UNIVERSIDADE FEDERAL DE SÃO CARLOS
DEPARTAMENTO DE COMPUTAÇÃO
BACHARELADO EM ENGENHARIA DA COMPUTAÇÃO

Vitor Henrique Bormio Nunes

Análise de artigos científicos sobre COVID-19: Uma perspectiva usando redes complexas

São Carlos - SP
2022

Vitor Henrique Bormio Nunes

Análise de artigos científicos sobre COVID-19: Uma perspectiva usando redes complexas

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Engenharia da Computação da Universidade Federal de São Carlos, como requisito parcial para a obtenção do título de Bacharel em Engenharia da Computação.

Orientação Prof. Dr. Alan Demétrius Baria Valejo

São Carlos - SP
2022

Dedico este trabalho aos meus pais, que sempre me apoiaram na minha trajetória estudantil e acadêmica durante toda a minha vida.

Agradecimentos

Primeiramente agradeço aos meus pais e avós, que sempre me incentivaram e deram suporte para que eu me graduasse.

Agradeço à minha irmã, ao meu cunhado e à minha namorada, que me ouviram e se interessaram por este trabalho durante sua execução.

Agradeço ao Prof. Dr. Alan Demétrius Baria Valejo pela oportunidade de realizar este trabalho e pela dedicação no processo de orientação.

Por fim, agradeço também aos meus colegas de turma, que sempre me ajudaram a querer e aprender cada vez mais, além da convivência e trabalho em equipe durante toda a trajetória no curso.

Você não pode esperar construir um mundo melhor sem melhorar os indivíduos. Para esse fim, cada um de nós deve trabalhar para o seu próprio aperfeiçoamento e, ao mesmo tempo, compartilhar uma responsabilidade geral por toda a humanidade.

(Marie Curie)

Resumo

A pandemia da COVID-19 foi um dos maiores eventos do século XXI, que teve impacto na vida de todos os habitantes do planeta. Do ponto de vista científico, tratou-se de um período acelerado para as áreas da medicina, farmacêutica e bioquímica, devido à busca por medidas para lutar contra a doença. Com o intuito de acelerar o combate à doença e disponibilizar dados para a comunidade científica, grupos de pesquisa compilaram e disponibilizaram uma base de dados que contém todos os artigos referentes aos estudos sobre COVID-19, SARS-CoV-2 e coronavírus, com o intuito de facilitar o acesso às publicações e acelerar o desenvolvimento de novas pesquisas para agilizar o combate ao vírus. Do ponto de vista computacional, existem diferentes modelos para melhorar a visualização e análise de um conjunto de dados. Dentre eles, destacam-se as redes complexas, pela sua facilidade em descrever suas unidades (vértices) e relações (arestas), bem como auxiliar a entender o comportamento de sistemas complexas. O objetivo deste trabalho é estudar os conceitos e ferramentas acerca das redes complexas para gerar visualizações e análises sobre redes de coautorias e de citações de artigos científicos sobre o coronavírus.

Palavras-chave: Redes Complexas; Redes de Grande Escala; Redes Sociais; Redes de Cooperação Científica; Redes de Citações Científicas; COVID-19; SARS-CoV-2; Coronavírus.

Abstract

The COVID-19 pandemic was one of the 21st century's biggest events, impacting the lives of all the people on the planet. From a scientific perspective, it was a very accelerated growth period for the medical, pharmaceutical and biochemistry areas, which developed research to fight the disease. In order to help the fight against the disease, scientific research groups compiled and made available a database containing articles of studies on COVID-19, SARS-CoV-2 and coronavirus, aiming to facilitate access to publications and the development of new research against the virus. From a computational perspective, we have different models to improve the visualization and analysis of datasets. One of these models is the complex networks, which describe their units (vertex) and relationships (edges) to help understand the behavior of complex systems. The objective of this work is to study the concepts and tools of complex networks to generate visualizations and analyses of co-authorship and citation networks of scientific articles on coronavirus.

Keywords: Complex Networks; Large Scale Networks; Social Networks; Scientific Cooperation Networks; Scientific Citation Networks; COVID-19; SARS-CoV-2; Coronavirus.

Lista de ilustrações

Lista de tabelas

Sumário

1 Introdução

A pandemia da COVID-19 impactou todo o mundo no início de 2020. Na comunidade científica, pesquisas sobre a doença se intensificaram em todas as áreas do conhecimento, com maior foco principalmente nas áreas de farmácia, medicina e bioquímica, para desenvolver medidas preventivas contra a disseminação do vírus, como por exemplo a vacinação e tratamentos para as pessoas já infectadas.

Este crescimento acelerado é benéfico para o avanço da ciência, mas por outro lado dificulta a organização das informações da situação atual das comunidades de pesquisadores em relação a determinadas áreas de estudo, pois o alto volume de trabalhos aumenta a complexidade de todo o ecossistema de cooperação científica. Nesse sentido, é interessante pensar em estudos sobre aplicação de modelagens de dados que facilitem a gestão e visualização dos dados envolvidos, para facilitar o acesso às informações pela comunidade científica, e a aplicação de algoritmos de aprendizado de máquina para encontrar padrões que possam auxiliar especialistas no desenvolvimento de políticas públicas de saúde.

Estes tipos de estudo são denominados "ciência pela ciência" ou, do inglês, "*science of science*". Tratam-se de trabalhos que buscam extrair, de conjuntos de dados, informações relevantes que possam alavancar outras áreas da ciência e acelerar o avanço científico. Retomando o conceito dos modelos de dados supracitados, as redes complexas vêm adquirindo cada vez mais relevância do ponto de vista de aplicação de *science of science*.

Do ponto de vista do combate à COVID-19, redes complexas já foram utilizadas, por exemplo, para estudos sobre a propagação de epidemias, baseando-se na dinâmica da mobilidade entre agentes infectados e não infectados pela doença (VENTURA et al., 2022). Além disso, também foram desenvolvidos sistemas em tempo real para monitoramento da doença entre pacientes e funcionários dentro de hospitais, baseando-se em análise de redes (PRICE et al., 2021).

Do ponto de vista de análises sobre cooperação científica, redes complexas já foram utilizadas em trabalhos para quantificar a interdisciplinariedade entre revistas científicas a partir de redes de citações (SILVA et al., 2013), e também para produzir *surveys* ou pesquisas que resumem o estado da arte para tópicos da ciência (SILVA et al., 2016).

Apesar de se ter estudos de aplicação de redes complexas em bases de dados sobre a COVID-19, como também pesquisas que abordam redes complexas relacionadas ao tema de *science of science*, ainda é pouco explorada na literatura a interseção entre essas duas áreas, ou seja, a aplicação de redes complexas para extrair informações sobre a comunidade científica que estuda a COVID-19.

Alguns aspectos interessantes sobre as redes complexas são as medidas de centralidade

associadas às suas unidades e a aplicação de algoritmos de aprendizados de máquina para detecção de comunidades de vértices, ou seja, elementos agrupados igualmente dentro da rede, devido ao conjunto de suas conexões. As medidas de centralidade são propriedades da estrutura topológica da rede, que ao serem analisadas quantitativamente, podem indicar características qualitativas de seus vértices e arestas. Já a detecção de comunidades tem como objetivo revelar estruturas modulares locais, constituídas por conjuntos de vértices densamente conectados entre si e esparsamente conectados com o restante da rede (VALEJO, 2014).

Nesse trabalho, irá se relacionar os aspectos descritos acima às redes de cooperação e citação de artigos científicos, a fim de realizar análises interessantes sobre o contexto atual de algumas áreas de estudo da COVID-19. Por exemplo, identificar as principais comunidades de autores responsáveis pelo desenvolvimento das vacinas, para o caso das redes de coautorias, ou determinar os artigos mais citados e relevantes para diversas áreas de atuação contra o combate da doença, para o caso das redes de citações. Será utilizada a base de dados *"COVID-19 Open Research Dataset Challenge"* (WANG et al., 2020), que atualmente possui mais de 900 mil artigos científicos sobre COVID-19, SARS-CoV-2 e assuntos correlacionados ao coronavírus, para se sintetizar informações relevantes para a comunidade científica.

1.1 Objetivos

1.1.1 Objetivos gerais

O principal objetivo deste trabalho é criar e analisar redes complexas, com o intuito de gerar novos conhecimentos sobre uma base de dados composta por artigos científicos da COVID-19, trazendo informações relevantes sobre o estado da arte, como por exemplo as principais comunidades de autores e artigos citados das áreas relacionadas ao combate à doença.

1.1.2 Objetivos específicos

- Extrair tópicos a partir dos resumos dos artigos para entender áreas de pesquisa para se aprofundar;
- Criar redes de coautoria e citação científica sobre áreas de pesquisa específicas, selecionadas utilizando os tópicos extraídos e conhecimentos gerais;
- Analisar os dados, medidas de centralidade e características das redes para gerar visualizações que tragam informações interessantes sobre os principais pesquisadores e artigos das áreas de pesquisa.

1.2 Organização do trabalho

O presente trabalho se organiza da seguinte maneira:

- No Capítulo 1, introduz-se a proposta do trabalho, bem como seus objetivos;
- No Capítulo 2, aborda-se a fundamentação teórica sobre redes complexas, modelagem de tópicos e demais assuntos correlacionados;
- No Capítulo 3, apresenta-se a metodologia de pesquisa e configuração dos experimentos acerca da construção das redes complexas;
- No Capítulo 4, ilustra-se e se discute o conjunto de resultados obtidos após a análise das redes complexas;
- No Capítulo 5, discorre-se sobre as conclusões do trabalho, suas possíveis contribuições para a ciência e também sugestões para futuras pesquisas.

2 Fundamentação Teórica

2.1 Redes Complexas

Sistemas complexos são conjuntos de unidades que interagem e se organizam, gerando conexões de comportamento não linearizado. Nesse sentido, as redes complexas são modelos de representação e análise de tais sistemas, com capacidade de descrever as unidades (vértices) e relações (arestas) para entender melhor a dinâmica dos dados envolvidos na modelagem (WATTS; STROGATZ, 1998; NEWMAN, 2001). Exemplificando-se as redes, é possível citar as redes de relacionamentos sociais, redes biológicas cerebrais, redes metabólicas, a World Wide Web e a Internet (ALBERT; JEONG; BARABASI, 1999; SCOTT, 2000). Redes complexas são modelos interessantes para dividir partes de um todo (vértices) e relacionar tais unidades por meio de conexões (arestas).

Dois conceitos importantes para as redes criadas no presente trabalho são relacionados às arestas. O primeiro deles é o peso. As arestas demonstram relações entre os vértices e estas relações podem ter maior relevância entre determinados vértices. Podemos pensar, por exemplo, na rede de coautorias de artigos científicos, em que um autor pode ter coautoria em vários artigos diferentes com um segundo autor, mas cooperação em apenas um trabalho com outro terceiro. Neste caso, é possível atribuir pesos diferentes às arestas, para indicar que algumas conexões são mais relevante que outras.

O segundo conceito é relacionado à direção e sentido das arestas, que podem ou não ser direcionadas. Para arestas não direcionadas, ambos os vértices enxergam o relacionamento da mesma forma. Já para arestas direcionadas, este comportamento não acontece, porque um dos vértices aponta em direção a outro, que consequentemente é apontado. É possível que um vértice aponte para o outro e vice-versa, mas neste caso estariam envolvidas duas arestas direcionadas.

2.1.1 Medidas de Centralidade

2.1.1.1 Grau

O grau é um atributo dos vértices de uma rede complexa, que indica o número de arestas conectadas, ou seja, indica quantas conexões determinado vértice tem com outros vértices. Em redes não direcionadas, o grau é um atributo de valor único. Já para o caso de redes direcionadas, o grau de um vértice é dividido entre grau de entrada e saída. Além disso, para redes em que as arestas têm pesos, é possível fazer uma ponderação do grau, de forma que arestas com maior peso tenham maior contribuição para o grau do vértice conectado.

Se pensarmos por exemplo numa rede direcionada de citações científicas, em que os vértices representam os artigos e as arestas relações de citação, o grau de entrada representa o número de vezes que um artigo foi citado por outros, o que pode indicar uma maior relevância ou não dele dentro da área ou assunto discorrido, sendo assim utilizado como uma métrica para julgar o impacto da pesquisa científica (NEWMAN, 2010).

2.1.1.2 Centralidade do Autovetor

A centralidade do autovetor é uma extensão do conceito do grau, que basicamente busca considerar, além do número de conexões, a "importância" da conexão perante à rede, dando a cada vértice uma pontuação ou "valor de importância" proporcional à soma das pontuações de seus vizinhos (NEWMAN, 2010). A Equação 2.1 define a expressão para o cálculo da centralidade do autovetor (BONACICH, 1987).

Essa medida indica que a centralidade de um vértice x_i é proporcional à soma das centralidades de seus vizinhos, onde A é a matriz de adjacências, k_L é o maior autovalor de A e x_j são as centralidades dos vizinhos. Desta forma, a centralidade de autovetor de um vértice é influenciada tanto pelo número de arestas j quanto pelos valores de centralidade x_j de seus vizinhos dentro da rede como um todo.

$$x_i = k_L^{-1} \sum_j A_{ij} x_j \tag{2.1}$$

Apesar de ser um método funcional tanto para redes direcionadas quanto para não direcionadas, a centralidade de autovetor tem alguns problemas no primeiro caso. Em redes direcionadas, a matriz de adjacências é em geral assimétrica, de forma que se tem dois autovetores, o da direita, referente às arestas que entram no vértice, e o da esquerda, refentes às arestas que saem do vértice. Assim, no momento do cálculo da centralidade do autovetor, é preciso escolher com qual dos autovetores se deseja trabalhar. Usualmente se escolhe o da direita, porque um vértice apontado por muitos outros costuma ser mais relevante do que um vértice que aponta para muitos outros, para a grande maioria das aplicações. Além deste caso, a centralidade de autovetor também se depara com um outro problema em redes direcionadas que são acíclicas, pois cada vértice não tem mais de um outro vértice fortemente conectado a ele, o que resulta que a centralidade de autovetor seja nula para todos os vértices da rede (NEWMAN, 2010).

2.1.1.3 Centralidade de Katz

A centralidade de Katz é uma medida que busca solucionar o problema da centralidade de autovetor para o caso das redes direcionadas. A Equação 2.2 ilustra a expressão que fornece a centralidade de Katz x_i de um vértice i. Basicamente, em comparação à expressão que calcula a centralidade de autovetor (Figura 2.2), há adição de um termo constante β_i.

Com β_i não nulo, para uma rede direcionada, até os vértices que não têm arestas de entrada podem ter centralidade não nula, e assim os vértices apontados por eles também são valorizados (NEWMAN, 2010). A partir disso, um vértice apontado por muitos outros poderá ter uma alta centralidade, o que resolve o problema da centralidade de autovetor para redes acíclicas.

$$x_i = \alpha \sum_i A_{ij} x_j + \beta_i \tag{2.2}$$

O termo α tem a função de balancear o valor do termo relacionado ao autovetor com a constante β_i. No entanto, há um limite para o valor de α, para que a expressão da centralidade de Katz possa convergir, que é $\alpha <= k_L$, em que k_L é o maior autovalor de A, assim como na expressão da centralidade de autovetor da Equação 2.1.

Contudo, ainda podemos identificar alguns problemas na centralidade de Katz. Se um vértice com alta centralidade apontar para muitos outros, esses outros também recebem altas centralidades. Se numa rede, a maioria dos vértices tiver altas centralidades, perde-se o sentido da métrica, que é identificar vértices com maior relevância para a rede (NEWMAN, 2010).

2.1.1.4 *PageRank*

O problema da centralidade de Katz veio a ser resolvido com uma uma nova medida, chamado *PageRank*, nome dado comercialmente pelo Google, por fazer parte do sistema de classificação do seu motor de buscas na *web*. A Equação 2.3 ilustra a expressão fornece o valor de *PageRank* x_i de um vértice i. Em comparação à centralidade de Katz, a diferença é que as centralidades dos vizinhos x_j são divididas pelo grau de saída do vértice i. Assim, tem-se uma solução para o problema de supervalorização em cadeia de todos os vértices da rede complexa (NEWMAN, 2010).

$$x_i = \alpha \sum_i \frac{A_{ij} x_j}{k_j^{out}} + \beta_i \tag{2.3}$$

2.1.1.5 Centralidade de proximidade

Centralidade de proximidade (ou *closeness centrality*) é uma medida de centralidade que mede a proximidade de um vértice a todos os outros da rede (NEWMAN, 2010). Sua expressão é indicada por C_i na Equação 2.4, onde l_i é a distância média de um vértice i em relação aos outros vértices de uma rede complexa de tamanho n.

Como C_i é inversamente proporcional a l_i, quanto maior a distância média de um vértice aos outros, menos ele se encontra numa posição central na rede, situando-se então numa

posição mais periférica. Analogamente, quanto menor l_i, maior C_i e mais o vértice se encontra numa posição central na rede.

$$C_i = \frac{1}{l_i} = \frac{n}{\sum_j d_{ij}} \tag{2.4}$$

2.1.1.6 Centralidade de intermediação

Centralidade de intermediação (ou *betweenness centrality*) é uma medida de centralidade que mede o quanto um vértice se localiza em caminhos entre outros vértices, ou seja, o quanto ele tem um papel de intermediar dentro da rede. Assim, quanto mais caminhos passam por um vértice, maior sua centralidade de intermediação (NEWMAN, 2010).

A Equação 2.5 (FREEMAN, 1977) ilustra a expressão que descreve, para uma rede não direcionada, a centralidade de intermediação x_i de um vértice i, em que n^i_{st} é o número de caminhos geodésicos entre os vértices s e t que passam por i, e g_{st} representa todos os caminhos geodésicos entre s e t.

Do ponto de vista de análise das redes complexas, a métrica de centralidade de intermediação traz uma noção interessante do fluxo de informação dentro da rede e da influência que cada um dos vértices pode exercer sobre esta característica (NEWMAN, 2010).

Pensando por exemplo numa rede de coautorias de artigos científicos, vértices com alto valor de centralidade de intermediação podem indicar autores que interligam diferentes grupos e comunidades científicas.

$$x_i = \sum_i \frac{n^i_{st}}{g_{st}} \tag{2.5}$$

2.1.2 Detecção de Comunidades

A detecção de comunidades é uma área extensa e importante no contexto de redes complexas. Ela se baseia na ideia de buscar grupos de vértices que ocorrem naturalmente em uma rede, ou seja, grupos de vértices com grande número de arestas entre si, mas com menos conexões aos outros vértices da rede. A Figura 1 ilustra uma rede com três comunidades, coloridas em amarelo, azul e verde. Assim, o objetivo dos algoritmos que realizam processos como este é separar a rede em grupos de vértices com muitas conexões entre eles (NEWMAN, 2010).

Para uma rede de coautorias de artigos científicos, por exemplo, este tipo de análise é interessante para entender quais os principais grupos científicos existentes e identificar possíveis novas colaborações que sejam interessantes para o estudo de determinada área.

Existem diferentes tipos de algoritmos com distintas abordagens que realizam detecção em comunidades em redes. Neste trabalho, não irá se discorrer sobre todos os tipos, mas vale citar os três principais tipos: *Fastgreedy*, *Infomap* e *Leading Eigenvector*.

O algoritmo aplicado para detecção de comunidades neste trabalho é o *Louvain*, tratando-se de um método heurístico que se baseia na otimização da modularidade, e que se mostrou extremamente adequado pela sua capacidade de processar rapidamente redes de grande escala (BLONDEL et al., 2008).

A Modularidade é uma medida de qualidade que quantifica uma estrutura de comunidades em uma rede. Essa medida compara a densidade real de arestas intra-comunidade e inter-comunidade em relação a uma rede aleatória com características semelhantes, uma vez que redes com estruturas aleatórias não possuem estrutura de comunidades.

O algoritmo Louvain agrupa vértices de modo a maximizar a Modularidade. A estratégia desse algoritmo é otimizar localmente a divisão em comunidades até que a modularidade global não consiga mais ser melhorada. Em outras palavras, o algoritmo utiliza uma estratégia multinível, ou seja, primeiro o algoritmo encontra uma estrutura de comunidades na rede através da maximização da Modularidade. Em seguida cada comunidade encontrada no passo anterior é contraída em um único vértice, chamado super-vértice. Além disso, arestas incidentes em super-vértices também são contraídas formando as chamadas super-arestas. Desse modo, o algoritmo gera uma rede reduzida com menos vértices e arestas. Essas duas etapas são repetidas, nível a nível, até que a modularidade máxima seja atingida. A Figura 2 ilustra esse processo.

Figura 1 – Exemplo - Comunidades, de (VALEJO et al., 2020)

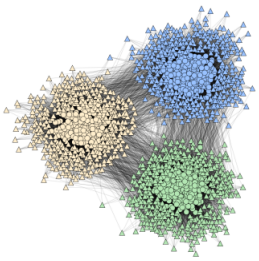
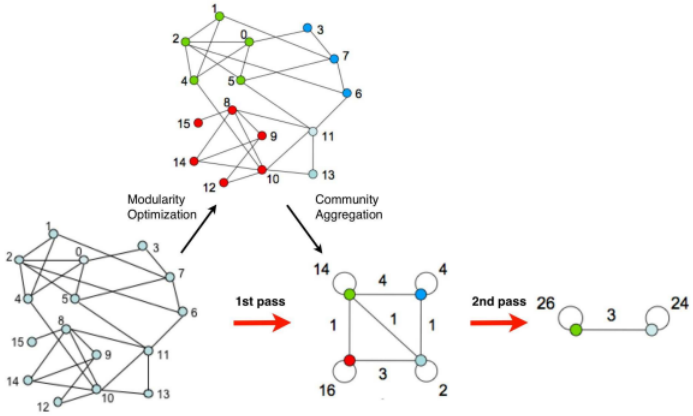


Figura 2 – Representação do algoritmo de Louvain (BLONDEL et al., 2008)



2.1.3 Algoritmos de Visualização de Rede

Algoritmos de visualização de redes são modelos que têm como objetivo encontrar uma melhor distribuição dos vértices e arestas para uma melhor análise visual da rede em questão. Existem diversas estratégias para distribuir os elementos ao longo da rede, de forma que é importante escolher o melhor algoritmo que se aplica a cada caso.

Neste trabalho, serão estudados os algoritmos Force Atlas, Force Atlas 2, Fruchterman-Reingold e OpenOrd (GEPHI, 2008), que são os modelos mais alinhados às características e os objetivos de análise das redes de coautorias e citações de artigos científicos.

2.1.3.1 Fruchterman-Reingold

O algoritmo Fruchterman-Reingold foi desenvolvido em 1991 e publicado no artigo *"Graph Drawing by Force-directed"* (FRUCHTERMAN; REINGOLD, 1991). Sua estratégia é de simular a rede como se ela fosse um sistema físico, em que cada vértice é uma partícula com massa específica, e as arestas são molas entre tais partículas. O objetivo do modelo é minimizar a energia desse sistema, buscando deixar os vértices mais espaçados e melhorar a visualização.

Devido à alusão ao sistema físico, sua abordagem é dita *force-directed*. A principal limitação deste algoritmo é relacionada à sua complexidade, $O(n^2)$, em que n é o número de vértices da rede. Por isso, o algoritmo é recomendado para redes de pequena escala, com até 1000 vértices. Além disso, ele não considera a relevância do peso das arestas no modelo.

2.1.3.2 OpenOrd

O algoritmo OpenOrd foi baseado no Frutcherman-Reingold e sua estratégia parte do princípio de *simulated annealing* (BERTSIMAS; TSITSIKLIS, 1993), um modelo probabilístico para resolver problemas de otimização de funções, principalmente com característica discreta. Seu nome vem da metalurgia, mais especificamente do conceito de termodinâmica que envolve aquecer e resfriar um material de forma controlada para modificar suas propriedades físicas. Além disso, outra parte crucial do modelo desse algoritmo é que ele corta arestas muito longas dentro da redes para permitir que os agrupamentos se separem, melhorando a visualização dos diferentes grupos e comunidades.

A grande vantagem desta algoritmo é que ele possui uma complexidade de $O(n \log n)$, o que permite seu bom funcionamento em redes com até um milhão de vértices. O modelo também considera como relevante o peso das arestas nas conexões entre os vértices. Por conta destas características, trata-se de um algoritmo interessante para ser utilizado na rede de coautorias de artigos científicos.

