



SimBraRedes

III Simpósio Brasileiro de Teoria  
e Ciência de Redes

## REDE SEMÂNTICA DE RESUMOS DA PRODUÇÃO CIENTÍFICA SOBRE SARS-CoV-2 DA REDE DE VIGILÂNCIA GENÔMICA DA FIOCRUZ

**Luiz Antonio de Assis Ferreira<sup>1</sup>**; Idê Gomes Dantas Gurgel<sup>2</sup>; Hernane B B Pereira<sup>3</sup>; Juliano de Carvalho Lima<sup>4</sup>

<sup>1</sup> Doutorando em Gestão de Saúde Pública, Instituto Aggeu Magalhães – Fiocruz, luiz.assis@fiocruz.br

<sup>2</sup> Doutora em Saúde Pública, Instituto Aggeu Magalhães – Fiocruz, ide.gomes@fiocruz.br

<sup>3</sup> Doutor em Engenharia Multmídia, Centro Universitário SENAI CIMATEC, Programa de Modelagem Computacional; Universidade do Estado da Bahia, Departamento de Educação, hbbpereira@gmail.com

<sup>4</sup> Doutor em Saúde Pública, Fundação Oswaldo Cruz, juliano.lima@fiocruz.br

### RESUMO

Sistemas complexos, como os que surgem da associação de palavras para a construção de um texto científico, podem ser estudados com o uso de redes sociais complexas. Este artigo tem por objetivo analisar a rede semântica de cliques construída a partir das temáticas abordadas e sintetizadas nos resumos dos trabalhos científicos sobre o SARS-CoV-2 produzidos pela Rede de Vigilância Genômica da Fiocruz, no período de março de 2020 a maio de 2024. Foram analisados 68 resumos de artigos elaborados por pesquisadores da rede da Fiocruz. A análise dos resumos foi realizada com técnicas da ciência de redes e teoria das redes. Os principais resultados apontam para uma estrutura de rede do tipo mundo pequeno, com características de uma rede livre de escala. Os resultados indicam que a produção proveniente da colaboração científica dos pesquisadores da Rede de Vigilância Genômica Fiocruz está fortemente centrada no objetivo desta rede de vigilância.

**PALAVRAS-CHAVE:** Redes Semânticas, Redes Sociais, Vigilância Genômica

<http://doi.org/10.55664/simbraredes2024.016>

### 1 INTRODUÇÃO

As redes complexas são caracterizadas como um tipo de sistema complexo e estão associadas a redes com topologias mais sofisticadas<sup>1</sup>, não triviais, composta por um conjunto de vértices interligados por um conjunto de arestas. Os sistemas complexos possuem propriedades raramente encontradas em modelos mais simples. De modo geral, as redes complexas envolvem inúmeros elementos que caracterizam arquiteturas mais complexas do que os clássicos gráficos aleatórios.<sup>2-4</sup> Por consequência, se representarmos sistemas complexos na forma de redes, podemos usar a estrutura conceitual de redes para explicar características que aparentam caráter universal.<sup>5</sup>

Na literatura científica, a teoria dos grafos, a análise de redes sociais e a teoria das redes complexas são usadas para compreender o comportamento e as estruturas de redes de diversos tipos. Podemos elencar estudos de redes tecnológicas em Watts e Strogatz<sup>6</sup> e em Ma, Sun e Zhang<sup>7</sup>; de redes de produção de conhecimento em equipes de pesquisa, como em Stvilia et al<sup>8</sup>; de redes de características de plantas, como os estudos de He et al<sup>9</sup>; e na biologia, identificamos pesquisas de redes de interação proteína-proteína, como a conduzida por Wilhelm, Nasheuer e Huang<sup>10</sup>. Wulff, Hills e Mata<sup>11</sup> investigaram diferenças individuais e de idade nas redes semânticas de adultos mais jovens e mais velhos.

Redes semânticas são um modo de representar informações e conhecimento com uma intenção objetiva a partir de determinado contexto. Essas redes são compostas por palavras ou conceitos com significado semântico e suas relações.<sup>12</sup>

O elemento básico em uma rede de cliques é o conjunto de  $n$  vértices, mutuamente conectados ou, simplesmente, clique. Como característica, temos a justaposição e/ou sobreposição de cliques. Quando a conexão entre duas cliques é realizada por um vértice, temos uma justaposição. Porém, quando a conexão entre duas cliques ocorre por meio de dois ou mais vértices, temos uma sobreposição.<sup>3</sup>

Analisar a rede semântica no âmbito da colaboração científica entre pesquisadores é fundamental para compreender o processo de produção do conhecimento e sua potencialidade para contribuir com políticas específicas. As redes semânticas são uma forma de pensar e estruturar o conhecimento existente em determinada comunidade<sup>13</sup>. Optamos por analisar os resumos dos trabalhos no lugar dos títulos ou palavras-chave, pelo fato de apresentarem uma sequência de períodos com mais informação capaz de representar o conteúdo do texto, contribuindo, desta forma, para o entendimento da produção científica na RVGF.



