

Universidade Federal de São Carlos (UFSCar)
Centro de Educação e Ciências Humanas (CECH)
Departamento de Ciências Sociais (DCSo)

A CIÊNCIA POLÍTICA E A CIÊNCIA SOCIAL COMPUTACIONAL:

Uma Revisão Sistemática

Trabalho de Graduação apresentado ao curso de Ciências Sociais da Universidade Federal de São Carlos como parte dos requisitos para obtenção do título de bacharel em Ciências Sociais.

Orientadora: Profa. Dra. Gleidylucy Oliveira da Silva

Lourdes Maria Porto Rosseto

São Carlos, Janeiro de 2024

LISTA DE SIGLAS

API	Application Programming Interface
CPC	Ciência Política Computacional
CSC	Ciências Sociais Computacionais
IA	Inteligência Artificial
LDA	Latent Dirichlet Allocation
PLN	Processamento de Linguagem Natural

LISTA DE GRÁFICOS

Gráfico 1 - Frequência de ano de publicação.....	17
Gráfico 2 - Frequência de palavras-chave em inglês.....	19
Gráfico 3 - Frequência de palavras-chave em português.....	20
Gráfico 4 - Frequência de palavras-chave em espanhol.....	22
Gráfico 5 - Palavras-chave por ano de publicação.....	23
Gráfico 6 - Coocorrência de métodos e palavras-chave.....	28
Gráfico 7 - Categorias de métodos por ano de publicação.....	32
Gráfico 8 - Categorias de áreas por ano de publicação.....	33
Gráfico 9 - Linguagens de programação por ano de publicação.....	35
Gráfico 10 - Grafo de referências.....	36

SUMÁRIO

INTRODUÇÃO.....	5
CAPÍTULO 1: O que são as Ciências Sociais Computacionais?.....	9
CAPÍTULO 2: Mapeando a produção do campo.....	14
2.1 Primeiros insights sobre os dados.....	17
2.2 Palavras-chaves por língua.....	18
2.3 Frequências cruzadas.....	27
2.3.1 Os métodos e as técnicas relacionadas.....	30
2.4 Análise de redes de referências.....	36
CONSIDERAÇÕES FINAIS.....	41
REFERÊNCIAS BIBLIOGRÁFICAS.....	46

RESUMO

O estudo apresenta uma análise na interseção da Ciência Política e as Ciências Sociais Computacional. O crescente volume de dados políticos disponíveis online que permitiram emergir o universo do *Big Data* bem como o desenvolvimento de ferramentas computacionais - como aprendizagem de máquina - capazes de auxiliar, modificar e testar a análise em ciências sociais e seus conceitos tem se desenvolvido de forma rápida nos últimos anos. Com isso, neste trabalho objetivamos analisar como técnicas computacionais têm sido aplicadas na literatura nacional e internacional das Ciências Sociais Computacional, com foco na Ciência Política e seus objetos.

Palavras-chave: Ciência Política Computacional, Big Data, Métodos Computacionais, Análise Cientométrica, Ciência Social Computacional.

INTRODUÇÃO

A emergência da Ciência Social Computacional (CSC) e da Ciência Política Computacional (CPC) representa uma revolução nas ciências sociais, marcando a transição de métodos analíticos tradicionais para abordagens avançadas que exploram o vasto potencial do *big data*. Este desenvolvimento reflete as limitações enfrentadas na era pré-digital, onde os dados eram escassos e as técnicas de análise não conseguiam capturar a complexidade das interações e comportamentos humanos. A CSC e a CPC possibilitam análises mais precisas e abrangentes, permitindo que os pesquisadores acessem e interpretem dados em uma escala e profundidade sem precedentes, fomentando uma compreensão mais rica das dinâmicas sociais e políticas. Autores fundamentais no campo têm enfatizado a importância de utilizar a teoria formal para guiar a seleção e interpretação na ciência política, destacando o potencial de enriquecer o desenho de experimentos e a observação de comportamentos teoricamente significativos. A CPC, especificamente, ilustra como a integração de teorias tradicionais com novas técnicas computacionais pode melhorar a compreensão das complexidades do sistema político.

Assim, o impacto da CSC e da CPC vai além da metodologia, reformulando o tecido das teorias sociais e políticas ao integrar novas formas de dados digitais. Estes dados estão desafiando as concepções tradicionais de interação social e política, permitindo a criação de teorias mais sofisticadas. Com isso, apesar dos avanços, a incorporação de métodos computacionais na ciência política não substitui as abordagens tradicionais, mas as complementa e enriquece. A análise de grandes conjuntos de dados oferecidos por redes sociais e pesquisas na internet abrem novos caminhos para entender as motivações e estratégias políticas, revelando padrões e relações que permaneciam ocultos. Na medida em que esses campos continuam a evoluir, sua aplicação e importância são projetadas para crescer em nossa sociedade cada vez mais digitalizada.

Esta pesquisa é fundamentada no reconhecimento do papel emergente dos métodos computacionais na ciência política como parte de uma tendência mais ampla em direção a práticas orientadas por dados nas ciências sociais, o que é impulsionado pelos avanços tecnológicos e pelo acesso a extensivos bancos de dados. Presume-se que esses métodos ofereçam uma compreensão aprimorada dos fenômenos políticos, trazendo à tona percepções e descobertas que não seriam possíveis através de abordagens convencionais.

Além disso, a CPC é vista como uma área com grande potencial para enriquecer o entendimento de uma gama variada de tópicos políticos, incluindo discursos, política externa, pesquisas de opinião e previsões, ao combinar análises precisas e abrangentes com conhecimentos das ciências sociais. O objetivo geral deste estudo é realizar uma análise cientométrica da produção bibliográfica internacional em Ciência Política Computacional para identificar as técnicas e os temas predominantes na literatura. Especificamente, busca-se compreender como as técnicas computacionais são aplicadas ao estudo do comportamento, comunicação e previsão política.

Para alcançar tais objetivos, foram coletados artigos científicos de portais online e plataformas de Ciência Política, usando palavras-chave relevantes para mapear os temas e metodologias recorrentes. As técnicas identificadas na análise de dados políticos serão categorizadas para facilitar a compreensão e sistematização das abordagens adotadas. Os objetivos específicos incluem investigar os métodos computacionais empregados na ciência política, particularmente aqueles relacionados à manipulação de grandes volumes de dados sobre fenômenos políticos, e refletir sobre as oportunidades que a Ciência Política Computacional apresenta para o estudo de tais fenômenos. Isso inclui contemplar como esses métodos podem contribuir para a compreensão, previsão de tendências e enfrentamento de desafios políticos, fornecendo insights valiosos para teorias e previsões políticas.

A metodologia empregada nesta pesquisa para atingir os objetivos descritos envolveu uma análise bibliográfica da produção científica na interseção da Ciência Política e das Ciências Sociais Computacionais, adotando uma abordagem combinada quantitativa e qualitativa. A coleta de dados foi realizada através de pesquisa por palavras-chave em repositórios acadêmicos e periódicos internacionais e regionais, selecionando artigos relevantes em três idiomas para garantir uma cobertura ampla e diversificada. Para a organização e análise dos dados, integramos ferramentas computacionais e análises qualitativas manuais, como o Zotero, para gestão bibliográfica, e o Google Sheets, para estruturação dos metadados. Scripts em R automatizaram a extração de palavras-chave e o MaxQDA foi utilizado para análise qualitativa do conteúdo. Visualizações gráficas foram criadas com o ggplot2 em R e análises de rede realizadas em Python e Gephi para mapear conexões entre os artigos. Esse processo metodológico assegurou a qualidade dos dados

coletados, permitindo uma análise detalhada das tendências na Ciência Política Computacional.

Se faz extremamente relevante a emergência da CSC na era digital como uma revolução na forma como os fenômenos sociais e políticos são estudados. Os avanços na tecnologia de dados e métodos analíticos têm permitido aos pesquisadores superar as restrições impostas por abordagens de coleta de dados tradicionais, como pesquisas e estudos de caso, e abordar questões complexas com um novo grau de precisão e profundidade. A CSC tem demonstrado ser eficaz em diversas áreas das ciências sociais, desde a sociologia até a economia, adaptando suas técnicas computacionais para atender às necessidades específicas e nuances de cada campo (Conte, 2013).

A CPC, um subcampo da CSC, é particularmente interessante, já que ela pressupõe a utilização das técnicas e ferramentas computacionais para analisar fenômenos políticos, incluindo o uso de Aprendizado de Máquina, PLN e análise de redes (Lazer et al., 2009). Graças ao aumento das redes sociais e outras formas de comunicação digital, agora há uma grande quantidade de dados disponíveis que podem esclarecer questões políticas, sendo fundamental o uso das técnicas computacionais em conjunto com a teoria política para garantir que as descobertas sejam não apenas tecnicamente corretas, mas também teoricamente significativas.

Além disso, a CPC tem mostrado ser útil na modelagem de sistemas sociais complexos, prevendo resultados e impactos de várias intervenções sociais e políticas. Por meio do uso de simulações computacionais, por exemplo, os pesquisadores podem testar diferentes cenários e políticas, ajudando a desvendar a complexa teia de causalidade nos sistemas sociais e fornecendo informações valiosas para a tomada de decisões políticas eficazes (Conte, 2013). Ademais, em contextos de crises sociopolíticas, mudanças populacionais, evolução dos meios de comunicação e epidemias/pandemias, a CPC possui o potencial da utilização de ferramentas essenciais para a compreensão e resposta a esses desafios complexos. A capacidade de coletar, analisar e interpretar grandes volumes de dados em tempo real significa que os pesquisadores agora têm ferramentas capazes de acompanhar esses eventos conforme eles se desdobram, fornecendo *insights* em tempo real que podem ajudar a informar as respostas políticas e sociais (Beieler et al., 2016).

Assim, os principais achados desta pesquisa destacam uma crescente atenção à Ciência Política Computacional nos últimos anos, com um notável aumento de publicações a partir de 2018, atingindo um pico em 2020. As palavras-chave analisadas revelam uma convergência metodológica global e generalista, mas também apontam para especificidades regionais nas abordagens de pesquisa. Adicionalmente, a frequência das palavras-chave ao longo dos anos mostra como as técnicas de aprendizado de máquina estão influenciando subáreas da ciência política, como política externa e direitos humanos. A análise de coocorrência, especialmente, demonstra a relação entre métodos e temas abordados nos estudos.

A pesquisa também ressalta um interesse crescente pelas linguagens de programação Python e R, sublinhando sua importância nas metodologias de pesquisa atuais. Por fim, a rede de citações dentro do campo enfatiza a influência de certos trabalhos na comunidade científica, com um foco deliberado nas conexões mais significativas para evidenciar as tendências e influências predominantes na área.

Assim, este trabalho está organizado, inicialmente, com a apresentação da fundamentação teórica das Ciências Sociais Computacionais, delineando a terminologia e os conceitos-chave que informam a pesquisa subsequente. A análise prossegue com um mapeamento da produção acadêmica no domínio da Ciência Política Computacional, explorando tendências e padrões identificados na literatura ao longo do tempo. A discussão se aprofunda na frequência e na distribuição das palavras-chave utilizadas nos artigos, considerando variações linguísticas e suas implicações metodológicas. Este estudo também enfoca a inter-relação entre os métodos empregados e os tópicos investigados, identificando como diferentes técnicas são aplicadas a questões específicas da ciência política.

Além disso, o texto aborda a crescente adoção de linguagens de programação na pesquisa em ciência política, refletindo sobre o papel dessas ferramentas na evolução do campo. Adicionalmente, uma análise de rede das referências bibliográficas fornece insights sobre as conexões e a influência de trabalhos significativos na área, destacando a interdisciplinaridade e o dinamismo do campo. Com isso, o estudo conclui com uma síntese dos principais achados, ponderando sobre o impacto destes na pesquisa em Ciências Sociais Computacionais e delineando possíveis trajetórias para futuras investigações. Ao longo do texto, a discussão é ancorada em evidências coletadas e analisadas, demonstrando a relevância e a aplicabilidade dos métodos computacionais na pesquisa acadêmica contemporânea.

CAPÍTULO 1: O que são as Ciências Sociais Computacionais?

Na era pré-digital, os pesquisadores das ciências sociais frequentemente trabalhavam com dados limitados pela capacidade humana de coleta e armazenamento analógico para compreender as interações humanas e comportamentos políticos. Como destacado por Lazer et al. (2009), esta abordagem apresentou várias limitações: potencial viés de autorrelato, restrições de tamanho da amostra, e dificuldade para captar a dinâmica das interações ao longo do tempo. Além disso, estes métodos tradicionais muitas vezes não conseguiam lidar com a crescente quantidade e complexidade dos dados disponíveis na era digital, deixando de capitalizar em oportunidades para extrair insights valiosos.

Nesse contexto, a consolidação da Ciência Social Computacional (CSC) surge como um marco na maneira como os pesquisadores das ciências sociais lidam com a coleta e análise de dados na era do *big data*. E, como aponta González-Bailón (2013) o termo "big data" enfatiza não apenas o volume puro de dados em si, mas também a riqueza de detalhes e amplitude dos níveis de análise. A autora destaca a importância da resolução espacial e temporal, que proporciona uma compreensão mais refinada das dinâmicas sociais, do individual ao coletivo, apontando como estes registram a dinâmica da comunicação e das interações sociais, diferenciando-os dos registros transacionais. Neste sentido, é inegável a complexidade que a Internet trouxe para as relações e trocas sociais, permitindo compreender de forma notável esses fenômenos, com o potencial de avanços teóricos nas ciências sociais (González-Bailón, 2013).

Portanto, autores, como Monroe et al. (2013), ressaltam o papel emergente das ciências sociais na era do *big data*, enfatizando como os dados sociais constituem grande parte desses conjuntos de dados vastos. Os autores defendem que os cientistas sociais devem desempenhar um papel ativo na definição das perguntas a serem feitas ao *big data* e determinar quais contribuições são valiosas para respostas convincentes.

Eles ilustram ainda que, no campo da ciência política, o *big data* pode aprimorar o desenho de experimentos, a precisão das comparações entre populações de interesse e a observação de comportamentos sociais e políticos que são teoricamente significativos, mas que antes eram difíceis de detectar. Através da teoria formal, discute-se como o processo de seleção desses dados pode tornar certos comportamentos observáveis em conjuntos de dados menos restritivos, ao contrário dos dados convencionais. Isso é particularmente relevante na

teoria da escolha social, que frequentemente lida com a onipresença e instabilidade das preferências multidimensionais (Monroe et al., 2014).

Sendo assim, a CSC representa tal iniciativa de combinar técnicas avançadas de coleta e processamento de dados com teorias e conceitos das ciências sociais, abrindo um novo espectro de possibilidades para a compreensão dos fenômenos sociais. Essa abordagem representa um salto qualitativo na forma como são estudadas as interações humanas e comportamentos políticos, pois permite a coleta e análise de dados em tempo real, potencialmente minimizando os vieses e limitações associados aos métodos tradicionais. Mais do que isso, como já colocado, a CSC expande significativamente o escopo de fenômenos que podem ser estudados, abrindo novos caminhos para o entendimento da dinâmica social.

