

ARTIGOS RETRATADOS SOBRE COVID-19: UMA ANÁLISE SOBRE SUA REDE DE CITAÇÕES*

Retracted articles on COVID-19: an analysis of their citation network

Hugo Souza Duca Aguiar¹ , Ingrid Quintanilha Pacheco² ,
Giseli Rabello Lopes³ , Maria Luiza Machado Campos⁴ , Jonice Oliveira⁵ 

Resumo: Para auxiliar no enfrentamento da pandemia da COVID-19, pesquisas foram aceleradas, acarretando em muitos artigos sendo publicados apressadamente, alguns destes sofrendo retratações. Neste trabalho são utilizadas técnicas de análise de redes sociais e métricas de centralidade e intermediação para avaliar a influência dos trabalhos retratados sobre os que os referenciam. Como resultado, foi verificado que o trabalho retratado mais referenciado teve 2.745 citações, 648 artigos fazem referência a mais de um retratado e sete retratados referenciam outros na mesma situação. Além disso, foi constatado que 82,11% das referências foram feitas após a retratação e que as palavras mais usadas nos títulos eram associadas aos possíveis tratamentos da doença.

Palavras-chave: COVID-19. Redes Sociais. Retratação. Artigos Retratados.

Abstract: To help fight the COVID-19 pandemic, some research studies were accelerated, resulting in many articles being hastily published, some of which were retracted. In the present work, social network analysis techniques and centrality and betweenness metrics were used to assess the influence of retracted articles on those that referred to them. As a result, it was found that the most referred retracted work had 2,745 citations, 648 articles refer to more than one retracted article and seven retracted works refer to others in the same situation. In addition, it was found that 82.11% of the references were made after the retraction and that the words most used in the titles were associated with possible treatments for the disease.

Keywords: COVID-19. Social Network. Retraction. Retracted Articles.

*Este artigo é uma reimpressão do original publicado no X Brazilian Workshop on Social Network Analysis and Mining - BRASNAM 2021, autorado pelos mesmos contribuidores e com o mesmo título.

1. Capitão-Tenente (T). Mestre em Informática pelo Programa de Pós-Graduação em Informática da Universidade Federal do Rio de Janeiro. Comando da Marinha, Brasília, DF - Brasil. E-mail: hugo.duca@marinha.mil.br

2. Mestranda em Ciência da Computação do Programa de Pós-Graduação em Informática da Universidade Federal do Rio de Janeiro, Rio de Janeiro - RJ, Brasil. E-mail: ingridpacheco@ic.ufrj.br

3. Professora da Universidade Federal do Rio de Janeiro, Rio de Janeiro - RJ, Brasil. E-mail: giseli@ic.ufrj.br

4. Professora da Universidade Federal do Rio de Janeiro, Rio de Janeiro - RJ, Brasil. E-mail: mluiza@ppgi.ufrj.br

5. Professora da Universidade Federal do Rio de Janeiro, Rio de Janeiro - RJ, Brasil. E-mail: jonice@ic.ufrj.br

1. INTRODUÇÃO

No início de 2020, a Organização Mundial da Saúde (OMS) declarou a pandemia da *Coronavirus Disease 2019* (COVID-19), uma doença relativamente nova. Embora o SARS-CoV já fosse conhecido, o SARS-CoV2 mostrou-se muito mais contagioso, e a doença mais agressiva com sintomas e características epidemiológicas ainda parcialmente desconhecidas. Assim, com a rápida disseminação da doença, as pesquisas foram aceleradas e muitos de seus resultados rapidamente divulgados, na tentativa de auxiliar o enfrentamento da crise (SOLTANI; PATINI, 2020). Ao mesmo tempo, considerando-se a velocidade com que as informações são divulgadas atualmente, nem sempre com sua veracidade ou fontes verificadas, o efeito de resultados imprecisos ou mal interpretados pode ser crítico.

A corrida para a publicação de artigos sobre a COVID-19 trouxe preocupações acerca da qualidade metodológica e da validade científica deles. Em junho de 2020, em uma pesquisa no *Retraction Database* usando as palavras-chave “COVID-19”, “coronavirus disease 2019”, “coronavirus 2019”, “SARS-CoV-2” e “2019nCoV”, foi possível encontrar 26 artigos que haviam sido removidos ou retratados (SOLTANI; PATINI, 2020).

A retratação de artigos pode ser definida como um mecanismo para corrigir a literatura e alertar os leitores sobre artigos que contenham conteúdos ou dados erroôneos e cujas descobertas ou conclusões possam ser falhas. O seu principal propósito é garantir a integridade e corrigir a literatura, e não punir os autores (BROWMAN *et al.*, 2019).