SimBraRedes

III Simpósio Brasileiro de Teoria  
e Ciência de Redes

Nesse sentido, neste trabalho, buscamos discutir a contribuição da colaboração científica durante o período da pandemia de covid, a partir das temáticas apresentadas e sintetizadas nos resumos de artigos científicos publicados por pesquisadores da Rede de Vigilância Genômica da Fiocruz (RVGF). Para alcançar tal objetivo, valemo-nos da análise de redes semânticas.

As redes de vigilância genômica têm como finalidade identificar novas variantes e linhagens de um patógeno logo após seu surgimento e, então, agir junto à saúde pública para impedir sua disseminação. Tal objetivo é muito mais abrangente do que o monitoramento da quantidade de variantes em cada região.

A RVGF realiza a vigilância genômica do SARS-CoV-2 no Brasil. A rede destina-se à pesquisa e à produção de conhecimento em colaboração entre pesquisadores, além de subsidiar o Ministério da Saúde na proposição de políticas públicas em Saúde.

## 2 METODOLOGIA

A base de dados usada para esta pesquisa foram os resumos (*abstracts*) dos artigos científicos publicados e disponíveis no *site* (<https://www.genomahcov.fiocruz.br/>) da Rede de Vigilância Genômica da Fiocruz (RVGF). Os 68 trabalhos analisados foram publicados no período de março de 2020 a maio de 2024. A rede semântica apresentada neste trabalho foi baseada nos resumos dos artigos, conforme descrito acima.

Uma rede pode ser representada por um grafo  $G = (V, E)$  que consiste em um conjunto de vértices  $V$  (finito e não vazio) e um conjunto de arestas  $E$  (que podem conectar os elementos de  $V$ )<sup>4,14</sup>. As métricas empregadas e calculadas neste trabalho basearam-se nos estudos de Nascimento, Pereira e Moret<sup>15</sup> e Fadigas e Pereira<sup>3</sup>. Os vértices da rede semântica de resumos são palavras lexicais presentes nos períodos dos resumos que são conectadas mutuamente quando aparecem no mesmo período ou pelos processos de justaposição e/ou sobreposição.

Para a construção do *corpus*, apoiamo-nos nas regras gerais apresentadas por Pereira, Fadigas, Senna e Moret<sup>4</sup> para processamento de títulos na construção de redes semânticas. As orientações apresentadas pelos autores foram aplicadas para todos os resumos. É importante observar que o idioma adotado nos resumos é o inglês. De modo a refletir a estrutura dos conceitos de significado contido nos textos, foram mantidas apenas as palavras com significado intrínseco, dispensando-se aquelas que possuem função meramente gramatical, como artigos, pronomes, preposições, conjunções, abreviações e interjeições. O passo seguinte foi reduzir as palavras restantes às suas formas canônicas, isto é, desconsiderar as flexões como formas plurais e formas verbais. As palavras desconhecidas foram preservadas.

Após esse tratamento manual, foi necessário dispor de um arquivo de texto (.txt), no qual cada resumo correspondia a uma linha do arquivo. Dessa forma, os vértices correspondem às palavras de cada *abstract*, e os laços são as conexões entre todas as palavras do mesmo período. Como resultado, temos que cada período é uma rede onde todos os vértices (palavras) estão interconectados, formando uma *clique* ou pelos processos de justaposição e sobreposição, ao surgirem em outros períodos. Considerando que, em uma rede de cliques, o principal elemento não é o vértice, mas sim um conjunto de  $n$  vértices ligados entre si<sup>3</sup>, temos que as palavras que aparecem em mais de um resumo são os vértices de conexão entre os resumos.

Para analisar uma rede, são propostos vários quantificadores que permitem analisar as diversas propriedades da rede. Neste estudo, usaremos os indicadores mais comumente citados: número de vértices  $n$ , número de arestas  $m$ , grau médio ou conectividade média  $\langle k \rangle$ , densidade  $\Delta$ , comprimento médio mínimo do caminho ou distância geodésica  $L$ , diâmetro  $D$ , distribuição de graus  $P(k)$  e o agrupamento de Watts e Strogatz<sup>6</sup>, a saber, coeficiente de aglomeração de um vértice  $C_i$  e coeficiente de aglomeração médio  $C_{WS} = \frac{1}{n} \sum_{i=1}^n C_i$ . Para a caracterização da rede semântica, usamos o cálculo das medidas de centralidades dos vértices: centralidade de grau  $C_D$ , a centralidade de proximidade  $C_C$  e a centralidade de intermediação  $C_B$ . Cabe-nos comentar que, se as distribuições de graus dos vértices seguirem uma lei de potência na forma  $P(k) \sim k^{-\gamma}$ , o expoente  $\gamma$  torna-se um índice importante para a caracterização da rede.<sup>5,15</sup>