Em suma, a era digital viu a emergência da CSC, como possibilidade para que os pesquisadores manipulassem e analisassem grandes conjuntos de dados de uma maneira que não era possível anteriormente. A capacidade de analisar números de dados massivos transformou muitos campos de estudo, embora a emergência das ciências sociais computacionais, guiada por esses dados, tenha sido reportada como muito mais lenta que tais campos (Lazer et al., 2009). Ainda assim, o avanço da interdisciplinaridade entre os campos computacionais e nas teorias do comportamento humano continua impulsionando as Ciências Sociais à adoção de métodos computacionais para a manipulação e análise de grandes conjuntos de dados. Dessa maneira, a consolidação da CSC representa um marco importante no avanço das ciências sociais. Por meio do uso de técnicas computacionais avançadas, os pesquisadores das ciências sociais agora têm a capacidade de coletar e analisar uma quantidade e variedade de dados sem precedentes, proporcionando novas maneiras de entender e responder aos desafios sociais e políticos complexos de nosso tempo. O impacto dessa mudança na pesquisa em ciências sociais é imenso, prometendo transformar a forma como entendemos e respondemos às complexidades da vida social e política.

Isso tem sido particularmente evidente no campo da Ciência Política Computacional (CPC), um subcampo da ciência política que usa métodos e técnicas computacionais, como Aprendizado de Máquina, Processamento de Linguagem Natural (PLN) e análise de rede, para estudar fenômenos políticos. O advento da CPC e da CSC transformou o panorama da pesquisa em ciências sociais e políticas. Agora, os pesquisadores podem lidar com conjuntos de dados de grande escala, diversificados e dinâmicos que eram anteriormente inacessíveis ou incontroláveis, sem desperdiçar potenciais insights possibilitados pela manipulação eficiente

de grandes conjuntos de dados (Elliott et al., 2009). Isso permite a captura de interações humanas e comportamentos políticos contínuos, diferenciados e em evolução que seriam difíceis, se não impossíveis, de capturar usando métodos tradicionais.

Os métodos computacionais também oferecem aos pesquisadores as ferramentas para modelar sistemas sociais complexos e dinâmicos, prevendo resultados e impactos de várias intervenções sociais e políticas. Isso pode ajudar a desvendar a complexa teia de causalidade nos sistemas sociais, levando a decisões e estratégias políticas mais eficazes. Nesse sentido, a mudança para aproveitar o poder dos métodos computacionais significa não apenas uma evolução da metodologia, mas também uma mudança fundamental em nossa capacidade de entender e responder a fenômenos sociais e políticos complexos (Lazer et al., 2020).

Em contextos de crises sociopolíticas, econômicas, mudanças populacionais, evolução dos meios de comunicação e sistemas informacionais, ou epidemias/pandemias, os cientistas sociais enfrentam a necessidade de organizar e classificar inúmeros fenômenos característicos de sistemas sociais complexos, que são difíceis de prever e tratar (Conte, 2013). O desenvolvimento tecnológico gerou novos meios de comunicação que criaram novos padrões de comportamento social, influenciando em fenômenos como opiniões públicas, movimentos sociopolíticos e personalidades globais e locais, em uma velocidade exponencial. Além disso, sistemas sociais complexos frequentemente apresentam interdependência e interferências entre suas propriedades e processos de transformações. Dessa forma, o amplo acesso à internet impactou significativamente a frequência, o alcance e a forma de comunicação e interação entre humanos, possibilitando discussões, rumores, teorias e posições políticas manifestadas em redes sociais, assim como preocupações do eleitorado manifestadas através de pesquisas na internet: dados que possibilitam entender o que as pessoas estão falando e como elas estão se conectando no âmbito político

A relevância dos métodos computacionais na ciência política é evidenciada por pesquisas recentes. De acordo com um estudo realizado pelo Pew Research Center (2021), cerca da metade dos adultos americanos obtêm suas notícias por meio das redes sociais, e a discussão política e atividade nas redes sociais aumentou significativamente nos últimos anos. Com isso, a Wankmüller (2022) discorre sobre a aderência de cientistas políticos aos métodos computacionais para analisar e estudar questões proeminentes na área, aproveitando a abundância de dados disponíveis nas redes sociais. No geral, a CPC procura usar métodos e

técnicas computacionais para melhorar nossa compreensão dos sistemas e processos políticos e desenvolver estratégias mais precisas e eficazes para enfrentar os desafios políticos.

Apesar de inúmeras vantagens, a ascensão da CSC e da CPC também trouxe novos desafios. A necessidade de lidar com a complexidade e a crescente quantidade de dados exige o desenvolvimento contínuo de novas técnicas e metodologias. É um equilíbrio delicado, mas quando alcançado, pode levar a insights poderosos, inovadores e éticos. Assim, à medida que esses campos continuam a se desenvolver e a evoluir, é provável que veremos cada vez mais a sua aplicação e importância em nossa sociedade digitalmente conectada (Lazer et al. 2013).

Posto isso, é importante notar que a Ciência Social Computacional (CSC) e a Ciência Política Computacional (CPC) não estão simplesmente remodelando os métodos de investigação nas ciências sociais e políticas, mas também reformulando o próprio tecido das teorias sociais e políticas (Edelmann et al., 2020). Por exemplo, novas formas de dados digitais, como postagens de mídia social e registros administrativos, estão desafiando as concepções tradicionais de interação social e política. Esses dados digitais estão permitindo a criação de teorias mais sofisticadas sobre como as relações sociais se formam e evoluem em diferentes contextos (Edelmann et al., 2020).

Também é crucial considerar o papel da Inteligência Artificial (IA) e do Aprendizado de Máquina nesse campo. De acordo com Edelmann et al. (2020), a IA, os bots e a realidade virtual ainda não tiveram seu "momento crucial" na sociologia. No entanto, essas tecnologias já estão oferecendo novas técnicas para estudar como as relações sociais se formam e evoluem em diferentes contextos (Alegre, 2020). Assim, é razoável antecipar que as técnicas de IA e Aprendizado de Máquina provavelmente terão um papel cada vez mais importante na CPC no futuro. Com isso, da mesma forma, a CSC e a CPC estão criando novas maneiras de construir a própria teoria.

Ainda, embora o uso de Aprendizado de Máquina possa não nos permitir decifrar o comportamento coletivo, a cultura ou a tomada de decisões humanas, podemos usar essas ferramentas para identificar novas dimensões do comportamento social ou testar a robustez de teorias existentes (Edelmann et al., 2020). Por exemplo, podemos comparar a lacuna entre as previsões produzidas por esses modelos e a realidade subjacente, permitindo-nos identificar a complexidade imprevista do comportamento humano ou identificar nossos pontos cegos.

Em especial, o subcampo da comunicação política se destaca pela valorização dada a fontes textuais de informação, como as redes sociais, refletindo na necessidade de aprimoramento de técnicas automatizadas para a análise textual política. Embora tenha começado de forma primariamente manual, o avanço do Aprendizado de Máquina e do PLN abre um leque de possibilidades para a aplicação dessa metodologia. Isso viabiliza análises mais profundas com enfoque em extração de opinião política (Ayuso, 2019; Jungherr et al., 2017; Marques, 2022; Zhu, 2010), fact-checking (Dayanik, 2022), identificação de tendências e padrões de comportamento (McCabe, 2022; Ramey et al., 2019), e monitoramento de narrativas políticas em tempo real (Arcila-Calderón et al., 2017).

Apesar do avanço tecnológico, é importante notar que o avanço da área não implica necessariamente na substituição das abordagens tradicionais das ciências sociais, mas sim na complementação e enriquecimento delas (Monroe et al. 2014). Esta integração da teoria política tradicional com as novas técnicas da CPC tem o potencial de melhorar a nossa compreensão das complexidades do sistema político. As redes sociais, em particular, oferecem uma riqueza de dados sobre as interações políticas que podem ajudar a lançar luz sobre as dinâmicas do comportamento político, os processos eleitorais e as tendências políticas (Stieglitz et al. 2013). A análise desses dados pode revelar padrões e relações previamente ocultos, contribuindo para uma compreensão mais profunda das motivações e estratégias políticas.

Posto isso, o mapeamento das metodologias computacionais, bem como os temas analisados com elas, permite entender os métodos mais adotados e as tarefas relacionadas, além das temáticas em construção nesse subcampo. Isso possibilita o aprofundamento de análises e, conseqüentemente, destaca as metodologias interdisciplinares usadas pelos cientistas sociais para a realização da análise. Dessa forma, o foco dessa sistematização é promover uma visualização das possibilidades de análises e conclusões mais robustas e abrangentes.

CAPÍTULO 2: Mapeando a produção do campo

Para atingir os objetivos estabelecidos de refletir sobre a produção da Ciência Política dentro das Ciências Sociais Computacionais, bem como a apropriação das ferramentas computacionais com teorias e pesquisas da área, optamos por analisar a produção bibliográfica ligada ao tema. Para tanto, fizemos uma análise combinada - quantitativa automatizada e qualitativa - no sentido de garantir um levantamento amplo da produção, mas, também um corpus pertinente ao objetivo.

Assim, o primeiro passo foi a coleta ampla e diversificada do material num recorte internacional e regional (América Latina). Para atingir esse objetivo, empregamos a pesquisa por palavras-chave em 4 repositórios: JSTOR, Web of Science, Google Acadêmico e Scielo. A escolha deste último se deve ao fato de reunir muitos artigos da produção nacional e regional em relação a outras bases com perfil mais internacional. Além disso, acrescentamos três revistas internacionais destacadas pelo foco em artigos metodológicos e com rigorosa revisão por pares. Foram elas: American Politics Research, American Journal Of Political Science, Political Science Research And Methods. Por fim, pensando ainda no caso brasileiro, recentemente houve a publicação de um dossiê em língua portuguesa sobre o tema, na Revista Simbiótica. Alguns artigos deste dossiê entraram no conjunto geral de dados a serem investigados e para respaldarem a pesquisa.

As palavras utilizadas na busca foram “ciência social”, “computacional”, “política”, “ciência política”, “modelagem preditiva”, “big data”, “processamento de linguagem natural” e “aprendizado de máquina” nos idiomas inglês, português e espanhol. Esse enfoque multilíngue visou também o objetivo de ser abrangente em termos de produções científicas globais e regionais.

A partir dos resultados preliminares das buscas, foi adotada uma abordagem sistemática para a seleção de artigos. Nos sites que utilizam rolagem infinita, foram considerados os primeiros 15 resultados mais relevantes. Essa escolha se baseou na premissa de que os algoritmos desses sites tendem a priorizar os trabalhos mais citados e relevantes no topo dos resultados de busca. Para plataformas que apresentavam resultados em páginas limitadas, a atenção se voltou para as duas primeiras páginas, presumindo que estas incluíam os artigos mais pertinentes e alinhados com os critérios de pesquisa. Em ambos os casos, a ordenação por relevância foi utilizada como critério de seleção, assegurando a escolha dos artigos mais

apropriados para os objetivos do estudo. Assim, ao todo, foram considerados 165 textos inicialmente no período de 20 anos - 2003 a 2023 - dos quais, 60 seguiram para a análise mais detalhada. Essa faixa temporal de duas décadas serviu para examinar o desenvolvimento da Ciência Política Computacional, permitindo uma análise das mudanças ocorridas ao longo dos anos nas metodologias empregadas e nos temas de interesse.

Dos 60 artigos selecionados para uma análise mais aprofundada, 25 foram redigidos em inglês, 22 em português e 13 em espanhol. Sozinhos, os textos em língua inglesa representam 41,7% do corpus, o que demonstra, de fato, uma força dos países centrais ou publicações em língua inglesa sobre o tema. Importante dizer que, juntos, a produção regional - Português e Espanhol - representa mais da metade do trabalho.

Essa etapa de seleção foi assim conduzida de forma qualitativa e centrada na leitura e avaliação do título, resumo, palavras-chave e metodologia adotada. Esse processo permitiu a identificação de artigos que não apenas se enquadram no contexto da Ciência Política, mas que também empregam abordagens computacionais, garantindo a consonância com os objetivos desta pesquisa. Dessa forma, foi possível agrupar artigos que abordam não apenas temas relevantes para a Ciência Política, mas que também aplicam técnicas como PLN e Aprendizado de Máquina em suas análises. Essa abordagem metódica para a seleção de artigos culminou na criação de um conjunto de dados abrangente e de alta qualidade, que reflete as mais recentes e relevantes pesquisas na intersecção da Ciência Política e abordagens computacionais. Essa seleção criteriosa assegura que a análise subsequente esteja embasada em evidências sólidas e proporciona uma base sólida para a pesquisa proposta.

2.1 Abordagem computacional combinada para análise dos dados

Na realização deste estudo, optamos por uma combinação de ferramentas computacionais fechadas com scripts e análise qualitativa, como já inicialmente apontado, permitindo assim o um desenho metodológico diversificado no uso de ferramentas para análise de dados. Isso se torna interessante porque permitiu que o pesquisador e o analista pudessem usar também de criatividade metodológica para a realização da pesquisa, contribuindo assim para um caminho que combinasse efetivamente ferramentas e abordagens qualitativas e quantitativas. Assim, foram utilizadas as ferramentas fechadas Zotero, Maxqda, Google Sheets (equivalente ao Excel, do pacote Office) juntamente com as linguagens R (via R Studio) e Python.

O primeiro passo foi organizar os artigos selecionados no Zotero, *software* destinado à gestão bibliográfica, facilitando o gerenciamento das informações bibliográficas. Por meio da funcionalidade de exportação do Zotero, os metadados foram convertidos para um arquivo CSV, que foi importado para o Google Sheets. Esse procedimento automatizou a organização dos dados em colunas específicas, como Ano de Publicação, Título, Resumo, Autores, Idioma, Local de Publicação, DOI, URL e Extras.

Em seguida, um *script* elaborado no ambiente do RStudio foi empregado para automatizar a extração e o preenchimento das palavras-chave de cada artigo. De forma paralela, uma análise manual foi conduzida utilizando o MaxQDA, *software* para análise de dados qualitativos e métodos mistos em pesquisas acadêmicas, para identificar as metodologias utilizadas nos trabalhos acadêmicos. O uso do MAXQDA se deu porque nem todos os textos trazem uma seção metodológica de forma explícita, o que dificulta a busca textual automatizada destes trechos. Assim, com a análise humana foi possível selecionar todas as partes onde havia a descrição metodológica dos textos para, posteriormente, rodar uma análise lexicométrica através de um *script* do RStudio. Assim, as metodologias foram classificadas em quatro categorias que serão expostas posteriormente neste artigo.

Adicionalmente, as funcionalidades nativas do Google Sheets possibilitaram a geração de gráficos que exibem a distribuição dos anos de publicação e das palavras-chave, proporcionando um entendimento visual das tendências nos dados coletados.