2.1.3.3 Force Atlas

Assim como o Fruchterman-Reingold, o Force Atlas é um algoritmo do tipo *force-directed*, inspirado em sistemas físicos que envolvem atração e repulsão de partículas. Foi criado em 2007 e é nativo do *software* Gephi (BASTIAN; HEYMANN; JACOMY, 2009), famoso na área de visualização e análise de redes.

Foi desenvolvido para distribuição de redes de pequena escala, com até dez mil vértices, por se tratar de um algoritmo de alta complexidade, $O(n^2)$, que preza principalmente pela boa qualidade do agrupamento e visualização dos vértices, e considera a relevância do peso das arestas na rede.

2.1.3.4 Force Atlas 2

Em 2011, foi lançada a segunda versão e evolução do algoritmo do tópico anterior, nomeado de Force Atlas 2 por seus autores. Este algoritmo busca manter as boas qualidade do seu precursor, mas funcionando também para redes de larga escala, com até um milhão de vértices. A redução de sua complexidade é devida à nova forma de cálculo do grau de atração e repulsão dos vértices, baseada agora na simulação de Barnes-Hut, que possui complexidade na ordem de $O(n \log n)$, que é também a complexidade do algoritmo.

Além disso, ele continua adotando a estratégia *force-directed* e atribuindo relevância para o peso das arestas na hora de fazer a distribuição da rede no espaço. Tais características tornam este algoritmo interessante para ser utilizado tanto na rede de coautorias quanto na rede de citações de artigos científicos.

2.2 Modelagem de Tópicos

A modelagem de tópicos é um tema secundário do ponto de vista dos objetivos do presente trabalho. Seu intuito foi gerar tópicos relevantes para filtrar a base de dados para gerar redes que tratassem dos temas extraídos. Por isso, irá se discorrer brevemente sobre este assunto.

A mineração de dados textuais é uma ampla área de estudo da computação, com alto número de pesquisas na área acadêmica e alto grau de aplicabilidade no mercado e setor empresarial. Dentro dessa área existe a modelagem de tópicos. De acordo com a definição dada por Blei (2012), um tópico pode ser definido como um conjunto de palavras que ocorre frequentemente em documentos semanticamente relacionados.

A extração de tópicos por sua vez é uma área que tenta identificar e descobrir padrões latentes nas relações entre documentos e termos, tal que tais padrões sejam significativos para o entendimento das relações entre tais documentos e termos (BLEI; NG; JORDAN, 2003; BLEI, 2012).

A extração de tópicos possui grande aplicação na área de linguagem natural e retirada de informações. Isto pode ser usado como, por exemplo, faz Bun e Ishizuka (2002) ao extrair tópicos de um determinado texto para análise de tópicos semanais de notícias ou ainda como feito, junto de outras técnicas, em Dong et al. (2013) para a classificação e recomendação de análises feitas na internet.

Este capítulo descreve dois conceitos utilizados para extração de tópicos em coleções de documentos. O primeiro deles é referente às técnicas de pré-processamento de linguagem natural, e o segundo é relacionado à modelagem de tópicos por meio do uso da técnica LDA (*Latent Dirichlet Allocation*) (BLEI; NG; JORDAN, 2003).

2.2.1 Pré-Processamento de Textos

Processamento de Linguagem Natural (PLN) é a área de estudo que busca trabalhar o entendimento das máquinas sobre a linguagem humana. Para arquivos de texto, na maioria das vezes se trabalha com dados não estruturados, ou seja, que não podem ser processados pelo computador. Por isso, sempre que se deseja realizar o processamento de um texto em busca de informações relevantes e sintetizadas, é preciso antes fazer o pré-processamento dos dados, para posteriormente aplicar o modelo desejado (LINHARES; LIMA, 2021).

O pré-processamento de dados textuais envolve usualmente as seguintes etapas:

- Remoção de *stopwords*: retirada de palavras irrelevantes para o significado do texto. Exemplos de palavras deste tipo no português são os artigos e preposições. Há bibliotecas disponíveis com listas prontas de *stopwords* para diversas línguas;

- Remoção de caracteres especiais: retirada de recursos textuais de pontuação, como por exemplo vírgula, ponto final, ponto de interrogação, entre outros;

- Tokenização: processo que divide uma grande parte de texto em subpartes, chamadas *tokens*. Exemplos de tokenização são separar as frases de um parágrafo ou então separar as palavras de uma frase.

- Lematização: transformação das palavras em seu lema, por meio da retirada da flexão, que é a modificação das palavras com afixos ou desinências que exprimem determinadas informações gramaticais. Um exemplo de lematização é a transformação do verbo "programou", conjugado, em "programar", no infinitivo;

- "Stematização": transfomação de palavras em seu radical. A palavra "computacional", por exemplo, seria modificada para "computacion" ao ser "stematizada". Este processo é geralmente utilizado quando não há uma preocupação com o contexto da frase em que a palavra está inserida;

- Vetorização: consiste na transformação das palavras em vetores, para que elas possam ser processadas posteriormente por algoritmos de aprendizado de máquina. Há diferentes modelos para vetorizar palavras. Dois deles serão discutidos nas subseções a seguir: *Bag of Words* e *TF-IDF*.

2.2.1.1 *Bag of Words*

O modelo *Bag of Words* ou "Bolsa de Palavras" se preocupa exclusivamente pela unicidade e cálculo da frequência das palavras, que são as informações presentes no vetor gerado pelo modelo, não havendo consideração do contexto em que as palavras estão inseridas ou atribuição de importância para cada palavra dentro das sentenças (LINHARES; LIMA, 2021).

2.2.1.2 *TF-IDF*

O TF-IDF é um modelo que traz consigo dois principais parâmetros portadores de informações aos vetores gerados a partir das palavras, os quais estão presentes em seu próprio nome: o *Term Frequency* (TF) e o *Inverse Term Frequency* (IDF).

O TF é calculado de forma similar ao que ocorre no *bag of words*, representando a frequência da palavra no documento, ou seja, o número de repetições de uma mesma palavra dividido pelo número total de palavras presentes dentro de um documento. No entanto, o grande diferencial deste modelo está no IDF, que é calculado pela expressão ilustrada pela Equação 2.6.

$$IDF = \log\left(\frac{n_D}{n_{Dw}}\right) \tag{2.6}$$

Na expressão, n_D representa o número total de documentos, e n_{Dw} o número de documentos contendo a palavra em questão. Assim, palavras que aparecem em muitos documentos têm altos valores de n_{Dw} e um menor IDF, porque são palavras mais gerais, que não trazem consigo uma informação diferencial para a maioria dos documentos. Já palavras mais específicas de alguns documentos, com baixo valor de n_{Dw}, podem trazer um significado mais importante ao contexto geral das informações que se quer extrair.

Assim, o modelo TF-IDF consegue equilibrar a frequência das palavras com seu grau de relevância dentro dos textos, entendendo quais palavras são importantes em um documento específico e quais palavras são importantes no contexto de todos os documentos (MADAN, 2019).

2.2.2 *Latent Dirichlet Allocation* (LDA)

Diversas vezes, trabalha-se com grande quantidade de documentos, e um dos principais objetivos do processamento de textos é sintetizar as informações mais importantes em

estruturas menores, por exemplo, em tópicos. Um dos principais algoritmos para modelagem de tópicos é o *Latent Dirichlet Allocation* ou LDA.

O LDA é um algoritmo de redução de dimensionalidade de dados, com características que o tornam ideal para o processamento de dados textuais. Seu funcionamento para o processamento de textos se baseia inicialmente em duas afirmações: documentos são compostos por tópicos, e tópicos são compostos por palavras.

Na primeira iteração do algoritmo, ele associa aleatoriamente um tópico para cada palavra no documento, e uma composição destes tópicos é utilizada para gerar os documentos. A partir disso, são geradas duas matrizes: a primeira de palavras por tópicos, e a segunda de tópicos por documentos.

Em seguida, o algoritmo itera por cada uma das palavras e tenta associá-las ao tópico correto, assumindo que todos os tópicos estão associados corretamente, exceto o tópico da palavra vigente. Em sequência, ele utiliza todas essas atribuições corretas de palavras a tópicos para tentar ajustar a palavra vigente ao tópico correto. Para isso, ele itera por cada documento e cada palavra, calculando duas distribuições de probabilidades:

- P1: proporção de palavras em um documento D que estão associadas atualmente a um tópico T;
- P2: proporção de associações da palavra vigente ao tópico T em relação a todos os documentos em que a palavra vigente está presente.

Por fim, a partir da multiplicação destas probabilidades, o algoritmo estima qual é o tópico mais provável de ser relevante para a palavra vigente. Após isso, o LDA itera múltiplas vezes, repetindo este mesmo processo, até atingir um estado estacionário, ou seja, em que não há mais mudanças de palavras para novos tópicos, porque todas as palavras já se encontram no seu tópico ótimo (SETHNEHA, 2021).

3 Metodologia de Pesquisa

Este capítulo descreve as ferramentas, dados e procedimentos utilizados para desenvolver os experimentos, que seguiram as seguintes etapas:

- Escolha das ferramentas de modelagem
- Análise do conjunto de dados
- Filtragem da base de dados
- Construção das redes
- Cálculo das medidas de centralidade
- Visualização das redes

3.1 Ferramentas

A linguagem de programação utilizada para a preparação dos dados das redes complexas e para a extração de tópicos foi Python, juntamente com uma série de bibliotecas para auxiliar os processos desenvolvidos.

Para todas as implementações, utilizou-se das bibliotecas *pandas*, *os* e *numpy*. Para o caso da preparação dos dados das redes de coautorias, além das já citadas, utilizou-se também *ast* e *csv*. No caso das redes de citações, as biblioteca utilizadas adicionalmente foram *json* e *networkx*. Por fim, para o caso da extração de tópicos, foram utilizadas também as bibliotecas *nltk*, *textlob*, *string*, *gensim* e *pprint*.

Para a criação das redes, utilizou-se as saídas dos algoritmos de preparação dos dados das redes como entrada para o *software* Gephi (BASTIAN; HEYMANN; JACOMY, 2009), amplamente reconhecido na área de redes complexas e recomendado para se trabalhar com redes de larga escala, que foi o principal motivo de sua escolha para este trabalho. Os cálculos das medidas de análise das redes também foram feitos com os algoritmos do Gephi, assim como a geração das visualizações a partir da aplicação dos algoritmos de visualização.

3.2 Base de dados

O conjunto de dados utilizado para desenvolver as modelagens do trabalho foi o *COVID-19 Open Research Dataset Challenge - CORD-19*, composto atualmente por mais de 900.000 artigos científicos sobre COVID-19, SARS-CoV-2 e coronavírus relacionados, tratando-se da

maior coleção de literatura de coronavírus legível por máquina disponível para mineração de dados até o momento. Ele é disponibilizado gratuitamente na plataforma *Kaggle* [1], e foi preparado pela Casa Branca juntamente com uma coalização de grupos de pesquisa.

Inicialmente, para se ter uma boa visualização do conteúdo da base de dados de artigos científicos da COVID-19 e começar a planejar as etapas do trabalho, utilizou a biblioteca *pandas* para criação de uma estrutura de *dataframe*, ilustrada pela Figura 3.

Figura 3 – Base de Dados - COVID

	cord_uid	sha	source_x	title	doi	pmcid	pubmed_id	license	abstr
0	ug7v899j	d1aafb70c066a2068b02786f8929fd9c900897fb	PMC	Clinical features of culture-proven Mycoplasma...	10.1186/1471-2334-1-6	PMC35282	11472636	no-cc	OBJE This retro chart des...
1	02tnwd4m	6b0567729c2143a66d737eb0a2f63f2dce2e5a7d	PMC	Nitric oxide: a pro-inflammatory mediator in l...	10.1186/rr14	PMC59543	11667967	no-cc	Inflan disea respi tract.
2	ejv2xln0	06ced00a5fc04215949aa72528f2eeaae1d58927	PMC	Surfactant protein-D and pulmonary host defense	10.1186/rr19	PMC59549	11667972	no-cc	Surfa prote D) pa in th.
3	2b73a28n	348055649b6b8cf2b9a376498df9bf41f7123605	PMC	Role of endothelin-1 in lung disease	10.1186/rr44	PMC59574	11686871	no-cc	Endo (ET-1 amin pepti
4	9785vg6d	5f48792a5fa08bed9f56016f4981ae2ca6031b32	PMC	Gene expression in epithelial cells in respons...	10.1186/rr61	PMC59580	11686888	no-cc	Respi syncy (RSV) pneu

Primeiramente, é possível notar que cada linha da tabela representa um artigo científico, e cada coluna, das 19 existentes, um tipo de característica atribuída à publicação. O interesse deste trabalho foi restrito apenas às colunas referentes aos autores, de onde se pôde retirar as relações de coautoria, aos resumos ou *abstracts*, de onde se pode extrair tópicos, e aos nomes dos arquivos *pdf-json-files*, arquivos externos que continham outras informações sobre os artigos, como por exemplo os artigos referenciados em suas citações, incluindo tanto artigos da base de dados quanto outros externos a ela.

Além disso, percebeu-se logo que se trata de uma base de dados que geraria redes com alto volume de dados. Neste sentido, viu-se então a necessidade de se realizar alguns tipos de filtragem de artigos para algumas datas e assuntos específicos, com o objetivo de viabilizar a geração de redes nas quais seria possível calcular as medidas de centralidade, que são algoritmos de alta complexidade computacional em alguns casos.

3.3 Filtragem dos dados

A primeira etapa de filtragem de dados foi a extração de tópicos, baseada na publicação *Topic Modeling and Latent Dirichlet Allocation (LDA) in Python* (LI, 2018).

Primeiramente, foram processados os dados dos resumos ou *abstracts* da base de dados de artigos científicos sobre COVID-19, seguindo as seguintes etapas:

- *Stopwords* foram importadas usando o módulo da *nltk*;

[1] Kaggle is a website that offers Jupyter notebooks environment, access free GPUs and repository of community published data and code.

- Todos os caracteres do texto foram convertidos em minúsculos;
- Caracteres como números ou pontos foram removidos por meio de expressão regular;
- Dígitos foram removidos dos textos;
- As *stopwords* foram removidas dos textos;
- Foi feita a lematização de todas as palavras.

Em seguida, criou-se uma lista com todas as palavras dos *abstracts* e uma estrutura de dicionário, com uma entrada para cada palavra. Após isso, converteu-se cada uma das entradas do dicionário para a estrutura de dados de *bag of words*, aplicando posteriormente o TF-IDF no conjunto de vetores. Ambos os modelos foram derivados do módulo *models* da biblioteca *gensim*.

Por fim, aplicou-se o modelo de LDA Multicore, do mesmo módulo e biblioteca supracitados, à saída gerada pelo modelo TF-IDF, e se imprimiu alguns dos principais tópicos gerados, estando quatro deles ilustrados pela Tabela 1.

A partir dos tópicos gerados e também do conhecimento dos principais assuntos acerca de COVID-19, aferiu-se que dois possíveis bons filtros para extrair subáreas de estudo das redes seriam: "Pneumonia" e "Comorbidities". O primeiro termo se refere a uma doença que é uma das possíveis complicações do coronavírus, e o segundo se refere às características que levam um paciente infectado a ter maiores chances de complicação para um estado grave.

Tabela 1 – Tópicos LDA

Tópico
Topic - 0.019*patient + 0.010*common + 0.009*comorbidities + 0.009*pneumoniae + 0.008*pneumonia + 0.008*infection + 0.007*died + 0.006*infant + 0.006*child + 0.006*identified
Topic - 0.013*et1 + 0.010*role + 0.010*disease + 0.008*contribution + 0.008*presumed + 0.008*dependent + 0.008*although + 0.007*evidence + 0.006*inflammatory + 0.006*lung

Além desses filtros, foram pré-definidos alguns outros termos que seriam também interessantes para filtragem do ponto de vista de segmentação de áreas de estudo. Os principais escolhidos foram os nomes científicos e comerciais das principais vacinas desenvolvidas para combater à COVID-19, usualmente conhecidas no Brasil pelos nomes de: CoronaVac, Pfizer, Astrazeneca, Janssen e Moderna. A escolha das vacinas foi baseada na ideia de se obter a partir delas os principais autores e artigos referentes aos seus desenvolvimentos. Dois outros termos que foram utilizados juntamente para a filtragem foram as palavras em inglês *"Elderly"* e *"Aged"*, utilizados com o intuito de se buscar artigos relacionados à análise do comportamento da doença conforme a idade, principalmente para o caso de pacientes idosos.

Finalmente, o último filtro planejado foi *"Complex Network"*, com o intuito de se buscar pesquisas sobre COVID-19 em que se há a aplicação de redes complexas. Também foi criada a rede a partir do filtro *"Network"*, mas por os resultados obtidos não fizeram parte do presente trabalho, pois a genericidade do termo, que também é usado em outras áreas do conhecimento, fez com que as redes não fossem interessantes ao objetivo desejado.

Por fim, os procedimentos executados para de fato realizar a filtragem da base de dados foram os seguintes:

- Remoção das linhas que continham campos nulos nas colunas de autores, resumos ou datas de publicação;
- Transformação das datas de publicação de texto para Timestamp;
- Filtragem para coletar apenas artigos publicados a partir de 2020, ano de início da pandemia e onde os estudos começaram a se acelerar;
- Filtragem para coletar apenas artigos contendo as palavras de interesse em seus respectivos resumos ou *abstracts*.

3.4 Construção das redes

3.4.1 Redes de Coautorias

O primeiro passo para construção das redes de coautorias foi identificar todos os autores únicos presentes em cada conjunto de artigos científicos filtrados. Para isso, implementou-se a seguinte rotina:

- Criar uma lista com todos os autores de todos os artigos científicos;
- Através do *pandas*, criar um *dataframe* com todos os autores e eliminar linhas de autores duplicadas;
- Criar um dicionário, de forma a atribuir um identificador único (número inteiro) para cada autor;
- Exportar o dicionário para um arquivo *csv*.

A partir disso, todos os vértices da rede coautorias, correspondentes aos autores, foram obtidos. Em seguida foi criado um arquivo *csv* contendo todas as relações de coautoria, ou seja, as arestas. Para isso, aplicou-se a seguinte rotina:

- Criar um *dataframe* em que cada linha corresponda a um artigo científico, que possui a lista de autores participantes de sua elaboração;

- Utilizar o dicionário, criado na etapa anterior, para substituir o nome dos autores por seus identificadores únicos;
- Para cada conjunto de autores em cada artigo científico, extrair todas as combinações únicas de pares entre eles e criar vetores de relação de coautoria [id-autor1, id-autor2] e adicioná-los a uma lista de arestas;
- A partir da lista de arestas, gerar um *dataframe* com duas colunas que contenham identificadores de autores, com cada linha representando, portanto, uma relação de coautoria;
- Para cada linha, contar o número de ocorrências de linhas que contenham os mesmos dois autores e atribuir o valor à uma terceira coluna de pesos;
- Eliminar duplicatas de relações de coautoria;
- Exportar o *dataframe* de arestas para um arquivo *csv*.

Por fim, utilizou-se a ferramenta de importação de planilha do software Gephi (BASTIAN; HEYMANN; JACOMY, 2009) para criação da rede de coautorias não direcionada e com pesos nas arestas. Este processo foi realizado para cada uma das bases geradas a partir das filtragens.

3.4.2 Redes de Citações

O processo para a criação das redes de citação foi baseado na rotina implementada no *notebook* II-COVID19-Citation Network (VASUJI, 2020), presente no *Kaggle* [2].

O processamento dos dados seguiu o seguinte conjunto de etapas:

- Carregar os arquivos *json* da coluna *pdf-json-files*;
- Coletar os dados título, resumo e identificadores únicos de cada artigo;
- Para cada artigo, criar uma lista de referências, agrupando o *DOI (Digital Object Identifier)*, utilizado como identificador dos artigos citados, além do título, ano e jornal de publicação das citações;
- Por meio da biblioteca *networkx*, adicionar vértices e arestas para criar a rede de citações dos artigos únicos e das referências citadas por cada um deles;
- Utilizar a função *to-pandas-edgelist* para criar um *dataframe* de representação da lista de arestas da rede;

[2] Kaggle is a website that offers Jupyter notebooks environment, access free GPUs and repository of community published data and code.

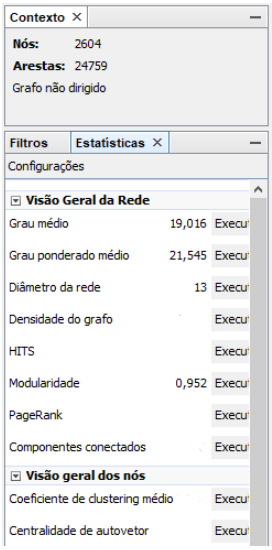
- Criar um dicionário para todos os vértices únicos da rede, atribuindo um número inteiro como novo identificador;
- Exportar o dicionário para um arquivo *csv*, obtendo assim uma lista de vértices, que são os artigos, e seus respectivos identificadores;
- Utilizar o dicionário para substituir os identificadores originais dos artigos por seus novos identificadores únicos (números inteiros);
- Exportar o *dataframe* de arestas para um arquivo *csv*.

Por fim, assim como no caso da rede de coautorias, utilizou-se a ferramenta de importação de planilha do software Gephi (BASTIAN; HEYMANN; JACOMY, 2009) para criação da rede de citações direcionada e sem pesos nas arestas. Novamente, o processo foi realizado para cada uma das bases geradas a partir das filtragens.

3.5 Cálculo de medidas de centralidade

O cálculo das medidas de centralidade das redes foi feito a partir da ferramenta de Estatísticas do Gephi (BASTIAN; HEYMANN; JACOMY, 2009), ilustrada pela Figura 4.

Figura 4 – Medidas de Centralidade - Gephi



De todas as medidas disponíveis, calculou-se as mais interessantes para a análise e visualização das redes, todas abordadas na seção 2, que são:

- Grau médio, que calcula o grau de cada vértice e em seguida a média da rede;
- Grau ponderado médio, que calcula o grau ponderado de cada vértice e em seguida a média da rede;

- Centralidades de proximidade e de intermediação dos vértices (BRANDES, 2001);

- Modularidade, que aplica um algoritmo de detecção de comunidades (BLONDEL et al., 2008) e calcula a resolução de modularidade da rede (LAMBIOTTE; DELVENNE; BARAHONA, 2008);

- PageRank, que calcula o valor do parâmetro deste mesmo nome para cada vértice da rede (BRIN; PAGE, 1998);

- Centralidade de autovetor, que calcula o valor do parâmetro deste mesmo nome para cada vértice da rede (BASTIAN; HEYMANN; JACOMY, 2009).

3.6 Visualização das Redes

3.6.1 Caracterização a partir das medidas de centralidade

Com o intuito de melhorar a visualização das redes do ponto de vista de entendimento dos vértices de maior importância, utilizou-se das medidas de centralidade para alterar a aparência (cores e tamanho) dos vértices. Foram utilizadas abordagens diferentes entre as redes de coautorias e citações, já que elas possuíam características distintas.

Para o caso das redes de coautorias, por se tratarem de redes não direcionadas, utilizou-se a centralidade de autovetor para variar o tamanho dos vértices, ou seja, vértices com centralidades maiores ficaram visualmente maiores na rede também. Desta forma, vértices com muitas conexões e vértices com vizinhos fortemente conectados foram valorizados, conforme estudado na teoria. Em alguns casos, também utilizou-se o grau ponderado para variar o tamanho dos vértices, para comparar os resultados obtidos em relação à centralidade do autovetor.

Para o caso das redes de citações, utilizou-se o parâmetro *PageRank* para variar o tamanho dos vértices, ou seja, vértices com valores de *PageRank* maiores ficaram visualmente maiores na rede também. Escolheu-se essa métrica por se tratar de uma rede direcionada, que apresenta alguns problemas para os casos das centralidades de autovetor ou Katz, solucionados pelo *PageRank*.

Para ambas as redes, de coautorias e de citações, também se utilizou a medida de centralidade de intermediação para variar o tamanho dos vértices, com o intuito de identificar vértices relevantes do ponto de vista de fluxo de informação dentro da rede.

Quanto à coloração dos vértices, para ambos os tipos de rede, a abordagem utilizada foi colorir os vértices de acordo com as comunidades em que foram inseridos automaticamente através do método de Louvain.