## 3 RESULTADOS E DISCUSSÃO

As propriedades da rede semântica dos resumos dos trabalhos publicados pelos pesquisadores da RVGF (Figura 1) são apresentadas na

Tabela 1. O número de vértices representa o vocabulário resultante após o tratamento do texto, conforme descrito no método.



SimBraRedes

III Simpósio Brasileiro de Teoria e Ciência de Redes

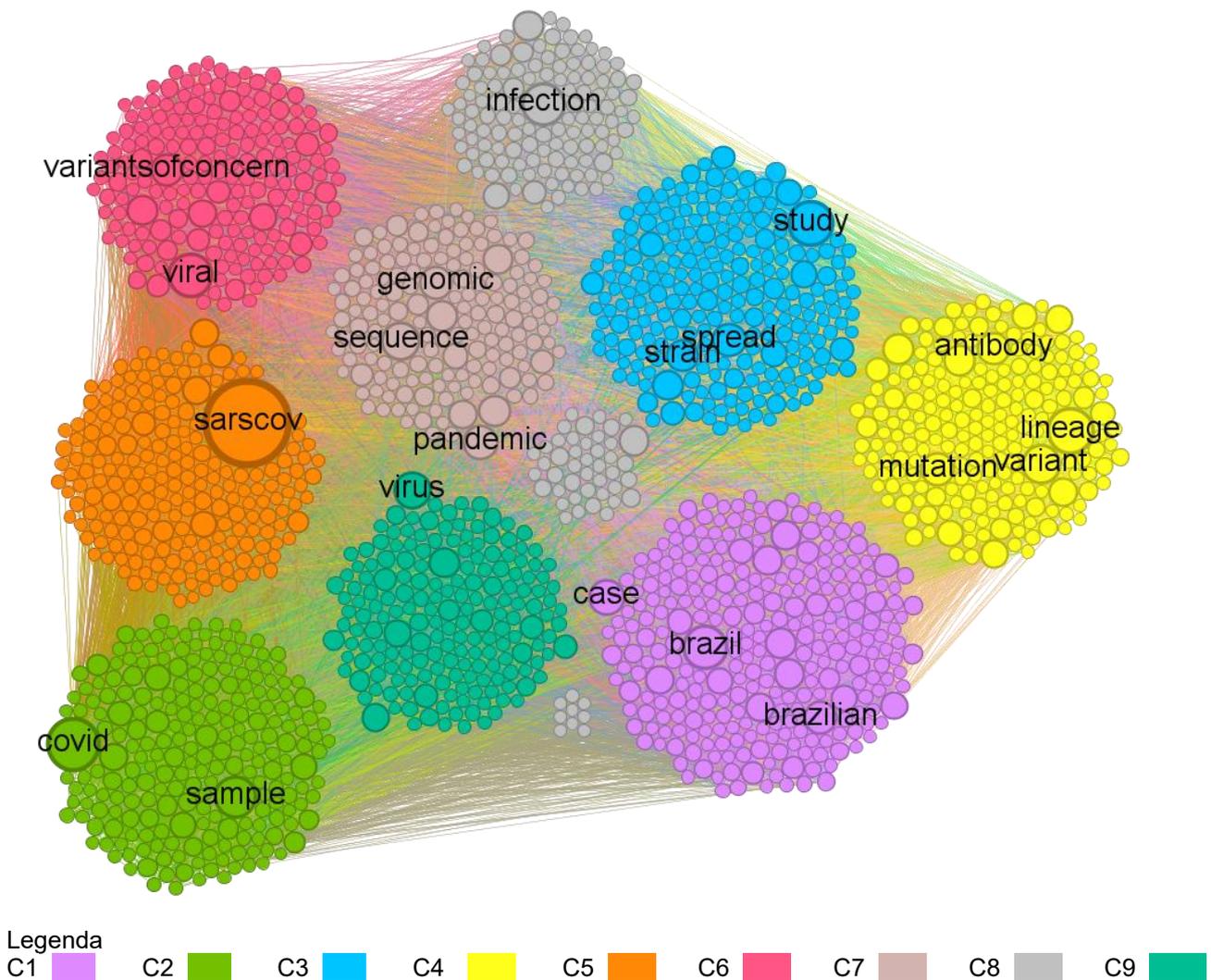
**Tabela 1** - Rede semântica de resumos

