	
OK	AC	
8:55Fernanda Ricardo CamposBoa tarde!	FRC	
FRC ...que bom que deu para você vir...Um beijo:)	IC	
verdade	EV	

Editar

A mostrar 1 a 10 de 107 registros

<

1

2

3

4

5

...

11

>

 Análise de discurso no chat

O discurso presente no chat da live sobre emoções no ensino de línguas é marcado por uma atmosfera de **entusiasmo, engajamento e colaboração** entre os participantes. Algumas características notáveis do discurso incluem:

- **Informalidade e emotividade:** A linguagem utilizada é informal, com abreviações, emoticons e expressões coloquiais, refletindo a natureza espontânea e interativa do chat. Os participantes expressam suas emoções de forma aberta, utilizando exclamações e adjetivos como “sensacional” e “maravilhosa” para demonstrar seu entusiasmo pela palestra.
- **Foco na interação:** O chat é um espaço de diálogo e troca de ideias, com os participantes interagindo diretamente com a palestrante e entre si. As perguntas, comentários e sugestões demonstram um interesse genuíno pelo tema e um desejo de aprofundar a discussão.
- **Comunidade de prática:** O chat evidencia a formação de uma comunidade de prática, com os participantes compartilhando suas experiências, dúvidas e perspectivas sobre o tema. A troca de informações e o apoio mútuo criam um senso de pertencimento e colaboração.
- **Valorização da pesquisa:** Os participantes demonstram apreço pela pesquisa acadêmica, reconhecendo a importância dos estudos sobre emoções no ensino de línguas. As perguntas sobre pesquisas futuras e indicações de leitura revelam um interesse em aprofundar seus conhecimentos sobre o tema.
- **Busca por soluções práticas:** Além do interesse teórico, os participantes também buscam soluções práticas para lidar com as emoções em sala de aula. As perguntas sobre o impacto da pandemia e as sugestões para futuras lives evidenciam a necessidade de ferramentas e estratégias para lidar com os desafios emocionais do contexto educacional atual.

Em suma, o discurso no chat reflete um ambiente de aprendizado colaborativo e engajado, onde os participantes se sentem à vontade para expressar suas opiniões, compartilhar suas experiências e buscar soluções para os desafios do ensino de línguas. A atmosfera positiva e o foco na interação contribuem para a construção de uma comunidade de prática que valoriza a pesquisa e busca aprimorar sua atuação profissional.

 Quais os principais assuntos abordados ?

 Análise de discurso no chat

 Qual a temática do chat ?

A live apresentada pela professora Ana Maria foi muito elogiada pelos participantes, que destacaram a excelência da sua apresentação sobre o tema “Afeto na aprendizagem de línguas” de Jane Arnold.

Os participantes fizeram perguntas relevantes, como o impacto da pandemia nas emoções de professores e alunos, e a relação entre os estudos sobre emoções e a perspectiva neoliberal na educação refletida em documentos como a BNCC.

A professora respondeu de forma satisfatória, sugerindo a necessidade de mais pesquisas e lives sobre como lidar com as emoções em sala de aula, tanto de professores quanto de alunos.

Os participantes agradeceram a live e demonstraram interesse em receber indicações de livros sobre o assunto

1: CHAT	2: INICIAIS
Estou cada dia mais fã da RAPLE	IC
Ana Maria é Show!!	IC
verdade	EV
siim	EV
Eu a turma da Let 615 rsrs	IC
que lindo	EV
OK	AC
FRC ...que bom que deu para você vir...Um beijo.)	IC
verdade	EV
Sim, acontece mesmo	TC

Editar

Análises e Resumos por modelos de IA: Gemini Advanced e MAXQDA
Transcrição da live de Henrique Rodrigues Leroy: "Corazonar, sentipensar, sulear e
esperançar em língua portuguesa adicional e materna"



Transcrição Online do vídeo
MAXQDA



Arquivo PDF com transcrição da live

As palavras e dinâmica da live são analisadas através da transcrição, gerando nuvem de palavras, tabelas de frequência e dimensão, análise semântica, de sentimentos e discurso através de programação em Python e IA generativa via Gemini Advanced e MAXQDA.