A linguagem R, através de seu pacote de visualização de dados, ggplot2, foi utilizada para elaborar gráficos de coocorrência, que serão examinados em seções subsequentes deste trabalho, a partir dos dados estruturados no arquivo CSV proveniente do Google Sheets.

Numa etapa posterior, a biblioteca em formato CSV exportada do Zotero foi utilizada para a execução de *scripts* na linguagem de programação Python para a leitura e a extração de referências bibliográficas cruzadas de cada um dos artigos. O objetivo desse processo foi a criação, visualização e análise de um gráfico de redes, visando conectar os artigos de maior relevância presentes na base de dados deste estudo através do Gephi, uma das mais populares ferramentas de análise e visualização de redes.

Este procedimento detalhado de coleta e análise de dados garantiu, não apenas a integridade do corpus da pesquisa, mas também possibilitou uma discussão alicerçada sobre as interconexões e influências mútuas entre os artigos, oferecendo uma perspectiva abrangente do campo investigado.

Importante ressaltar que, alguns dos métodos utilizados, não apenas exemplificam a aplicação, mas também ilustram a viabilidade de métodos computacionais para facilitar a

análise e a manipulação de dados em pesquisas acadêmicas. Essa abordagem demonstra, mesmo que de forma não exaustiva, como ferramentas e a familiarização com *scripts* computacionais podem ser empregados eficientemente por pesquisadores para extrair *insights* valiosos de conjuntos de dados complexos.

2.1 Primeiros *insights* sobre os dados

A primeira informação relevante sobre o material apareceu já na organização dos textos por ano de publicação. Como dissemos, a faixa temporal ampla abarcando quase toda a produção do século XXI permitiu identificar mudanças ocorridas ao longo dos anos. Sendo assim, um aspecto notável dentro dessa linha temporal é o incremento da produção a partir de

Gráfico 1 - Frequência de ano de publicação



Fonte: Elaboração Própria

2018, com um primeiro pico no ano de 2020, com a publicação de 8 artigos, como pode ser observado no **gráfico 1**.

Esse aumento pode ser indicativo de um crescente interesse e investimento no campo, possivelmente impulsionado por eventos globais ou avanços tecnológicos. A tendência de crescimento prossegue nos anos subsequentes, com 2021, 2022 e 2023 apresentando, respectivamente, 7, 11, sendo esse o máximo, e 8 artigos dentro do conjunto de dados

selecionado para esta pesquisa. Notavelmente, 2022 se destaca como o ano de maior relevância em termos de publicações, dentro dos critérios estabelecidos, possivelmente refletindo um vigoroso e engajamento na área. Essa trajetória ascendente na quantidade de publicações serve como um indicativo robusto do dinamismo da Ciência Política Computacional, mas também do caráter recente desse interesse. As reflexões sobre tais tendências temporais potencialmente sinalizam a influência de fatores externos que podem ter estimulado a pesquisa e a produção de conhecimento nessa área.

Posto isso, esta distribuição temporal somada à distribuição linguística dos artigos selecionados oferece uma oportunidade única de comparar abordagens e perspectivas de diferentes contextos culturais e acadêmicos. Isso é particularmente relevante em um campo como o da Ciência Política Computacional, onde os contextos políticos e sociais podem influenciar significativamente a pesquisa e suas conclusões. Portanto, o conjunto de dados compilado proporciona uma base rica e diversificada para análises subsequentes, oferecendo insights valiosos sobre as tendências globais e regionais na interseção da ciência política e da tecnologia computacional.

2.2 Palavras-chaves por língua

Um segundo passo interessante no escrutínio da produção selecionada foi a análise das palavras-chave. Embora todos os artigos incluídos apresentem métodos claramente definidos, uma parcela considerável, 19 deles, não especificava palavras-chave, o que pode ter sido reflexo de práticas de catalogação ou editoriais distintas. Apesar disso, consideramos relevante avaliar quais termos chaves predominaram entre os artigos selecionados e que as traziam definidas. A análise destas palavras-chave sugere tendências em termos de tópicos e metodologias utilizadas nos artigos selecionados para a base de dados.

No **gráfico 2**, temos que, no idioma inglês, por exemplo, termos como "computational social science" e "social media" com uma predominância significativa, indicando uma forte interseção entre o uso de mídias sociais e a análise computacional em estudos das ciências sociais.

Gráfico 2 - Frequência de palavras-chave em inglês



Fonte: Elaboração própria.

Como exemplo da intersecção das duas palavras-chave destacadas, se sobressai o artigo de Jungherr, Jürgens e Schoen (2012), intitulado "Why the Pirate Party Won the German Election of 2009 or The Trouble With Predictions: A Response to Tumasjan et al. 'Predicting Elections With Twitter: What 140 Characters Reveal About Political Sentiment'", que responde ao estudo de Tumasjan et al. (2010) sobre a previsibilidade dos resultados eleitorais alemães com base em dados do Twitter. Tumasjan et al. (2010) argumentaram que a frequência relativa das menções dos partidos políticos no Twitter refletia os resultados eleitorais. Jungherr et al. (2012), no entanto, contestam essa alegação, demonstrando que os resultados eram contingentes a escolhas arbitrárias e que a frequência das menções no Twitter não permitia uma previsão precisa dos resultados eleitorais.

Na metodologia, os autores raspam (coleta de dados de um site) e analisaram publicações do Twitter (*tweets*), concentrando-se nas menções de partidos políticos e em palavras-chave relacionadas à política. Eles compararam esses dados com os resultados reais das eleições, incluindo uma análise detalhada do Partido Pirata. Para isso, empregaram técnicas de ciências sociais computacionais para extrair, processar e analisar grandes conjuntos de dados de redes sociais, sublinhando a importância de uma metodologia bem fundamentada na análise de dados de redes sociais, especialmente ao coletar esses dados para prever ou interpretar fenômenos políticos.

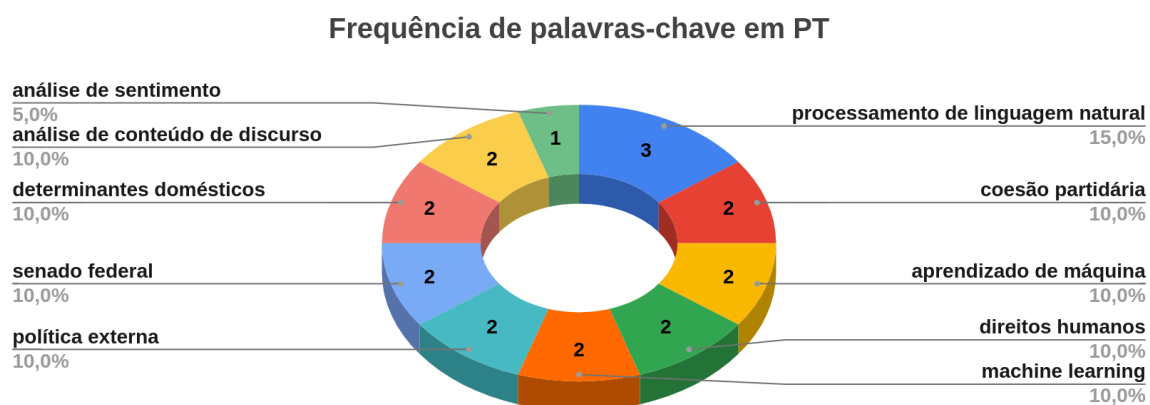
Ainda, a comparação entre os termos mais frequentes nos três idiomas mostra que, enquanto alguns tópicos como "machine learning" são comuns em múltiplos contextos

linguísticos, no entanto, a língua inglesa também possui destaques específicos, como "democratic deliberation" e "political polarization", como demonstrado no **gráfico 2**. Isso pode refletir não apenas as diferenças nas práticas de pesquisa, mas também nas agendas políticas e sociais predominantes em diferentes contextos linguísticos e culturais.

Para ilustrar tal destaque, o estudo “Political Polarization on the Digital Sphere: A Cross-platform, Over-time Analysis of Interactional, Positional, and Affective Polarization on Social Media” de Moran Yarchi, Christian Baden e Neta Kligler-Vilenchik (2020) realiza uma análise de polarização política em várias redes sociais (Facebook, Twitter e WhatsApp) estabelecendo um banco de dados através da raspagem de publicações e analisando por meio de métodos de PLN. No que se refere ao enfoque em deliberação democrática, o artigo “Leveraging AI for democratic discourse: Chat interventions can improve online political conversations at scale” de Lisa P. Argyle e colaboradores (2023) conduz um experimento sobre o uso de modelos de linguagem de IA como um assistente de chat para melhorar a qualidade das conversas políticas online.

Já em português, no **gráfico 3**, observa-se uma ênfase em “processamento de linguagem natural” e “aprendizado de máquina”/“machine learning”, refletindo o foco em metodologias de análise de texto e dados.

Gráfico 3 - Frequência de palavras-chave em português



Fonte: Elaboração própria.

A partir desses dados, é possível inferir que certos termos são comuns e podem indicar um consenso sobre abordagens importantes na área. A predominância de termos como “aprendizado de máquina” e "processamento de linguagem natural" em diferentes idiomas

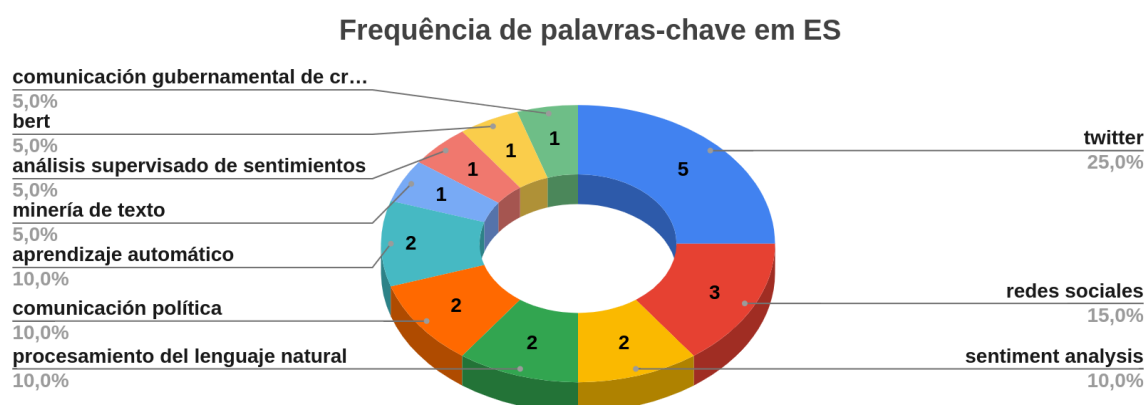
reforça a relevância destas técnicas computacionais na pesquisa política contemporânea, dando foco a dados que tratam da comunicação humana - textos e falas - nessas pesquisas.

Se faz relevante representar tal predominância com a pesquisa de Matheus Felipe dos Anjos, intitulada "Análise de Sentimentos em Textos Curtos Dirigidos à Eleição Presidencial do Brasil em 2018", que focou na utilização de técnicas de Aprendizado de Máquina, especificamente redes neurais, para a análise de sentimentos em textos relacionados à eleição presidencial brasileira de 2018. O objetivo era classificar esses textos, obtidos do Twitter, como positivos ou negativos, fornecendo insights sobre a percepção pública dos candidatos e a dinâmica política durante o período eleitoral.

Para atingir esse objetivo, foram empregadas técnicas avançadas de PLN e diferentes arquiteturas de redes neurais. O estudo envolveu a coleta de dados do Twitter usando palavras-chave relacionadas aos candidatos, seguida de um processo de pré-processamento dos dados, que incluía a tokenização, remoção de *stopwords* e vetorização dos textos. Em seguida, esses dados foram usados para treinar e validar os modelos de redes neurais, com o objetivo de otimizar a classificação de sentimentos. As ferramentas utilizadas para desenvolver e implementar os modelos incluíam bibliotecas integradas na linguagem de programação Python. Este estudo exemplifica a intersecção entre as ciências sociais computacionais e o uso de redes sociais, demonstrando como as técnicas de PLN e Aprendizado de Máquina podem ser aplicadas para analisar e interpretar grandes volumes de dados textuais de redes sociais, fornecendo insights valiosos sobre o comportamento e as preferências políticas dos usuários.

Posto isso, enquanto alguns termos são universais e indicam uma convergência metodológica, outros refletem especificidades regionais ou linguísticas que podem sinalizar abordagens distintas para problemas semelhantes. Isso também pode sugerir as áreas que estão ganhando tração ou que são de particular interesse para comunidades de pesquisa em diferentes regiões. Por exemplo, como é possível observar no **gráfico 4**, a ênfase em "twitter" nos artigos em espanhol pode indicar uma tendência de estudos na comunicação política desta plataforma específica.

Gráfico 4 - Frequência de palavras-chave em espanhol



Fonte: Elaboração própria.

O estudo de Carlos Arcila-Calderón e colaboradores, intitulado "Análisis supervisado de sentimientos políticos en español: clasificación en tiempo real de tweets basada en aprendizaje automático" e publicado em 2017, ilustra tal ênfase, abordando a aplicação de técnicas de aprendizagem automática para análise de sentimentos em tweets políticos em espanhol. O foco principal é desenvolver um classificador em tempo real que possa analisar e classificar opiniões políticas em tweets, utilizando técnicas avançadas de aprendizagem automática tanto em um ambiente computacional local como em sistemas de computação distribuída para lidar com grandes volumes de dados (*big data*).