Redes	$n$	$m$	$\langle k \rangle$	$D$	$L$	$C_{ws}$	$\Delta$	$Comp$
RS Resumos	1848	38997	42,205	5	2,33	0,726	0,023	1
Rede aleatória	1848	38997	42,205	3	2,35	0,023	0,023	1

Legenda – RS: Rede Semântica;  $n$ : cardinalidade do conjunto de vértices;  $m$ : cardinalidade do conjunto de arestas;  $\langle k \rangle$ : grau médio;  $D$ : diâmetro;  $L$ : caminho mínimo médio;  $C_{ws}$ : coeficiente de aglomeração médio;  $\Delta$ : densidade;  $Comp$ : quantidade de componentes.

Para representar a rede semântica, usamos o *software* Gephi e escolhemos o algoritmo *Circle Pack Layout* e na primeira hierarquia usamos a modularidade. Calculamos a modularidade e as comunidades, aplicando o algoritmo de Blondel, Guillaume, Lambiotte e Lefebvre<sup>16</sup>. Na representação da rede semântica (Figura 1), o diâmetro dos vértices está associado ao seu grau (maior número de conexões com outros vértices).

**Figura 1** - Rede semântica de resumos





SimBraRedes

III Simpósio Brasileiro de Teoria  
e Ciência de Redes

A partir da análise dos dados, foram sugeridas 11 comunidades que, para o nosso estudo, são comunidades temáticas associadas à vigilância genômica do SARS-CoV-2. A Figura 1 revela as comunidades temáticas e as palavras que se destacam (maior grau), sugerindo o que se discute em cada uma das comunidades.

A modularidade ( $M$ ) é uma medida que indica a qualidade da partição de uma comunidade em comparação a outra forma de partição e pode variar entre  $0 < M < 1$ . Com base nos valores de  $M$ , Barabási<sup>17</sup> propôs que uma ótima partição possua  $M \geq 0,41$ ; uma subótima partição é encontrada quando temos  $0,22 \leq M < 0,41$ ; se  $M = 0$  temos uma comunidade única e se  $M$  for negativo, temos que cada nó é uma comunidade. O valor  $M = 0,303$  encontrado na rede semântica indica, consoante o mesmo autor, uma partição subótima para ótima.

A maior comunidade (C1), representando 15,04%, indica que o tema principal abordado está vinculado aos casos de covid-19 no Brasil, ou seja, trata aspectos regionais, o que pode ser evidenciado pelas palavras “brazilian” (brasileiro), “brazil” (Brasil) e “case” (caso). A segunda comunidade (C2) mais expressiva, com 12,45%, apresenta expressões como “covid”, “sample” (amostra), sugerindo que o enfoque é o estudo de amostras do vírus causador da covid-19 e a identificação de novas variantes. Um terceiro grupo expressivo (C3), representando 12,39%, realça palavras como “study” (estudo), “spread” (espalhamento) e “strain” (cepa, descendência), mostrando que o ponto de interesse é a disseminação e o monitoramento do patógeno causador da covid-19. A quarta comunidade relevante (C4), com 11,15%, enfatiza vocábulos que transmitem a preocupação com a mutação (“mutation”), a linhagem (“lineage”), anticorpo (“antibody”) e variantes (“variant”) do vírus. A palavra “sarscov” (SARS-CoV-2), vértice com o maior grau da rede semântica, está presente em uma comunidade (C5) que representa 11,04% denotando o interesse na proliferação do vírus. As expressões “variantsofconcern” (variante de preocupação) e “viral” (viral) estão presentes em uma comunidade (C6) que versa sobre as variantes virais e seu potencial de infecção de humanos. As expressões “sequence” (sequenciamento), “pandemic” (pandêmico) e “genomic” (genômica) fazem parte de uma comunidade (C7) que representa 8,55% que discute o sequenciamento genético, pilar da vigilância genômica.

Analisando as 11 comunidades, é possível depreender que a temática de cada uma apresenta coerência interna e se complementam entre si. O conjunto expressa claramente o sentido da vigilância genômica de um patógeno, do sequenciamento genético e monitoramento das variantes do vírus.

A ideia de coesão está também relacionada com a quantidade de arestas e com a forma da rede a partir das conexões estabelecidas, e permite avaliar a intensidade do relacionamento entre os vértices. As medidas de coesão mais comumente usadas são a densidade ( $\Delta$ ), o diâmetro ( $D$ ) e o grau médio  $\langle k \rangle$ .<sup>18</sup>

O valor do diâmetro encontrado para a rede semântica é  $D = 5$ . Este valor indica que a maior distância geodésica entre um par de vértices (i.e., duas palavras do discurso) é cinco e aponta para o fato de que as palavras mais distantes dos períodos de cada um dos resumos estão conectadas por meio de quatro outras palavras.