Tabela de frequência e nuvem de palavras da live
MAXQDA



Mostrar 10 registros

Pesquisar:

Palavra	Comprimento da palavra	Frequência	%	Ranking
ser	3	660	8.27	1
né	2	532	6.67	2
gente	5	320	4.01	3
aqui	4	113	1.42	4
estar	5	92	1.15	5
aí	2	91	1.14	6
trazer	6	90	1.13	7
falar	5	83	1.04	8
falir	5	67	0.84	9
idear	5	65	0.81	10

Editar



Transcrição: Análise e resumo IA
Gemini Advanced



Transcrição: Análise e resumo IA
MAXQDA

 Quais os principais assuntos da transcrição ?

 Pode fazer um resumo detalhado ?

 O que o texto fala sobre descolonização ?

O texto discute a **descolonização** como um processo crucial na educação e na linguagem. Enfatiza a necessidade de **desconstruir** as ideias coloniais que perpetuam hierarquias e desigualdades, defendendo a **valorização da diversidade** e do conhecimento de comunidades marginalizadas, como as indígenas e quilombolas.

A descolonização é apresentada como uma forma de **romper com o pensamento eurocêntrico**, que coloca a Europa como centro do conhecimento e da cultura, e de **abraçar a pluralidade de saberes e cosmovisões**. Isso implica em questionar as relações de poder existentes e promover uma educação mais justa e inclusiva.

No contexto do ensino de línguas, a descolonização envolve a **valorização das diferentes variedades linguísticas** e o reconhecimento da influência de outras línguas, como as africanas e indígenas, na formação do português brasileiro. O palestrante defende a importância de **trazer para a sala de aula** essa diversidade, utilizando materiais autênticos e abordagens que promovam o respeito e a inclusão.

Em suma, a descolonização é apresentada como um processo contínuo de **questionamento e transformação**, que busca romper com as estruturas de poder coloniais e promover uma educação e uma linguagem mais justas, inclusivas e representativas da diversidade do mundo.

 Resumo detalhado da transcrição

O texto começa apresentando a rede de apoio à professores de línguas estrangeiras, seus objetivos de troca de experiências e materiais, e convidando os participantes a seguir nas redes sociais. Em seguida, a palestrante agradece a sua equipe e colaboradores que tornam possível a realização dessas atividades.

Então, ela apresenta o convidado da live, o professor Henrique Leroy, que é professor adjunto da UFMG na área de linguística aplicada e trabalha com temas como o giro decolonial, práticas translingues e transculturais na formação de professores de língua portuguesa.

O professor Leroy então inicia sua fala explicando os conceitos de colonialismo, colonialidade e decolonialidade 176-225. Ele detalha que a colonialidade é uma estrutura maior que abrange dimensões como a colonialidade do poder, do ser, do saber e de gênero 239-429.

Ao abordar a colonialidade da linguagem, Leroy enfatiza a necessidade de desconstruir a visão eurocêntrica e hegemônica da língua, trazendo para a sala de aula outras variedades linguísticas, como as línguas indígenas, africanas e de comunidades quilombolas 907-1135. Ele destaca a importância de uma pedagogia decolonial que valorize os saberes e letramentos outros 1136-1215.

O professor também discute a questão da neutralidade e imparcialidade nos discursos acadêmicos, argumentando que todo discurso carrega a subjetividade de quem o produz 1948-1999. Nesse sentido, ele apresenta conceitos como "sentir-pensar" e "sulear", que buscam romper com a racionalidade ocidental 1983-2003.

Por fim, Leroy indica algumas referências bibliográficas importantes para aprofundar os temas abordados 2080-2125, e comenta sobre a fala de Paulo Freire sobre a criança já ler o mundo antes de chegar à escola 2150-2207.

Análise dos comentários do video no Youtube



Tabela de frequência e nuvem de palavras dos comentários
MAXQDA



Mostrar 10 registros

Pesquisar:

Palavra	Comprimento da palavra	Frequência	%	Ranking
ser	3	37	5.79	1
língua	6	14	2.19	2
estar	5	9	1.41	3
yes	3	9	1.41	3
colonial	8	7	1.10	5
dizer	5	5	0.78	6
cultura	7	4	0.63	7
das	3	4	0.63	7
leroy	5	4	0.63	7
live	4	4	0.63	7

Editar

A mostrar 1 a 10 de 429 registros

<

1

2

3

4

5

...