3.6.2 Aplicação dos algoritmos de visualização

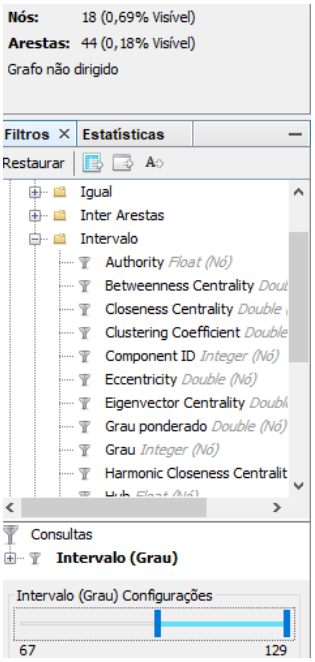
Os dois algoritmos escolhidos para a aplicação nas redes complexas foram o OpenOrd e o Force Atlas 2, por se tratarem de algoritmos com bom desempenho em evidenciar a presença de comunidades distintas dentro da rede, e também por serem otimizados para lidar com redes de larga escala, ou seja, com milhares ou até milhões de vértices e arestas.

O algoritmo de visualização OpenOrd foi aplicado nas redes de coautorias, por apresentar melhor desempenho para redes não direcionadas, segundo um material do próprio Gephi sobre os algoritmos de visualização disponíveis na ferramenta (GEPHI, 2008). Além disso, o algoritmo considera o peso das arestas para gerar as visualizações, o que também é importante para a rede de coautorias, que possui arestas com tal característica variável. Já o algoritmo de visualização Force Atlas 2 foi utilizado nas redes de citações, por ser adequado à aplicação em redes direcionadas.

Devido à alta quantidade de elementos nas redes geradas, mesmo com a aplicação dos algoritmos para melhorar a distribuição dos vértices nas visualizações das redes, houve dificuldade em se identificar quais os principais vértices das redes, ou seja, quais os principais autores ou artigos de maior relevância das áreas de pesquisa filtradas. Por isso, aplicou-se filtros de vértices para reduzir o número de elementos e gerar novas visualizações dos mais importantes vértices presentes em cada uma delas.

Para isso, utilizou-se a ferramenta de filtros do Gephi (BASTIAN; HEYMANN; JACOMY, 2009). Com ela, foi possível filtrar elementos específicos a partir de características da rede, por exemplo para se obter os vértices com maiores valores de grau, como ilustrado pela Figura 5.

Figura 5 – Filtros - Gephi



4 Análise e Discussão dos Resultados

Neste capítulo, serão mostradas as visualizações das redes obtidas e também as principais informações descobertas a partir da análise de cada uma das redes filtradas. Discutirá-se, em cada uma das seções, as características das redes geradas pelos filtros utilizados, nas quais se abordará tanto a questão de coautorias como citações dos artigos científicos. Vale salientar que as redes apresentaram tamanhos distintos, o que em alguns casos foi um fator facilitador ou dificultador das análises. Além disso, as redes de coautorias foram maiores que as redes de citações, no geral.

A Tabela 2 ilustra as quantidades de vértices e arestas de cada uma das redes filtradas. Nota-se que as redes filtradas por "CoronaVac" e *"Complex Network"* são as que possuem menores quantidades de vértices e arestas. As redes de "Astrazeneca", "Moderna" e "Janssen" têm proporções semelhantes, com "Pfizer" possuindo uma escala três vezes maior que elas em relação a número de vértices. Já as redes de filtros *"Comorbidities"*, "Pneumonia" e *"'Aged | Elderly"* geraram redes mais robustas, com cerca de 100 mil vértices.

Tabela 2 – Tamanho das Redes

Tipo de rede	Filtro	Quantidade de vértices	Quantidade de arestas
Coautorias	CoronaVac	2.621	24.759
	Astrazeneca	8.080	92.408
	Moderna	7.944	101.266
	Janssen	9.062	132.807
	Pfizer	23.786	273.268
	Comorbidities	89.498	927.833
	Pneumonia	113.781	916.819
	Aged | Elderly	120.440	1.049.001
	Network	117.714	1.949.488
	Complex Network	1.203	3.812
Citações	CoronaVac	1.646	1.940
	Astrazeneca	4.624	6.038
	Moderna	4.652	5.891
	Janssen	2.415	2.834
	Pfizer	11.007	17.140
	Comorbidities	72.131	108.643
	Pneumonia	97.334	152.815
	Aged | Elderly	143.227	184.207
	Network	184.474	213.278
	Complex Network	5.293	5.245

Um fator relevante para a análise das redes são suas comunidades, que foram geradas automaticamente a partir do método de Louvain (BLONDEL et al., 2008). Para todas as

redes, atribuiu-se cores características apenas para as principais comunidades presentes. Por isso, na maioria das visualizações completas das redes, as comunidades mais relevantes ficam coloridas ao centro e as menores comunidades com cor cinza ao redor. A Tabela 3 indica o número de comunidades gerado para cada uma das redes elaboradas. Nota-se que os números de comunidades são diretamente proporcionais aos tamanhos das redes.

Tabela 3 – Quantidades de Comunidades

Tipo de Rede	Filtro	Quantidade de comunidades
Coautorias	Coronavac	166
	Astrazeneca	570
	Moderna	458
	Janssen	390
	Pfizer	1201
	Pneumonia	7411
	Comorbidities	6043
	Aged | Elderly	8770
	Network	14275
	Complex Network	216
Citações	Coronavac	46
	Astrazeneca	75
	Moderna	75
	Janssen	63
	Pfizer	151
	Pneumonia	3569
	Comorbidities	2179
	Aged | Elderly	1156
	Network	6216
	Complex Network	98

Por fim, outra informação relevante é que nas redes de citações, há artigos que datam anteriormente a 2020, pois, apesar de se ter filtrado para se obter apenas publicações posteriores a esse ano, as publicações podiam citar artigos com datas anteriores a 2020, que também entraram na composição da rede.

4.1 Análise das redes sobre as vacinas

Primeiramente, realizou-se a análise das redes geradas a partir de termos relacionados aos nomes da vacinas desenvolvidas contra o coronavírus. Das redes filtradas para as cinco vacinas, as análises referentes à CoronaVac e à Astrazeneca se encontram nas próximas subseções. Já as análises das redes das outras três vacinas, Moderna, Janssen e Pfizer, encontram-se no Apêndice. Essa divisão foi devida ao fato de que as vacinas, em sua maioria, tiveram resultados

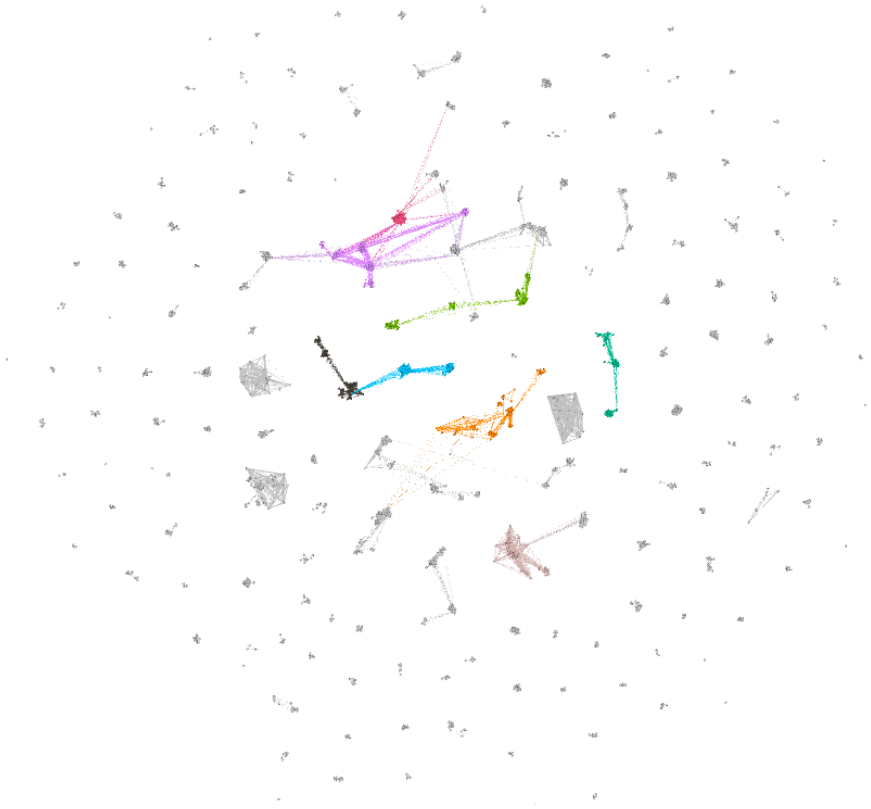
obtidos semelhantes, de forma que dois exemplos já ilustram as informações relevantes sobre o tema.

4.1.1 CoronaVac

4.1.1.1 Rede de Coautorias

A partir da otimização da modularidade, o método de Louvain identificou automaticamente as comunidades distintas para a rede. Após a execução do algoritmo de visualização OpenOrd, foi possível identificar algumas das principais comunidades, como ilustra a Figura 6.

Figura 6 – Comunidades de Autores - CoronaVac



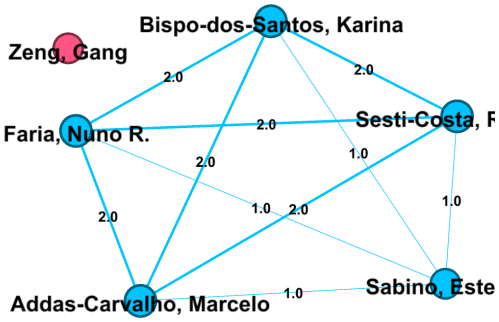
É possível ver muitas comunidades pequenas nas regiões periféricas e algumas comunidades maiores ao centro, estando algumas delas mais próximas, como a preta com a azul ou a lilás com a vermelha. No entanto, algumas delas estão totalmente isoladas, como é o caso da laranja, indicando um grupo científico que pouco interage com outros.

A Figura 7 ilustra a visualização dos top 6 vértices com maiores valores de centralidades de autovetor, característica que indica vértices de importância na rede. Pode-se notar que dos seis principais autores, cinco deles são da mesma comunidade (azul claro) e representam pesquisadores brasileiros. Vale destacar os nomes de Ester Sabino, imunologista e pesquisadora na Universidade de São Paulo (USP), e Nuno Faria, pesquisador na Universidade de Oxford

(Reino Unido), que são os dois principais responsáveis por liderar o sequenciamento do genoma do coronavírus.

O autor "Zeng, Gang" faz parte de outra comunidade (vermelha), que não tem ligação de com a comunidade azul claro. Trata-se do diretor clínico da Sinovac, empresa farmacêutica desenvolvedora da vacina CoronaVac. Vale ressaltar que ele é o pesquisador com maior grau ponderado dentro da rede, ou seja, que mais se conecta a outro autores.

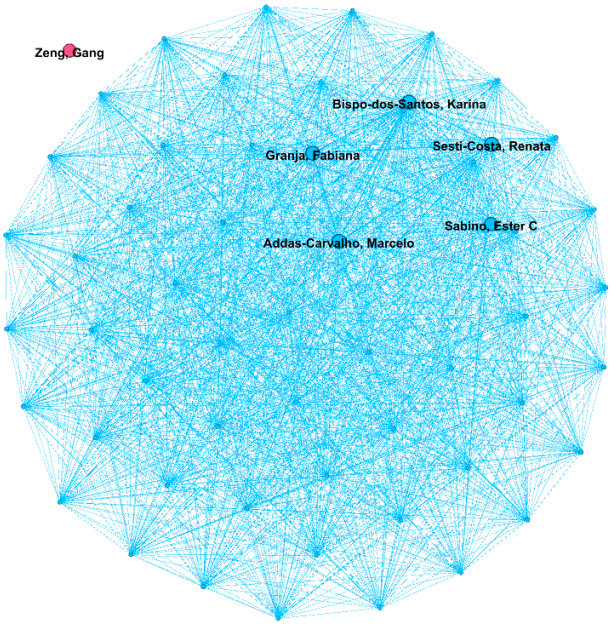
Figura 7 – Top 6 Autores - Centralidade de Autovetor - CoronaVac



Como os principais autores da rede são de fato pesquisadores de grande influência no contexto da CoronaVac, isso significa que esta e as outras redes de coautorias desenvolvidas neste trabalho atingiram seu objetivo, que era gerar a visualização dos principais autores envolvidos em algumas áreas de pesquisa.

Gerou-se também uma visualização com os 57 vértices com maiores valores de centralidade de autovetor da rede, como ilustra a Figura 8.

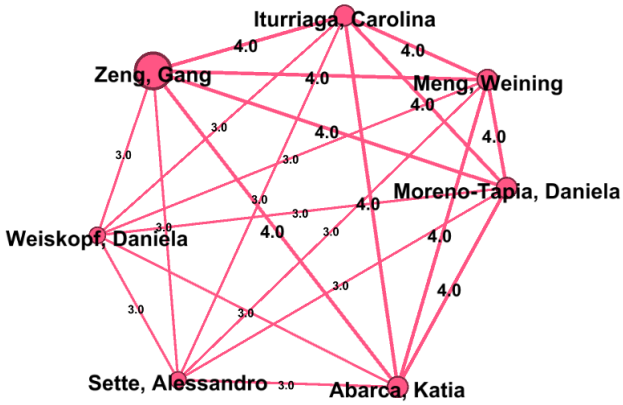
Figura 8 – Top 57 Autores - Centralidade de Autovetor - CoronaVac



Pode-se ver na Figura 8 que o vértice que representa o autor “Zeng, Gang” continua sem conexões, e todos os outros vértices fazem parte da mesma comunidade (azul claro). Nota-se também uma alta densidade dentro desta comunidade, visto que se tem milhares de arestas para algumas dezenas de vértices, indicando um grupo fortemente conectado. Dado o conhecimento sobre os principais autores da comunidade, conclui-se então que ela referencia um grupo de pesquisa majoritariamente brasileiro, que colabora entre si em muitos artigos relacionados à vacina, o que é coerente, visto que a CoronaVac foi uma vacina frequentemente estudada e aplicada em larga escala no Brasil.

Outra visualização que se traz para este tópico é dos principais autores da comunidade do vértice que representa “Zeng, Gang”, como ilustra a Figura 9.

Figura 9 – Comunidade de “Zeng, Gang” - CoronaVac



É possível identificar que de fato se tratam de autores influentes na área e ligados à farmacêutica Sinovac, como por exemplo Katia Abarca, que é especialista em doenças contagiosas e diretora médica dos estudos da vacina da Sinovac no Chile.

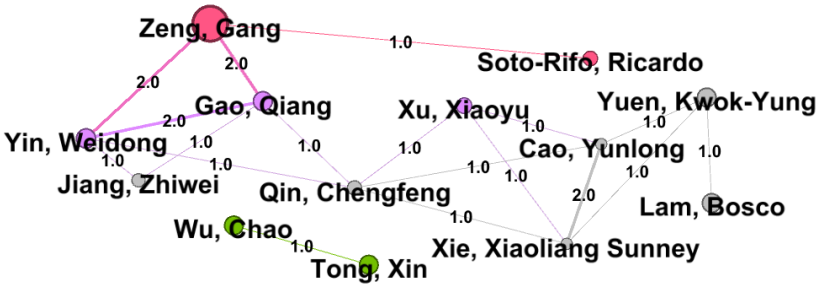
Uma outra visualização está representada na Figura 10, a qual ilustra os vértices com maiores valores de centralidade de intermediação, ou seja, vértices que mais intermedeiam caminhos entre outros vértices da rede. Do ponto de vista da rede de cooperação científica, isso significa que são vértices que podem fazer uma ponte entre diferentes comunidades de pesquisadores. Tais vértices podem ser importantes para a recomendação de parcerias científicas entre grupos distintos de pesquisadores.

Do ponto de vista de análise de uma rede de coautorias, pode-se dizer que vértices com elevadas centralidades de intermediação representam autores com influência em diferentes subáreas de estudo, e que, portanto, conectam grupos mais especializados em conteúdos específicos.

Ao se comparar Figura 7 com a Figura 10, é interessante notar que há apenas um vértice em comum nas duas redes, o autor “Zeng, Gang”. Isso indica que diferentes medidas de centralidade podem trazer conclusões distintas sobre a importância dos vértices em uma rede,

então é importante saber quais informações cada medida traz consigo no seu conceito para buscar o que se quer enxergar a partir da rede.

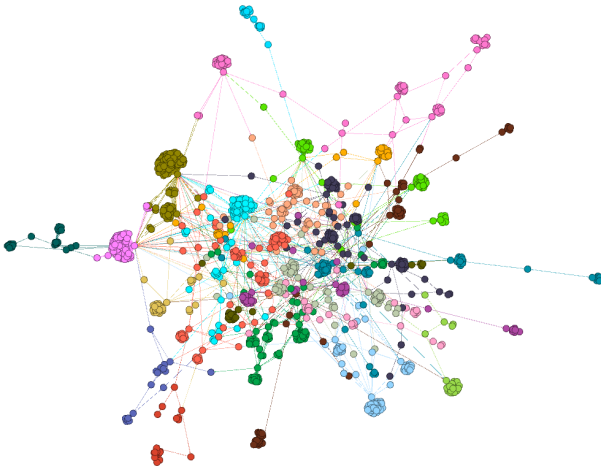
Figura 10 – Top 13 Autores - Centralidade de Intermediação - CoronaVac



4.1.1.2 Rede de Citações

A partir da otimização da modularidade, o método de Louvain identificou automaticamente as comunidades distintas para a rede. Em seguida, aplicou-se o algoritmo de visualização Force Atlas 2, de forma que se obteve a visualização da rede de citações, ilustrada pela Figura 11:

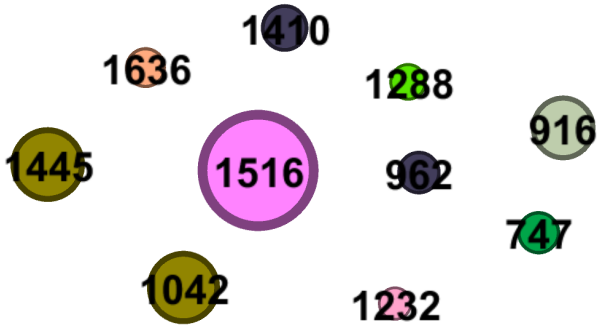
Figura 11 – Rede de Citações - CoronaVac



Podemos identificar na Figura 11 a presença de vários grupos de artigos científicos, que representam algumas subáreas dentre os artigos selecionados para esta rede, havendo conexões pontuais entre essas comunidades, que podem representar a interseção entre diferentes setores de estudo.

Em seguida, variou-se o tamanho dos vértices proporcionalmente às suas medidas de PageRank, que segundo o livro *"Networks: An Introduction"* (NEWMAN, 2010), mostrou-se uma das melhores métricas para avaliar a importância de um vértice dentro da rede para o caso de redes direcionadas. Com isso, obteve-se a rede ilustrada na Figura 12.

Figura 12 – Top 10 Artigos - *PageRank* - CoronaVac



A Tabela 4 ilustra os títulos de cada um dos artigos da Figura 13. O artigo com maior valor *PageRank* é indicado pelo vértice com ID 1042, intitulado *"Single-cell RNAseq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019- nCoV infection"* (ZOU et al., 2020), que estuda os riscos da doença para diferentes órgãos do corpo humano. De acordo com a plataforma Google Scholar, atualmente o artigo tem 1874 citações, evidenciando que se trata de um artigo extremamente relevante para a área de infectologia no geral, não somente para o desenvolvimento da vacina CoronaVac.

Tabela 4 – Top 10 Artigos - *PageRank* - CoronaVac

ID	Título
747	Efficacy and Safety of a COVID-19 Inactivated Vaccine in Healthcare Professionals in Brazil: The PROFISCOV Study
916	Omicron variant showed lower neutralizing sensitivity than other SARS-CoV-2 variants to immune sera elicited by vaccines after boost
962	COVID-19 vaccination: What's the evidence for extending the dosing interval?
1042	Single-cell RNAseq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection
1232	Immunogenicity after two doses of inactivated virus vaccine in healthcare workers with and without previous COVID-19 infection: prospective observational study
1288	Omicron SARS-CoV-2 variant: What we know and what we don't. Anaesth Crit Care Pain Med.
1410	A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker)
1445	A universal design of betacoronavirus vaccines against COVID-19, MERS, and SARS
1516	Therapeutic potential of medicinal plants against COVID-19: The role of antiviral medicinal metabolites
1636	Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations

Assim como para as redes de coautorias, outra métrica utilizada para a rede de citações é a centralidade de intermediação. Assim como foi feito para o caso do PageRank, filtrou-se os vértices com os 10 maiores valores de centralidade de intermediação, como ilustra a Figura 13, juntamente de seus títulos, presentes na Tabela 5.

Figura 13 – Top 10 Artigos - Centralidade de Intermediação - CoronaVac

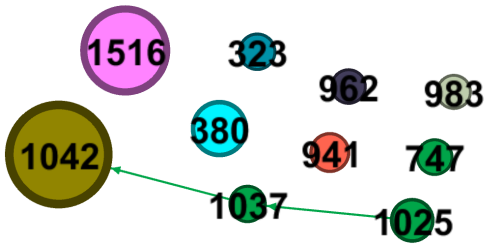


Tabela 5 – Top 10 Artigos - Centralidade de Intermediação - CoronaVac

ID	Título
323	T cell immune responses to SARS-CoV-2 and variants of concern (Alpha and Delta) in infected and vaccinated individuals
380	Covid-19 vaccines and variants of concern: A review
747	Efficacy and Safety of a COVID-19 Inactivated Vaccine in Healthcare Professionals in Brazil: The PROFISCOV Study
941	Single-dose administration and the influence of the timing of the booster dose on immunogenicity and efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine: A pooled analysis of four randomised trials
962	COVID-19 vaccination: What's the evidence for extending the dosing interval?
983	Surrogate Virus Neutralization Test Based on Antibody-Mediated Blockage of ACE2-Spike Protein-Protein Interaction
1027	MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization
1035	SARS-CoV-2 variants show resistance to neutralization by many monoclonal and serum-derived polyclonal antibodies
1042	Single-cell RNAseq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection
1516	Therapeutic potential of medicinal plants against COVID-19: The role of antiviral medicinal metabolites

Do ponto de vista desta rede de citações, que se mostra modularizada e com grupos bem definidos, vértices com alta centralidade de intermediação indicam artigos que conectam comunidades, ou seja, artigos que têm conteúdo que aborda parte de cada uma das comunidades que interliga. Um exemplo interessante de trabalho com tais características é o artigo *"Covid-19 vaccines and variants of concern: A review"*, que faz uma revisão completa de todas as vacinas e variantes da doença. Destaca-se também o artigo *Efficacy and Safety of a COVID-19*

Inactivated Vaccine in Healthcare Professionals in Brazil: The PROFISCOV Study, estudo do Instituto Butantan para entender a eficácia e segurança da CoronaVac em profissionais de saúde no Brasil, que ficaram na linha de frente nos tratamentos contra a COVID-19.

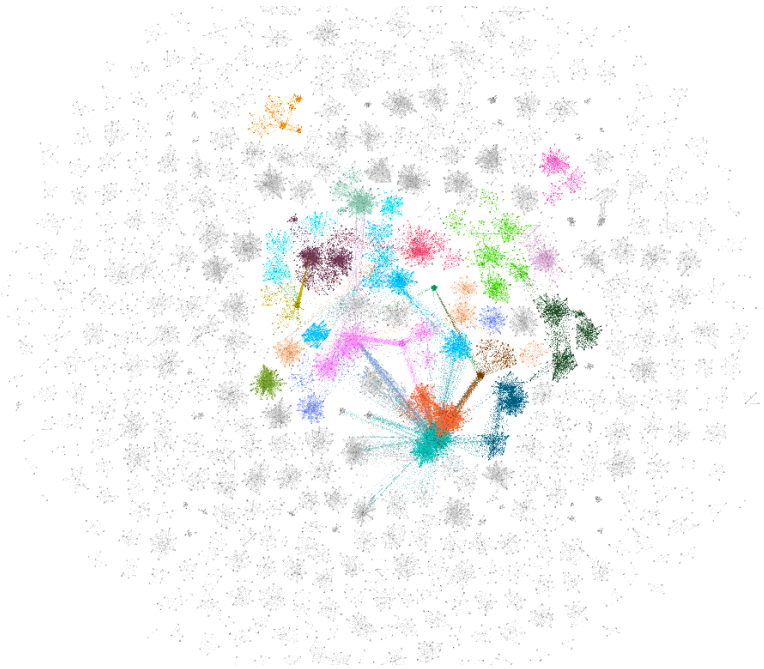
Outro fator interessante de ser analisado é a presença de quatro dos artigos em ambos os top 10 gerados, PageRank e centralidade de intermediação, o que indica que tratam-se de artigos relevantes tanto nas suas respectivas áreas quanto em outras áreas correlatas, sendo assim importantes no contexto da rede de citações.