A densidade descreve a conectividade entre os vértices sendo expressa como a proporção entre os laços existentes e os laços possíveis na rede. A densidade é a probabilidade de existir um laço entre qualquer par de nós escolhidos aleatoriamente e é dada por  $\Delta = n(n - 1)/2$ . Pode variar de 0 (zero), onde não existe nenhum laço, até 1, onde todos os laços possíveis estão presentes.<sup>19,20</sup> Em uma rede semântica de resumos, a densidade indica a quantidade de ligações entre as palavras, isto é, a tendência de os resumos estarem ligados por meio de um número expressivo de palavras. A densidade da rede semântica foi de  $\Delta = 2,3\%$ , indicando o quão conectados estão os resumos dos trabalhos estudados. Esse valor de densidade informa que a rede semântica estudada é esparsa.<sup>18</sup>

O valor encontrado para o grau médio  $\langle k \rangle$  na rede semântica de resumos sugere que uma palavra (vértice) se conectou em média a 42 outros vértices, aproximadamente.

Ainda em relação à conectividade, temos que a rede em tela possui um único componente. Uma rede formada por um único componente indica haver um caminho entre qualquer par de vértices. Portanto, a abordagem dos resumos dos trabalhos publicados pelos pesquisadores da RVGF apresenta coesão, por consequência se valem de palavras similares. Em que pese a existência de um vocabulário de 1848 palavras estas se repetem nos distintos resumos, denotando serem termos característicos da área de vigilância genômica, conquanto tenhamos identificado 11 comunidades conforme relatado anteriormente.

Em relação à caracterização topológica da rede em “aleatória”, “mundo pequeno” ou “livre de escalas”, o coeficiente de aglomeração médio  $C_{WS}$ , o caminho mínimo médio  $L$  e a distribuição de graus  $P(k)$  são medidas fundamentais, considerando o método proposto por Watts e Strogatz<sup>6</sup>. O  $C_{WS}$  usado é o proposto por Watts e Strogatz<sup>6</sup> que retrata a adjacência imediata de um vértice  $i$  em uma dada rede. O valor encontrado para o coeficiente de aglomeração médio  $C_{WS} = 0,726$  sugere uma alta conectividade entre as palavras dos resumos.

O valor encontrado para o caminho mínimo médio foi de  $L = 2,33$ , o que equivale aproximadamente a duas arestas para que duas palavras estejam conectadas.

Com o auxílio do *software* Gephi, construímos uma rede aleatória, baseada no modelo de Erdős e Renyi<sup>21</sup>, com o mesmo número de vértices e grau médio que a rede semântica de resumos. De acordo com Watts e Strogatz<sup>6</sup>, uma rede apresenta características de mundo pequeno se  $C_{ws} \gg C_a$  e se  $L_{ws} \approx L_a$ . Pelos resultados obtidos, podemos concluir que a rede semântica assume características de uma rede mundo pequeno. Em relação a uma rede semântica de resumos, podemos dizer que as palavras, em termos geodésicos, estão muito próximas umas das outras. O coeficiente de aglomeração médio  $C_{ws}$  sugere haver uma alta probabilidade de que duas palavras ligadas a uma terceira estejam elas mesmas ligadas entre si.

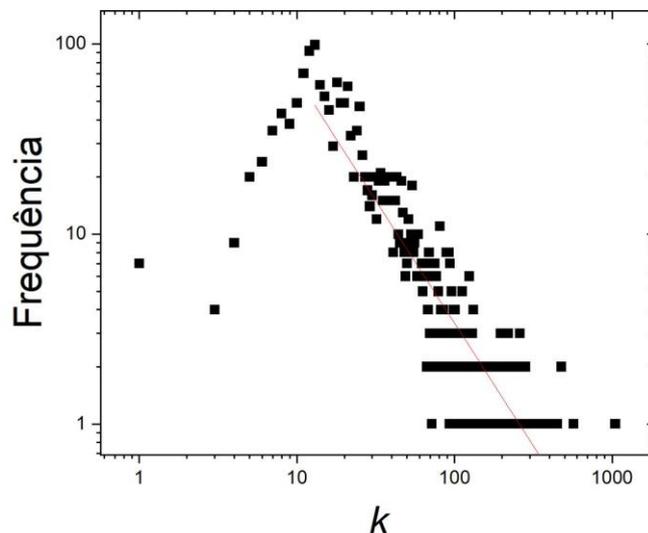
Em redes complexas, é importante observar a distribuição de graus para caracterizar a sua topologia. Quando a distribuição de graus se aproxima de uma lei de potência, estamos diante de uma rede livre de escala. A alta conectividade percebida na rede semântica indica existir uma atração por alguns vértices, o que nos levou a verificar a distribuição de graus da rede.