43

>



Comentários: Análise e resumo IA
Gemini Advanced

 Resumo detalhado dos comentários

O texto discute a importância da linguagem e da descolonização do pensamento, abordando como as línguas coloniais são usadas como ferramentas de dominação e a internalização da opressão colonial através da adoção dessas línguas.

Ele destaca a necessidade de valorizar e revitalizar as línguas indígenas e locais, promovendo uma mudança de pensamento nas universidades em relação à ética da palavra, apesar dos desafios enfrentados em regiões marcadas pela colonização.

O texto conclui com um apelo por mais pessoas engajadas na luta pela descolonialidade da linguagem e do pensamento.

 Quais os principais assuntos abordados ?



Comentários: Análise e resumo IA
MAXQDA

 Resumo detalhado dos comentários

O texto apresenta uma discussão sobre a importância da linguagem e da descolonização do pensamento. Ele aborda a obra de Ngũgĩ wa Thiong'o, que discute como as línguas coloniais são usadas como ferramentas de dominação, e a obra de Frantz Fanon, que analisa as formas pelas quais a adoção de línguas coloniais pode levar à internalização da opressão colonial.

O texto também menciona a obra de Achille Mbembe, que examina criticamente como a violência colonial se manifestou através da linguagem. Há uma reflexão sobre a imposição de línguas coloniais e a supressão de línguas africanas, bem como a necessidade de valorizar e revitalizar as línguas indígenas e locais.

Aborda ainda a questão da mudança de pensamento e postura dentro das universidades em relação à ética da palavra, destacando os desafios enfrentados nesse processo, especialmente em regiões marcadas pela colonização. Há um apelo para que haja mais pessoas engajadas nessa luta pela descolonialidade da linguagem e do pensamento.

Mostrar 2 registros

Pesquisar:

Iniciais	2: Comentário	8: Publicado em
ME	Feliz de poder acompanhar!	29/04/2022 20:13
FE	Dois palhaços que se dizem "letrados" em Língua Portuguesa, todos é "meuzovo". Querem que preparemos as pessoas para prestarem concurso e vestibulares dessa forma, se recolham às suas bolhas, rasas, medíocres, sorrateiras. Nada de novo, afinal é o modus operandi dos mamadores do Estado!	31/05/2024 23:18

Editar

A mostrar 1 a 2 de 10 registros

<

1

2

3

4

5

>



Arquivo PDF com transcrição da live

Clique para abrir arquivo em nova aba do navegador



Mostrar registros

Pesquisar:

Palavra	Comprimento da palavra	Frequência	%	Ranking
gente	5	118	3.12	1
estar	5	61	1.62	2
cultura	7	57	1.51	3
aqui	4	55	1.46	4
lingua	6	54	1.43	5
falar	5	51	1.35	6
ai	2	43	1.14	7
pessoa	6	43	1.14	7
ir	2	41	1.09	9
formar	6	40	1.06	10

Editar

A mostrar 1 a 10 de 310 registros

<

1

2

3

4

5

...

31

>

 Quais os principais assuntos da transcrição ?

A transcrição aborda principalmente os seguintes temas:

- **A importância da cultura na interação em língua estrangeira:** O texto enfatiza que a cultura é um elemento fundamental na comunicação em língua estrangeira, influenciando tanto a produção quanto a interpretação de mensagens.
- **O conceito de “sotaque cultural”:** O texto introduz o termo “sotaque cultural” para descrever as diferenças culturais que podem levar a mal-entendidos na interação em língua estrangeira.
- **A necessidade de sensibilizar os alunos para as diferenças culturais:** O texto argumenta que é crucial que os professores de línguas estrangeiras sensibilizem seus alunos para a possibilidade de diferenças culturais na interação, a fim de evitar mal-entendidos e promover uma comunicação eficaz.
- **O uso de vídeos e outros materiais autênticos para ensinar cultura:** O texto sugere o uso de vídeos, charges e outros materiais autênticos para explorar a perspectiva intercultural em sala de aula, proporcionando aos alunos exemplos concretos de como a cultura se manifesta na comunicação.
- **A importância de evitar estereótipos culturais:** O texto destaca a importância de não essencializar ou estereotipar culturas, reconhecendo a diversidade e a complexidade das formações socioculturais.
- **A aprendizagem intercultural como uma oportunidade de autoconhecimento:** O texto conclui que aprender sobre outras culturas através do estudo de línguas estrangeiras também nos permite aprender mais sobre nós mesmos e nossa própria cultura.