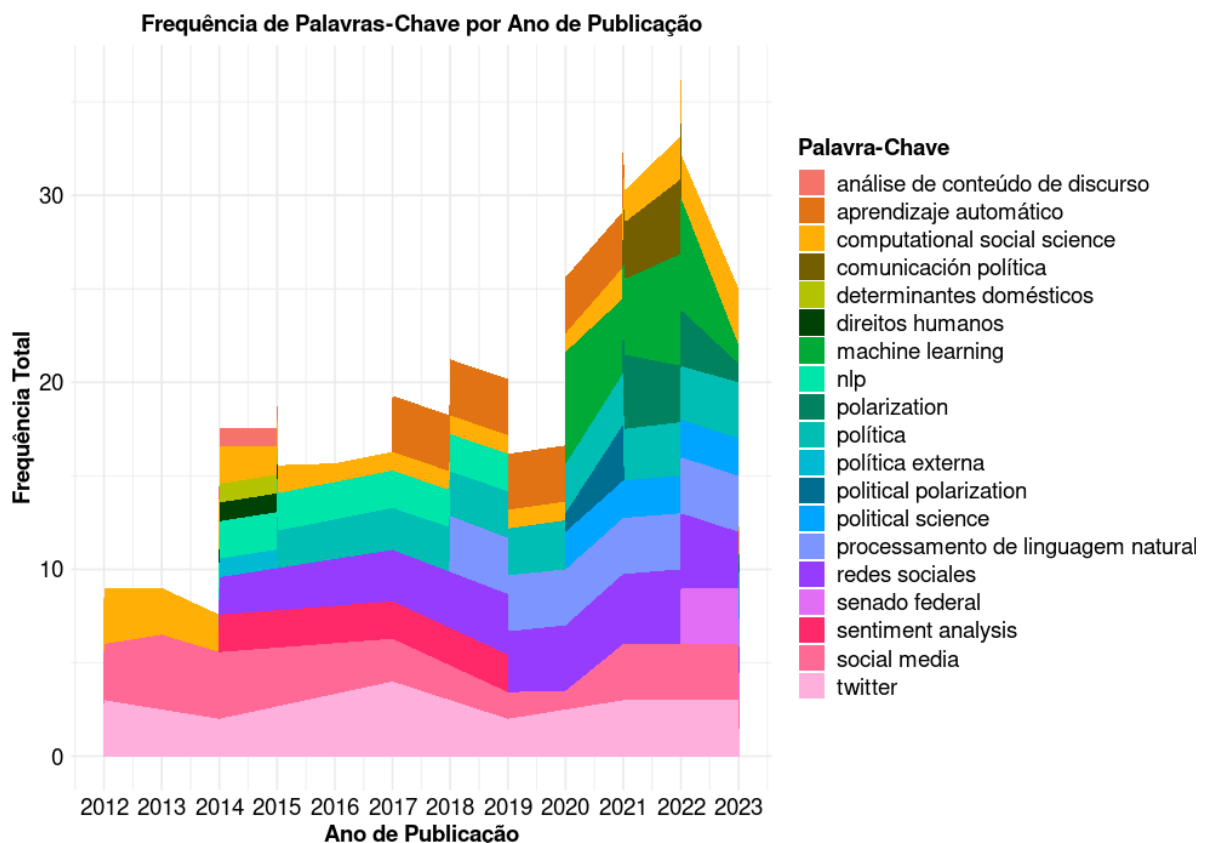
A metodologia empregada neste estudo envolve a implementação de algoritmos de aprendizagem automática supervisionada para classificar textos políticos em língua espanhola. Os autores utilizaram algoritmos de classificadores probabilísticos e técnicas de PLN para analisar os sentimentos expressos nos tweets. Este processo inclui a coleta de dados do Twitter em tempo real, através da Application Programming Interface (API) do Twitter, o que permite a análise imediata e contínua dos sentimentos expressos nas mensagens. Adicionalmente, o estudo também se beneficia da integração de ferramentas e bibliotecas integradas na linguagem Python, que são fundamentais no treinamento e implementação dos modelos de classificação. Sendo assim, essa é apenas uma das pesquisas que representam uma intersecção significativa entre as ciências sociais computacionais, PLN e Aprendizado de Máquina, destacando o papel crucial do Twitter e das redes sociais na análise política contemporânea, em especial, em trabalhos de língua espanhola.

Essas observações reforçam a importância de uma abordagem multilíngue na Ciência Política Computacional, uma vez que permite uma compreensão mais ampla e diversificada das tendências globais. Ao mesmo tempo, uma padronização mais rigorosa poderia também ajudar a mitigar o problema identificado, onde uma parcela considerável de artigos não especifica palavras-chave, o que desafia a análise de tendências em grandes conjuntos de dados.

2.3 Palavras-chaves em geral

O **gráfico 5**, por sua vez, revela tendências interessantes na interseção da ciência política computacional com outros domínios específicos, apresentando a frequência de todas as palavras-chaves por ano. Ao se concentrar na frequência de termos como "aprendizaje automático" (aprendizagem de máquina), "comunicación política" (comunicação política), "política externa", "polarização", "direitos humanos" e "análise de conteúdo de discurso", podemos inferir como a adoção de técnicas de aprendizagem de máquina está potencialmente transformando a pesquisa em várias subáreas da ciência política.

Gráfico 5 - Palavras-chave por ano de publicação



Fonte: Elaboração própria.

Outro aspecto importante é a interdisciplinaridade inerente a essas técnicas. A integração de métodos de Aprendizado de Máquina e PLN na Ciência Política não apenas enriquece a análise política, mas também fomenta colaborações entre campos como ciência da computação, estatística e ciências sociais, como já apontado por Lazer et al. (2009). Isso leva a uma melhor compreensão dos dados e a insights mais robustos, pois combina a expertise técnica com a compreensão contextual e teórica dos fenômenos políticos.

A interdisciplinaridade inerente a essas técnicas também é outro elemento interessante. A integração de métodos de Aprendizado de Máquina e PLN na Ciência Política não apenas enriquece a análise política, mas também fomenta colaborações entre campos como ciência da computação, estatística e ciências sociais, como já apontado por Lazer et al. (2009). Isso leva a uma melhor compreensão dos dados e a insights mais robustos, pois combina a expertise técnica com a compreensão contextual e teórica dos fenômenos políticos.

A frequência crescente do termo "aprendizaje automático" sugere um maior interesse e integração de métodos de aprendizagem de máquina na análise política. Este aumento pode ser explicado pela necessidade de processar volumes substanciais de dados qualitativos e quantitativos, característicos da era digital em que vivemos. Como já apontado, muitos autores afirmam que a aprendizagem de máquina oferece ferramentas para decifrar padrões complexos, desde tendências de voto até a disseminação de informações, que são inerentes à comunicação política e à política externa.

Por outro lado, a relação entre Aprendizado de Máquina com várias áreas, entre elas a comunicação política revela-se particularmente relevante e multifacetada. As técnicas de Aprendizado de Máquina, juntamente com outras abordagens computacionais, oferecem ferramentas poderosas para analisar extensos conjuntos de dados de comunicação. Isso inclui uma variedade de conteúdos, desde discursos políticos, propaganda e o dinâmico fluxo de informações nas redes sociais, até interações mais sutis e discursos diplomáticos. O uso dessas técnicas avançadas permite uma compreensão mais aprofundada e matizada de como as mensagens políticas são cuidadosamente formuladas, disseminadas e recebidas pelo público.

No contexto da política externa, os métodos computacionais também assumem um papel crucial. Eles são capazes de analisar discursos diplomáticos e prever tendências nas

relações internacionais, oferecendo insights estratégicos que podem ser fundamentais para informar decisões políticas e estratégias de negociação.

Por sua vez, a utilização de *softwares* como o NVivo, que oferecem abordagens semi-automatizadas para a análise de discursos, complementa essas técnicas mais avançadas. (Silva, 2015). Isso é particularmente útil para desvendar as camadas de significado e as intenções subjacentes em discursos políticos. Por exemplo, a análise computacional pode identificar nuances em discursos diplomáticos que sinalizam mudanças nas políticas ou nas relações entre países, o que é vital para a interpretação do âmbito político externo.

A pesquisa "Política Externa Brasileira de Direitos Humanos do Governo Dilma Rousseff (2011-2014): Análise de Conteúdo de Discursos Oficiais", realizada por Danielle Costa da Silva em 2015, serve como um exemplo ilustrativo de como softwares como o NVivo podem ser utilizados para análises qualitativas detalhadas em ciências sociais. A metodologia da pesquisa baseia-se no método proposto por Bardin ([1977] 2011, apud SILVA, 2015), que considera a análise de conteúdo dos discursos como um conjunto de técnicas sistemáticas e objetivas para descrever o conteúdo das mensagens. Neste estudo, o software NVivo 10 foi utilizado para analisar discursos oficiais, focando especificamente em menções ao termo "direitos humanos". A partir do banco de dados do LABMUNDO, foi feita uma triagem, resultando em um total de 110 discursos selecionados para análise, que correspondem a 17,268% do total de discursos no banco de dados (Silva, 2015).

A "polarização" é outro termo que apresenta uma frequência significativa, destacando como a divisão política é uma área de interesse constante. A análise de dados obtidos por meio de raspagem de páginas da web, como perfis de redes sociais (Weber, 2009; Yarchi et al. 2021), pode revelar as dinâmicas subjacentes da polarização, identificando fatores contribuintes, como a segregação ideológica em plataformas *online*. A pesquisa "Medios de Comunicación y Polarización; Un Análisis del Feminismo en España mediante Procesamiento del Lenguaje Natural", realizada por Ignacio-Jesús Serrano-Contreras, foca na análise da polarização na mídia e nas redes sociais, especificamente em relação ao feminismo na Espanha. Utilizando métodos computacionais avançados, o estudo emprega técnicas de raspagem de dados para coletar informações de fontes online, incluindo redes sociais e sites de notícias.

A metodologia do estudo envolve a extração de dados da web, focando em coletar informações relevantes de fontes variadas, sem critérios padronizados, realizando esse processo por meio do uso de APIs facilitando a extração de dados. A técnica de *scraping*, especificamente, permite a geração de arquivos, geralmente CSV, que são então usados para análises posteriores. Após a coleta de dados, são aplicadas técnicas de mineração de texto e PLN para analisar os dados coletados, buscando padrões e informações relevantes.

O estudo se concentra na polarização semântica, que está intrinsecamente relacionada à polarização política. A polarização é operacionalizada como a distância entre o *ratio* calculado para cada célula de dados (como um artigo de imprensa, comentário em rede social ou discurso político) e a mediana do agregado dos *ratios*, obtendo um número que pode variar entre 0 (sem polarização) e 2 (polarização máxima). Este índice tem a capacidade de medir as divergências que podem surgir ao longo do tempo em um meio, um partido político ou entre usuários de uma rede social.

Portanto, o estudo exemplifica como a raspagem de dados e a análise computacional podem ser empregadas para entender e quantificar a polarização em debates sociais e políticos, particularmente em relação ao feminismo na Espanha, destacando a interseção entre ciências sociais computacionais, análise de redes sociais e técnicas de PLN.

A análise de conteúdo e de discurso, por fim, é uma área que naturalmente se beneficia do avanço em *softwares* de análise qualitativa, abordagens de PLN e aprendizagem de máquina, uma vez que são métodos capazes de processar e interpretar grandes conjuntos de dados textuais, proporcionando uma compreensão facilitada quantitativa e qualitativamente dos temas, estilos retóricos e estruturas de discurso utilizados em diferentes contextos políticos (Cardoso, 2022; Figueredo et al. 2022; Maguire 2021).

A pesquisa "A Natural Language Measure of Ideology in the Brazilian Senate" de Figueredo et al. (2022) se destaca significativamente na análise da polarização ideológica e do discurso político no Senado Brasileiro. Utilizando técnicas avançadas de PLN e Aprendizado de Máquina, os pesquisadores desenvolveram um método para medir o conteúdo ideológico dos discursos políticos, tratando a linguagem como unidade de análise e aplicando um algoritmo de classificação automática para representar dicionários que melhor explicam cada posição política.

O estudo analisa os discursos dos senadores brasileiros de várias legislaturas, de 1995 a 2018, empregando uma abordagem inovadora para a construção de um dicionário político que evolui ao longo do tempo. Em vez de usar um dicionário pré-definido, os autores construíram seu próprio dicionário político, permitindo analisar a estabilidade e a transição das crenças políticas. Utilizando dados de discursos políticos, eles foram capazes de capturar a polarização comparando a evolução da precisão de classificação do método, onde uma maior precisão indica discursos mais facilmente distinguíveis entre diferentes dimensões ideológicas. Isso indica posições mais polarizadas.

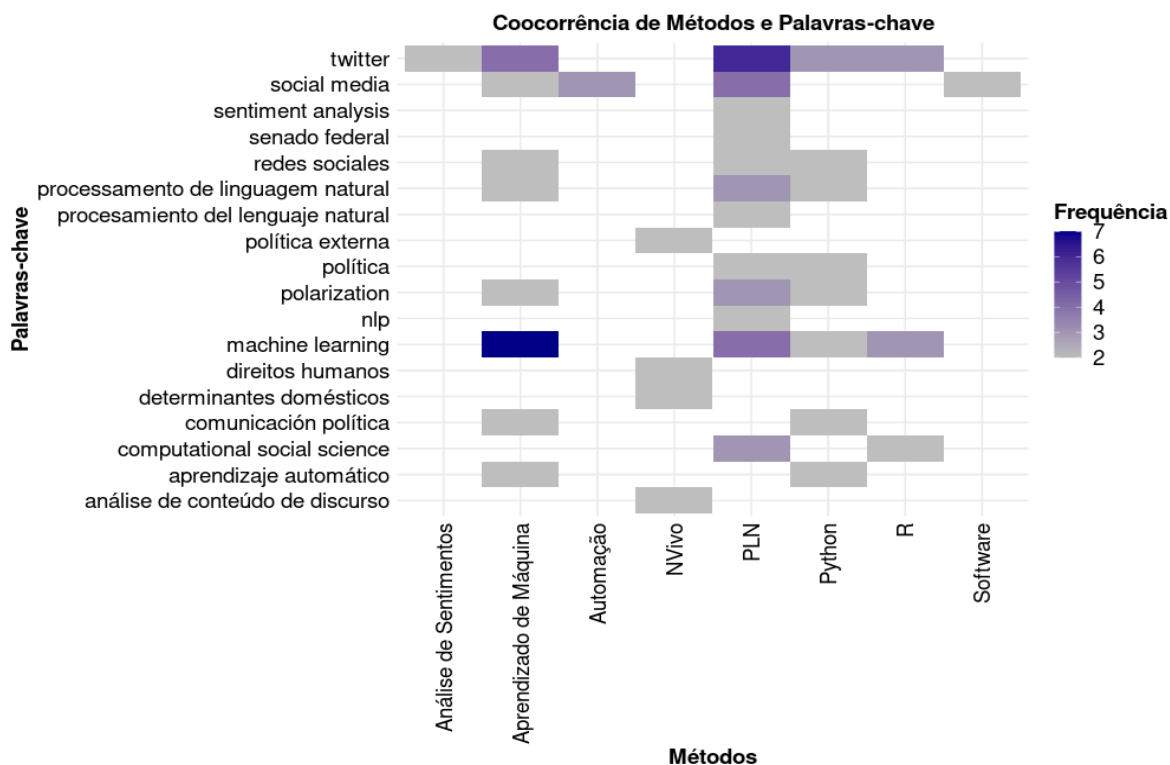
A pesquisa demonstra como as técnicas de PLN e Aprendizado de Máquina podem ser efetivamente aplicadas para analisar grandes volumes de dados textuais no contexto das ciências sociais computacionais. A análise dos discursos políticos fornece uma compreensão mais profunda dos padrões ideológicos e da dinâmica da polarização política, destacando o potencial dessas abordagens para revelar nuances e tendências significativas no discurso político brasileiro.

A correlação entre o aumento do uso de as abordagens computacionais de Aprendizado de Máquina, análise de discurso e PLN, e a frequência de termos como "comunicação política", "política externa", "polarização", e "direitos humanos" sugere que a ciência política computacional está cada vez mais aplicando métodos avançados para entender complexidades políticas em áreas mais aderentes ao tema. Além disso, essas técnicas estão permitindo aos pesquisadores abordar questões políticas com uma profundidade e escala antes inatingíveis, o que poderia explicar o aumento observado nas frequências desses termos. Como discutido por Lazer et al. (2020), as implicações dessas tendências para a pesquisa futura e a prática política são vastas, indicando o potencial de uma era de análise política enriquecida pela ciência de dados e aprendizagem computacional.