4.1.2 Astrazeneca

4.1.2.1 Rede de Coautorias

A rede de coautorias para o filtro "Astrazeneca" é ilustrada pela Figura 14. É possível observar grande proximidade entre as duas maiores comunidades (verde-água e laranja). Além disso, nota-se a presença de outros grupos com tamanho relevante, como o rosa e o verde claro.

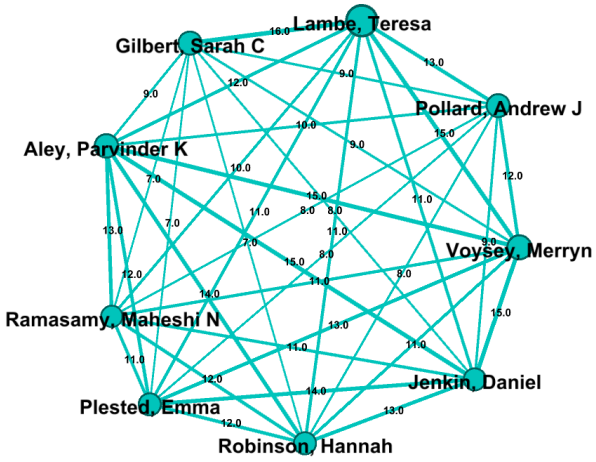
Figura 14 – Comunidades de Autores - Astrazeneca



A Figura 15 ilustra a rede com os nove vértices com maiores valores de centralidade de autovetor, com o tamanho dos vértices proporcional a estes valores. Destaca-se a pesquisadora "Lambe, Teresa", que se trata de uma pesquisadora da área de imunologia da universidade de Oxford, sendo também uma das principais responsáveis pelo desenvolvimento da vacina Astrazeneca.

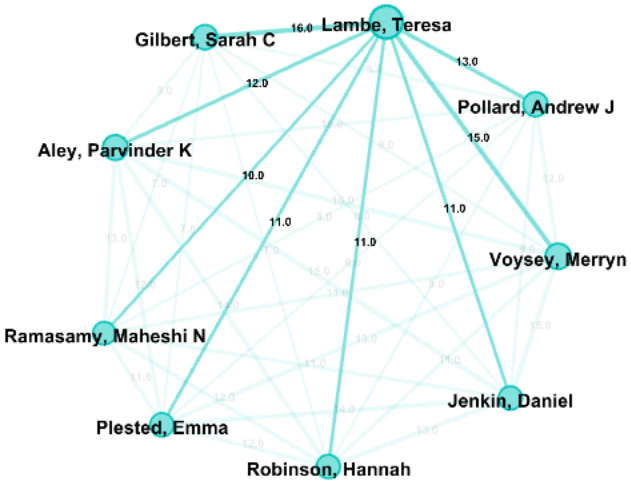
A Figura 16 ilustra todas as conexões de Teresa Lambe com os outros pesquisadores, bem como o peso das arestas, que representa o número de relações de coautoria com cada autor.

Figura 15 – Top 9 Autores - Centralidade de Autovetor - Astrazeneca



Nota-se que se trata em geral de uma rede fortemente conectada, tratando-se do principal grupo de pesquisa responsável pelo desenvolvimento da vacina.

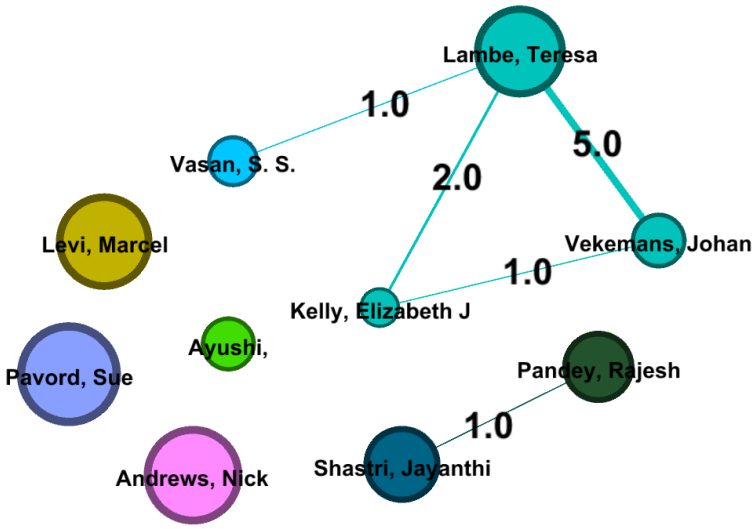
Figura 16 – Conexões Teresa Lambe - Astrazeneca



A Figura 17 ilustra os dez pesquisadores com maiores centralidades de intermediação, com seus tamanhos proporcionais a tais valores. Nota-se que o vértice correspondente à Teresa Lambe também tem grande relevância neste caso, o que indica que além de ter forte presença na sua própria comunidade, ela também é central do ponto de vista de conexão com outros grupos, sendo portanto, uma pesquisadora importante para promover a colaboração entre diferentes grupos de pesquisa.

O vértice com maior centralidade de autovetor é o representado por “Pavord, Sue”, uma das autoras do artigo “Clinical Features of Vaccine-Induced Immune Thrombocytopenia and Thrombosis” (PAVORD et al., 2021), que estuda uma nova síndrome desenvolvida por efeitos colaterais à vacina da Astrazeneca, o que é uma área de estudo diferente de Teresa Lambe, que tem foco no desenvolvimento da vacina.

Figura 17 – Top 10 Autores - Centralidade de Intermediação - Astrazeneca



4.1.2.2 Rede de Citações

A Figura 18 ilustra a rede de citações para os artigos filtrados com os nomes da vacina Astrazeneca. É possível notar a presença de algumas comunidades principais, com maiores quantidades de vértices, como a azul claro, a verde claro e a rosa. Além disso, nota-se que dentro das próprias comunidades, há pontos de concentração de alguns grupos de vértices também.

Figura 18 – Rede de Citações - Astrazeneca

A Figura 19 e a Tabela 6 ilustram, respectivamente, a representação dos vértices com maiores valores de *PageRank* e o índice e título de cada um dos artigos representados pelos vértices. Dos títulos, pode-se notar a presença de assuntos recorrentes nas áreas de pesquisa, como o caso da síndrome estudada por Sue Pavord na seção anterior, *"Vaccine-Induced*

Immune Thrombocytopenia and Thrombosis", mostrando que trata-se de um assunto de alta relevância na comunidade científica. Além disso, outros assuntos interessantes são referentes à resposta das vacinas às variantes da doença e também efetividade sobre aplicação de doses extras.

Figura 19 – Top 11 Artigos - *PageRank* - Astrazeneca

Tabela 6 – Top 11 Artigos - *PageRank* - Astrazeneca

ID	Título
697	Adenovirus-mediated overexpression of novel mutated IkappaBalpha inhibits nuclear factor kappaB activation in endothelial cells
3121	Treatment of ChAdOx1 nCoV-19 vaccine-induced immune thrombotic thrombocytopenia related acute ischemic stroke
3589	Effectiveness of a third dose of the BNT162b2 mRNA COVID-19 vaccine for preventing severe outcomes in Israel: An observational study
3714	Effectiveness of COVID-19 Vaccines against the B.1.617.2 (Delta) Variant
3831	Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations
3897	Platelet factor 4 binds to bacteria, inducing antibodies cross-reacting with the major antigen in heparin-induced thrombocytopenia
3994	Laboratory testing for suspected COVID-19 vaccine-induced (immune) thrombotic thrombocytopenia
4109	Addressing the vaccine confidence gap
4391	Two doses of the SARS-CoV-2 BNT162b2 vaccine enhances antibody responses to variants in individuals with prior SARS-CoV-2 infection 2021
4405	Therapeutic potential of medicinal plants against COVID-19: The role of antiviral medicinal metabolites
4557	SARS-CoV-2 Vaccine ChAdOx1 nCoV-19 Infection Of Human Cell Lines Reveals Low Levels of Viral Backbone Gene Transcription Alongside Very High Levels of SARS-CoV-2 S Glycoprotein Gene Transcription

Analogamente para o caso do *PageRank*, a Figura 20 e a Tabela 7 ilustram, respectivamente, a representação dos vértices com maiores valores de centralidade de intermediação e o índice e título de cada um dos artigos representados pelos vértices. Nota-se que três desses artigos também estavam presentes para o caso dos maiores *PageRank*, o que indica artigos que além de importantes em suas respectivas áreas, também são centrais para a rede como um todo. É possível destacar o vértice com identificador 697, que tem os maiores valores tanto em *PageRank* quanto em centralidade de intermediação. Esse vértice representa o

artigo *"Adenovirus-mediated overexpression of novel mutated IkappaBalpha inhibits nuclear factor kappaB activation in endothelial cells"* (WRIGHTON et al., 1996), que possui data de publicação consideravelmente anterior ao surgimento da COVID-19, indicando que a rede complexa é capaz de mapear a contribuição de estudos de diferentes épocas para o desenvolvimento de novas áreas dentro da ciência.

Figura 20 – Top 12 Artigos - Centralidade de Intermediação - Astrazeneca

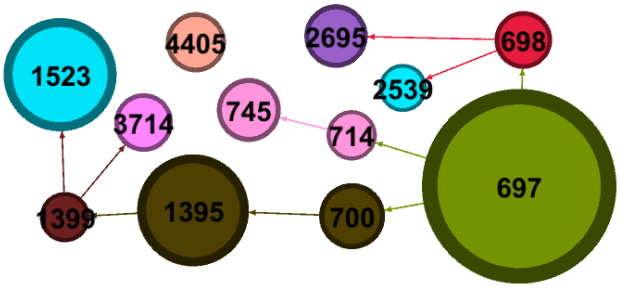


Tabela 7 – Top 12 Artigos - Centralidade de Intermediação - Astrazeneca

ID	Título
697	Adenovirus-mediated overexpression of novel mutated IkappaBalpha inhibits nuclear factor kappaB activation in endothelial cells
698	Safety and Efficacy of Single-Dose Ad26.COV2.S Vaccine against Covid-19
700	Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneu-monia in
714	Syndromes of Thrombotic Microangiopathy
745	Lessons Learnt from COVID
1395	COVID-19 Vaccines vs. Variants-Determining How Much Immunity Is Enough
1399	A pneumonia outbreak associated with a new coronavirus of probable bat origin
1523	How to cite this article: Hadj Hassine I. Covid-19 vaccines and variants of concern: A review
2539	Preferences for COVID-19 vaccine distribution strategies in the US: A discrete choice survey
2695	Monitoring of COVID-19 medicines
3714	Effectiveness of COVID-19 Vaccines against the B.1.617.2 (Delta) Variant
4405	Therapeutic potential of medicinal plants against COVID-19: The role of antiviral medicinal metabolites

4.2 Análise das redes sobre tópicos obtidos pelo LDA

Nesta seção, analisou-se as redes filtradas a partir dos tópicos encontrados pelo algoritmo LDA, que foram "Pneumonia" e *"Comorbidities"*.

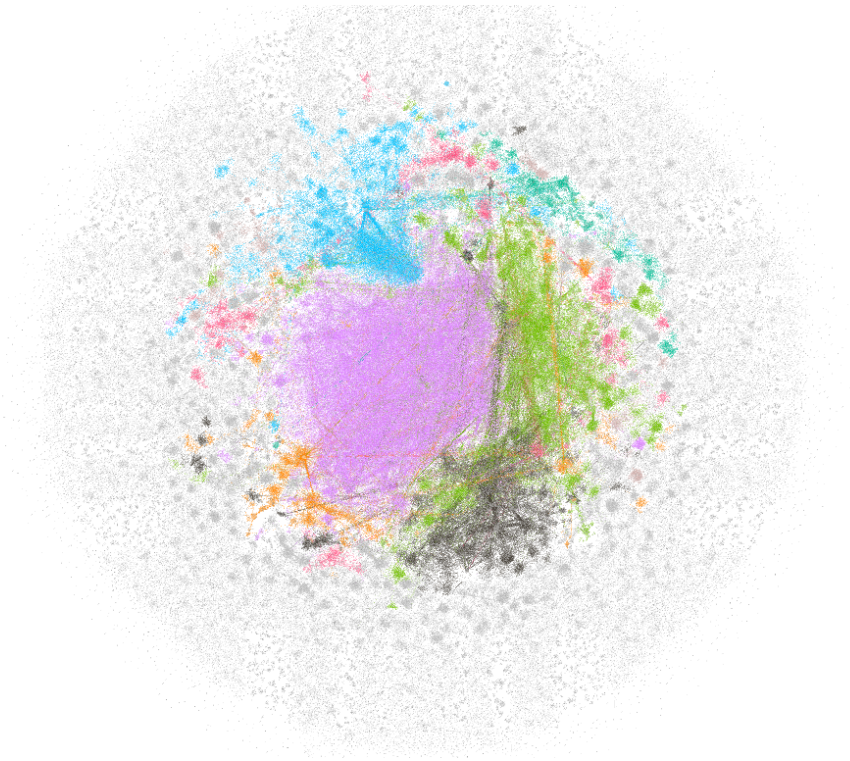
4.2.1 Pneumonia

A pneumonia é uma doença respiratória que antecede a existência da COVID-19, mas que também pode vir a ser um de seus possíveis casos de complicação. Serão abordadas a seguir as redes referentes a este tópico.

4.2.1.1 Rede de Coautorias

A Figura 21 ilustra a rede de coautorias filtrada a partir do termo pneumonia. Em comparação com as redes de vacinas, nota-se a presença de principais comunidades maiores em termos do percentual total de vértices da rede, como é o exemplo das comunidades lilás, verde claro e azul claro.

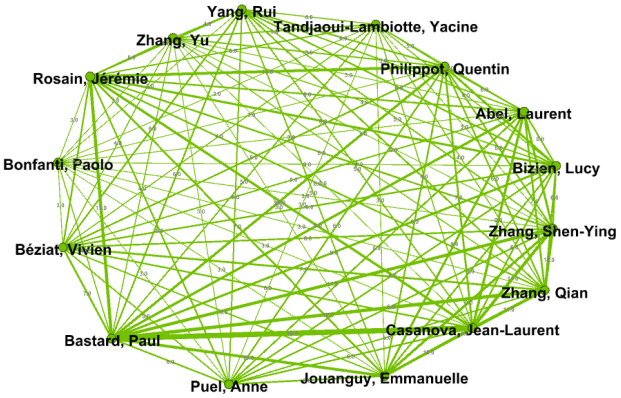
Figura 21 – Comunidades de Autores - Pneumonia



A Figura 22 ilustra os principais quinze vértices do ponto de vista da métrica de centralidade de autovetor, em que todos fazem parte da mesma comunidade (verde claro). O vértice com maior centralidade de autovetor representa *"Rosain, Jérémie"*, um pesquisador em imunologia

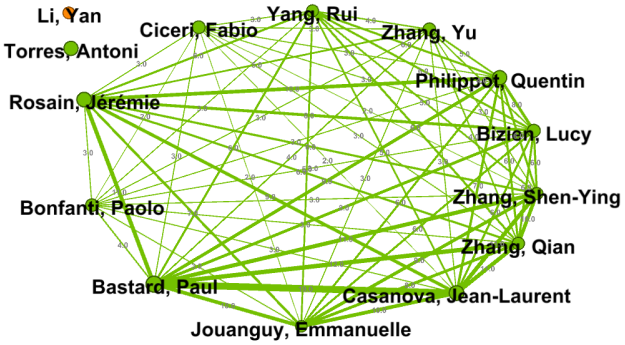
do *Institut Imagine*, que analisa os casos graves de COVID-19 e busca associações da doença a fatores genéticos.

Figura 22 – Top 15 Autores - Centralidade de Autovetor - Pneumonia



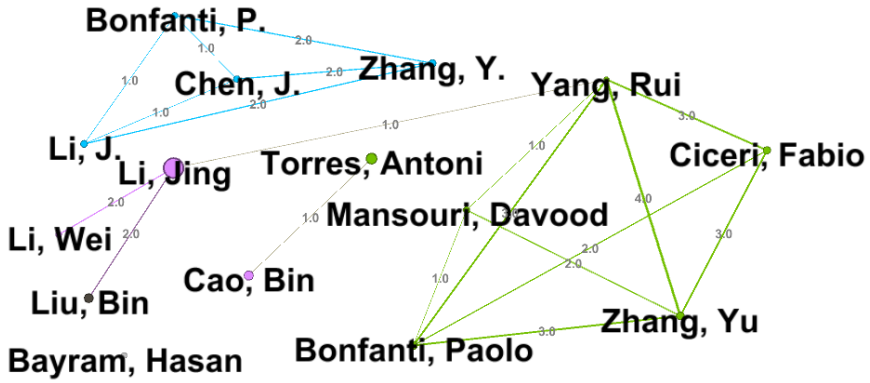
A Figura 23 ilustra os principais quinze vértices do ponto de vista da métrica de grau ponderado, em que quase todos fazem parte da mesma comunidade (verde claro) encontrada com a centralidade de autovetor, havendo também a presença de um vértice da comunidade laranja. O vértice com maior valor para essa métrica representa o autor *"Bastard, Paul"*, que também faz parte do *Institut Imagine* e é o principal autor do artigo *"Preexisting autoantibodies to type I IFNs underlie critical COVID-19 pneumonia in patients with APS-1"* (BASTARD et al., 2021), que estuda a associação de fatores genéticos a pacientes com complicação de pneumonia após se infectarem com a COVID-19. Na lista de coautores deste artigo em específico, é possível ver vários dos nomes presentes na rede em questão.

Figura 23 – Top 15 Autores - Grau Ponderado - Pneumonia



A Figura 24 ilustra os vértices com os maiores valores de centralidade de intermediação da rede, havendo a presença de três das principais comunidades da rede (lilás, verde claro e azul claro). O vértice com maior valor para esta métrica representa o pesquisador "Li, Jing", que conecta a comunidade lilás aos principais pesquisadores da comunidade verde claro, através da relação de coautoria com "Yang, Rui".

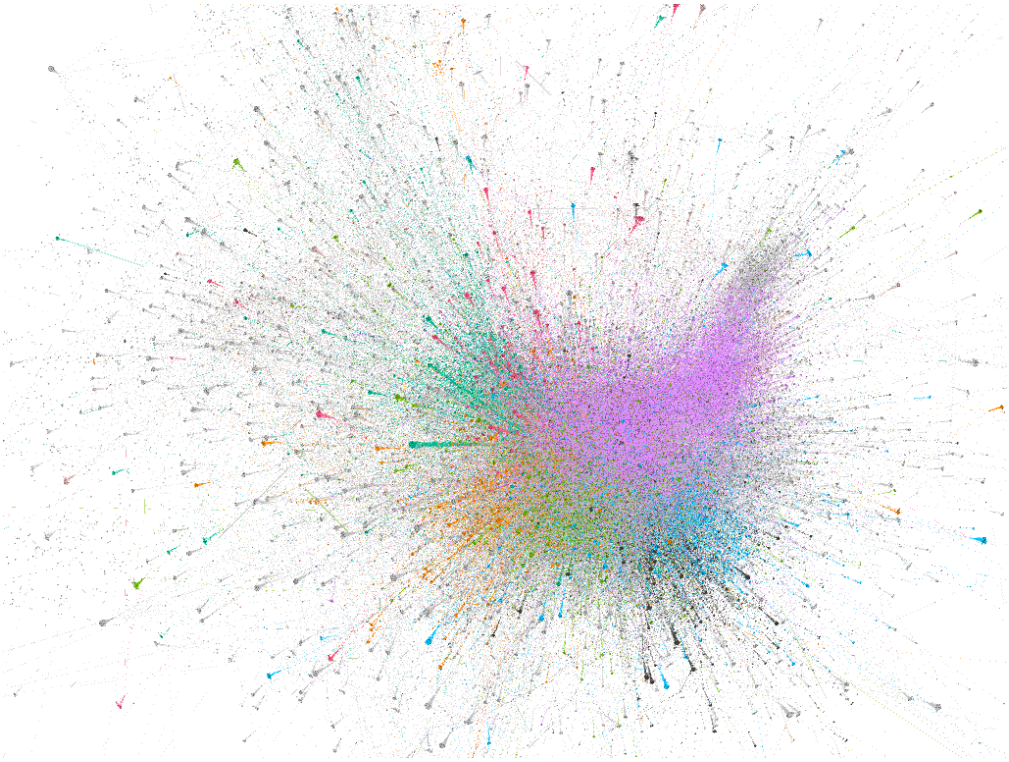
Figura 24 – Top 15 Autores - Centralidade de Intermediação - Pneumonia



4.2.1.2 Rede de Citações

A Figura 25 ilustra a rede de citações gerada para o filtro da palavra "pneumonia". Em comparação com as redes geradas para as vacinas, nota-se a maior dificuldade de se detectar polos dentro das comunidades, devido ao maior volume de vértices. Por outro lado, pode-se ver a divisão das comunidades de forma mais uniforme, com a comunidade lilás sendo a principal da rede, envolta de outras de tamanho também relevante, como por exemplo a laranja e a verde claro.

Figura 25 – Rede de Citações - Pneumonia



A Figura 26 e a Tabela 8, ilustram, respectivamente, os vértices com maiores valores de *PageRank* e o índice e título de cada um dos artigos representados pelos vértices. Desta vez,

não vemos mais interseção com artigos mais citados para o caso das vacinas, indicando que de fato trata-se de uma rede que aborda outras áreas de estudo, pelo menos na visão de seus principais vértices. No geral, ao se analisar o conteúdo dos títulos dos artigos, é possível constatar que de forma geral a maioria deles aborda áreas de estudo relacionadas à como a doença afeta diferentes tipos de pessoas, envolvendo variáveis como sexo, idade e também fatores genéticos.

Figura 26 – Top 12 Artigos - *PageRank* - Pneumonia

Tabela 8 – Top 12 Artigos - *PageRank* - Pneumonia

ID	Título
48543	Kynurenic acid underlies sex-specific immune responses to COVID-19
50058	COVID-19 Treatment Guidelines
76072	Thrombocytopathy and endotheliopathy: crucial contributors to COVID-19 thromboinflammation
80086	The omicron variant of SARS-CoV-2: Understanding the known and living with unknowns
81176	Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant
84802	Swine acute diarrhea syndrome coronavirus replication in primary human cells reveals potential susceptibility to infection
85535	Effectiveness of Dexmedetomidine Combined with High Flow Nasal Oxygen and Long Periods of Awake Prone Positioning in Moderate or Severe COVID-19 Pneumonia
85760	Coronavirus global pandemic: An overview of current findings among pediatric patients
89884	Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study
92168	Biochemical and structural insights into the mechanisms of SARS coronavirus RNA ribose 2'-O-methylation by nsp16/nsp10 protein complex
95257	Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations
96736	The griffithsin dimer is required for high-potency inhibition of HIV-1: evidence for manipulation of the structure of gp120 as part of the griffithsin dimer mechanism

A Figura 27 e a Tabela 9, ilustram respectivamente, os vértices com maiores valores de centralidade de intermediação e o índice e título de cada um dos artigos representados por eles.

Comparando essa tabela com a Tabela 8, nota-se uma interseção de apenas um vértice, de índice 48.543, que representa o artigo de título "Kynurenic acid underlies sex-specific immune responses to COVID-19" (CAI et al., 2021). Esse artigo estuda a relação entre a metabolização de ácido quinurênico com as diferenças da taxa de mortalidade entre sexos perante à infecção do coronavírus, e foi publicado na *Science*, uma das revistas mais importantes do mundo.