Em resumo, a transcrição discute a importância da cultura na comunicação em língua estrangeira, destacando a necessidade de sensibilizar os alunos para as diferenças culturais e evitar estereótipos, a fim de promover uma comunicação intercultural eficaz e reflexiva.

 Resumo detalhado da transcrição

O professor Nelson apresentou uma discussão abrangente sobre a importância da cultura e da interculturalidade no ensino de línguas estrangeiras. Ele destacou que a cultura permeia e envolve os processos de ensinar, aprender e usar línguas, sendo um constituinte relevante. O conceito de cultura é complexo e multifacetado, o que se reflete em diferentes desdobramentos na prática pedagógica.

Nelson enfatizou a necessidade de sensibilizar aprendizes e professores quanto à constituição sociocultural dos processos de interação e comunicação. Ele explicou que, na interação em língua estrangeira, há a produção de formas simbólicas a partir de uma perspectiva sociocultural, que podem ser interpretadas de maneira distinta pelo interlocutor, devido a suas próprias estruturas sociais. Isso pode gerar mal-entendidos e ruídos na comunicação.

O professor apresentou diversos exemplos ilustrativos dessa dinâmica, como o caso da estudante brasileira na Argentina que utilizava formas de tratamento inadequadas, e o funcionário irlandês que cumprimentava diariamente seus colegas de trabalho com um aperto de mão, causando estranhamento. Esses episódios revelam o que Nelson denominou de “sotaque cultural”, em oposição ao “sotaque linguístico”.

Diante dessa complexidade, Nelson argumentou que não é possível ensinar todas as possíveis circunstancialidades da interação intercultural. Porém, é fundamental sensibilizar os aprendizes e professores para essa realidade, evitando estereótipos e essencializações. Nesse sentido, ele sugeriu o uso de materiais autênticos, como propagandas, vídeos humorísticos e produções artísticas, que permitam a discussão de aspectos culturais e interculturais.

Por fim, Nelson destacou a importância de não apenas conhecer o outro, mas também de se conhecer melhor a si mesmo por meio do olhar do outro. Essa perspectiva reflexiva é essencial no ensino e aprendizagem de línguas estrangeiras, pois permite uma compreensão mais ampla e profunda das dinâmicas culturais envolvidas.

Análise dos comentários do video no Youtube





Comentários: Análise e resumo IA
Gemini Advanced

 Quais sentimentos estão presentes nos comentários ?

Os comentários expressam os seguintes sentimentos:

1. Entusiasmo e admiração:

“Que bacana essa iniciativa!” demonstra entusiasmo pela iniciativa da RAPLE.

“Nelson é um encanto :)” expressa admiração pela pessoa mencionada.

2. Desejo de participação:

“Queria ter visto ao vivo!!” e “Quero fazer parte desta rede!!!!” indicam um forte desejo de se envolver e participar da rede e seus eventos.

3. Apreciação e gratidão:

“profunda teórica e analiticamente” demonstra apreciação pela qualidade da apresentação de Nelson.

“Obrigada, Nelson, Cibebe e equipe” expressa gratidão às pessoas envolvidas.

4. Acolhimento e reciprocidade:

“é um prazer ter você conosco em nossa rede, Idalena!” expressa o acolhimento à nova integrante da rede.

“A RAPLE te agradece, Nildi!” e “obrigada, Nildi querida!!” demonstram reciprocidade à apreciação expressada por Nildi.

Em resumo, os comentários refletem uma atmosfera positiva, com sentimentos de entusiasmo, admiração, desejo de participação, apreciação e gratidão sendo compartilhados entre os participantes da rede e seus eventos.

<https://g.co/gemini/share/3f08d2472ff3>

 Resumo do conteúdo



Comentários: Análise e resumo IA
MAXQDA

 Quais sentimentos estão presentes nos comentários ?