2.3 Frequências cruzadas

Outro instrumento analítico empregado neste estudo foi identificar as coocorrências de métodos e palavras-chave na literatura selecionada. Para tanto, utilizamos um mapa de calor, ilustrado no **gráfico 6**. Estruturado com os métodos dispostos no eixo horizontal (x) e as palavras-chave no eixo vertical (y), cada interseção no gráfico, ou célula, indica a frequência com que um método específico é mencionado juntamente com uma determinada palavra-chave.

Gráfico 6 - Coocorrência de métodos e palavras-chave



Fonte: Elaboração própria.

A variação de cores, que vai de tons mais claros de cinza até o azul escuro serve como um indicador da coocorrência entre método e tema de forma forte (alta frequência/cor mais escura) ou fraca (baixa frequência/cor mais clara), sugerindo uma força de associação entre os aspectos investigados.

Isso indica então uma forte associação entre “Twitter” e “Social Media” como temas e “PLN”, “Aprendizado de Máquina” e “Automação” como métodos. A utilização de dados em texto de redes sociais para fins de pesquisa, em especial nas Ciências Sociais, permite aos pesquisadores interessados nas relações e fenômenos políticos testarem suas hipóteses a partir de dados de vastas redes sociais com detalhes qualitativos e longitudinais sem precedentes (Bail, 2015). Isto porque durante décadas a análise política de texto e discurso foi tradicionalmente realizada de forma manual devido à falta de ferramentas automatizadas disponíveis aos pesquisadores (Dayanik, 2022). É com isso que os avanços recentes em Aprendizado de Máquina e PLN ganham espaço, possibilitando a automatização dessa análise.

A partir disso, a aderência a técnicas como as citadas, permite uma abordagem mais eficiente e abrangente na análise de grandes volumes de dados textuais, como os gerados em

plataformas de mídia social. Essa evolução metodológica é crucial, pois transforma a maneira como os pesquisadores podem abordar e interpretar interações sociais e discursos políticos em ambientes digitais. Com o uso de Aprendizado de Máquina e PLN, torna-se possível processar e analisar conjuntos de dados de tamanho e complexidade que seriam impraticáveis por métodos manuais.

Com isso, tais técnicas automatizadas não apenas ampliam o escopo da pesquisa em Ciência Política Computacional, mas também aumentam a precisão e a profundidade das análises. Por exemplo, a análise de sentimentos e a identificação de padrões e tendências em discussões políticas nas redes sociais tornam-se mais acessíveis e detalhadas. Além disso, a capacidade de analisar rapidamente grandes quantidades de dados permite aos pesquisadores responderem a questões emergentes em tempo real, um aspecto particularmente relevante em estudos que lidam com eventos políticos dinâmicos e em constante mudança.

Um exemplo disso é a tese de mestrado "Aproximación al análisis pragmático en redes sociales mediante el uso de técnicas de enmarcado y Transformers" de Raúl Álvarez Prieto, publicada em 2021, apresenta uma abordagem inovadora para a análise de discursos políticos em redes sociais, especificamente no Twitter. Combinando técnicas avançadas de PLN e Aprendizado de Máquina, usando a linguagem de programação Python para desenvolver um modelo baseado em Transformers, o estudo concentra-se em identificar e analisar as estratégias de enquadramento (framing) utilizadas por diferentes partidos políticos na Espanha, explorando como esses enquadramentos influenciam a percepção pública.

O modelo proposto é treinado para classificar tweets de acordo com diferentes categorias de enquadramento político, como "Monarquia", "República", "Plurinacional" e "Uninacional". A pesquisa analisa uma grande quantidade de tweets para mapear como diferentes partidos políticos e figuras públicas posicionam-se em relação a essas categorias, oferecendo um panorama detalhado do discurso político na Espanha.

O estudo de Prieto (2021) representa uma interseção significativa entre as ciências sociais computacionais e as redes sociais, demonstrando como a combinação de PLN, Aprendizado de Máquina e técnicas computacionais pode fornecer insights profundos sobre padrões de comunicação política e estratégias de enquadramento. Além disso, o uso de modelos baseados em Transformers, uma técnica de ponta em PLN, ilustra o potencial dessas

tecnologias para analisar e interpretar grandes conjuntos de dados textuais, como os gerados em plataformas de mídia social.

Ao considerarmos isoladamente cada eixo do gráfico 5 insights valiosos também surgem. No caso do eixo vertical (y), palavras-chave como "twitter", "social media" e "machine learning" emergem como áreas de foco significativo, indicando os temas que estão atualmente no centro das atenções acadêmicas, mas também como tendências emergentes e possíveis novas direções na pesquisa da Ciência Política Computacional.

No caso do eixo horizontal (x), percebemos a concentração dos métodos em apenas 8 descritores e englobando tanto ferramentas, quanto áreas ou metodologias amplas. Isso se deu porque nos textos analisados, a primeira categorização dos métodos foi baseada nas descrições fornecidas pelos próprios autores em seus trabalhos. Assim, nos artigos que mencionam campos como o PLN, como sua metodologia principal, não era possível inferir com certeza qual linguagem de programação específica foi empregada ou se outras técnicas, como Aprendizado de Máquina ou Análise de Sentimentos, foram integradas. Portanto, a categorização dos métodos acabou trazendo mais descritores amplos do que técnicas de áreas particulares, aplicação de *softwares* específicos e emprego de técnicas de automação, agrupando, inicialmente, todos os métodos declarados sob a mesma categoria.

2.3.1 Os métodos e as técnicas relacionadas

Assim, para investigar mais a fundo como as ferramentas, técnicas e linguagens computacionais estão sendo integradas à Ciências Sociais Computacional e à Ciência Política fizemos uma análise mais pormenorizada dos métodos utilizados ao longo dos anos, o que além de permitir tornar mais nítida as diferenças, permitiu insights valiosos sobre a evolução e a adaptação das práticas na Ciência Política Computacional.

Como mencionado anteriormente, a primeira classificação em cada artigo baseou-se nas descrições fornecidas nos textos. Depois, organizamos cada método em quatro grupos distintos de categorias refletindo diferentes aspectos e abordagens metodológicas.

- a. **Áreas:** Esta categoria abrange campos específicos de estudo e aplicação, como PLN, Aprendizado de Máquina, Análise de Sentimentos e Latent Dirichlet Allocation (LDA). O LDA, em particular, é um modelo estatístico avançado no campo do PLN, utilizado para identificar e categorizar estruturas de tópicos em grandes conjuntos de

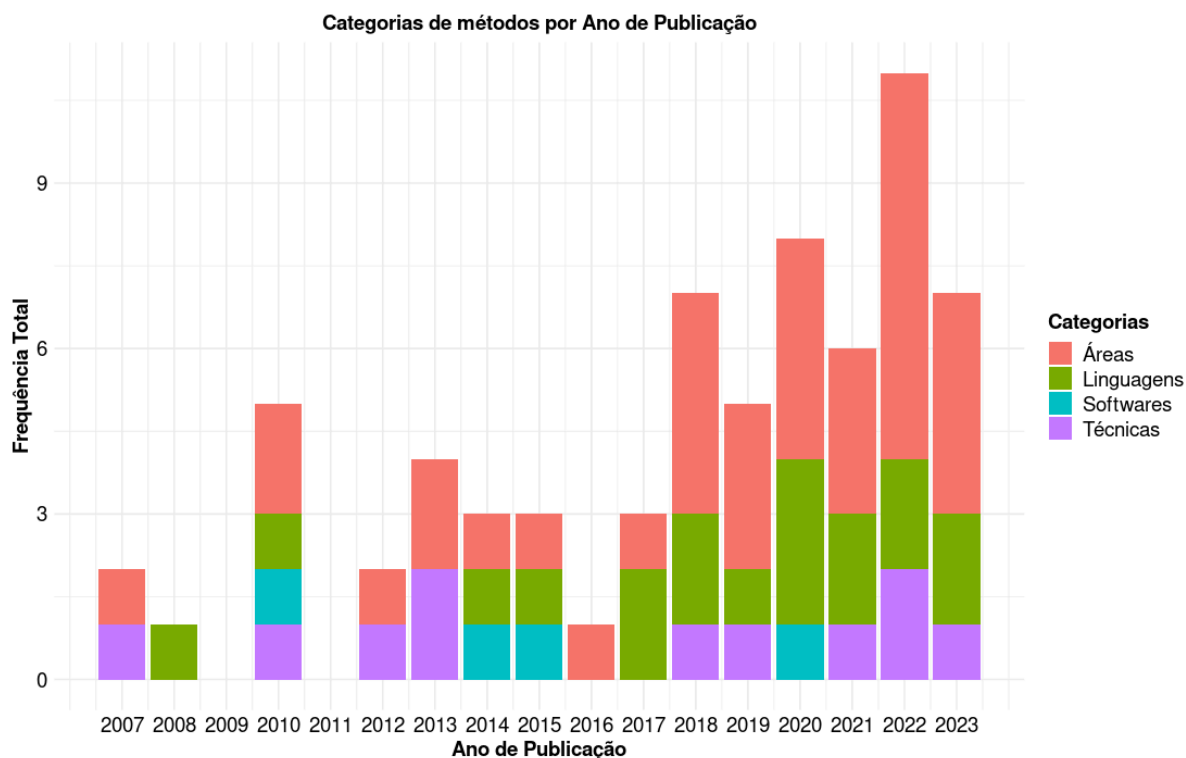
dados textuais. A inclusão dessas áreas permite uma visão detalhada de como abordagens teóricas e conceituais específicas são empregadas ao longo do tempo.

- b. **Linguagens de Programação:** Nesta categoria, são consideradas as linguagens de programação explicitamente mencionadas nos artigos, como R, Python e C++. A presença e a popularidade de determinadas linguagens ao longo dos anos podem indicar tendências na preferência de ferramentas de programação e refletir a evolução tecnológica no campo.
- c. **Softwares:** Inclui ferramentas de *software* específicas mencionadas nos artigos, como Gephi, utilizado para visualização e análise de redes, e NVivo, aplicado na análise de dados qualitativos. A análise da utilização desses *softwares* ao longo do tempo pode revelar preferências por certas ferramentas e a emergência de novas tecnologias no campo.
- d. **Técnicas:** Esta categoria abrange processos computacionais específicos para análise, processamento ou coleta de dados, como web scraping e métodos de automação. A inclusão dessas técnicas oferece uma perspectiva sobre a aplicação prática de métodos computacionais na pesquisa e como eles evoluíram para atender às necessidades de coleta e análise de dados cada vez mais complexos.

Esta abordagem permitiu ainda compreender como as mudanças tecnológicas e as inovações no campo da ciência da computação influenciam diretamente as práticas e as abordagens na pesquisa política.

O **gráfico 7** oferece uma visão detalhada e reveladora da frequência com que elementos das quatro categorias previamente delineadas são mencionados ao longo do tempo. Entretanto, é importante ressaltar que 2009 e 2011 não possuem dados, uma vez que não foram coletados artigos ou pesquisas referentes a esses períodos, limitando a possibilidade de traçar as tendências para nesses períodos específicos.

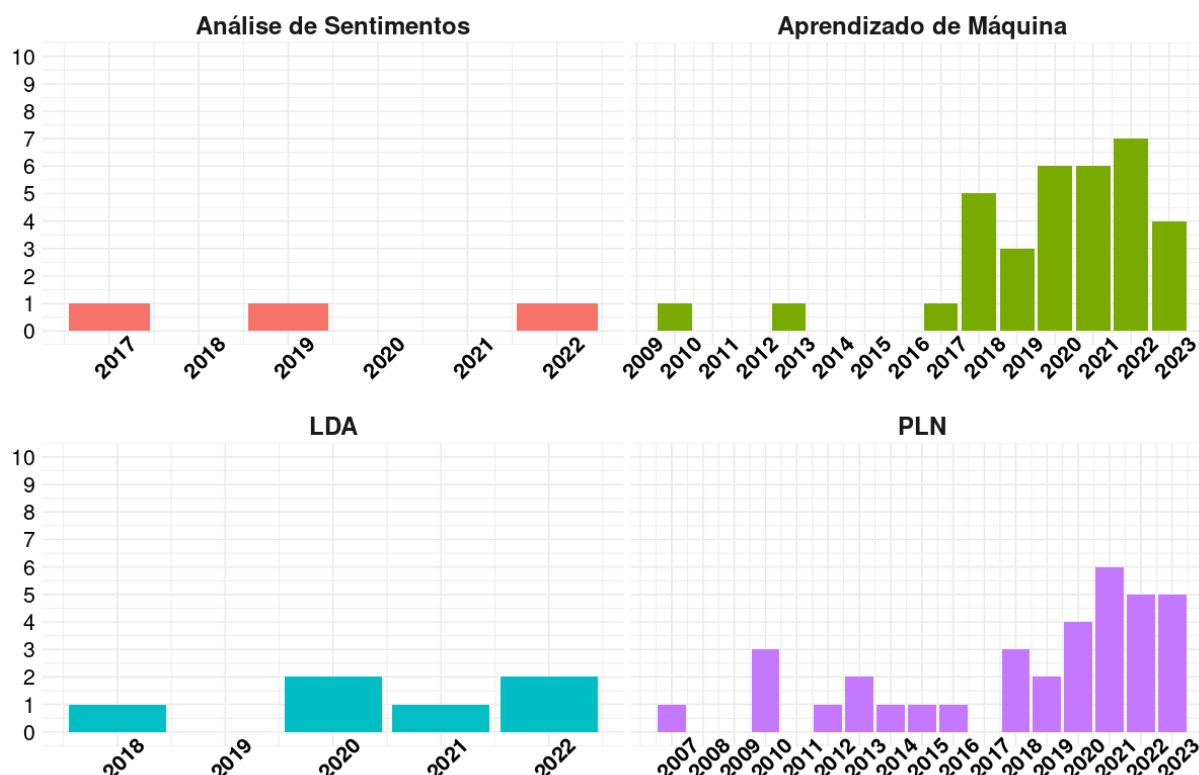
Gráfico 7 - Categorias de métodos por ano de publicação



Fonte: Elaboração própria.

Este gráfico não é apenas uma representação visual, mas também um testemunho da crescente influência das áreas computacionais nas pesquisas em ciência política, com ênfase em campos como PLN e Aprendizado de Máquina. A análise deste gráfico desvenda uma tendência marcante: a ascensão e a consolidação das tecnologias computacionais como ferramentas indispensáveis na pesquisa contemporânea em ciências sociais. Em especial, podemos analisar mais detalhadamente, no **gráfico 8**, a predominância de cara área ao longo dos anos, podendo inferir a popularidade de PNL e Aprendizado de Máquina para tais pesquisas.

Gráfico 8 - Categorias de áreas por ano de publicação



Fonte: Elaboração própria.