Figura 27 – Top 13 Artigos - Centralidade de Intermediação - Pneumonia

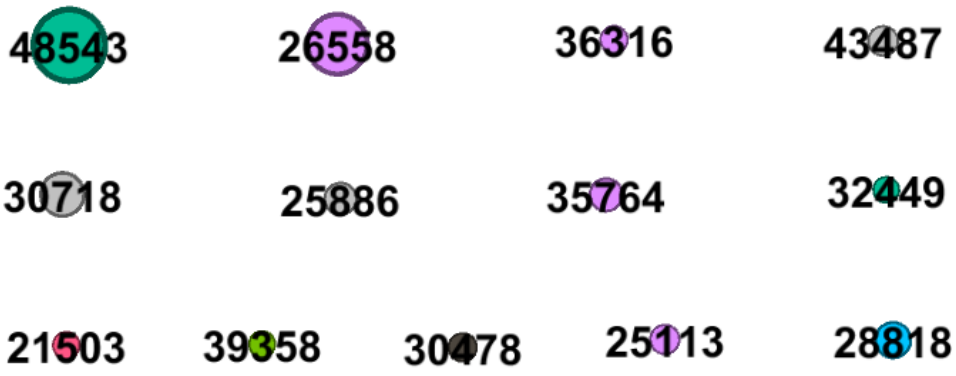


Tabela 9 – Top 13 Artigos - Centralidade de Intermediação - Pneumonia

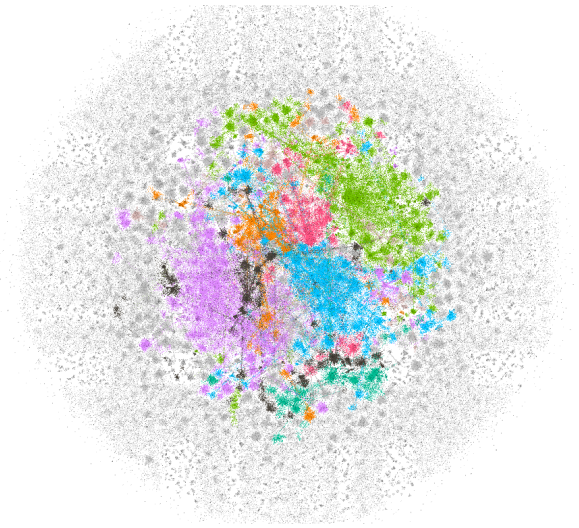
ID	Título
21503	Coronaviruses-drug discovery and therapeutic options
25113	Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection?
25886	Human recombinant soluble ACE2 in severe COVID-19
26558	COVID-19 infection in children
28818	Complement as a target in COVID-19?
30478	Single-cell RNAseq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection
30718	An in-silico evaluation of dietary components for structural inhibition of SARS-Cov-2 main protease
32449	Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations
35764	A novel coronavirus from patients with pneumonia in China
36316	Lung cavitation in COVID-19: Co-infection complication or rare evolution?
39358	The Possible Role of Vitamin D in Suppressing Cytokine Storm and Associated Mortality in COVID-19 Patients
43487	A framework for One Health research
48543	Kynurenic acid underlies sex-specific immune responses to COVID-19

4.2.2 *Comorbidities*

4.2.2.1 Rede de Coautorias

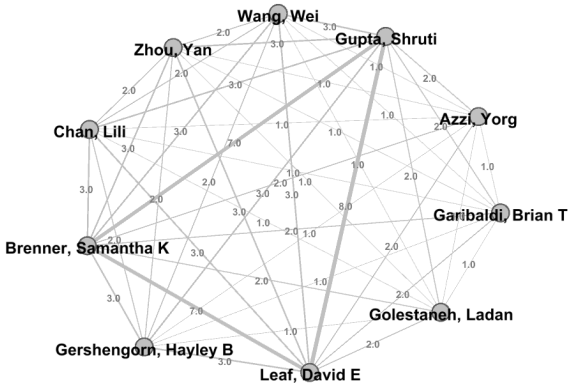
A Figura 28 ilustra a rede de coautorias filtrada a partir do termo *comorbidities* ou "comorbidades", que indica características que possam ser agravantes para os pacientes em caso de infecção por COVID-19. Nota-se a presença de três comunidades principais: a lilás e a verde claro, com a azul claro entre elas.

Figura 28 – Comunidades de Autores - Comorbidities



A Figura 29 ilustra os vértices com maiores valores de centralidade de autovetor, em que todos fazem parte de uma comunidade que não é das mais volumosas da rede. Os vértices com maiores centralidades de autovetor representam os pesquisadores "Wang, Wei" e "Gupta, Shruti", que são coautores por exemplo do artigo "Association Between Early Treatment With Tocilizumab and Mortality Among Critically Ill Patients With COVID-19" (GUPTA et al., 2021), um estudo sobre a associação entre o tratamento precoce com o anticorpo tocilizumabe e a taxa de mortalidade de pacientes infectados com COVID-19.

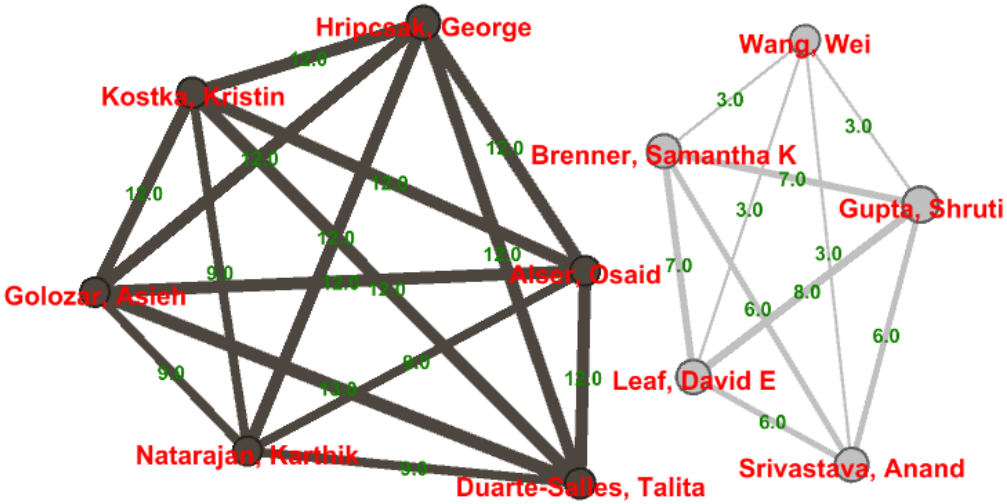
Figura 29 – Top 10 Autores - Centralidade de Autovetor - Comorbidities



A Figura 30 ilustra os vértices com maiores valores de grau ponderado, incluindo alguns dos autores presentes na Figura 29, mas também uma outra comunidade de autores, de cor cinza escuro. O principal autor para esta métrica é George Hripcsak, pesquisador na área informática aplicada à biomedicina na *Columbia University - Data Science Institute*, tendo contribuições para os estudos sobre a COVID-19 acerca das áreas de ciência e análise de dados. É possível encontrar alguns dos autores desta visualização no artigo *"Characteristics and Outcomes of Over 300,000 Patients with COVID-19 and History of Cancer in the United States and Spain"* (ROEL et al., 2021), que aborda características de pacientes infectados com COVID-19 que tiveram câncer em seu histórico médico.

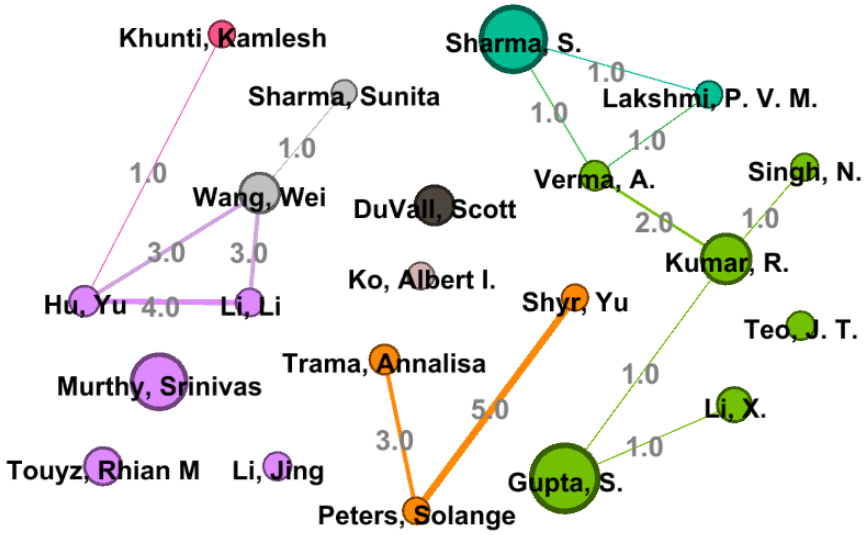
Outro artigo em que temos a presença de George Hripcsak é "Observational Study of Hydroxychloroquine in Hospitalized Patients with Covid-19" (GELERIS et al., 2020), um estudo sobre o uso de hidroxicloroquina como medicação para pacientes com COVID-19, tema polêmico no Brasil e no mundo desde o começo da pandemia.

Figura 30 – Top 11 Autores - Grau Ponderado - Comorbidities



A Figura 31 ilustra os vértices com os maiores valores de centralidade de intermediação da rede, havendo a presença de diversas comunidades da rede. É interessante ressaltar a presença do pesquisador "Wang, Wei", que também é representado por um dos vértices com maior centralidade de autovetor. Além disso, esta visualização indicou um problema no método de criação das redes, que é a presença de diferentes formas de escrita de nome para um mesmo autor, o que resulta na duplicação de um mesmo pesquisador dentro da rede. O exemplo que ilustra este comportamento é o caso do vértice rotulado "Gupta, S.", que também aparece como "Gupta, Shruti" na Figura 29.

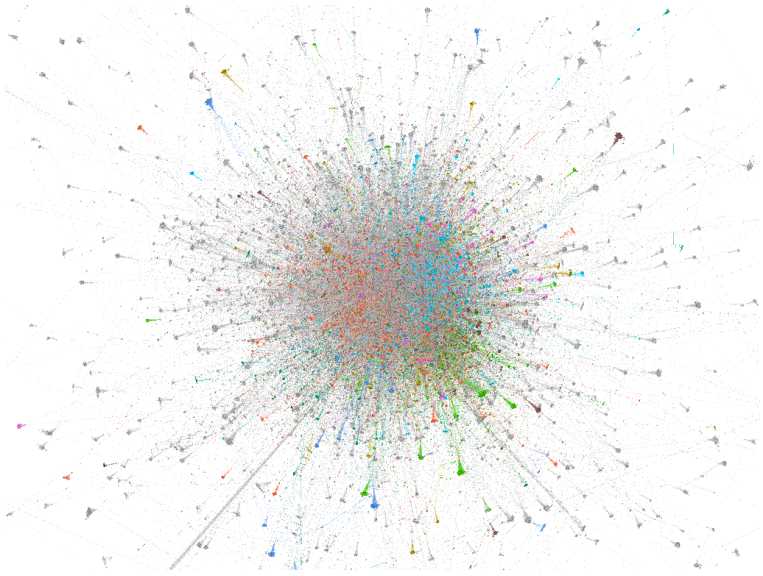
Figura 31 – Top 21 Autores - Centralidade de Intermediação - Comorbidities



4.2.2.2 Rede de Citações

A Figura 32 ilustra a rede de citações gerada para o filtro *"comorbidities"*. Visualmente, é difícil identificar as principais comunidades existentes a partir das cores, visto que seus vértices acabaram ficando muito próximos, o que indica uma rede fortemente conectada entre as diferentes comunidades existentes.

Figura 32 – Rede de Citações - Comorbidities



A Figura 33 e a Tabela 10 ilustram os vértices com maiores valores de *PageRank* e o índice e título de cada um dos artigos representados pelos vértices, respectivamente. Ao se analisar o conteúdo dos títulos dos artigos, é possível constatar que de forma geral a maioria deles aborda relações entre características de pacientes com infecções de COVID-19, o que de

fato compactua com a definição de comorbidade dentro do contexto do estudo de doenças. Vale destacar o artigo intitulado *"A 21st Century Evil: Immunopathology and New Therapies of COVID-19"* (SILVA et al., 2020), que tem como coautores pesquisadores brasileiros e aborda a interação entre o SARS-CoV-2 e o sistema imunológico do hospedeiro do paciente durante a infecção, além de discutir os principais mecanismos imunopatológicos envolvidos no desenvolvimento da doença e possíveis novas abordagens terapêuticas.

Figura 33 – Top 11 Artigos - *PageRank* - Comorbidities

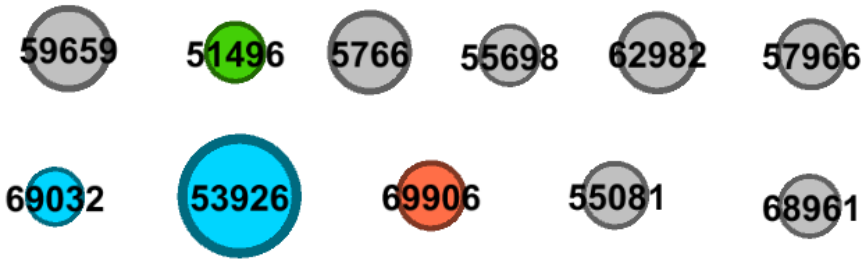


Tabela 10 – Top 11 Artigos - *PageRank* - Comorbidities

ID	Título
5766	Umbilical cord mesenchymal stem cell treatment for Crohn's disease: a randomized controlled clinical trial
51496	Antiplatelet drug therapy moderates immune-mediated liver disease and inhibits viral clearance in mice infected with a replication-deficient adenovirus
53926	A 21st Century Evil: Immunopathology and New Therapies of COVID-19
55081	Pathogenic T cells and inflammatory monocytes incite inflammatory storm in severe COVID-19 patients
55698	Development and delivery of a real-time hospital-onset COVID-19 surveillance system using network analysis
57966	Protecting people with multiple sclerosis through vaccination
59659	Pituitary Disorders and COVID-19, Reimagining Care: The Pandemic A Year and Counting
62982	The Nrf2-ARE pathway: An indicator and modulator of oxidative stress in neurodegeneration
68961	Neutralization of variant under investigation b.1.617 with sera of bbv152 vaccinees
69032	Indian Genome Variation Consortium; Dash, D. IGVBrowser-A genomic variation resource from diverse Indian populations
69906	SARS to COVID-19: what we have learned about children infected with COVID

A Figura 34 e Tabela 11 ilustram os vértices com maiores valores de centralidade de intermediação e o índice e título de cada um dos artigos representados por eles, respectivamente. Ao se analisar os títulos da tabela, bem como o tamanho dos vértices, é interessante ressaltar

o artigo intitulado *"Comorbidities and the risk of severe or fatal outcomes associated with coronavirus disease 2019: A systematic review and meta-analysis"* (ZHOU et al., 2020b), que traz consigo uma revisão avaliativa sobre a associação entre diferentes comorbidades e a gravidade da COVID-19, que é um assunto central tendo em vista o filtro utilizado para criação desta rede.

Figura 34 – Top 12 Artigos - Centralidade de Intermediação - Comorbidities

Tabela 11 – Top 12 Artigos - Centralidade de Intermediação - Comorbidities

ID	Título
12582	Single-cell RNAseq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection
16445	Angiotensinconverting enzyme 2 protects from lethal avian influenza A H5N1 infections
16854	Immunometabolic Dysregulation at the Intersection of Obesity and COVID-19
18250	Self-testing for the detection of SARS-CoV-2 infection with rapid antigen tests for people with suspected COVID-19 in the community
24070	Clinical characteristics of COVID-19 patients with digestive symptoms in Hubei, China: a descriptive, cross-sectional, multicenter study
26139	Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations
28308	Coronavirus disease (COVID-19) current status and future perspectives: A narrative review
36973	Comorbidities and the risk of severe or fatal outcomes associated with coronavirus disease 2019: a systematic review and meta-analysis
43928	Clinical and pathological findings in SARS-CoV-2 Disease Outbreaks in Farmed Mink
46371	CDC. Symptoms of Coronavirus
51496	Antiplatelet drug therapy moderates immune-mediated liver disease and inhibits viral clearance in mice infected with a replication-deficient adenovirus
53926	A 21st Century Evil: Immunopathology and New Therapies of COVID-19

4.3 Análises de termos relevantes

Nesta seção, foram analisadas redes filtradas a partir de termos relevantes dos pontos de vista de áreas de estudo sobre o coronavírus e também em relação a redes complexas. Os filtros utilizados foram *"Aged / Elderly"* e *"Complex Network"*

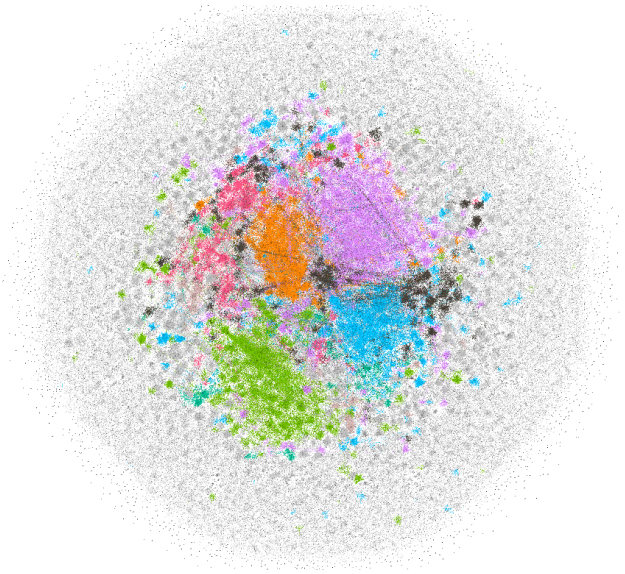
4.3.1 *Aged / Elderly*

As redes elaboradas nesta seção foram filtradas a partir dos termos *"aged"* e *"elderly"*, com o intuito de se buscar artigos relacionando o comportamento da COVID-19 com a idade dos pacientes idosos. No entanto, após se analisar alguns dos artigos filtrados, notou-se que a maioria deles enquadrava apenas o termo *"aged"* em seus resumos, e não somente para se referir a pessoas idosas, mas no sentido de atribuir idade a pacientes analisados, como por exemplo na frase retirada de um dos artigos: *"...from two Canadian hospitals of consecutive ICU patients aged equal or higher than 18 years"*. Por isso, pode-se dizer que a rede desta seção tem como foco determinar quais os pesquisadores e artigos mais importantes nas pesquisas da COVID-19 envolvendo a idade dos pacientes em geral.

4.3.1.1 Rede de Coautorias

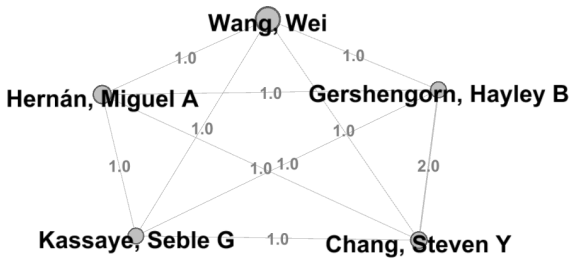
A Figura 35 ilustra a rede de coautorias filtrada a partir dos termos referentes *"aged"* e *"elderly"*. Nota-se a presença de boa parte das principais comunidades bem nichadas dentro da rede, ou seja, com a maioria de seus vértices próximos e interligados, como é o caso das comunidades lilás, verde claro, azul claro e laranja. Já a comunidade preta, que também é uma das principais, já é mais dispersa, permeando entre as demais.

Figura 35 – Comunidades de Autores - *Aged / Elderly*



A Figura 36 ilustra os vértices com maiores valores de centralidade de autovetor, que não pertencem às comunidades mais volumosas da rede. O vértice com maior centralidade de autovetor representa o pesquisador "Wang, Wei", que já apareceu na rede sobre comorbidades. Como boa parte dos autores desta visualização também foram coautores em um artigo relacionado a comorbidades (GUPTA et al., 2021), é possível constatar a proximidade dos resultados obtidos entre o filtro *"comorbidities"* e *"aged"*, o que faz sentido, visto que a idade é um fator que se enquadra como possível comorbidade para a doença.

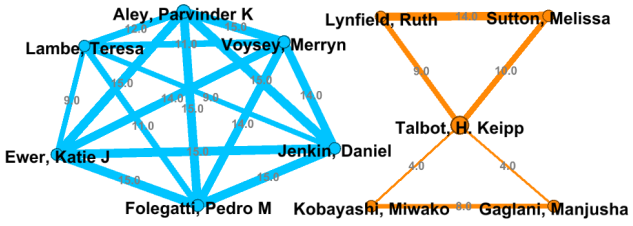
Figura 36 – Top 5 Autores - Centralidade de Autovetor - *Aged | Elderly*



A Figura 37 ilustra os vértices com maiores valores de grau ponderado., O resultado obtido foi diferente em comparação com a métrica de centralidade de autovetor. Desta vez, tem-se duas comunidades principais, nas quais se vê alguns nomes já citados neste trabalho. Na comunidade laranja, temos Manjusha Gagliani, que trabalha na área de pediatria, correlacionada também aos estudo sobre COVID-19 em pacientes crianças. Além disso, na comunidade azul claro, temos Teresa Lambe, que já vimos ser uma das principais desenvolvedoras da vacina Oxford Astrazeneca.

Tais resultados fazem pensar que esta se trata de uma rede com autores que abordam áreas distintas sobre o estudo de COVID-19, diferentemente do que o ocorreu para os casos da maioria das vacinas, em que as visualizações ilustraram seus desenvolvedores na maioria dos casos.

Figura 37 – Top 11 Autores - Grau Ponderado - *Aged | Elderly*



A Figura 38 ilustra os vértices com maiores valores de centralidade de intermediação, com a presença de alguns pesquisadores presentes também nas visualizações das outras seções. Um destes autores é Fabio Ciceri, que é coautor juntamente com alguns nomes já citados anteriormente, Paul Bastard e Jerémie Rosain, em um artigo com conteúdo que aborda o

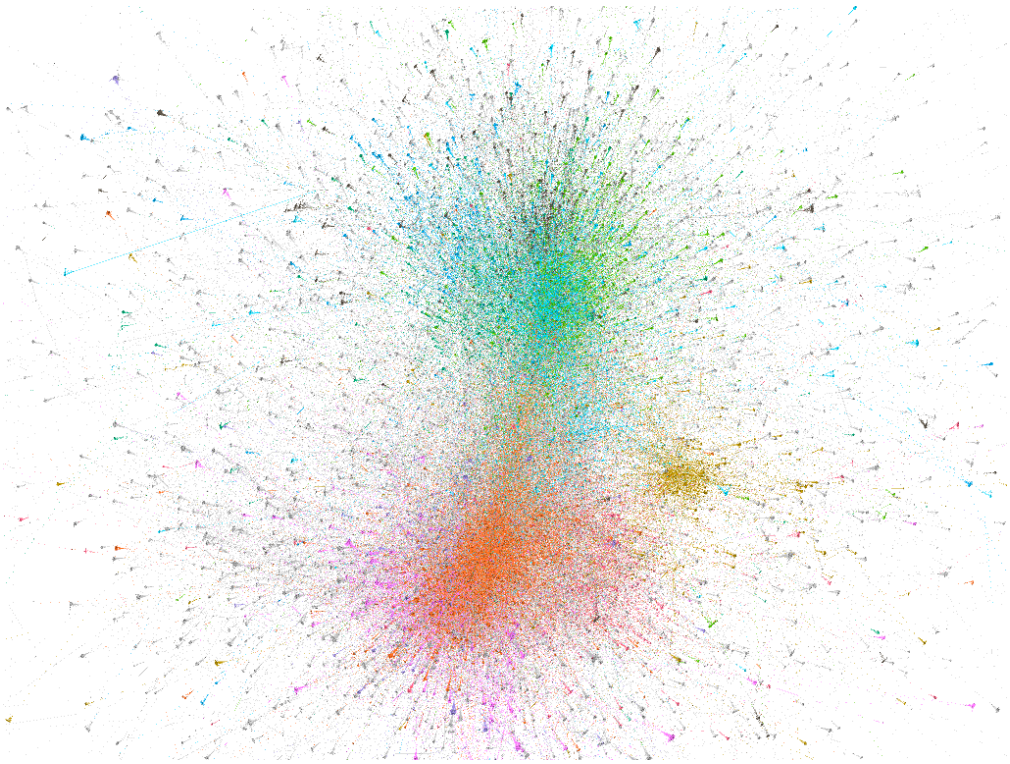
interesse desta seção, que é relacionar a taxa de mortalidade da doença às idades dos pacientes: *"The risk of COVID-19 death is much greater and age-dependent with type I IFN autoantibodies"* (MANRY et al., 2022).

Figura 38 – Top 12 Autores - Centralidade de Intermediação - *Aged | Elderly*

4.3.1.2 Rede de Citações

A Figura 39 ilustra a rede de citações gerada para os filtros *"aged"* e *"elderly"*. Visualmente, identifica-se dois pares de comunidades principais que estão próximas: a laranja com a lilás, e a azul clara com a verde clara.