Os comentários expressam os seguintes sentimentos:

1. Entusiasmo e admiração:

“Que bacana essa iniciativa!” demonstra entusiasmo pela iniciativa da RAPLE.

“Nelson é um encanto :)” expressa admiração pela pessoa mencionada.

2. Desejo de participação:

“Queria ter visto ao vivo!!” e “Quero fazer parte desta rede!!!!” indicam um forte desejo de se envolver e participar da rede e seus eventos.

3. Apreciação e gratidão:

“profunda teórica e analiticamente” demonstra apreciação pela qualidade da apresentação de Nelson.

“Obrigada, Nelson, Cibebe e equipe” expressa gratidão às pessoas envolvidas.

4. Acolhimento e reciprocidade:

“é um prazer ter você conosco em nossa rede, Idalena!” expressa o acolhimento à nova integrante da rede.

“A RAPLE te agradece, Nildi!” e “obrigada, Nildi querida!!” demonstram reciprocidade à apreciação expressada por Nildi.

Em resumo, os comentários refletem uma atmosfera positiva, com sentimentos de entusiasmo, admiração, desejo de participação, apreciação e gratidão sendo compartilhados entre os participantes da rede e seus eventos.

<https://g.co/gemini/share/3f08d2472ff3>

 Resumo do conteúdo

Mostrar 10 registros

Pesquisar:

2: Comentário	3: Autor	8: Publicado em
Que bacana essa iniciativa da RAPLE!!! Queria ter visto ao vivo!! Nelson é um encanto :) Quero fazer parte desta rede!!!!	@IC	29/08/2021 23:06
é um prazer ter você conosco em nossa rede, Idalena!	@CR	29/10/2021 20:27
Estou vendo novamente a live do Nelson e verifico como é profunda teórica e analiticamente sua fala. Obrigada, Nelson, Cibebe e equipe.	@NR	30/03/2023 00:56
A RAPLE te agradece, Nildi!	@MP	19/09/2023 17:54
obrigada, Nildi querida!!	@CR	26/11/2023 14:08

Editar

Análise do chat durante a live



Tabela de frequência e nuvem de palavras do chat ao vivo



Mostrar registros

Pesquisar:

Palavra	Comprimento da palavra	Frequência	%	Ranking
bom	3	6	2.87	1
db	2	6	2.87	1
cn	2	4	1.91	3
ouvir	5	4	1.91	3
palestrar	9	4	1.91	3
raple	5	4	1.91	3
vp	2	4	1.91	3
boa	4	3	1.44	8
ir	2	3	1.44	8
língua	6	3	1.44	8

Editar

A mostrar 1 a 10 de 149 registros

<

1

2

3

4

5

...

15

>



Chat: Análise e resumo IA

Gemini Advanced

 Qual o conteúdo do chat ?

O texto parece ser a transcrição de um bate-papo online (chat) relacionado a uma palestra ou apresentação conduzida pelo Professor Nelson Viana. O tema central gira em torno da **interculturalidade**, com discussões sobre como as pessoas se veem e se entendem em diferentes contextos culturais.

Alguns pontos específicos mencionados incluem:

- A dificuldade de enxergar o outro por completo devido às suas múltiplas influências culturais.
- A transversalidade cultural no contexto brasileiro, reconhecendo tanto características comuns quanto particularidades individuais.
- O papel da imersão na revelação de aspectos interculturais.
- A tendência de usar rótulos como "brasileiro" ou "russo" para explicar comportamentos, simplificando a complexidade individual.

A palestra parece ter sido bem recebida, com muitos elogios e agradecimentos ao Professor Nelson e aos organizadores do evento.

 Resumo do conteúdo

O professor Nelson Viana conduziu uma palestra enriquecedora sobre a importância de se considerar os aspectos interculturais no ensino de línguas estrangeiras. Ele destacou que muitas vezes as pessoas tendem a generalizar e rotular comportamentos como sendo "típicos" de determinada nacionalidade, sem levar em conta a diversidade cultural existente dentro de um mesmo país.

O professor enfatizou a necessidade de os futuros professores de línguas estrangeiras irem além do ensino meramente teórico do idioma, buscando abordar elementos culturais que permitam aos alunos desenvolver um olhar mais amplo e diversificado. Ele sugeriu que a imersão no contexto da língua-alvo pode ser uma estratégia valiosa para revelar esses aspectos interculturais.