Radford e Joseph (2020) exploram esta tendência com profundidade, argumentando persuasivamente que o Aprendizado de Máquina não é apenas uma adição valiosa, mas uma necessidade iminente para as ciências sociais. Eles vão além, sugerindo que as ciências sociais, por sua vez, têm muito a contribuir para o campo do Aprendizado de Máquina. Esta interação bidirecional não é apenas frutífera, mas essencial, pois a teoria social pode enriquecer significativamente a pesquisa em Aprendizado de Máquina, orientando o desenvolvimento de modelos mais precisos e preditivos. A integração da teoria social não é um mero complemento; ela tem o potencial de elevar o desempenho e a relevância dos modelos de Aprendizado de Máquina, tornando-os mais alinhados com as complexidades e nuances dos fenômenos sociais.

Em paralelo, a era atual, descrita por Grimmer et al. (2020) como uma era de abundância de dados, representa um convite e um desafio para os cientistas sociais. Eles são incentivados a repensar e remodelar suas abordagens metodológicas, aproveitando a vasta quantidade de dados disponíveis. Esta era de dados volumosos e acessíveis permite que métodos dedutivos tradicionais sejam complementados, e até mesmo superados, por abordagens mais dinâmicas, sequenciais e interativas. A capacidade dos métodos de

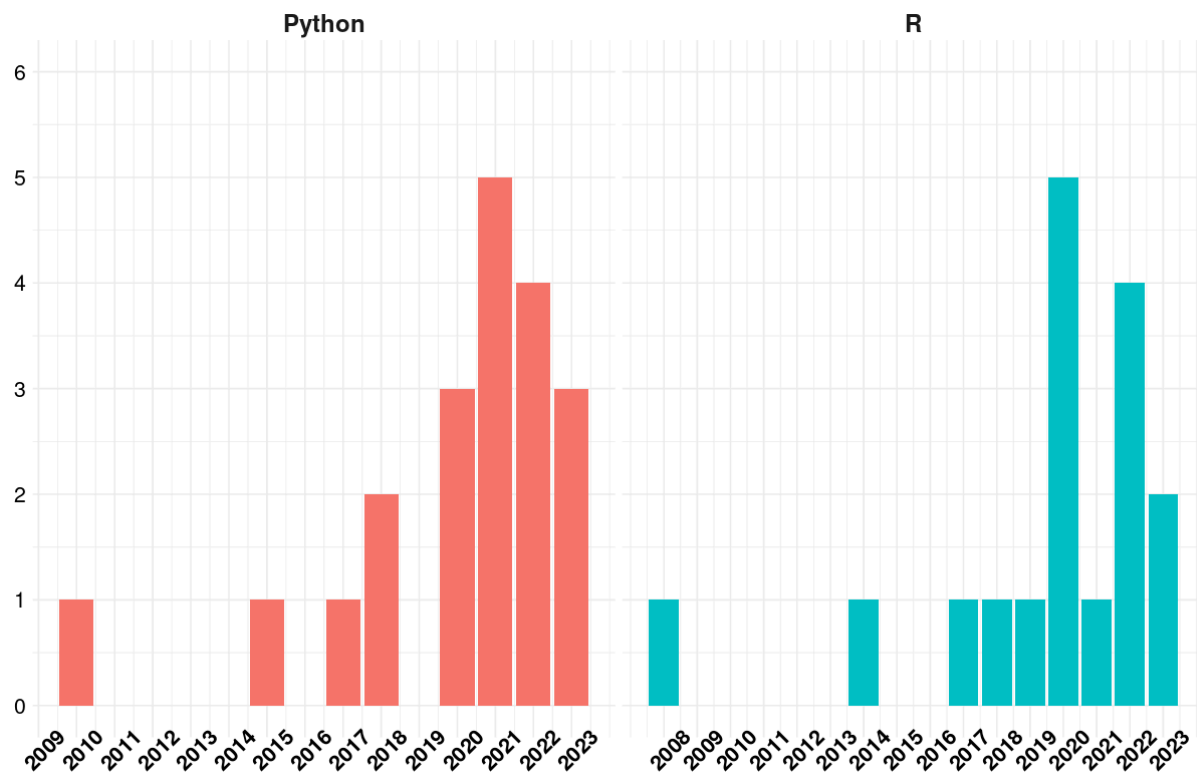
Aprendizado de Máquina de processar e analisar esses grandes conjuntos de dados oferece oportunidades sem precedentes para insights mais profundos e análises mais rigorosas. Além disso, a teoria formal, longe de ser relegada a um segundo plano, assume um papel ainda mais crucial, fornecendo um quadro essencial para a concepção e interpretação de análises complexas, um papel que os dados brutos, por si só, não podem desempenhar.

Neste cenário evolutivo, Chatsiou e Mikhaylov (2020) destacam a trajetória do PLN, um campo que se tornou um pilar na análise de textos em ciências políticas. Desde a década de 1980, impulsionado por avanços no Aprendizado de Máquina e em tecnologias de hardware, o PLN tem se concentrado cada vez mais no uso de modelos estatísticos para decifrar e interpretar padrões na linguagem humana. Esta abordagem, aplicada a vastos conjuntos de textos, tem permitido progressos notáveis em tarefas como classificação de documentos e redução de dimensionalidade, fundamentais para a análise de dados em ciência política.

A incorporação de desenvolvimentos recentes em aprendizado profundo e Aprendizado de Máquina no PLN abriu portas para novas possibilidades de pesquisa, permitindo aos cientistas sociais realizar análises mais profundas e abrangentes de dados textuais. Estes avanços não apenas expandem as fronteiras do conhecimento, mas também redefinem as metodologias e as ferramentas à disposição dos pesquisadores em ciências sociais, marcando uma era de inovação e interdisciplinaridade.

Posto isso, se faz interessante analisar a crescente adesão às linguagens de programação, especificamente o aumento da adesão a linguagens como Python e R ao longo dos anos. No **gráfico 9** é possível observar uma tendência geral de aumento na frequência de menção de ambas as linguagens ao longo dos anos, com um pico significativo em 2021 para Python e em 2022 para R. A popularidade dessas linguagens nas pesquisas do campo da Ciência Política, conforme ilustrado no gráfico, reflete os papéis cada vez mais centrais desempenhados por PLN e Aprendizado de Máquina. A adoção crescente dessas linguagens de programação está intrinsecamente ligada à sua adequação e eficácia em lidar com as demandas específicas dessas áreas tecnológicas avançadas.

Gráfico 9 - Linguagens de programação por ano de publicação



Fonte: Elaboração própria.

Python, por exemplo, destaca-se por sua vasta coleção de bibliotecas especializadas que oferecem acessibilidade para a implementação de modelos complexos de aprendizado de máquina e para a análise avançada de dados textuais, facilitando o desenvolvimento de pesquisas sofisticadas, como as apresentadas neste texto. Por sua vez, R também viu um aumento significativo na sua adoção, particularmente pela sua força estatística e pacotes extensivamente utilizados em análise de dados e modelagem preditiva. A preferência por R em certos círculos acadêmicos deve-se à sua capacidade de realizar análises estatísticas complexas, que são fundamentais para a validação e interpretação de modelos de aprendizado de máquina e para a estruturação de conjuntos de dados em PLN.

A progressão vista nos anos mais recentes do gráfico não apenas denota uma inclinação para a utilização de Python e R, mas também sublinha a sinergia entre as capacidades dessas linguagens e as necessidades das ciências sociais. A capacidade do Python e do R de oferecer soluções robustas e uma gama de funcionalidades para o tratamento e análise de dados volumosos é um dos principais catalisadores para sua adoção. A aderência a essas linguagens é uma resposta direta à busca por ferramentas que não apenas suportem, mas também

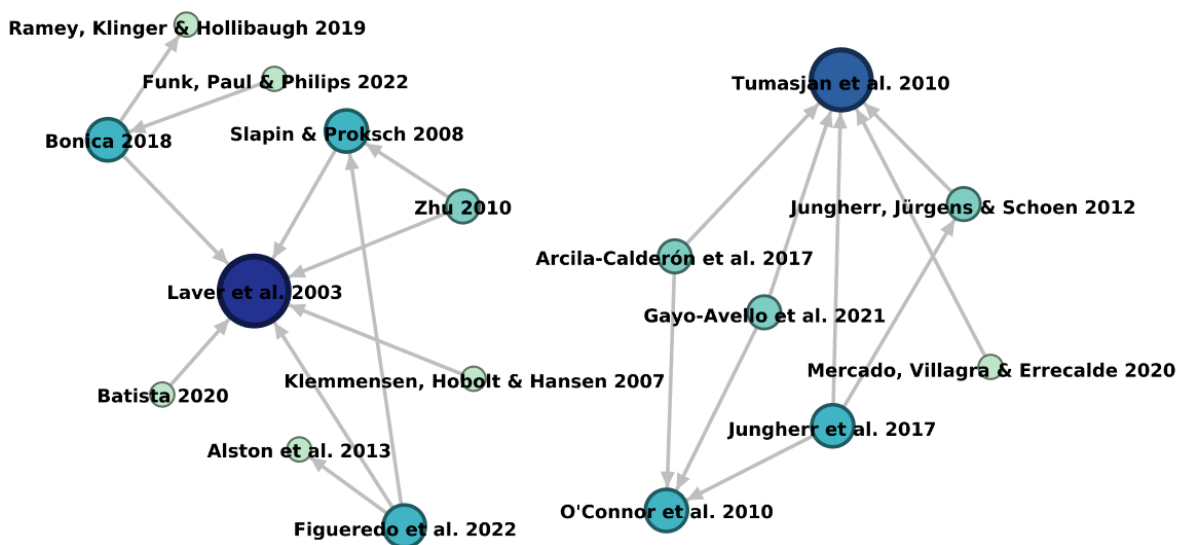
aprimorem o trabalho com PLN e ML, refletindo a integração indispensável de técnicas computacionais avançadas na análise de fenômenos sociais complexos.

2.4 Análise de redes de referências

A análise de redes de citações bibliográficas, nos serve como uma ferramenta valiosa na revisão sistemática de literatura, particularmente no campo da Ciência Política Computacional como uma subárea emergente em pesquisas acadêmicas. A representação gráfica dessas redes permite visualizar as conexões entre trabalhos acadêmicos e identificar tanto os artigos centrais, que exercem uma influência significativa sobre o campo, quanto as relações entre autores e escolas de pensamento.

O **gráfico 10** em questão destaca, por meio de diferentes tamanhos de nós, a frequência com que determinados artigos são citados, sugerindo sua relevância dentro da rede científica. Além disso, ele prioriza a visualização das interações mais significativas, observa-se a escolha deliberada de apresentar 17 nós centrais. Esta seleção foi realizada com o intuito de refinar a rede, excluindo nós que não possuíam conexões com outros, e assim, focalizando nas relações mais robustas e nos artigos mais referenciados. Tal escolha metodológica visa a clareza visual e a relevância analítica, permitindo uma interpretação mais precisa das influências e tendências dentro da Ciência Política Computacional.

Gráfico 10 - Grafo de referências



Fonte: Elaboração Própria

Posto isso, temos como um dos destaques o trabalho de Laver et al. (2003), que aparece com um nó de tamanho e tom de azul proeminente, o que implica uma centralidade e,

possivelmente, um papel pioneiro e fundamental na consolidação da Ciência Política Computacional, dado sua data de publicação mais antiga. Isso pode ser indicativo de que o artigo estabeleceu metodologias ou teorias que são amplamente referenciadas e utilizadas em estudos subsequentes. O trabalho supracitado, “Extracting Policy Positions from Political Texts Using Words as Data”, publicado em 2003 pela revista acadêmica *American Political Science Review*, foi elaborado por Michael Laver, Kenneth Benoit and John Garry para introduzir uma nova metodologia para extrair posições políticas de textos, tratando-os não como discursos a serem interpretados, mas como dados na forma de palavras (Laver et al., 2003). A inovação deste estudo reside em sua abordagem "cega para a linguagem" que quantifica a frequência das palavras para estimar posições políticas, diferindo de métodos anteriores que exigiam interpretação humana ou algorítmica baseada em dicionários de codificação pré-definidos.

O estudo expande a aplicabilidade de sua abordagem para além da análise de manifestos partidários, abordando também a estimativa de posições políticas a partir de discursos legislativos. Essa técnica de pontuação de palavras, segundo os autores, replica com sucesso as estimativas de políticas publicadas sem os custos substanciais de tempo e mão-de-obra que esses métodos anteriores exigiam. Com isso, o avanço metodológico proposto foi viabilizado pelo desenvolvimento de um *software* específico, intitulado WORDSCORES, concebido para processar e analisar grandes volumes de texto de maneira eficiente. O *software* possibilita a contagem de palavras e a realização de cálculos subsequentes de forma rápida e simultânea, permitindo assim a replicação e a aplicação dos métodos delineados pelos autores com maior agilidade e precisão. A utilização do WORDSCORES, que é descrito como de fácil uso e livremente acessível (Laver et al., 2003), reflete a preocupação dos autores com a replicabilidade e a disseminação de sua técnica inovadora na comunidade científica. Essas escolhas tecnológicas e metodológicas demonstram o potencial de integração entre ciência política e computação, e a capacidade de trazer insights quantitativos robustos para o campo das ciências sociais.

Da mesma forma, Tumasjan et al. (2010) também se destaca, sugerindo que este trabalho pode ter sido um ponto de inflexão ou uma contribuição significativa para a pesquisa na área. Analisando em detalhes, o trabalho em questão, “Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment”, é um estudo de 2010, conduzido por Andranik Tumasjan, Timm Sprenger, Philipp Sandner e Isabell Welpe, que teve como objetivo primordial investigar a potencialidade da plataforma X, antes conhecida como

Twitter, como um preditor de resultados eleitorais, fundamentando-se na premissa de que as menções e o sentimento político expressos na rede social em questão poderiam refletir as tendências eleitorais. A metodologia empregada focou na análise de 104.003 *tweets* políticos veiculados antes das eleições nacionais na Alemanha de 2009, utilizando-se do *software* LIWC2007 para aferição automática do sentimento. Este *software*, especializado na análise de componentes emocionais, cognitivos e estruturais dos textos, baseia-se em um dicionário interno psicometricamente validado para medir a prevalência de palavras-chave que se alinham a dimensões psicológicas específicas.

A conclusão do estudo apontou para a correspondência entre a frequência de menções dos partidos no Twitter e as preferências eleitorais, sugerindo que dados de mídias sociais podem efetivamente antecipar comportamentos eleitorais. Este achado é de significativa relevância para a Ciência Política Computacional, pois demonstra como métodos computacionais e a análise de grandes volumes de dados podem ser utilizados para entender e prognosticar dinâmicas políticas. Dessa forma, tal pesquisa estabeleceu um marco ao evidenciar a aplicabilidade de técnicas computacionais na análise de fenômenos sociais e políticos, incentivando o desenvolvimento de um novo campo interdisciplinar que une análises quantitativas a estudos sociais e políticos, como demonstrado ao se destacar como o segundo nó melhor referenciado dentro da base de artigos do atual artigo.