Figura 39 – Rede de Citações - *Aged | Elderly*



A Figura 40 e a Tabela 12 ilustram os vértices com maiores valores de *PageRank* e o índice e título de cada um dos artigos representados por eles, respectivamente. Ao analisar os

títulos de alguns deles, nota-se a presença de diversas áreas de estudo sobre a COVID-19, desde análise de redes genéticas (KARAMI et al., 2021), que é também o vértice com maior valor de *PageRank* da rede, até saúde mental relacionada à pandemia da COVID-19 (ZHOU et al., 2020a), o que leva a crer, assim como no caso da rede de coautorias, que se trata de uma rede que aborda diferentes conceitos sobre a doença.

Figura 40 – Top 12 Artigos - *PageRank* - *Aged* / *Elderly*

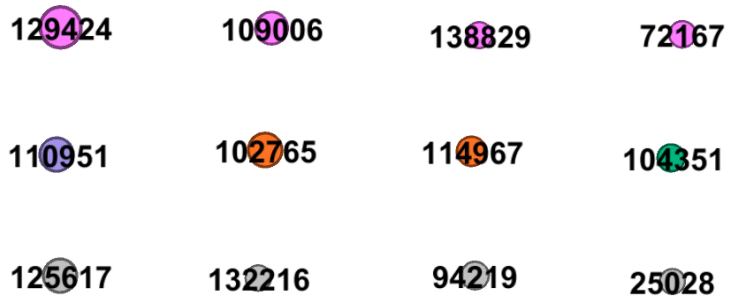


Tabela 12 – Top 12 Artigos - *PageRank* - *Aged* / *Elderly*

ID	Título
25028	Association of Solid Fuel Use With Risk of Cardiovascular and All-Cause Mortality in Rural China
72167	Covid-19 and co-morbidities: a role for Dipeptidyl Peptidase 4 (DPP4) in disease severity?
94219	Stapled Peptides-A Useful Improvement for Peptide-Based Drugs
102765	Diet-Induced Type II Diabetes in C57BL/6J Mice
104351	Unraveling the Invisible but Harmful Impact of COVID-19 on Deaf Older Adults and Older Adults with Hearing Loss
109006	Pathogenic T cells and inflammatory monocytes incite inflammatory storm in severe COVID-19 patients
110951	Mental health response to the COVID-19 Outbreak in China
114967	NA: % change in mean was not available due to extremely low levels of pre-treatment methylation
125617	The Nrf2-ARE pathway: An indicator and modulator of oxidative stress in neurodegeneration
129424	Weighted Gene Co-Expression Network Analysis Combined with Machine Learning Validation to Identify Key Modules and Hub Genes Associated with SARS-CoV-2 Infection
132216	Risk Factors for Severe Respiratory Syncytial Virus Lower Respiratory Tract Infection
138829	A clinically meaningful metric of immune age derived from high-dimensional longitudinal monitoring

A Figura 41 e a Tabela 13 ilustram, respectivamente, os vértices com maiores valores de centralidade de intermediação e o índice e título de cada um dos artigos por eles representados.

Destaca-se o vértice com maior valor para esta métrica, que representa o artigo intitulado *"COVID-19 infection in children"* (SINHA et al., 2020), que estuda o comportamento da COVID-19 em crianças e adolescentes, tratando-se de um artigo central para relacionar a doença às idades dos pacientes infantis.

Figura 41 – Top 12 Artigos - Centralidade de Intermediação - *Aged* / *Elderly*

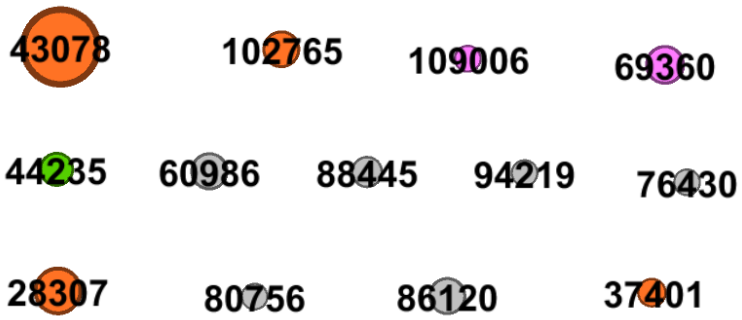


Tabela 13 – Top 12 Artigos - Centralidade de Intermediação - *Aged* / *Elderly*

ID	Título
28307	Emerging pandemic diseases: how we got to COVID-19
37401	SARS-CoV-2 and the possible connection to ERs, ACE2, and RAGE: Focus on susceptibility factors
43078	COVID-19 infection in children
44235	Meta-Analysis of Psychosocial Interventions
60986	What the COVID-19 pandemic is telling humanity
69360	Gut microbiota, inflammation, and molecular signatures of host response to infection
76430	COVID-19 -Recent advancements in identifying novel vaccine candidates and current status of upcoming SARS-CoV-2 vaccines
80756	Quantifying risks and interventions that have affected the burden of lower respiratory infections among children younger than 5 years: An analysis for the Global Burden of Disease Study
86120	Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations
88445	Clinical and pathological findings in SARS-CoV-2 Disease Outbreaks in Farmed Mink
94219	Stapled Peptides-A Useful Improvement for Peptide-Based Drugs
102765	Diet-Induced Type II Diabetes in C57BL/6J Mice
109006	Pathogenic T cells and inflammatory monocytes incite inflammatory storm in severe COVID-19 patients

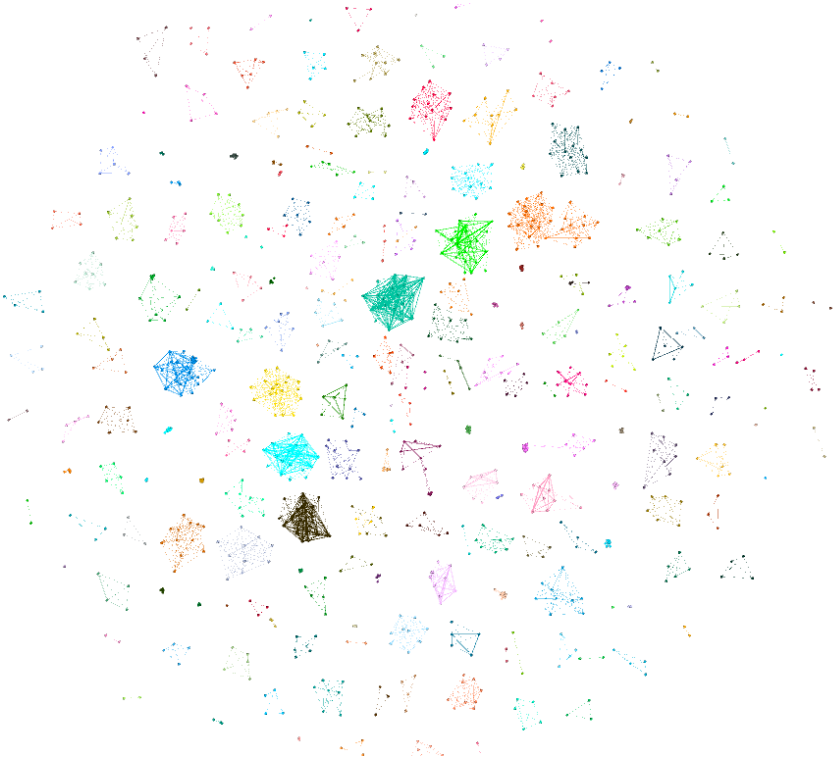
4.3.2 *Complex Network*

Finalmente, as últimas redes foram criadas com o filtro *"complex network"*, com o intuito de se obter modelagens que tragam informações sobre o uso de redes complexas dentro do estudo do coronavírus. Vale destacar que também foram geradas as redes apenas com o termo *"network"* como filtro, mas os resultados não atingiram o objetivo desejado: as redes geradas abordavam artigos de diversas áreas, pelo fato da palavra ser utilizada para diferentes assuntos.

4.3.2.1 Rede de Coautorias

A rede de coautorias para o filtro *"complex network"* é ilustrada pela Figura 42. Nota-se que as principais comunidades, que possuem maiores densidades de arestas e vértices, são bem definidas, havendo poucas conexões entre elas.

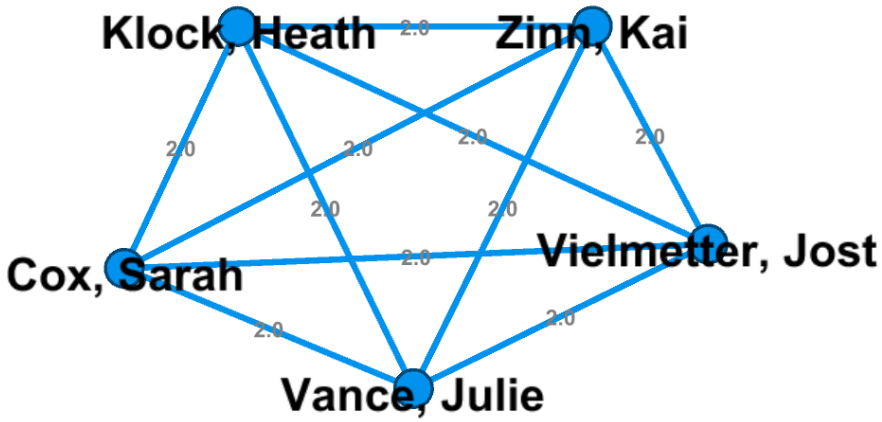
Figura 42 – Comunidades de Autores - *Complex Network*



A Figura 43 ilustra os cinco principais vértices da rede em termos de centralidade do autovetor, sendo todos pertencentes à comunidade azul escuro. O principal artigo que une estes autores é *"A human IgSF cell-surface interactome reveals a complex network of protein-protein interactions"* (WOJTOWICZ et al., 2020), trazendo uma aplicação de redes complexas na área de bioquímica. Trata-se de uma rede de autores norte-americanos. Por exemplo, Jost Vielmetter e Kai Zinn são pesquisadores na universidade Caltech, um dos principais centros

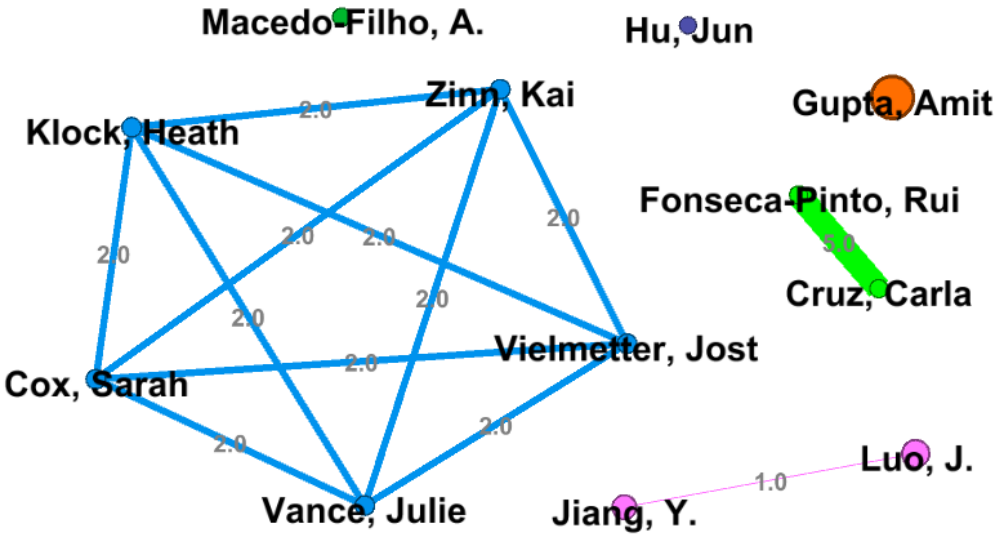
de pesquisa em ciência e engenharia do mundo, e Sarah Cox é pesquisadora na Novartis, empresa com polo em San Diego e referência na área médica.

Figura 43 – Top 5 Autores - Centralidade de Autovetor - *Complex Network*



A Figura 44 ilustra os doze principais vértices da rede em termos de centralidade de intermediação. Nota-se que a comunidade da Figura 36 está totalmente presente nesta visualização também. Além disso, a presença de alguns nomes de origem na língua portuguesa trouxe a ideia de se fazer uma análise mais profunda das comunidades destes autores.

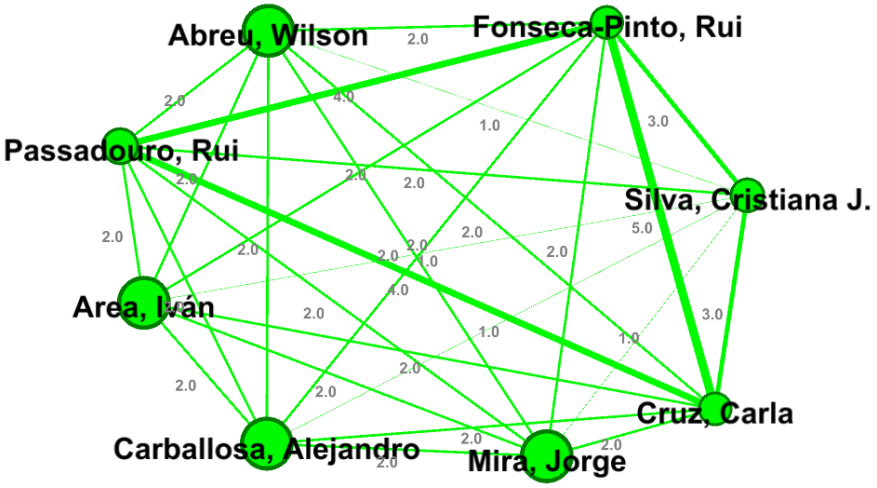
Figura 44 – Top 12 Autores - Centralidade de Intermediação - *Complex Network*



A Figura 45 ilustra os oito principais pesquisadores da comunidade verde claro em termos de centralidade do autovetor. Averiguou-se que se trata de um grupo de pesquisa de Portugal, responsável pelo artigo *"Optimal control of the COVID-19 pandemic: controlled sanitary deconfinement in Portugal"* (SILVA et al., 2021b), publicado na revista *Nature* e que propõe um modelo preditivo para antecipar as consequências de decisões políticas sobre o nível de isolamento da população portuguesa, e pelo artigo *"Complex network model for COVID-19:*

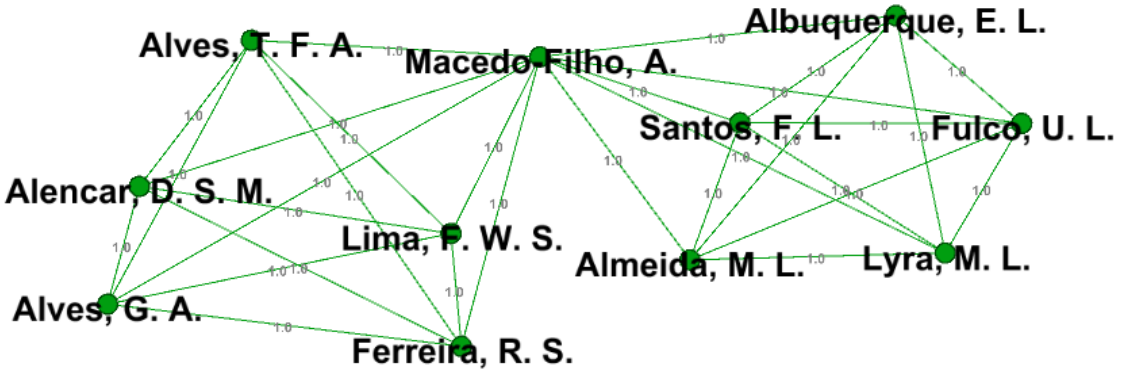
Human behavior, pseudo-periodic solutions and multiple epidemic waves" (SILVA et al., 2021a), que propõe um modelo de propagação da doença entre os indivíduos da sociedade.

Figura 45 – Comunidade Portugal - *Complex Network*



A Figura 46 ilustra uma comunidade brasileira de autores. No entanto, não se trata de uma das principais comunidades da rede de coautorias. Na verdade, esta rede engloba apenas dois artigos da base de dados, nos quais o pesquisador "Macedo-Filho, A." está presente como autor em ambos, fazendo a ligação entre as duas subcomunidades, representadas respectivamente por cada um dos artigos. Os títulos dos artigos são *"Modified Epidemic Diffusive Process on the Apollonian Network"* (ALENCAR et al., 2021) e *"Critical properties of the SIS model on the clustered homophilic network"* (SANTOS et al., 2020), e ambos abordam sobre modelos de propagação de doenças em epidemias. Neste caso, nota-se uma oportunidade de mais colaborações surgirem entre estes grupos, visto que estudam assuntos semelhantes.

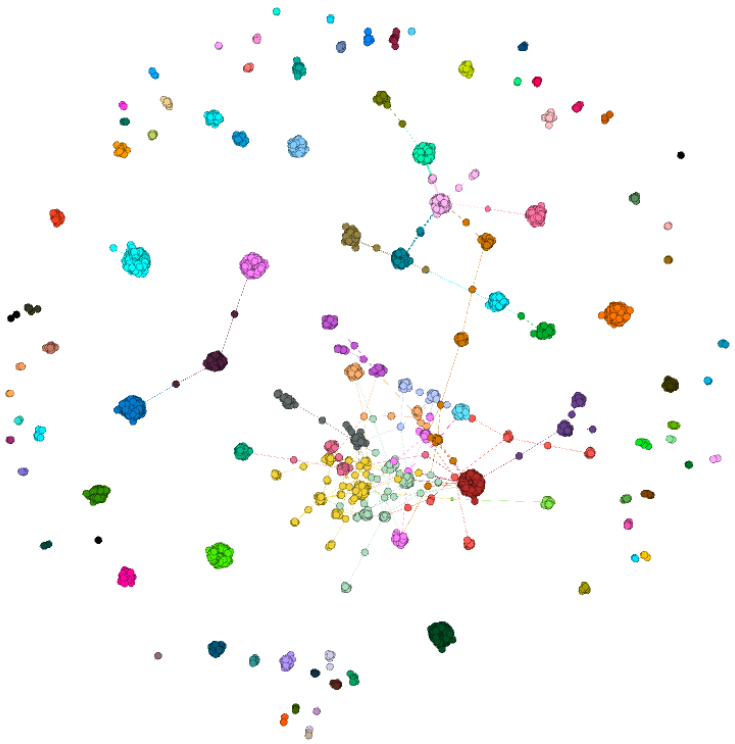
Figura 46 – Comunidade Brasil - *Complex Network*



4.3.2.2 Rede de Citações

A Figura 47 ilustra a rede de citações para os artigos filtrados a partir do filtro *"complex networks"*. Nota-se que é uma rede com a maioria de suas comunidades muito bem definidas e fortemente conectadas internamente. A comunidade amarela, por outro lado, parece ser a mais esparsa e que mais tem ligações com outros grupos de vértices.

Figura 47 – Rede de Citações - *Complex Network*



A Figura 48 e a Tabela 14 ilustram, respectivamente, os vértices com maiores valores de centralidade de intermediação e o índice e título de cada um dos artigos por eles representados.

Figura 48 – Top 15 Artigos - *PageRank* - *Complex Network*

Tabela 14 – Top 15 Artigos - *PageRank* - *Complex Network*

ID	Título
661	Exposure to air pollution and COVID-19 mortality in the United States: A nationwide cross-sectional study
1171	Learning to discover social circles in ego networks
1260	Network clique cover approximation to analyze complex contagions through group interactions
1340	Cytoscape: a software environment for integrated models of biomolecular interaction networks
1622	Structure and overlaps of communities in networks
1732	Identifying influential spreaders in complex multilayer networks: A centrality perspective
1809	Modeling and forecasting the COVID-19 temporal spread in Greece: an exploratory approach based on complex network defined splines
2939	Neural networks and the bias/variance dilemma
3282	Quantifying the association between domestic travel and the exportation of novel coronavirus (2019-ncov) cases from wuhan, china in 2020: a correlational analysis
3737	The concept and computation method of grey absolute correlation degree
4248	Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression
4462	COVID-19 control in China during mass population movements at New Year
5188	An information flow model for conflict and fission in small groups
5220	Clinical characteristics of 2019 novel coronavirus infection in China
5240	A new measure of identifying influential nodes: efficiency centrality

Analogamente, a Figura 49 e a Tabela 15 ilustram, respectivamente, os vértices com maiores valores de centralidade de intermediação e o índice e título de cada um dos artigos por eles representados.

Sob a perspectiva de que boa parte dos principais vértices das duas visualizações são os mesmos, será feita desta vez uma análise em conjunto dos principais artigos presentes na rede de citações, abordando tanto o conceito de *PageRank* quanto o de centralidade de intermediação.

Figura 49 – Top 15 Artigos - Centralidade de Intermediação - *Complex Network*

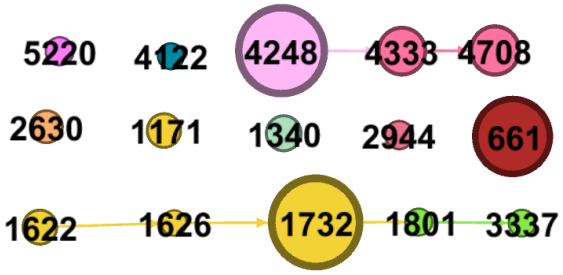


Tabela 15 – Top 15 Artigos - Centralidade de Intermediação - *Complex Network*

ID	Título
661	Exposure to air pollution and COVID-19 mortality in the United States: A nationwide cross-sectional study
1171	Learning to discover social circles in ego networks
1340	Cytoscape: a software environment for integrated models of biomolecular interaction networks
1622	Structure and overlaps of communities in networks
1626	Systematic comparison between methods for the detection of influential spreaders in complex networks
1732	Identifying influential spreaders in complex multilayer networks: A centrality perspective
1801	The spread of obesity in a large social network over 32 years
2630	Olopatadine Hydrochloride Ophthalmic Solution FDA Approval
2944	Dynamic Connectivity in a Financial Network Using Time-Varying DCCA Correlation Coefficients
3337	Trade liberalization and income inequality: The case for Pakistan
4122	Caspase-8 is the molecular switch for apoptosis, necroptosis and pyroptosis
4248	Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression
4333	Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication
4708	Coronaviruses drug discovery and therapeutic options
5220	Clinical characteristics of 2019 novel coronavirus infection in China

Nota-se primeiramente a grande relevância de dois artigos. O primeiro deles se intitula: *"Identifying influential spreaders in complex multilayer networks: A centrality perspective"* (BASARAS et al., 2017), tratando-se de um estudo para identificar propagadores em redes complexas sob a perspectiva de medidas de centralidade. O segundo é intitulado: *"Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression"* (QI et al., 2013), que estuda um método para controlar a expressão de genes, baseado também em redes complexas.

Um outro artigo interessante, principalmente para do ponto de vista do presente trabalho, é intitulado *"A new measure of identifying influential nodes: Efficiency centrality"* (WANG; DU; DENG, 2017). Ele aborda a proposta de uma nova medida de centralidade, denominada centralidade de eficiência. O método basicamente parte da ideia de ordenar os vértices propagadores de toda uma rede, remover cada um deles e considerar o nível de eficiência da rede após a remoção.

5 Conclusão

Primeiramente, conclui-se que, no contexto geral, foi possível extrair informações relevantes sobre os principais autores de algumas áreas de estudo através de análises baseadas nas medidas de centralidade. Para o caso das redes sobre as vacinas contra a COVID-19, foi possível encontrar as comunidades dos principais pesquisadores envolvidos no desenvolvimento delas. Viu-se também que diferentes medidas de centralidade trouxeram diferentes informações sobre as redes. Enquanto medidas como centralidade de autovetor e grau ponderado evidenciaram vértices correspondentes a pesquisadores muito presentes nos artigos que geraram as redes, a centralidade de intermediação trouxe informações sobre autores que mais interligam diferentes comunidades, sendo bons conectores entre grupos de pesquisa.

Além disso, para outros temas mais gerais, foi possível detectar autores com alto grau de relevância associados aos assuntos relacionados aos tópicos filtrados, como por exemplo Jérémie Rosain e Paul Bastard, pesquisadores no *Institut Imagine*, organização que se mostrou referência para os estudos que relacionam as taxas de mortalidade do coronavírus a fatores genéticos. Outro comportamento interessante é que em algumas redes, todos os autores com maior relevância pertenciam a uma mesma comunidade, mas em outros casos havia mais de uma comunidade envolvida, o que é interessante do ponto de vista se entender novas colaborações que possam surgir entre essas principais comunidades.