A Figura 2 mostra a distribuição de graus da rede semântica de resumos (Figura 1). O gráfico sugere que a rede semântica segue uma lei de potência na forma  $P(k) \sim k^{-\gamma}$ , sendo os valores encontrados para  $\gamma = 1,29274$  e para o coeficiente de ajuste  $R^2 = 0,79348$ . Podemos dizer que estamos diante de uma rede livre de escala. Comparado ao trabalho de Barabási e Albert<sup>22</sup> os valores de  $\gamma$  devem permanecer entre 2,1 e 4, porém, o valor encontrado se alinha às redes descritas nas obras de Seyed-allaei<sup>23</sup>.

Deve-se ponderar que, do ponto de vista das redes complexas, as redes livres de escala apresentam maior robustez em relação à retirada de vértices aleatoriamente, sendo, porém, mais vulneráveis à retirada daqueles com maior grau.<sup>12,24</sup> Fizemos uma breve análise de robustez da rede semântica. A estratégia usada foi a remoção de vértices com maior grau de centralidade. Ao remover os 20 vértices com maior grau (respeitando o critério descrito a seguir), observamos que o ponto de corte acontece na palavra “lineage” (linhagem). Neste ponto, ocorre a desconexão da rede que passou a ter 2 componentes, o maior componente, contudo, representa 99,95% da rede. O percentual do maior componente, após a retirada dos vértices, resultou em uma rede com o número de vértices muito próximo ao da rede antes da retirada. Isso significa uma espécie de alinhamento do discurso, percebido na repetição das palavras lexicais. O menor componente apresenta um único vértice, de grau 1, ligado ao vértice “lineage”. Apesar dos estudos de sequenciamento genômico buscarem identificar a linhagem dos patógenos explicando a importância do termo “lineage” na rede semântica, fica evidenciado que a rede semântica apresenta robustez mesmo com a retirada de vértices com maior centralidade, ou seja, há indícios de que não apresenta “hubs” que a tornem vulnerável com a sua retirada.

Com base nos autores e na análise topológica, é possível dizer que a rede semântica estudada é uma rede mundo pequeno e uma rede livre de escala, podendo ambas as características ser concomitantes.

**Figura 2** - Distribuição de Graus





A aplicação de alguns indicadores comumente usados na análise de redes sociais permite algumas interpretações sobre o comportamento da rede. Os índices de centralidade podem ser relevantes para conhecer a importância de algumas palavras. Neste trabalho, usamos os índices de centralidade de grau  $C_D$ , centralidade de proximidade  $C_C$  e centralidade de intermediação  $C_B$ .

Em pesquisas de redes sociais, os principais índices usados são os de centralidade. O conceito de centralidade está associado ao vértice. De uma forma mais ampla, a centralidade de um vértice está diretamente relacionada à contribuição que aquele vértice tem para a estrutura da rede. Podemos, assim, pensar a centralidade como a importância estrutural de um vértice, uma vez que os atores centrais devem ser os mais ativos, pois possuem mais laços com outros atores da rede.<sup>14,24</sup>

Para a identificação das palavras mais importantes, adotamos como critério de corte a média do grau ( $\mu_c$ ) de todos os vértices somado a quatro desvios-padrão ( $\sigma$ ). Assim, foi possível destacar 20 palavras. Os índices de centralidade de cada palavra estão relacionados na Tabela 2.

A centralidade de grau de um vértice é definida como o número de vértices adjacentes a ele, ou o grau do vértice. Em redes semânticas, uma palavra terá mais importância em função do número de palavras adjacentes. A RVGF tem como finalidade identificar novas variantes do vírus causador da covid-19 logo após seu surgimento e, então, agir junto ao sistema de saúde pública para impedir sua disseminação. Observando as principais palavras dos resumos, é possível identificar um núcleo central voltado para o estudo do sequenciamento genômico do vírus. Com o maior número de conexões, temos a palavra “sarscov” (SARS-Cov-2 ou CoronaVírus da Síndrome Aguda Respiratória Grave 2), nome dado ao vírus causador da “covid” (covid-19), vértice que surge com o segundo maior número de conexões. A palavra “lineage” (linhagem) surge com o terceiro maior número de conexões e conceitualmente é entendida como o conjunto de vírus que descendem de um ancestral comum, com mutações que a diferenciem de outras variantes. Todas as linhagens e variantes do novo coronavírus pertencem à espécie SARS-CoV-2.<sup>25</sup> As demais palavras conduzem a uma convergência dos resumos para o estudo do sequenciamento genômico do vírus.