Ao longo da palestra, o professor Nelson Viana compartilhou diversas reflexões enriquecedoras, as quais foram muito elogiadas pelos participantes. Eles destacaram a importância desse tema para a formação de professores de línguas e agradeceram ao professor pela oportunidade de aprendizado.

Houve um grande entusiasmo e engajamento da audiência, que demonstrou ter sido uma palestra muito produtiva e inspiradora.



Transcrição Online do vídeo
MAXQDA



Arquivo PDF com transcrição da live

Clique para abrir arquivo em nova aba do navegador

Enseino de línguas na contempor...
Live da RAPLE
27/10 • 14h30
tema | Ensino de línguas na contemporaneidade: revisitando conceitos de abordagens, métodos, técnicas e outras cositas más

Assistir no YouTube

PELAS REDES SOCIAIS DA RAPLE
RAPLE.UNESP RAPLE

Assistir no YouTube

Profª Rosângela Santos de Oliveira
Profª Cibele Cecília de Faria Rosendahl



Tabela de frequência e nuvem de palavras da transcrição da live
MAXQDA



Mostrar 10 registros

Pesquisar:

Palavra	Comprimento da palavra	Frequência	%	Ranking
gente	5	165	3.71	1
ser	3	130	2.92	2
fazer	5	99	2.22	3
língua	6	79	1.77	4
falar	5	73	1.64	5
aí	2	67	1.51	6
método	6	66	1.48	7
coisa	5	64	1.44	8
aqui	4	59	1.33	9
achar	5	51	1.15	10

Editar

A mostrar 1 a 10 de 1.077 registros

< 1 2 3 4 5 ... 108 >

 Quais os principais assuntos da transcrição ?

• **O projeto RAPLE:** Um projeto de extensão da Faculdade de Ciências e Letras da UNESP em Araraquara, que visa promover o diálogo e a troca de experiências entre professores de línguas estrangeiras.

• **As lives da RAPLE:** Uma das ações do projeto, que já acontecem há algum tempo e proporcionam um espaço rico para discussões e interação entre os participantes.

• **Agradecimentos da palestrante:** Agradecimentos aos membros da equipe do projeto Raple, aos bolsistas, aos colaboradores e à instituição parceira CLDP pelo apoio e colaboração.

• **Ensino de línguas na contemporaneidade:** Reflexão sobre os conceitos de abordagens, métodos e técnicas de ensino de línguas, revisitando suas origens e discutindo sua aplicabilidade na atualidade.

• **A complexidade da aprendizagem de línguas:** Discussão sobre a natureza multifacetada da aprendizagem de línguas e a importância de considerar o aluno como um indivíduo em constante transformação, com necessidades e características diversas.

• **O papel do professor como mediador:** Ênfase na importância da mediação do professor no processo de ensino e aprendizagem de línguas, especialmente no contexto da utilização de materiais autênticos e da promoção da interação entre os alunos.

• **O uso da língua materna em sala de aula:** Discussão sobre o papel da língua materna no ensino de línguas estrangeiras, considerando sua relevância em determinados momentos e a importância de equilibrar seu uso com a exposição à língua alvo.

• **A importância do planejamento:** Destaque para o papel crucial do planejamento no ensino de línguas, incluindo a seleção de materiais autênticos e a criação de atividades que promovam a interação e o desenvolvimento da autonomia dos alunos.

 Pode fazer um resumo detalhado ?

 Análise do discurso

 Resumo detalhado da transcrição

 Quais os principais assuntos abordados ?



Não houve comentários ao video no Youtube

Análise do chat no Youtube



Mostrar registros

Pesquisar:

Palavra	Comprimento da palavra	Frequência	%	Ranking
obrigar	7	9	5.81	1
professorar	11	9	5.81	1
material	8	4	2.58	3
ser	3	4	2.58	3
autêntico	9	3	1.94	5
enriquecedor	12	3	1.94	5
interessante	12	3	1.94	5
tardar	6	3	1.94	5
boa	4	2	1.29	9
certificar	10	2	1.29	9

Editar

A mostrar 1 a 10 de 111 registros

<

1

2

3

4

5

...