Assim, a contribuição do trabalho para a Ciência Política Computacional é indiscutível, promovendo uma nova compreensão sobre como os sentimentos políticos e as tendências eleitorais podem ser monitorados e previstos através das mídias sociais. Tal abordagem inovadora nitidamente estimulou o avanço de pesquisas subsequentes na área, reforçando o valor da integração entre computação e ciências políticas e pavimentando o caminho para futuras investigações que possam explorar o vasto potencial dos dados digitais no entendimento do comportamento político e eleitoral.

Paralelamente, o artigo “A Scaling Model for Estimating Time-Series Party Positions from Texts”, de Jonathan B. Slapin e Sven-Oliver Proksch, publicado em 2008, se destaca por mencionar Laver et al. (2003) e por ser mencionado por dois artigos que também mencionam a pesquisa supracitada. Ele apresenta um modelo estatístico inovador para estimar posições políticas de partidos ao longo do tempo, a partir de textos políticos, como manifestos partidários. Este método, denominado WORDFISH, é baseado na análise de frequência de

palavras nos textos, partindo do pressuposto de que o uso relativo das palavras pelos partidos fornece informações sobre sua localização em um espaço político.

A principal inovação do WORDFISH reside em sua capacidade de produzir estimativas temporais das posições políticas sem a necessidade de textos de referência, assumindo uma distribuição estatística subjacente das contagens de palavras. Este método permite utilizar todas as palavras em cada documento e estimar a importância de cada uma delas, oferecendo uma visão detalhada de quais palavras são importantes para posicionar os partidos no espectro político esquerda-direita. A implementação do WORDFISH foi realizada por meio de um programa de computador escrito para a linguagem estatística R, facilitando assim sua aplicação e replicabilidade em pesquisas futuras. Assim, o trabalho de Slapin e Proksch (2008) se destaca por avançar no uso de métodos computacionais para análises políticas, complementando e expandindo abordagens anteriores, como a de Laver et al. (2003). Ele oferece um exemplo claro de como a interseção entre a Ciência Política Computacional e a análise de textos pode enriquecer o entendimento das posições e dinâmicas partidárias ao longo do tempo.

De forma similar, a pesquisa de O'Connor (2010) também merece atenção pela menção de três artigos em comum com Tumasjan et al. (2010). Publicado por Brendan O'Connor, Ramnath Balasubramanyan, Bryan R. Routledge e Noah A. Smith, o artigo "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series" aborda uma metodologia inovadora para analisar a relação entre a opinião pública e o sentimento expresso em textos, especificamente no Twitter. Este estudo destaca como a Ciência Política Computacional pode se beneficiar da análise de grandes volumes de dados textuais, oferecendo uma perspectiva única sobre a opinião pública.

A metodologia empregada envolveu a coleta de dados do Twitter, totalizando cerca de 1 bilhão de mensagens postadas durante os anos de 2008 e 2009. Utilizando a API do Twitter e o fluxo de dados em tempo real, "gardenhose", os pesquisadores coletaram um conjunto representativo de mensagens públicas, variando de 100.000 a 7 milhões de mensagens por dia. Para a análise de sentimento, empregaram um lexicon de subjetividade do OpinionFinder, que inclui palavras marcadas como positivas e negativas. Cada mensagem foi classificada como positiva ou negativa com base na presença dessas palavras.

Os autores compararam a frequência de palavras de sentimento relacionadas a tópicos específicos no Twitter com dados de pesquisas de opinião pública sobre confiança do consumidor e opinião política. Eles descobriram que a frequência de palavras de sentimento no Twitter correlacionava-se significativamente com as tendências observadas nas pesquisas de opinião pública. Além disso, a análise revelou que a informação de sentimento obtida do Twitter poderia até prever movimentos futuros nas pesquisas de opinião. Este estudo é mais um exemplo de técnicas computacionais avançadas, como a Análise de Sentimentos e o PLN, sendo aplicadas para entender e prever fenômenos sociais e políticos. A pesquisa de O'Connor et al. (2010) demonstra a viabilidade e o valor de usar dados de redes sociais para complementar e, em alguns casos, substituir métodos tradicionais de coleta de opinião pública, como pesquisas e enquetes.

Posto isso, a rede de citações fornece uma perspectiva única sobre o desenvolvimento intelectual da Ciência Política Computacional, permitindo que fossem identificadas duas pesquisas de mais de uma década atrás, influenciando e se conectando com pesquisas contemporâneas ou mais recentes, além das contribuições fundamentais quanto às emergentes no campo. A análise dessa rede informa sobre a trajetória da disciplina e as tendências atuais, haja vista as metodologias utilizadas pelas pesquisas em destaque no grafo de redes das referências e como ganharam atenção. Dessa forma, introduz-se uma narrativa coerente sobre a evolução do conhecimento na Ciência Política Computacional, destacando os trabalhos que ajudaram a formar a espinha dorsal do campo e os avanços que moldaram seu futuro.

CONSIDERAÇÕES FINAIS

A CSC e a CPC têm demonstrado um potencial significativo para a análise de dados complexos e a geração de insights relevantes. Nesse cenário, a introdução do *big data* no cenário acadêmico, como destacado por Monroe et al. (2014), representa uma mudança significativa e uma oportunidade notável para as ciências sociais. Adicionalmente, esta mudança reflete também as limitações dos métodos tradicionais, que muitas vezes não conseguiam captar a complexidade das interações humanas. Agora, com a CSC e a CPC, os pesquisadores podem realizar análises mais precisas e abrangentes, acessando e interpretando dados de maneiras antes inimagináveis. Isso não apenas fomenta uma compreensão mais profunda das dinâmicas sociais e políticas, mas também realça a importância de integrar a teoria formal na seleção e interpretação de dados em Ciência Política. Especificamente, a CPC demonstra como a combinação de teorias tradicionais com técnicas computacionais inovadoras pode enriquecer a análise das complexidades do sistema político.

Além disso, o impacto da CSC e da CPC estende-se para além da metodologia, remodelando o tecido das teorias sociais e políticas, uma vez que o influxo de novas formas de dados digitais está desafiando conceitos tradicionais e permitindo o desenvolvimento de teorias mais sofisticadas. Apesar dos avanços tecnológicos, a incorporação de métodos computacionais em ciências políticas não substitui, mas sim complementa e enriquece as abordagens tradicionais. O acesso a grandes conjuntos de dados, como aqueles oferecidos por redes sociais e pesquisas online, abre novos caminhos para compreender motivações e estratégias políticas. Na medida que esses campos evoluem, sua aplicação e importância na nossa sociedade digitalizada continuam a crescer, prometendo novas fronteiras para a pesquisa em ciências sociais e políticas. Em suma, a integração do *big data* nas ciências sociais representa uma oportunidade de expansão e transformação significativa para o campo (Monroe et al., 2014).

Posto isso, este estudo, abordando a influência crescente da CSC e da CPC, centrou-se na identificação das técnicas e temas mais recorrentes na área, com o objetivo de compreender como as técnicas computacionais são aplicadas ao estudo de fenômenos políticos. Uma das descobertas mais significativas, evidenciada pela análise temporal dos textos publicados, é o aumento substancial na produção de artigos a partir de 2018, com um pico notável em 2020, sugerindo um crescente interesse e investimento no campo, potencialmente impulsionado por eventos globais e avanços tecnológicos. A tendência ascendente continuou nos anos seguintes,

com 2022 emergindo como o ano mais prolífico em termos de publicações. Esses resultados quantitativos não apenas refletem um vigor na área, mas também indicam uma expansão e reconhecimento crescentes da CPC na esfera acadêmica internacional.

A análise das palavras-chave dos artigos revelou tendências interessantes em termos de tópicos e metodologias. Uma parcela considerável dos artigos não especificava palavras-chave, destacando variações nas práticas de catalogação. No entanto, entre os que as incluíam, termos como "computational social science" e "social media" em inglês, e "processamento de linguagem natural" e "aprendizado de máquina" em português, foram predominantes. Essas tendências refletem uma forte intersecção entre o uso de mídias sociais e análise computacional, bem como um foco em metodologias avançadas de análise de dados. Além disso, a ênfase em termos como "twitter" em artigos em espanhol aponta para uma concentração em plataformas específicas de mídia social. Este mosaico de termos não só sugere áreas de interesse crescente, mas também revela como abordagens e focos podem variar em contextos regionais e linguísticos diferentes.

Outro aspecto relevante do estudo foi o uso de um mapa de calor para identificar coocorrências de métodos e palavras-chave. Esta análise revelou fortes associações entre temas como "Twitter" e "Social Media" e métodos como "PLN", "Aprendizado de Máquina" e "Automação". Essas coocorrências sinalizam uma integração significativa entre métodos computacionais avançados e áreas temáticas dentro da CPC, indicando como técnicas como PLN e aprendizado de máquina estão transformando a pesquisa em subáreas da ciência política, como análise de sentimentos, identificação de padrões e tendências em discussões políticas nas redes sociais. Este estudo, ao detalhar a aplicação e a evolução dessas técnicas na CPC, não apenas fornece uma visão aprofundada das tendências atuais, mas também destaca o potencial de transformação dessas tecnologias na pesquisa política contemporânea.

Ademais, a metodologia adotada, uma análise cientométrica da literatura internacional em CPC, destacou a intersecção entre técnicas computacionais avançadas e abordagens tradicionais em ciências sociais. Foi possível observar como as técnicas computacionais estão sendo aplicadas ao estudo do comportamento, da comunicação e da predição política, no geral. Ferramentas como Zotero e RStudio, juntamente com a análise visual por meio de gráficos de frequências e coocorrência gerados pelo Google Sheets e pelo pacote ggplot2, além de análises de rede no Gephi, evidenciou a eficácia dos métodos computacionais

adotados na geração de percepções aprofundadas e descobertas inovadoras que seriam desafiadoras de se obter através de métodos convencionais.

A pesquisa revelou uma diversidade de campos nos artigos selecionados, refletindo que a presença internacional na CPC tem se espalhado em áreas mais aderentes -, bem como que já há contribuição das comunidades de pesquisa latino-americanas na crescente adesão aos métodos computacionais. Termos como “aprendizado de máquina” e "processamento de linguagem natural", presentes em diferentes idiomas, também enfatizaram a relevância dessas técnicas computacionais na pesquisa política contemporânea.

Os resultados analisados destacaram o dinamismo e a vitalidade da Ciência Política Computacional. A integração entre pesquisadores tradicionais e adeptos das técnicas computacionais mostrou-se fundamental para promover análises mais robustas e abrangentes, enriquecendo a pesquisa política com novas perspectivas e abordagens. E a ascensão constante nas publicações, identificada nesta pesquisa, sinaliza a influência de fatores externos, como mudanças no cenário político global ou inovações tecnológicas, que podem ter impulsionado a pesquisa e o conhecimento na área. Além disso, a necessidade de padronização na catalogação de artigos para aprimorar a precisão e utilidade das bases de dados, como a utilizada neste estudo, é crucial para construir a Ciência Política Computacional como uma comunidade coesa.

A análise de redes de citações bibliográficas, em especial, serviu como uma ferramenta valiosa na revisão sistemática da literatura, especialmente útil no campo emergente da CPC. Esta representação gráfica facilitou a visualização das conexões entre trabalhos acadêmicos, permitindo identificar tanto os artigos centrais, que exercem influência significativa no campo, quanto as relações entre autores e questões centrais da área. O gráfico destacou, por meio de diferentes tamanhos de nós, a frequência com que certos artigos são citados, sugerindo sua relevância dentro da rede científica. A seleção de 17 nós centrais priorizou as interações mais significativas, focando nas relações mais robustas e nos artigos mais referenciados para clareza visual e relevância analítica.

O trabalho de Laver et al. (2003) se destacou como um nó proeminente, indicando sua centralidade e possível papel pioneiro na consolidação da Ciência Política Computacional. Este artigo introduziu uma nova metodologia para extrair posições políticas de textos, utilizando uma abordagem inovadora que trata textos não como discursos para interpretação,

mas como dados na forma de palavras. O estudo de Tumasjan et al. (2010), por sua vez, também se destacou, sugerindo uma contribuição significativa para a pesquisa na área, ao explorar o Twitter como um potencial preditor de resultados eleitorais.

A integração de abordagens inovadoras e interdisciplinares na pesquisa política, evidenciada por esses estudos, sublinha o potencial dos dados digitais no entendimento do comportamento político e eleitoral. A rede de citações informa sobre a trajetória da disciplina e tendências atuais, introduzindo uma narrativa coerente sobre a evolução do conhecimento na Ciência Política Computacional e destacando os trabalhos que ajudaram a formar a espinha dorsal do campo.

Em conclusão, a CSC e a CPC têm o potencial de revolucionar a forma como as ciências sociais e políticas são conduzidas, equilibrando o valor das técnicas computacionais com a apreciação e respeito pelas tradições e princípios fundamentais das ciências sociais. O futuro destas disciplinas depende de uma abordagem equilibrada que reconheça tanto o valor das técnicas computacionais quanto a importância das tradições e princípios das ciências sociais, bem como o papel do cientista social na construção da pesquisa. A colaboração entre as ciências sociais tradicionais e as ciências da computação é essencial para enfrentar eficazmente os desafios emergentes, e a formação dos futuros cientistas sociais deve ser adaptada para incorporar habilidades computacionais e de codificação. Este treinamento permitirá aos pesquisadores manipular e analisar os grandes volumes de dados digitais disponíveis, proporcionando um entendimento mais profundo do contexto social subjacente a esses dados.

Finalmente, o estudo reitera a necessidade de maior engajamento com outras disciplinas, onde a CSC e a CPC continuam a prosperar. Com o avanço da pesquisa em campos como ciência da computação e engenharia, os sociólogos têm a oportunidade de desempenhar um papel central nessa conversa interdisciplinar. Este papel é crucial não apenas porque a CSC e a CPC têm muito a oferecer à sociologia, mas também porque muitos dos problemas estudados pelos cientistas sociais computacionais são inerentemente sociológicos. A CSC e a CPC, portanto, têm o potencial de trazer uma revolução significativa na maneira como as ciências sociais e políticas são abordadas e entendidas.

Embora tenham sido feitos progressos notáveis, há oportunidades claras para melhorias e direcionamentos futuros, já que o mapeamento e a categorização eficientes das áreas e

técnicas mais populares nessas disciplinas é fundamental para o avanço contínuo do campo. Isso não só facilitará a identificação de lacunas de pesquisa e áreas emergentes, mas também permitirá uma compreensão mais precisa das tendências e evoluções dentro desses dois campos. O desenvolvimento de um sistema de catalogação mais padronizado e abrangente ajudará na organização e no acesso aos dados, tornando mais fácil para os pesquisadores localizar e utilizar informações relevantes, além de consolidar ainda mais a área de pesquisa.

Como discutido, a implementação de ferramentas computacionais avançadas para a análise de grandes volumes de dados pode enriquecer significativamente as pesquisas em ciências sociais. Assim, técnicas como aprendizado de máquina e processamento de linguagem natural, que já se mostraram valiosas na pesquisa política contemporânea, podem ser exploradas ainda mais. A integração dessas técnicas com métodos tradicionais das ciências sociais proporcionará uma análise mais profunda e nuances mais ricas, oferecendo novas perspectivas e compreensões. Além disso, a promoção de uma abordagem interdisciplinar, combinando as competências e insights de várias disciplinas, pode impulsionar inovações e soluções criativas para questões complexas.