A centralidade de proximidade mede quanto o vértice está próximo dos demais vértices da rede. Esse índice amplia o entendimento da importância do vértice, posto que visa expressar não só a relevância local da palavra em relação aos vizinhos mais próximos, mas também em relação a toda a rede. Mais uma vez, destacam-se as palavras “sarscov” (SARS-Cov-2), “covid” (covid-19) e “lineage” (linhagem) como as mais importantes, tendo por base os dois índices de centralidade.

A centralidade de intermediação ( $C_B$ ) mede até que ponto um vértice se encontra como meio para alcançar outros vértices, ou seja, reflete o quanto a palavra conecta outras palavras que não estão conectadas entre si. Na rede semântica em estudo, observamos que as palavras “sarscov” (SARS-Cov-2) e “covid” (covid-19) possuem o maior índice para as três centralidades. A palavra “lineage” (linhagem) também se apresenta como importante conector entre outras palavras, juntamente com a palavra estudo.

**Tabela 2** - Medidas de Centralidade das principais palavras

Vértice	$C_D$	$C_C$	$C_B$	Número componentes com retirada	% do maior componente
SARScov	1047	0.695931	0.190403	1	100%
covid	568	0.589343	0.049824	1	100%
Lineage	478	0.572891	0.030752	2	99,95%
Study	478	0.571827	0.032822	2	99,95%
Viral	449	0.567086	0.031764	2	99,95%
Brazil	445	0.565697	0.02386	2	99,95%
infection	430	0.565005	0.02649	2	99,95%
sample	408	0.559358	0.037978	2	99,95%
variant	374	0.553823	0.017796	2	99,95%
sequence	362	0.550522	0.02079	2	99,95%
virus	350	0.547097	0.02099	2	99,95%



SimBraRedes

III Simpósio Brasileiro de Teoria  
e Ciência de Redes

case	332	0.543875	0.013944	2	99,95%
pandemic	314	0.541166	0.009285	2	99,95%
spread	306	0.536763	0.010283	2	99,95%
genomic	295	0.536451	0.010688	2	99,95%
mutation	292	0.538013	0.010166	2	99,95%
brazilian	290	0.539743	0.009854	2	99,95%
strain	288	0.540058	0.015859	2	99,95%
variantsofconcern	282	0.535984	0.008295	2	99,95%
antibody	282	0.535984	0.009689	2	99,95%

Legenda –  $C_D$ : Centralidade de Grau;  $C_C$ : Centralidade de Proximidade;  $C_B$ : Centralidade de Intermediação.

#### 4 CONSIDERAÇÕES FINAIS

O método usado para a análise da rede semântica, baseada nos resumos da produção científica dos pesquisadores da RVGF, mostrou-se eficaz para o alcance dos objetivos propostos.

Com fundamento no método de Watts e Strogatz<sup>6</sup>, é possível argumentar que a rede semântica dos resumos é uma rede que se pode caracterizar como de mundo pequeno. Alicerçado em Barabási e Albert<sup>22</sup> e de Seyed-Allaei<sup>23</sup>, concluímos que a rede semântica apresenta características de uma rede livre de escala, ou seja, segue uma lei de potência. A análise de robustez da rede apontou como ponto de corte o termo “lineage” (linhagem), comprovando a sua importância no conteúdo dos resumos, tendo em vista que o sequenciamento genômico estuda a linhagem dos patógenos. Tais achados sugerem que estamos diante de uma rede mundo pequeno, a qual apresenta características topológicas de uma rede livre de escala.

Os vértices que apresentam maior centralidade de grau, de proximidade e de intermediação são capazes de explicar que os resumos dos trabalhos desenvolvidos pelos pesquisadores estão em consonância com o objetivo da RVGF. As 20 palavras selecionadas a partir do ponto de corte possuem um papel estrutural na rede semântica e demonstram alinhamento com o objetivo da RVGF.

Por fim, podemos dizer que o estudo revelou que o uso da ciência e da teoria das redes permitiram compreender a colaboração científica entre pesquisadores a partir do discurso sintetizado em seus resumos.