12

>

🏠 Análise de discurso no chat

A partir da análise do chat, podemos observar que a discussão girou em torno de temas relacionados ao **ensino de línguas** e à **formação de professores**. O foco principal parece ter sido a **abordagem comunicativa** e o uso de **materiais autênticos** no ensino de línguas.

Alguns pontos-chave que emergem do diálogo:

- **Questionamento sobre o uso da repetição:** Um participante questiona a permanência da estratégia da repetição na abordagem comunicativa, considerando-a mecânica, e busca sugestões de alternativas menos mecânicas com o mesmo objetivo.
- **Importância do material autêntico:** A discussão também aborda a importância do material autêntico, mas com uma visão crítica, reconhecendo que nem sempre é possível para alunos de nível básico interagirem efetivamente com esse tipo de material.
- **Relação entre língua materna e língua estrangeira:** Outro ponto interessante é a menção da relação entre a língua materna e a língua estrangeira no processo de aprendizagem, destacando como essa relação pode levar a descobertas linguísticas interessantes e facilitar a aprendizagem de línguas próximas.
- **Cunho crítico na análise de materiais:** O chat evidencia a importância de se adotar um cunho crítico ao analisar materiais didáticos, incluindo os autênticos, para garantir que sejam adequados ao nível dos alunos.
- **Valorização da formação de professores:** A formação de professores e o papel do Treinamento de Professores também são mencionados, ressaltando a importância da reflexão crítica na prática docente.

Além disso, o chat revela:

- **Interação positiva e respeitosa entre os participantes:** O diálogo se desenrola de forma positiva e respeitosa, com os participantes expressando gratidão pela oportunidade de aprendizado e reconhecendo a contribuição da professora e da RAPLE.
- **Engajamento dos participantes:** Os participantes demonstram engajamento, fazendo perguntas, compartilhando reflexões e agradecendo pelas respostas recebidas.
- **Ambiente de aprendizado colaborativo:** O chat cria um ambiente de aprendizado colaborativo, no qual os participantes trocam ideias e conhecimentos sobre o ensino de línguas.

Em suma, a análise do chat evidencia um rico debate sobre questões relevantes para o ensino de línguas, com destaque para a abordagem comunicativa, o uso de materiais autênticos e a importância da formação de professores reflexivos e críticos. O diálogo se caracteriza pela interação positiva, pelo engajamento dos participantes e pela criação de um ambiente de aprendizado colaborativo.

🏠 Análise de discurso no chat

🔍 Qual a temática do chat ?

A live apresentada pela professora Ana Maria foi muito elogiada pelos participantes, que destacaram a excelência da sua apresentação sobre o tema “Afeto na aprendizagem de línguas” de Jane Arnold.

Os participantes fizeram perguntas relevantes, como o impacto da pandemia nas emoções de professores e alunos, e a relação entre os estudos sobre emoções e a perspectiva neoliberal na educação refletida em documentos como a BNCC.

A professora respondeu de forma satisfatória, sugerindo a necessidade de mais pesquisas e lives sobre como lidar com as emoções em sala de aula, tanto de professores quanto de alunos.

Os participantes agradeceram a live e demonstraram interesse em receber indicações de livros sobre o assunto

Mostrar 10 registros

Pesquisar:

♥♥♥♥

♥♥♥♥♥

♥♥♥

Time lindo! ♥

Boa tarde, um prazer estar aqui!

Boa tarde a todos!

Sim, muitas escolas de idiomas usam Treinamento de Professores.

E as vezes, também, fazer a relação com a língua materna, leva à conhecimentos linguísticos interessantes. De fato, talvez esse extremismo seja desnecessário.

Oi RG tudo bem? Parabéns pela apresentação! Eu lembro que você comentou sobre a estratégia da repetição permanecer mesmo dentro da abordagem comunicativa e agindo quase como uma incoerência.

Então eu gostaria de perguntar se você tem alguma sugestão de estratégia com a mesma finalidade que a repetição, mas menos mecânica 😊

A RAPLE também ajuda nesse sentido de descobrir materiais, com as indicações pedagógicas que são disponíveis no portal!

Editar

A mostrar 1 a 10 de 27 registros

<

1

2

3

>