Olhando para o futuro, é crucial que os programas de formação em ciências sociais incorporem treinamento em habilidades computacionais e de codificação. Isso permitirá que os futuros cientistas sociais aproveitem efetivamente as tecnologias emergentes e os vastos conjuntos de dados digitais disponíveis. Ao mesmo tempo, é importante manter um equilíbrio entre a adoção de novas tecnologias e a preservação dos princípios fundamentais das ciências sociais. Além disso, para garantir que as análises baseadas em big data sejam conduzidas com rigor intelectual, é necessário que se mantenha um engajamento mais profundo com teorias e métodos das ciências sociais. Isso ajudará a garantir que as conclusões derivadas dos dados sejam robustas e confiáveis. Assim, a colaboração contínua interdisciplinar será vital para enfrentar os desafios emergentes de um mundo cada vez mais orientado por dados.

Em conclusão, é imperativo continuar promovendo o diálogo interdisciplinar, haja vista que não apenas enriquecerá a pesquisa em CSC e CPC, mas também garantirá que essas disciplinas se mantenham relevantes e aplicáveis aos problemas sociais e políticos contemporâneos. A CSC e a CPC têm um grande potencial para transformar as ciências sociais e políticas, e a exploração plena desse potencial depende da capacidade dos pesquisadores de integrar eficazmente as técnicas computacionais com as ricas tradições teóricas e metodológicas das ciências sociais.

REFERÊNCIAS BIBLIOGRÁFICAS

ABADÍA, A.; MANFREDI, L. C.; SAYAGO, J. T. Comunicación de crisis durante la pandemia del Covid-19 y su impacto en los sentimientos de la ciudadanía. **Opinião Pública**, v. 29, n. 1, p. 199–225, abr. 2023.

ALEGRE, P. O USO DA INTELIGÊNCIA ARTIFICIAL NAS CIÊNCIAS SOCIAIS: o caso do patriotismo dos brasileiros. 2020.

ANJOS, M. F. D. ANÁLISE DE SENTIMENTOS EM TEXTOS CURTOS DIRIGIDOS À ELEIÇÃO PRESIDENCIAL DO BRASIL EM 2018. 2018.

ARCILA-CALDERÓN, C. et al. Análisis supervisado de sentimientos políticos en español: clasificación en tiempo real de tweets basada en aprendizaje automático. **El Profesional de la Información**, v. 26, n. 5, p. 973, 14 set. 2017.

ARGYLE, L. P. et al. Leveraging AI for democratic discourse: Chat interventions can improve online political conversations at scale. **Proceedings of the National Academy of Sciences**, v. 120, n. 41, p. e2311627120, 10 out. 2023.

AYUSO, L. Análisis de sentimientos aplicado a la opinión política en Twitter: un sistema de clasificación en tiempo real. 2019.

BAIL, C. A. Combining natural language processing and network analysis to examine how advocacy organizations stimulate conversation on social media. **Proceedings of the National Academy of Sciences**, v. 113, n. 42, p. 11823–11828, 18 out. 2016.

BAIL, C. A. et al. Assessing the Russian Internet Research Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017. **Proceedings of the National Academy of Sciences**, v. 117, n. 1, p. 243–250, 7 jan. 2020.

BAIL, C. A.; MERHOUT, F.; DING, P. Using Internet search data to examine the relationship between anti-Muslim and pro-ISIS sentiment in U.S. counties. **Science Advances**, v. 4, n. 6, p. eaao5948, jun. 2018.

BEIELER, John et al. Generating political event data in near real time. **Computational social science**, p. 98, 2016.

BATISTA, M. QUAIS POLÍTICAS IMPORTAM? Usando ênfases na agenda legislativa para mensurar saliência. **Revista Brasileira de Ciências Sociais**, v. 35, n. 104, p. e3510411, 2020.

CALVO, D.; CAMPOS-DOMÍNGUEZ, E.; DÍEZ-GARRIDO, M. Hacia una campaña computacional: herramientas y estrategias online en las elecciones españolas. **Revista Española de Ciencia Política**, v. 51, p. 123–154, 29 nov. 2019.

CANDIDO, M. R.; JÚNIOR, J. F.; CAMPOS, L. A. Desigualdades na elite da Ciência Política brasileira. *Civitas* - **Revista de Ciências Sociais**, v. 19, n. 3, p. 564, 13 dez. 2019.

CARDOSO, L. F. C. Análise de Sentimento usando Abordagem Léxica de Discursos do Senado Federal. 2022.

CONTE, Rosaria et al. Manifesto of computational social science. **The European Physical Journal Special Topics**, v. 214, p. 325-346, 2012.

EDELMANN, Achim et al. Computational social science and sociology. **Annual Review of Sociology**, v. 46, p. 61-81, 2020.

ELLIOTT, E. W.; HO, K.; HOLMES, J. S. Political Science Computing: A Review of Trends in Computer Evolution and Political Science Research. **Journal of Information Technology & Politics**, v. 6, n. 2, p. 166–175, 12 maio 2009.

FERETTI, V. E.; JUNCKES, I. J.; CLEMENTE, A. J. Ciência Política e análise de redes: uma metodologia para o mapeamento de comunidades temáticas. **Guaju**, v. 4, n. 2, p. 229, 18 dez. 2018a.

FIGUEREDO, F. C. D.; MUELLER, B.; CAJUEIRO, D. O. A natural language measure of ideology in the Brazilian Senate. **Revista Brasileira de Ciência Política**, n. 37, p. e246618, 2022.

GAYO-AVELLO, D.; METAXAS, P.; MUSTAFARAJ, E. Limits of Electoral Predictions Using Twitter. **Proceedings of the International AAAI Conference on Web and Social Media**, v. 5, n. 1, p. 490–493, 3 ago. 2021.

GENTZKOW, M.; SHAPIRO, J. M. Ideological Segregation Online and Offline *. **The Quarterly Journal of Economics**, v. 126, n. 4, p. 1799–1839, nov. 2011.

GONZÁLEZ-BAILÓN, Sandra. Social science in the era of big data. **Policy & internet**, v. 5, n. 2, p. 147-160, 2013.

INSTITUTO DE CIENCIAS DE LA COMPUTACIÓN, UNIVERSIDAD DE BUENOS AIRES, BUENOS AIRES, ARGENTINA et al. Aprendizaje automático para el análisis cross-plataforma de la comunicación política: Gobierno y oposición argentinos en Facebook, Instagram y Twitter. **Cuadernos.info**, n. 55, p. 256–280, 2023.

JACOBS, Jeff. **"Our Word Is Our Weapon": Text-Analyzing Wars of Ideas from the French Revolution to the First World War**. Columbia University, 2022.

JOSÉ MANUEL MAGALLANES. Reflexiones metodológicas para la Ciencia Política del siglo XXI. **Politai**, v. 1, n. 0, 1 jun. 2010.

JUNGHERR, A. et al. Digital Trace Data in the Study of Public Opinion: An Indicator of Attention Toward Politics Rather Than Political Support. **Social Science Computer Review**, v. 35, n. 3, p. 336–356, jun. 2017.

JUNGHERR, A.; JÜRGENS, P.; SCHOEN, H. Why the Pirate Party Won the German Election of 2009 or The Trouble With Predictions: A Response to Tumasjan, A., Sprenger, T. O., Sander, P. G., & Welpe, I. M. "Predicting Elections With Twitter: What 140 Characters Reveal About Political Sentiment". **Social Science Computer Review**, v. 30, n. 2, p. 229–234, maio 2012.

JUSTIN GRIMMER, MARGARET E. ROBERTS, BRANDON M. STEWART. Text as Data: A New Framework for Machine Learning and the Social Sciences. **Digital Scholarship in the Humanities**, v. 38, n. 1, p. 456–458, 3 abr. 2023.

KEUSCHNIGG, M.; LOVSJÖ, N.; HEDSTRÖM, P. Analytical sociology and computational social science. **Journal of Computational Social Science**, v. 1, p. 3–14, 2018.

LAVER, M.; BENOIT, K.; GARRY, J. Extracting Policy Positions from Political Texts Using Words as Data. **American Political Science Review**, v. 97, n. 02, maio 2003.

LAZER, David et al. Computational social science. **Science**, v. 323, n. 5915, p. 721-723, 2009.

LAZER, David MJ et al. Computational social science: Obstacles and opportunities. **Science**, v. 369, n. 6507, p. 1060-1062, 2020.

MAGUIRE, T. E. B. Aprendizaje automático y modelización de tópicos : un estudio de caso sobre la agenda mediática en contexto de las elecciones Argentina 2015. 2021.

MARQUES, T. L. Mineração de Textos no Twitter: Aprendizado de Máquina para Classificação da Opinião Política dos Usuários. 2022.

MCCABE, Stefan D. **Essays on the Measurement of Online Behavior**. 2022. Tese de Doutorado. Northeastern University.

MERCADO, V.; VILLAGRA, A.; ERRECALDE, M. Análisis Exploratorio de los Datos para la Identificación de la Alineación Política de Periodistas Argentinos. **Informes Científicos Técnicos - UNPA**, v. 12, n. 3, p. 203–218, 14 dez. 2020.

MONROE, Burt L. et al. No! Formal theory, causal inference, and big data are not contradictory trends in political science. **PS: Political Science & Politics**, v. 48, n. 1, p. 71-74, 2015.

MORENO-MERCADO, J. M.; GARCÍA-MARÍN, J.; LUENGO, Ó. G. Conflictos armados y la construcción de narrativas a través de Twitter. El caso de la guerra entre Armenia y Azerbaiyán. **Revista Española de Ciencia Política**, v. 59, p. 89–114, 29 jul. 2022.

MOTA, B. M. F. Métodos quantitativos em Ciência Política Quantitative methods in Political Science Métodos cuantitativos en Ciencia Política. **Sociedade e Cultura.**, 2022.

MOURA, T. Monitorando e entendendo a eleição brasileira por meio de Processamento de Linguagem Natural. 2019.

MUELLER, H.; RAUH, C. Reading Between the Lines: Prediction of Political Violence Using Newspaper Text. **American Political Science Review**, v. 112, n. 2, p. 358–375, maio 2018.

DE OLIVEIRA, Nicollas R. et al. Processamento de linguagem natural para identificação de notícias falsas em redes sociais: ferramentas, tendências e desafios. **Sociedade Brasileira de Computação**, 2020.

PAVESI, Patrícia P.; VALENTIM, Julio. Ciências Sociais Computacionais: um novo paradigma para as Ciências Sociais?. **Simbiótica. Revista Eletrônica**, v. 8, n. 4, p. 1-16, 2021.

PEW RESEARCH CENTER. News Use Across Social Media Platforms 2017. 2017. Disponível em: <https://www.pewresearch.org/journalism/2017/09/07/news-use-across-social-media-platforms-2017/>. Acesso em: 22 dez. 2023.

POLATO, R.; BALANIUK, R.; FERNEDA, E. Coesão partidária na câmara dos deputados: uma abordagem com ciência de dados. **Revista Eletrônica de Ciência Política**, v. 12, n. 2, 30 mar. 2022a.

POLATO, R.; BALANIUK, R.; FERNEDA, E. Coesão partidária na câmara dos deputados: uma abordagem com ciência de dados. **Revista Eletrônica de Ciência Política**, v. 12, n. 2, 30 mar. 2022b.

POUSHTER, Jacob; BISHOP, Caldwell; CHWE, Hanyu. Social media use continues to rise in developing countries but plateaus across developed ones. **Pew research center**, v. 22, p. 2-19, 2018./

PRIETO, R. A. Aproximación al análisis pragmático en redes sociales mediante el uso de técnicas de enmarcado y Transformers. 2021.

RIBEIRO, André Luís et al. Processamento de linguagem natural aplicado à classificação de decretos administrativos brasileiros. In: **Anais Estendidos do XXVI Simpósio Brasileiro de Sistemas Multimídia e Web**. SBC, 2020. p. 57-60.

ROSENDO, A. Partidos brasileiros refletem a opinião dos senadores? 2023.

SERRANO CONTRERAS, I. J. Medios de Comunicación y Polarización; Un Análisis del Feminismo en España mediante Procesamiento del Lenguaje Natural. Universidad de Granada, , 2021. Disponível em: <<http://hdl.handle.net/10481/71409>>

SILVA, D. C. D.; HERNÁNDEZ, L. G. Aplicação metodológica da análise de conteúdo em pesquisas de análise de política externa. **Revista Brasileira de Ciência Política**, n. 33, p. e218584, 2020.

SISKOU, Wassiliki et al. Automated Detection and Annotation for Calls to Action in Latin-American Social Media Postings. 2022.

SLAPIN, J. B.; PROKSCH, S. A Scaling Model for Estimating Time-Series Party Positions from Texts. **American Journal of Political Science**, v. 52, n. 3, p. 705–722, jul. 2008.

SOARES, K. T. Ciência de Dados Sociais: uma visão sociológica das Humanidades Digitais. **Simbiótica Revista Eletrônica**, v. 8, n. 4, p. 46–70, 15 jan. 2022.

TÖRNBERG, Petter et al. Simulating social media using large language models to evaluate alternative news feed algorithms. **arXiv preprint arXiv:2310.05984**, 2023.

AMORES, Javier J. et al. Detectando el odio ideológico en Twitter. Desarrollo y evaluación de un detector de discurso de odio por ideología política en tuits en español. **Cuadernos. info**, n. 49, p. 98-124, 2021.

VIEIRA, G. A. Uso de aprendizado de máquina na avaliação de políticas públicas: uma revisão de escopo. 2023.

VINHAS, Otávio Iost. Os sentidos da Facada em Jair Bolsonaro: uma Análise Sociocibernética de Redes Sociais no Twitter. **Simbiótica. Revista Eletrônica**, v. 8, n. 4, p. 153-190, 2021.

WANKMÜLLER, Sandra. **Advancing natural language processing in political science**. 2022. Tese de Doutorado. Imu.

WEBER, Ingmar. " Political polarization of web search queries and hashtags" by Ingmar Weber, with Martin Vesely as coordinator. **ACM SIGWEB Newsletter**, v. 2013, n. Summer, p. 1-10, 2013.

WOJCIESZAK, Magdalena et al. Non-news websites expose people to more political content than news websites: Evidence from browsing data in three countries. **Political Communication**, v. 41, n. 1, p. 129-151, 2024.

WORDLICZEK, Łukasz. Neural Networks and Political Science: Testing the Methodological Frontiers. **Empiria: Revista de metodología de ciencias sociales**, n. 57, p. 37-62, 2023.

YARCHI, Moran; BADEN, Christian; KLIGLER-VILENCHIK, Neta. Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. **Political Communication**, v. 38, n. 1-2, p. 98-139, 2021.

ZHU, Lei. Political Opinion Identification, Mining and Retrieval. 2010.