#### REFERÊNCIAS

- <sup>1</sup> Dorogovtsev SN, Mendes JFF. **The nature of complex networks**. Oxford, United Kingdom: Oxford university press; 2022.
- <sup>2</sup> Oliveira Do Nascimento J, Pereira HBB, Do Vale Cunha M, Albano Moret M. Sistemas complexos e ciências das redes: redes semânticas baseadas em abstracts e keywords do ensino de física nacional. **Cereus**. 31 de março de 2019;11(1):161–83.
- <sup>3</sup> Fadigas IS, Pereira HBB. A network approach based on cliques. **Physica A: Statistical Mechanics and its Applications**. maio de 2013;392(10):2576–87.
- <sup>4</sup> Pereira HBB, Fadigas IS, Senna V, Moret MA. Semantic networks based on titles of scientific papers. **Physica A: Statistical Mechanics and its Applications**. março de 2011;390(6):1192–7.
- <sup>5</sup> Caldeira SMG, Petit Lobão TC, Andrade RFS, Neme A, Miranda JGV. The network of concepts in written texts. **Eur Phys J B**. fevereiro de 2006;49(4):523–9.
- <sup>6</sup> Watts DJ, Strogatz SH. Collective dynamics of ‘small-world’ networks. **Nature**. 1998;393(6684):440–2.
- <sup>7</sup> Ma J, Sun Z, Zhang Y. Enhancing traffic capacity of multilayer networks with two logical layers by link deletion. **IET Control Theory & Appl**. janeiro de 2022;16(1):1–6.
- <sup>8</sup> Stvilia B, Hinnant CC, Schindler K, Worrall A, Burnett G, Burnett K, et al. Composition of scientific teams and publication productivity at a national science lab. **J Am Soc Inf Sci**. fevereiro de 2011;62(2):270–83.
- <sup>9</sup> He N, Li Y, Liu C, Xu L, Li M, Zhang J, et al. Plant trait networks: improved resolution of the dimensionality of adaptation. **Trends in Ecology & Evolution**. outubro de 2020;35(10):908–18.
- <sup>10</sup> Wilhelm T, Nasheuer HP, Huang S. Physical and functional modularity of the protein network in yeast. **Molecular & Cellular Proteomics**. maio de 2003;2(5):292–8.
- <sup>11</sup> Wulff DU, Hills TT, Mata R. Structural differences in the semantic networks of younger and older adults. **Sci Rep**. 12 de dezembro de 2022;12(1):21459.



SimBraRedes

III Simpósio Brasileiro de Teoria  
e Ciência de Redes

- <sup>12</sup> Grilo M, Fadigas IS, Miranda JGV, Cunha MV, Monteiro RLS, Pereira HBB. Robustness in semantic networks based on cliques. **Physica A: Statistical Mechanics and its Applications**. abril de 2017;472:94–102.
- <sup>13</sup> Hartley RT, Barnden JA. Semantic networks: visualizations of knowledge. **Trends in Cognitive Sciences**. agosto de 1997;1(5):169–75.
- <sup>14</sup> Borgatti SP, Everett MG, Johnson JC. **Analyzing social networks**. Los Angeles: SAGE; 2013. 296 p.
- <sup>15</sup> Oliveira Do Nascimento J, Pereira HBB, Albano Moret M. Grafos e Teoria de Redes: uma análise do Ensino de Física Brasileiro no período 1972-2006 por meio de cliques de palavras-chave. **Cereus**. 1º de agosto de 2018;10(2):315–39.
- <sup>16</sup> Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. **J Stat Mech [Internet]**. 9 de outubro de 2008 [citado 31 de julho de 2024];2008(10):P10008. Disponível em: <https://iopscience.iop.org/article/10.1088/1742-5468/2008/10/P10008>
- <sup>17</sup> Barabási AL, Pósfai M. **Network science**. Cambridge: Cambridge university press; 2016.
- <sup>18</sup> Fadigas IDS, Henrique T, Senna VD, Moret MA. Análise de redes semânticas baseada em títulos de artigos de periódicos científicos: o caso dos periódicos de divulgação em educação matemática. São Paulo. 2009;11(1).
- <sup>19</sup> Newman MEJ. **Networks: an introduction**. Oxford ; New York: Oxford University Press; 2010. 772 p.
- <sup>20</sup> Wasserman S, Faust K. **Social network analysis: methods and applications**. Cambridge ; New York: Cambridge University Press; 1994. 825 p. (Structural analysis in the social sciences).
- <sup>21</sup> Erdős PL, Rényi A. On the evolution of random graphs. **Transactions of the American Mathematical Society**. 1984;(286):257–257.
- <sup>22</sup> Barabási AL, Albert R. Emergence of scaling in random networks. 1999;286.
- <sup>23</sup> Seyed-allaei H, Bianconi G, Marsili M. Scale-free networks with an exponent less than two. **Phys Rev E**. 10 de abril de 2006;73(4):046113.
- <sup>24</sup> Latora V, Marchiori M. A measure of centrality based on network efficiency. **New J Phys**. 28 de junho de 2007;9(6):188–188.
- <sup>25</sup> Fiocruz. Rede Genômica Fiocruz [Internet]. [citado 5 de agosto de 2024]. Disponível em: <https://www.genomahcov.fiocruz.br/>