Em relação às análises das redes de citações, notou-se que em boa parte dos casos, os principais artigos do ponto de vista da métrica *PageRank* coincidiam com boa parte dos principais artigos com maiores centralidades de intermediação. Este comportamento faz sentido, visto que artigos que são resumos do estado da arte de alguma área, como por exemplo as revisões sobre as vacinas, costumam interligar áreas de estudo (centralidade de intermediação), sendo assim também amplamente citados (*PageRank*). Outro aspecto interessante encontrado nas redes de citações foi a presença de estudos anteriores às datas filtradas, como por exemplo o caso do artigo *"Adenovirus-mediated overexpression of novel mutated IkappaBalpha inhibits nuclear factor kappaB activation in endothelial cells"* (WRIGHTON et al., 1996). Este comportamento permite entender também como estudos anteriores podem colaborar para o desenvolvimento de novas áreas na ciência.

Outro fator que teve impacto relevante nas análises foi o tamanho das redes geradas. Para assuntos específicos, em que as redes tiveram escalas menores, como por exemplo no caso das vacinas ou do filtragem pelo tema de redes complexas, as comunidades de autores geradas foram melhor segmentadas e os principais artigos obtidos se mostraram mais assertivos em relação aos temas filtrados. Para as redes com maiores proporções e temas mais genéricos, como "pneumonia" ou "comorbidades", a análise precisou ser melhor direcionada, visto que

as redes englobavam uma diversidade maior de assuntos.

Em relação às redes de citações filtradas para o assunto de redes complexas, obteve-se resultados interessantes do ponto de vista dos tipos de artigos que utilizam redes complexas acerca do tema da COVID-19, com a maioria dos artigos abordando temas mais técnicos, relacionados a modelagens do comportamento de estruturas bioquímicas ou modelos de propagação de epidemias, por exemplo. Também se encontrou um artigo acerca de uma nova medida de centralidade, que pode ser interessante para a realização de estudos semelhantes ao presente trabalho. Além disso, nas redes de coautorias, foi possível encontrar comunidades de pesquisadores portugueses e brasileiros, que podem ser interessantes vínculos para futuras pesquisas que relacionem redes complexas ao coronavírus.

5.1 Trabalhos futuros

Para trabalhos futuros, sugere-se algumas possibilidades que se poderia abordar a fim de gerar novos estudos relacionados:

- Aplicar novas medidas de centralidade, como eficiência (WANG; DU; DENG, 2017), centralidade de proximidade ou autoridades e *hubs*, para detectar autores e artigos relevantes dentro de redes de coautorias e citações;
- Utilizar os anos de publicação dos artigos para elaborar redes de citações que mostrem a evolução temporal do estado da arte algumas áreas;
- Anexar a rede de filiação dos autores para fazer análises sobre as instituições e regiões geográficas;
- Realizar a predição de novas conexões dentro das redes de coautorias a partir de informações das estruturas de comunidades;
- Mapear as diferentes regiões das redes geradas através de extração de tópicos, atribuindo assuntos específicos a determinados autores e artigos.

Referências

ALBERT, R.; JEONG, H.; BARABASI, A.-L. Diameter of the world-wide web. *Nature*, v. 401, p. 130–131, 09 1999.

ALENCAR, D. et al. Modified epidemic diffusive process on the apollonian network. *arXiv preprint arXiv:2110.14141*, 2021.

BASARAS, P. et al. Identifying influential spreaders in complex multilayer networks: A centrality perspective. *IEEE Transactions on Network Science and Engineering*, IEEE, v. 6, n. 1, p. 31–45, 2017.

BASTARD, P. et al. Preexisting autoantibodies to type i ifns underlie critical covid-19 pneumonia in patients with aps-1. *Journal of Experimental Medicine*, The Rockefeller University Press, v. 218, n. 7, 2021.

BASTIAN, M.; HEYMANN, S.; JACOMY, M. Gephi: an open source software for exploring and manipulating networks. In: *Proceedings of the international AAAI conference on web and social media*. [S.l.: s.n.], 2009. v. 3, n. 1, p. 361–362.

BERNAL, J. L. et al. Effectiveness of covid-19 vaccines against the b. 1.617. 2 (delta) variant. *New England Journal of Medicine*, Mass Medical Soc, 2021.

BERTSIMAS, D.; TSITSIKLIS, J. Simulated annealing. *Statistical science*, Institute of Mathematical Statistics, v. 8, n. 1, p. 10–15, 1993.

BLEI, D. M. Probabilistic topic models. *Communications of the ACM*, ACM New York, NY, USA, v. 55, n. 4, p. 77–84, 2012.

BLEI, D. M.; NG, A. Y.; JORDAN, M. I. Latent dirichlet allocation. *Journal of machine Learning research*, v. 3, n. Jan, p. 993–1022, 2003.

BLONDEL, V. D. et al. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, IOP Publishing, v. 2008, n. 10, p. P10008, 2008.

BONACICH, P. Power and centrality: A family of measures. *American journal of sociology*, University of Chicago Press, v. 92, n. 5, p. 1170–1182, 1987.

BRANDES, U. A faster algorithm for betweenness centrality. *Journal of mathematical sociology*, Taylor & Francis, v. 25, n. 2, p. 163–177, 2001.

BRIN, S.; PAGE, L. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, Elsevier, v. 30, n. 1-7, p. 107–117, 1998.

BUN, K. K.; ISHIZUKA, M. Topic extraction from news archive using tf* pdf algorithm. In: IEEE. *Proceedings of the Third International Conference on Web Information Systems Engineering, 2002. WISE 2002.* [S.l.], 2002. p. 73–82.

CAI, Y. et al. Kynurenic acid may underlie sex-specific immune responses to covid-19. *Science Signaling*, American Association for the Advancement of Science, v. 14, n. 690, p. eabf8483, 2021.

DONG, R. et al. Topic extraction from online reviews for classification and recommendation. In: AAAI. *Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI 13), Beijing, China, 3-9 August 2013*. [S.l.], 2013. p. 1310–1316.

EDWARDS, L. J. et al. Research on the epidemiology of sars-cov-2 in essential response personnel (recover): Protocol for a multisite longitudinal cohort study. *JMIR Research Protocols*, JMIR Publications Inc., Toronto, Canada, v. 10, n. 12, p. e31574, 2021.

FREEMAN, L. C. A set of measures of centrality based on betweenness. *Sociometry*, JSTOR, p. 35–41, 1977.

FRUCHTERMAN, T. M. J.; REINGOLD, E. M. Graph drawing by force-directed placement. *Softw. Pract. Exper.*, John Wiley & Sons, Inc., v. 21, n. 11, p. 1129–1164, 1991.

GELERIS, J. et al. Observational study of hydroxychloroquine in hospitalized patients with covid-19. *New England Journal of Medicine*, Mass Medical Soc, v. 382, n. 25, p. 2411–2418, 2020.

GEPHI. Gephi tutorial layouts. 2008. Disponível em: <https://gephi.org/tutorials/gephi-tutorial-layouts.pdf>.

GUPTA, S. et al. Association between early treatment with tocilizumab and mortality among critically ill patients with covid-19. *JAMA internal medicine*, American Medical Association, v. 181, n. 1, p. 41–51, 2021.

KARAMI, H. et al. Weighted gene co-expression network analysis combined with machine learning validation to identify key modules and hub genes associated with sars-cov-2 infection. *Journal of clinical medicine*, Multidisciplinary Digital Publishing Institute, v. 10, n. 16, p. 3567, 2021.

KHAN, T. et al. Therapeutic potential of medicinal plants against covid-19: The role of antiviral medicinal metabolites. *Biocatalysis and Agricultural Biotechnology*, Elsevier, v. 31, p. 101890, 2021.

LAMBIOTTE, R.; DELVENNE, J.-C.; BARAHONA, M. Laplacian dynamics and multiscale modular structure in networks. *arXiv preprint arXiv:0812.1770*, 2008.

LI, S. Topic modeling and latent dirichlet allocation (lda) in python. *Towards Data Science*, v. 31, 2018.

LINHARES, T.; LIMA, M. de. Pln – processamento de linguagem natural para iniciantes. 2021. Disponível em: <https://insightlab.ufc.br/pln-processamento-de-linguagem-natural-para-iniciantes/>.

LOMBARDI, A. et al. Mini review immunological consequences of immunization with covid-19 mrna vaccines: preliminary results. *Frontiers in immunology*, Frontiers, v. 12, p. 677, 2021.

MADAN, R. Tf-idf/term frequency technique: Easiest explanation for text classification in nlp using python (chatbot training on words). 2019. Disponível em: <https://rb.gy/k7nvnp>.

MANRY, J. et al. The risk of covid-19 death is much greater and age-dependent with type i ifn autoantibodies. 2022.

NEWMAN, M. E. J. Scientific collaboration networks: I. network construction and fundamental results. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, v. 64, n. 1, p. 1–8, 2001.

NEWMAN, M. E. J. *Networks: An Introduction*. New York, NY, USA: Oxford University Press, Inc., 2010. ISBN 0199206651, 9780199206650.

OKAYASU, H. et al. Mucosal immunity and poliovirus vaccines: impact on wild poliovirus infection and transmission. *Vaccine*, Elsevier, v. 29, n. 46, p. 8205–8214, 2011.

PAVORD, S. et al. Clinical features of vaccine-induced immune thrombocytopenia and thrombosis. *New England Journal of Medicine*, Mass Medical Soc, v. 385, n. 18, p. 1680–1689, 2021.

PRICE, J. R. et al. Development and delivery of a real-time hospital-onset covid-19 surveillance system using network analysis. *Clinical Infectious Diseases*, Oxford University Press US, v. 72, n. 1, p. 82–89, 2021.

QI, L. S. et al. Repurposing crispr as an rna-guided platform for sequence-specific control of gene expression. *Cell*, Elsevier, v. 152, n. 5, p. 1173–1183, 2013.

ROEL, E. et al. Characteristics and outcomes of over 300,000 patients with covid-19 and history of cancer in the united states and spain. *Cancer Epidemiology and Prevention Biomarkers*, AACR, v. 30, n. 10, p. 1884–1894, 2021.

SANTOS, F. et al. Critical properties of the sis model on the clustered homophilic network. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 559, p. 125067, 2020.

SCOTT, J. *Social Network Analysis: A Handbook*. [S.l.]: SAGE Publications, 2000. ISBN 9780761963394.

SETHNEHA. Part 2: Topic modeling and latent dirichlet allocation (lda) using gensim and sklearn. 2021. Disponível em: <https://www.analyticsvidhya.com/blog/2021/06/part-2-topic-modeling-and-latent-dirichlet-allocation-lda-using-gensim-and-sklearn/>.

SILVA, C. J. et al. Complex network model for covid-19: human behavior, pseudo-periodic solutions and multiple epidemic waves. *Journal of mathematical analysis and applications*, Elsevier, p. 125171, 2021.

SILVA, C. J. et al. Optimal control of the covid-19 pandemic: controlled sanitary deconfinement in portugal. *Scientific reports*, Nature Publishing Group, v. 11, n. 1, p. 1–15, 2021.

SILVA, F. N. et al. Using network science and text analytics to produce surveys in a scientific topic. *Journal of Informetrics*, Elsevier, v. 10, n. 2, p. 487–502, 2016.

SILVA, F. N. et al. Quantifying the interdisciplinarity of scientific journals and fields. *Journal of Informetrics*, Elsevier, v. 7, n. 2, p. 469–477, 2013.

SILVA, T. F. et al. A 21st century evil: immunopathology and new therapies of covid-19. *Frontiers in Immunology*, Frontiers, p. 2763, 2020.

SINHA, I. P. et al. Covid-19 infection in children. *The Lancet Respiratory Medicine*, Elsevier, v. 8, n. 5, p. 446–447, 2020.

VALEJO, A. et al. A benchmarking tool for the generation of bipartite network models with overlapping communities. *Knowledge and Information Systems*, Springer, v. 62, n. 4, p. 1641–1669, 2020.

VALEJO, A. D. B. *Refinamento multinível em redes complexas baseado em similaridade de vizinhança*. Tese (Doutorado) — Universidade de São Paulo, 2014.

VASUJI. Ii-covid19-citation network. *Kaggle*, 2020. Disponível em: <https://www.kaggle.com/vasuji/ii-covid19-citation-network>.

VENTURA, P. C. et al. Epidemic spreading in populations of mobile agents with adaptive behavioral response. *Chaos, Solitons & Fractals*, Elsevier, v. 156, p. 111849, 2022.

WANG, L. L. et al. Cord-19: The covid-19 open research dataset. *ArXiv*, ArXiv, 2020.

WANG, S.; DU, Y.; DENG, Y. A new measure of identifying influential nodes: Efficiency centrality. *Communications in Nonlinear Science and Numerical Simulation*, Elsevier, v. 47, p. 151–163, 2017.

WATTS, D. J.; STROGATZ, S. H. Collective dynamics of 'small-world' networks. *Nature*, v. 393, n. 6684, p. 440–2, 1998.

WOJTOWICZ, W. M. et al. A human igsf cell-surface interactome reveals a complex network of protein-protein interactions. *Cell*, Elsevier, v. 182, n. 4, p. 1027–1043, 2020.

WRIGHTON, C. et al. Inhibition of endothelial cell activation by adenovirus-mediated expression of i kappa b alpha, an inhibitor of the transcription factor nf-kappa b. *The Journal of experimental medicine*, v. 183, n. 3, p. 1013–1022, 1996.

ZHOU, J. et al. Mental health response to the covid-19 outbreak in china. *American Journal of Psychiatry*, Am Psychiatric Assoc, v. 177, n. 7, p. 574–575, 2020.

ZHOU, Y. et al. Comorbidities and the risk of severe or fatal outcomes associated with coronavirus disease 2019: A systematic review and meta-analysis. *International Journal of Infectious Diseases*, Elsevier, v. 99, p. 47–56, 2020.

ZOU, X. et al. Single-cell rna-seq data analysis on the receptor ace2 expression reveals the potential risk of different human organs vulnerable to 2019-ncov infection. *Frontiers of medicine*, Springer, v. 14, n. 2, p. 185–192, 2020.

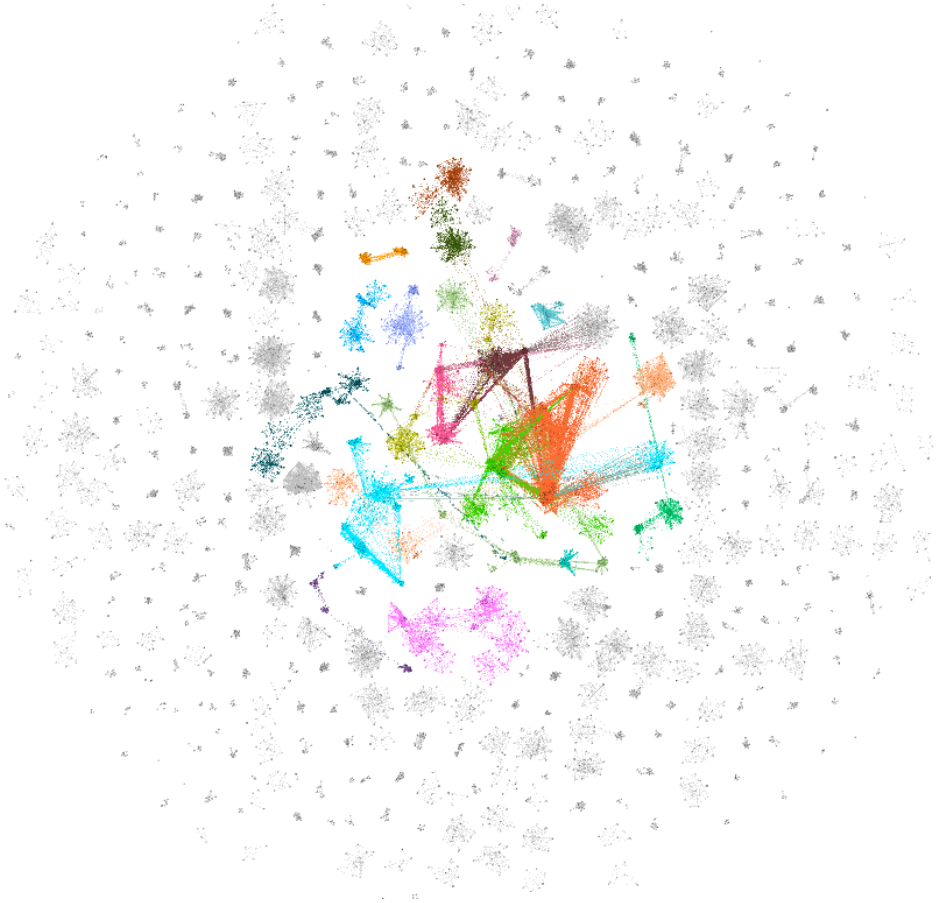
Apêndices

APÊNDICE A – Moderna

A.1 Rede de Coautorias

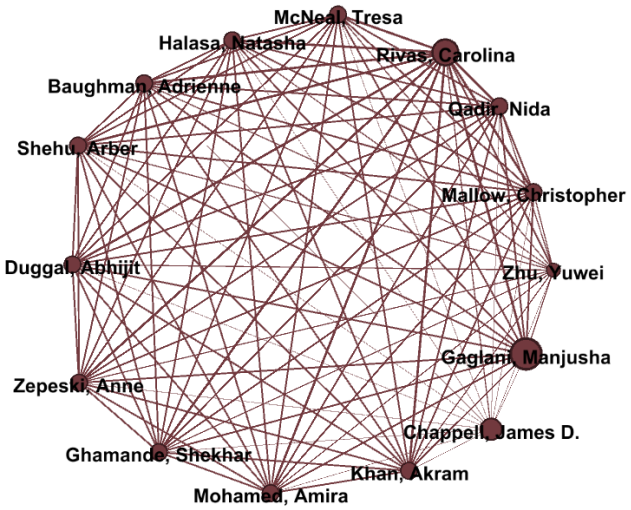
A rede de coautorias para o filtro “Moderna” é ilustrada pela Figura 50. Nota-se a presença de três grupos principais na rede: o laranja, o verde claro e o azul claro. Os dois primeiros se mostram concentrados e próximos na rede, enquanto o terceiro parece ter diferentes pontos de concentração e conexão com diferentes áreas da rede complexa.

Figura 50 – Comunidades de Autores - Moderna



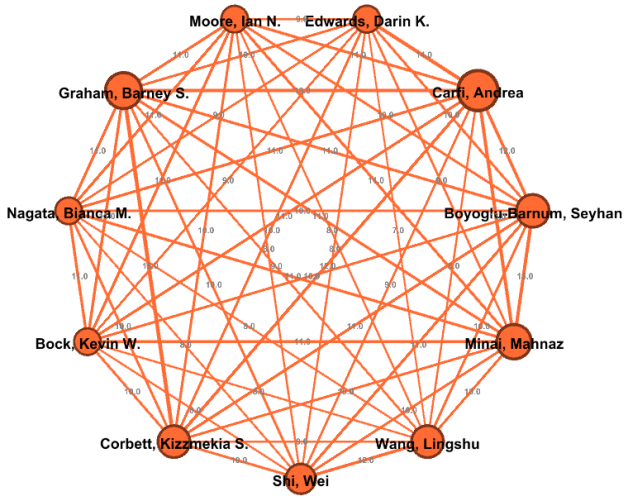
A Figura 51 ilustra os vértices com os maiores valores de centralidade de autovetor da rede, presentes na comunidade marrom, que é a quinta maior da rede. Ao se pesquisar sobre a principal autora, “Gaglani, Manjusha”, vê-se que ela faz parte da *Texas A&M Health Science Center*, estando envolvida principalmente na área da pediatria, correlacionada ao combate à COVID-19. Portanto, diferentemente dos casos da CoronaVac e Astrazeneca, a visualização não indicou desta vez os responsáveis de fato pelo desenvolvimento da vacina.

Figura 51 – Top 10 Autores - Centralidade de Autovetor - Moderna



Uma nova visualização que se idealizou, durante os experimentos com o a rede filtrada com os termos da vacina Moderna, foi dos autores com maiores valores de grau ponderado na rede, para entender se os vértices do caso da centralidade de autovetor se repetiriam novamente. A resposta obtida foi negativa, pois na Figura 52, é possível notar a presença de vértices da comunidade laranja, cujo autor com maior grau ponderado representa o pesquisador Andrea Carfi, que é o CSO (*Chief Scientific Officer*) da empresa Moderna, o que leva a entender que a rede representada desta vez representa alguns dos responsáveis pelo desenvolvimento da vacina.

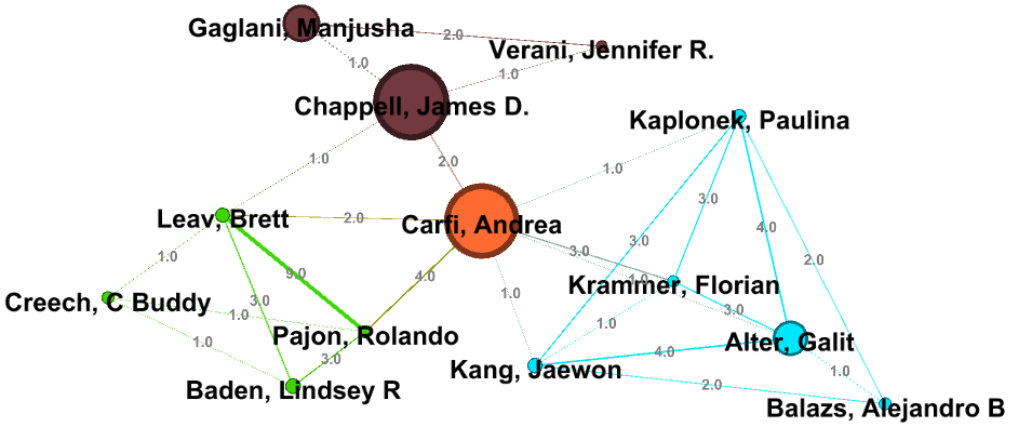
Figura 52 – Top 11 Autores - Grau Ponderado - Moderna



Visualizando agora a Figura 53, que contém os vértices da rede com maiores centralidades de intermediação, vê-se novamente a presença relevante de Andrea Carfi, que conecta três das principais comunidades da rede. Além disso, vê-se também a alta relevância de “Chappell,

James D.", que tem como principal área de estudo a pediatria, se conectando tanto com Carfi quanto com Manjusha Gagliani.

Figura 53 – Top 13 Autores - Centralidade de Intermediação - Moderna



A.2 Rede de Citações

A Figura 54 ilustra a rede de citações para os artigos filtrados com os nomes da vacina Moderna. Após executar o algoritmo Force Atlas 2, nota-se a presença de várias comunidades com considerável parcela de vértices, o que indica uma rede com diversos grupos de artigos de importância, mas é possível destacar principalmente as comunidades verde-água, rosa e laranja.

Figura 54 – Rede de Citações - Moderna

A Figura 55 e a Tabela 16 ilustram, respectivamente, a representação dos vértices com maiores valores de *PageRank* e o índice e título de cada um dos artigos representados pelos vértices. Dentre os títulos, é possível notar a presença de alguns artigos que também estiveram entre os de maior importância para as redes da CoronaVac e da Astrazeneca, como por

exemplo o artigo: *"Therapeutic potential of medicinal plants against COVID-19: The role of antiviral medicinal metabolites"* (KHAN et al., 2021), que trata do estudo de tratamentos com plantas medicinais contra a COVID-19.

Outro artigo que vale a pena se destacar é o *"Mini Review Immunological Consequences of Immunization With COVID-19 mRNA Vaccines: Preliminary Results"* (LOMBARDI et al., 2021), que se trata de uma revisão dos resultados das vacinas elaboradas com base na tecnologia de mRNA.

Figura 55 – Top 11 Artigos - *PageRank* - Moderna

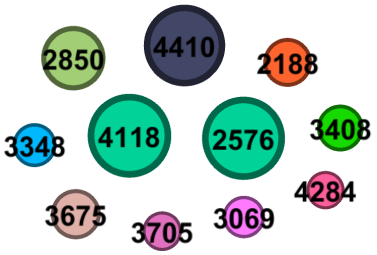


Tabela 16 – Top 11 Artigos - *PageRank* - Moderna

ID	Título
2188	Monitoring of COVID-19 medicines
2576	Research on the Epidemiology of SARS-CoV-2 in Essential Response Personnel
2850	Single-cell RNAseq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection
3069	Safety and Tolerability of the BNT162b2 mRNA COVID-19 Vaccine in Dialyzed Patients
3348	Effectiveness of a third dose of the BNT162b2 mRNA COVID-19 vaccine for preventing severe outcomes in Israel: An observational study
3408	Effectiveness of COVID-19 Vaccines against the B.1.617.2 (Delta) Variant
3675	Antibody Repertoires for Exploring Antibody Diversity and Predicting Antibody 917 Prevalence
3705	Interactive Tree Of Life (iTOL) v4: recent updates and new developments
4118	Mini Review Immunological Consequences of Immunization With COVID-19 mRNA Vaccines: Preliminary Results
4284	A novel coronavirus from patients with pneumonia in China
4410	Therapeutic potential of medicinal plants against COVID-19: The role of antiviral medicinal metabolites

A Figura 56 e a Tabela 17 ilustram, respectivamente, a representação dos vértices com maiores valores de centralidade de intermediação e o índice e título de cada um dos artigos representados pelos vértices. Vale destacar dois artigos da rede como um todo, que estão tanto entre os maiores *PageRank* quanto centralidade de intermediação da rede. O primeiro se

entitula *"Research on the Epidemiology of SARS-CoV-2 in Essential Response"* (EDWARDS et al., 2021), que aborda estimar a incidência de infecções sintomáticas e assintomáticas por SARS-CoV-2 em trabalhadores essenciais. O segundo tem como título *"Single-cell RNAseq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019- nCoV infection"* (ZOU et al., 2020), que foi também um dos principais artigos da rede de citações do filtro "CoronaVac".

Figura 56 – Top 12 Artigos - Centralidade de Intermediação - Moderna

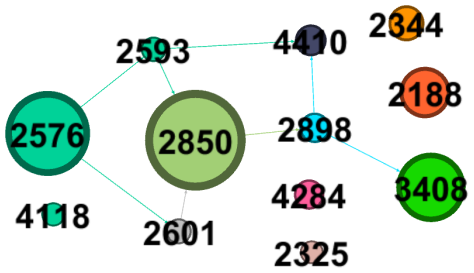


Tabela 17 – Top 12 Artigos - Centralidade de Intermediação - Moderna

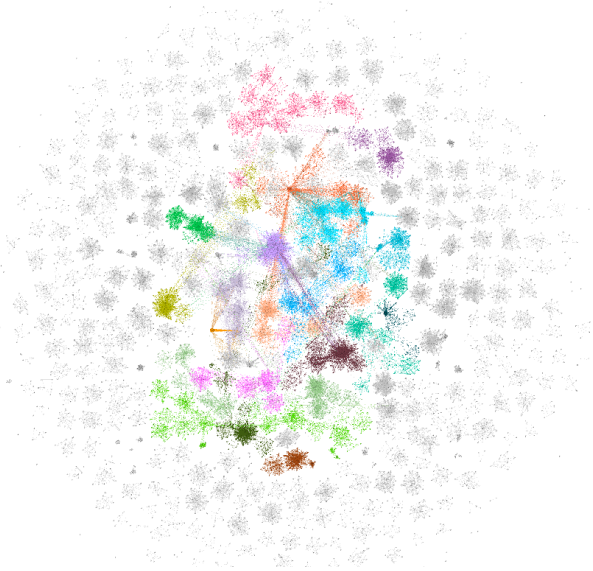
ID	Título
2188	Monitoring of COVID-19 medicines
2325	Stereotypic Neutralizing VH Antibodies Against SARS-CoV-2 Spike Protein Receptor Binding Domain in Patients With COVID-19 and Healthy Individuals
2344	SARS-CoV-2 genomic variations associated with mortality rate of COVID-19
2576	Research on the Epidemiology of SARS-CoV-2 in Essential Response Personnel
2593	Protective efficacy of in vitro synthesized, specific mRNA vaccines against influenza A virus infection
2601	Rna-Based COVID-19 vaccine BNT162b2 selected for a pivotal efficacy study
2850	Single-cell RNAseq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection
2898	Safety and efficacy of an rAd26 and rAd5 vector-based heterologous prime-boost COVID-19 vaccine: An interim analysis of a randomised controlled phase 3 trial in Russia
3408	Effectiveness of COVID-19 Vaccines against the B.1.617.2 (Delta) Variant
4118	Mini Review Immunological Consequences of Immunization With COVID-19 mRNA Vaccines: Preliminary Results
4284	A novel coronavirus from patients with pneumonia in China
4410	Therapeutic potential of medicinal plants against COVID-19: The role of antiviral medicinal metabolites

APÊNDICE B – Janssen

B.1 Rede de Coautorias

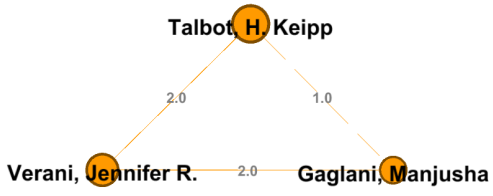
A Figura 57 ilustra a rede de coautorias gerada para a vacina Janssen. Nota-se a presença de diversos grupos de tamanho relevante, sendo difícil identificar visualmente quais são os principais, ou seja, com maior número de vértices. No entanto, na ferramenta Gephi, foi possível identificar que o lilás, o verde claro e o laranja são, respectivamente, os três com maiores quantidades de vértices.

Figura 57 – Comunidades de Autores - Janssen



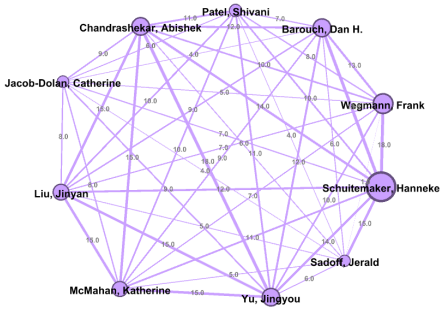
A Figura 58 ilustra a rede com os vértices com maiores centralidades de autovetor. Nota-se a presença novamente de Manjusha Gaglani, presente na seção de coautorias sobre a vacina Moderna. Além disso, ao se pesquisar sobre as duas outras pesquisadoras, nota-se que ambas têm trabalho nas áreas de infectologia relacionada à pediatria, o que indica novamente que se trata de uma área de estudo importante no contexto da COVID-19.

Figura 58 – Top 3 Autores - Centralidade de Autovetor - Janssen



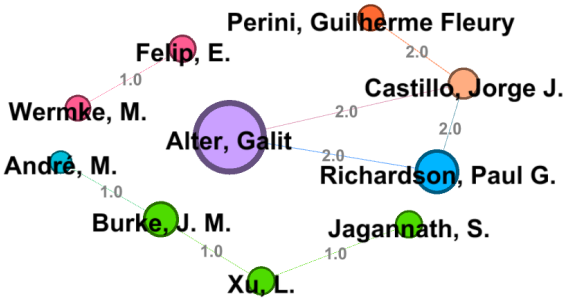
Com o intuito de buscar de fato os responsáveis pelo desenvolvimento da vacina Janssen, criou-se novamente um filtro para se obter os vértices com maiores valores de grau ponderado, ilustrada pela Figura 59. Destaca-se a pesquisadora holandesa Hanneke Schuitemaker, que tem como cargo *VP, Global Head of Viral Vaccine Discovery and Translational Medicine* na *Janssen Pharmaceuticals*, o que indica que as pessoas desta rede de colaboração científica são algumas das principais responsáveis pelo desenvolvimento da vacina.

Figura 59 – Top 10 Autores - Grau Ponderado - Janssen



A Figura 60 ilustra os principais vértices da rede no quesito de centralidade de intermediação, sendo notável a presença de pesquisadores de diversas comunidades. Vale ressaltar a presença do vértice rotulado *"Alter, Galit"*, que além de possuir a maior centralidade de intermediação desta rede, também está presente na Figura 53, sendo também um dos principais vértices para a rede da Moderna para esta métrica. Galit Alter é pesquisadora na *Harvard Medical School* e líder de grupo no *Ragon Institute of MGH, MIT e Harvard*, sendo também uma autora relevante para o contexto da vacina Janssen.

Figura 60 – Top 10 Autores - Centralidade de Intermediação - Janssen



B.2 Rede de Citações

A rede de citações para o filtro "Janssen" é ilustrada pela Figura 57. As três maiores comunidades, em termos de quantidade de vértices, são a verde, a rosa e a azul claro, respectivamente.

Figura 61 – Rede de Citações - Janssen

A Figura 62 e a Tabela 18 ilustram, representação dos vértices com maiores valores de *PageRank* e o índice e título de cada um dos artigos representados pelos vértices, respectivamente. Dentre os títulos, é possível notar a presença de alguns artigos que já apareceram dentre os principais das redes das vacinas anteriores. Destaca-se o artigo *"Effectiveness of Covid-19 vaccines against the B. 1.617. 2 (Delta) variant"* (BERNAL et al., 2021), que estuda a eficiência das vacinas contra a variante Delta do vírus.

Analogamente, a Figura 63 e a Tabela 19 ilustram a representação dos vértices com maiores valores de centralidade de intermediação e o índice e título de cada um dos artigos representados pelos vértices, respectivamente. Vale destacar que há a interseção de 6 artigos, comparando-se a Tabela 18 com a Tabela 19, o que indica que a maioria dos principais artigos da rede, também são artigos centrais, que interligam as áreas do conhecimento.

Figura 62 – Top 10 Artigos - *PageRank* - Janssen

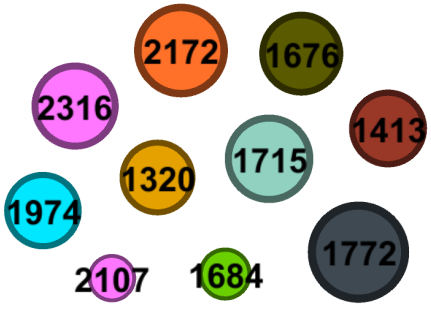


Tabela 18 – Top 10 Artigos - *PageRank* - Janssen

ID	Título
1320	Monitoring of COVID-19 medicines
1413	Compromised humoral functional evolution tracks 47 with SARS-CoV-2 mortality
1676	Factors That Influence the Immune Response to Vaccination
1684	Thrombocytopenia following Pfizer and Moderna SARS-CoV-2 vaccination
1715	Effectiveness of a third dose of the BNT162b2 mRNA COVID-19 vaccine for preventing severe outcomes in Israel: An observational study
1772	Effectiveness of COVID-19 Vaccines against the B.1.617.2 (Delta) Variant
1974	Addressing the vaccine confidence gap
2107	American Heart Association/American Stroke Association Stroke Council Leadership. Diagnosis and Management of Cerebral Venous Sinus Thrombosis with Vaccine-Induced Immune Thrombotic Thrombocytopenia
2172	Therapeutic potential of medicinal plants against COVID-19: The role of antiviral medicinal metabolites
2316	SARS-CoV-2 Vaccine ChAdOx1 nCoV-19 Infection Of Human Cell Lines Reveals Low Levels of Viral Backbone Gene Transcription Alongside Very High Levels of SARS-CoV-2 S Glycoprotein Gene Transcription

Figura 63 – Top 10 Artigos - Centralidade de Intermediação - Janssen

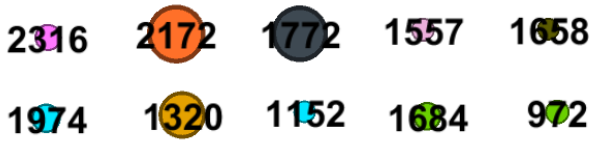


Tabela 19 – Top 10 Artigos - Centralidade de Intermediação - Janssen

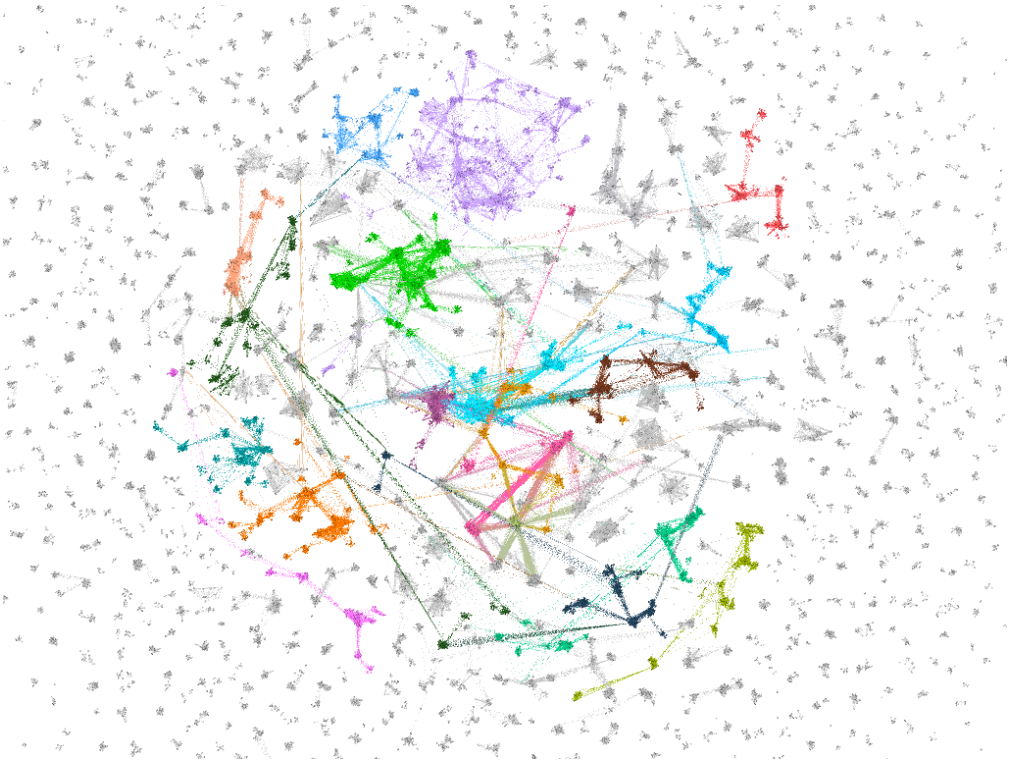
ID	Título
972	Clinical-Forensic Autopsy Findings to Defeat COVID-19 Disease: A Literature Review
1152	Preferences for COVID-19 vaccine distribution strategies in the US: A discrete choice survey
1320	Monitoring of COVID-19 medicines
1557	Infection and Drug Resistance
1658	A high-throughput, bead-based, antigen-specific assay to assess the ability of antibodies to induce complement activation
1684	Thrombocytopenia following Pfizer and Moderna SARS-CoV-2 vaccination
1772	Effectiveness of COVID-19 Vaccines against the B.1.617.2 (Delta) Variant
1974	Addressing the vaccine confidence gap
2172	Therapeutic potential of medicinal plants against COVID-19: The role of antiviral medicinal metabolites
2316	SARS-CoV-2 Vaccine ChAdOx1 nCoV-19 Infection Of Human Cell Lines Reveals Low Levels of Viral Backbone Gene Transcription Alongside Very High Levels of SARS-CoV-2 S Glycoprotein Gene Transcription

APÊNDICE C – Pfizer

C.1 Rede de Coautorias

A Figura 64 ilustra a rede de coautorias gerada para Pfizer. Trata-se de uma rede de uma escala um pouco maior em comparação com as outras redes sobre vacinas, que tinham todas menos de dez mil vértices, enquanto esta possui mais vinte mil vértices.

Figura 64 – Comunidades de Autores - Pfizer



A Figura 65 e a Figura 66 representam, respectivamente, a rede de coautorias filtrada para os vértices com maiores valores de centralidade do autovetor e grau ponderado. Nota-se que estas duas visualizações são parecidas, indicando vértices da mesma comunidade e também em comum nos dois casos.

Além disso, há a presença de alguns autores que já apareceram nas visualizações das seções anteriores, como é o caso, por exemplo, de Manjusha Gagliani. Tal fato, juntamente com o fato de se tratar de uma rede de maior escala, indica que boa parte das publicações filtradas para o caso das outras vacinas deve ter sido incluída dentro do filtro da Pfizer.

Diferentemente do que ocorreu nas outras vacinas, para o caso da rede de coautorias da Pfizer, as métricas de centralidade de autovetor ou grau ponderado não resultaram no grupo responsável pelo desenvolvimento da vacina.

Figura 65 – Top 11 Autores - Centralidade de Autovetor - Pfizer

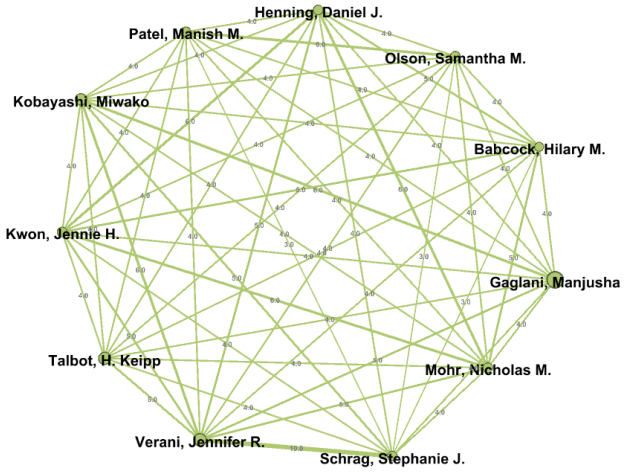
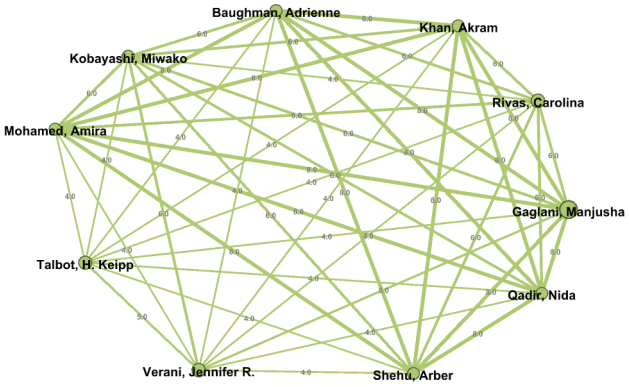
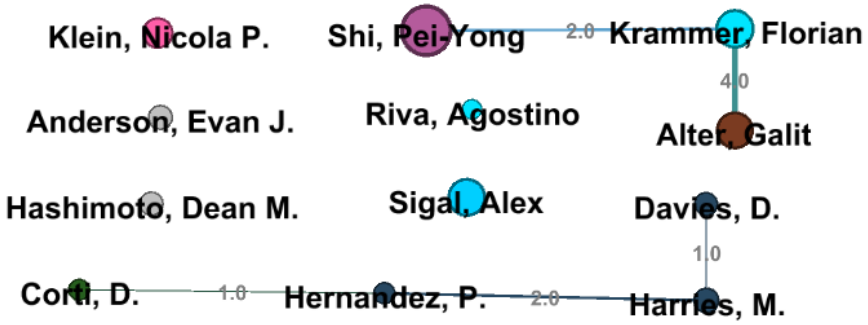


Figura 66 – Top 10 Autores - Grau Ponderado - Pfizer



A Figura 67 ilustra os principais autores do ponto de vista da métrica de centralidade de intermediação. Como nos outros casos, há representantes de diversas comunidades. Destaca-se novamente a presença da pesquisadora Galit Alter. Além dela, outros autores relevantes para esta rede do ponto de vista dessa métrica são: Pei-Yong Shi, vice-presidente de inovação em pesquisa em *The University of Texas Medical Branch*, e Florian Krammer, pesquisador em *Icahn School of Medicine at Mount Sinai*.

Figura 67 – Top 12 Autores - Centralidade de Intermediação - Pfizer



C.2 Rede de Citações

A Figura 68 ilustra a rede de citações para o caso da rede gerada pelos filtros do nome da vacina Pfizer. Em comparação com as redes anteriores, vê-se uma rede densa, devido ao maior volume de vértices, o que dificulta um pouco a determinação de quais são as principais comunidades da rede, mas é possível perceber boas parcelas de vértices lilás e verdes claro.

Figura 68 – Rede de Citações - Pfizer

A Figura 69 e a Tabela 20 ilustram os vértices com maiores valores de *PageRank* e o índice e título de cada um dos artigos representados pelos vértices, respectivamente. Dentre os títulos, há repetições de artigos que já apareceram para as vacinas anteriores, como é por exemplo o caso do artigo de índice 10.584, *"Therapeutic potential of medicinal plants against COVID-19: The role of antiviral medicinal metabolites"* (KHAN et al., 2021), que aborda o uso de plantas medicinais em tratamentos contra a COVID-19. No entanto, o artigo com maior *PageRank* é exclusivo desta rede, com o título: *"Mucosal immunity and poliovirus vaccines: impact on wild poliovirus infection and transmission"* (OKAYASU et al., 2011), um artigo que 2011 que estuda a imunidade gerada pela vacina da poliomelite, tratando-se de um

estudo que influenciou estudos sobre as vacinas contra o coronavírus.

Figura 69 – Top 12 Artigos - *PageRank* - Pfizer

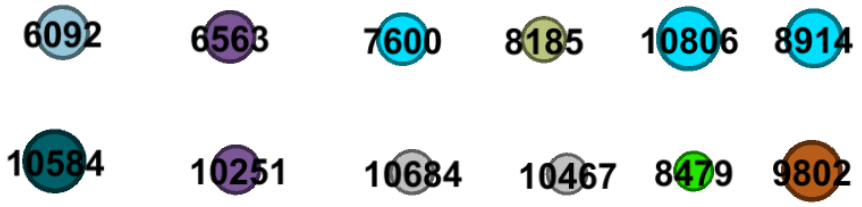


Tabela 20 – Top 12 Artigos - *PageRank* - Pfizer

ID	Título
6092	Monitoring of COVID-19 medicines
6563	Research on the Epidemiology of SARS-CoV-2 in Essential Response Personnel
7600	Single-cell RNAseq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection
8185	Adverse effects of COVID-19 mRNA-1273 vaccine: A randomized, cross-sectional study on healthcare workers with detailed self-reported symptoms
8479	Optimizing effectiveness of COVID-19 vaccination will laboratory stewardship play a role
8914	Effectiveness of COVID-19 Vaccines against the B.1.617.2 (Delta) Variant
9802	COVID-19 update: FDA issues policies to guide medical product developers addressing virus variants
10251	Mini Review Immunological Consequences of Immunization With COVID-19 mRNA Vaccines: Preliminary Results
10467	Breast Feeding and Respiratory Morbidity in Infancy: A Birth Cohort Study
10584	Therapeutic potential of medicinal plants against COVID-19: The role of antiviral medicinal metabolites
10684	The Molecular Signatures Database (MSigDB) hallmark gene set collection
10806	Mucosal immunity and poliovirus vaccines: Impact on wild poliovirus infection and transmission

A Figura 70 e a Tabela 21 ilustram os vértices com maiores valores de centralidade de intermediação e o índice e título de cada um dos artigos representados por eles. Comparando essa tabela com a Tabela 20, nota-se uma interseção de sete vértices, estando boa parte deles presentes nas redes das outras vacinas, o que fortalece a hipótese de que esta rede da Pfizer acabou englobando boa parte das publicações dos casos anteriores que envolvem vacinas em um contexto geral.

Figura 70 – Top 12 Artigos - Centralidade de Intermediação - Pfizer

Tabela 21 – Top 12 Artigos - Centralidade de Intermediação - Pfizer

ID	Título
1643	SARS-CoV-2 infection after vaccination in patients with inflammatory rheumatic and musculoskeletal diseases
2627	IDSA clinical practice guideline for vaccination of the immunocompromised host
3434	Covid-19 vaccines and variants of concern: A review
6092	Monitoring of COVID-19 medicines
6391	Proton Technologies AG General Data Protection Regulation (GDPR)
6563	Research on the Epidemiology of SARS-CoV-2 in Essential Response Personnel
6604	PEGylated liposomes: Immunological responses
7600	Single-cell RNAseq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection
8914	Effectiveness of COVID-19 Vaccines against the B.1.617.2 (Delta) Variant
9802	COVID-19 update: FDA issues policies to guide medical product developers addressing virus variants
10251	Mini Review Immunological Consequences of Immunization With COVID-19 mRNA Vaccines: Preliminary Results
10806	Mucosal immunity and poliovirus vaccines: Impact on wild poliovirus infection and transmission