Dessa forma, uma retratação demonstra que o trabalho em questão não deveria ter sido publicado, que suas conclusões e dados não devem ser usados para apoio a pesquisas futuras. Alguns dos motivos para a retratação de um artigo são: plágio, pesquisa não confiável, dados publicados anteriormente e pesquisa antiética. Quando ocorre uma retratação, significa que ocorreu algum problema tão significativo com a pesquisa que suas descobertas podem ser invalidadas (SHETH; THAKER, 2014).

Quando usada corretamente, uma citação serve como ferramenta valiosa para apoiar uma afirmação, método ou hipótese. Por outro lado, um trabalho que foi oficialmente marcado como retratado pelo editor foi, em essência, riscado na íntegra do registro acadêmico (SILVA; BORNEMANN-CIMENTI, 2017).

Entretanto, muitos são os casos em que os artigos, mesmo após sofrerem retratação, continuam sendo usados para apoiar novas descobertas científicas, o que pode colocar em risco o processo científico. No site *retractionwatch.com*, podem-se observar exemplos de artigos que possuem o número de citações ainda maior após a sua retratação, demonstrando a gravidade do problema mencionado.

Nesse contexto, visando analisar retratações e citações na temática da COVID-19, são formuladas as seguintes perguntas de pesquisa a serem respondidas pelo presente artigo:

- P1 Quais são os artigos retratados mais referenciados?;
- P2 Quais os principais temas e similaridades entre os artigos retratados?;
- P3 As citações ocorreram antes ou após o aviso de retratação?;
- P4 Existe uma cadeia de artigos retratados (que referenciam outros na mesma condição)?;
- P5 Quais são os artigos que mais referenciam artigos retratados (possivelmente diminuindo significativamente a sua qualidade)?

Visando responder às questões de pesquisa elencadas, e considerando-se a falta de um conjunto de dados único com as informações necessárias, foi construído, por meio de um *script* em Python, um conjunto de dados que continha 5.884 trabalhos que referenciavam 73 artigos retratados na temática de COVID-19, dados suficientes para as análises propostas.

Posteriormente, foi construída uma rede com base no conjunto encontrado, e foram empregadas diferentes técnicas de análise de redes sociais (ARS) (WASSERMAN; FAUST, 1994). Assim, foi possível encontrar os cinco artigos retratados mais referenciados, majoritariamente após o aviso de retratação, dentre os quais um é citado por 2.745 artigos e outro ainda referencia dois trabalhos na mesma situação, demonstrando a existência de uma cadeia de referências de retratados.

Ademais, foi possível descobrir que uma grande porção desses artigos mencionava remédios que eram citados como possíveis tratamentos para a COVID-19 no início da pandemia. Diante desses resultados ficou perceptível como são recorrentes as citações a artigos retratados, e o quanto isso pode impactar a confiabilidade das pesquisas.

O restante deste artigo está organizado da seguinte forma. Na seção 2, são discutidos os trabalhos relacionados a artigos, citações e retratações na temática da COVID-19. Na seção 3, são discutidas as métricas utilizadas na análise de redes sociais.

Na seção 4, são apresentadas as informações sobre a criação do conjunto de dados. Na seção 5, são apresentadas as análises realizadas e os resultados obtidos. Por fim, na seção 6, são apresentadas as conclusões e vias para trabalhos futuros.

2. TRABALHOS RELACIONADOS

Nesta seção, são discutidos alguns trabalhos relacionados especificamente a artigos, retratações e citações na temática da COVID-19/SARS-CoV-2.

2.1. PUBLICAÇÕES E CITAÇÕES

Diferentes estudos enfatizam o crescimento das publicações sobre COVID-19 ao longo da pandemia. No final de 2020 havia estimativas acerca da publicação de mais de 100 mil artigos e pré-prints sobre o assunto e da possibilidade de esse número ultrapassar os 200 mil em dezembro do mesmo ano (ELSE, 2020). Além disso, entre as análises apresentadas estão os principais tópicos abordados (modelagem epidemiológica, controle de propagação, saúde pública, diagnósticos e testes, saúde mental e mortalidade em hospitais) e os países dos autores centrais (China e Estados Unidos) (BOSCHIEROA *et al.*, 2021).

Em outro estudo, foram utilizados dados coletados da base de dados United States National Library of Medicine (PubMed), entre março e abril de 2020, sendo realizada uma classificação de acordo com o nível de evidências para uma avaliação quantitativa de qualidade metodológica e uma análise narrativa dos pontos fortes e fracos das publicações sobre COVID-19 (ZDRAVKOVIC *et al.*, 2020). Esse procedimento mostrou evidências de que a qualidade das publicações sobre o tema nos três periódicos da área médica mais bem classificados (*The New England Journal of Medicine [NEJM]*, *The Journal of the American Medical Association [JAMA]* e *The Lancet*) está abaixo da média de qualidade desses periódicos. Os autores relataram ainda que o número de publicações sobre COVID-19 é quase igual ao número de publicações sobre todos os outros tópicos, destacando a importância de fomentar um debate sobre o valor científico, a ética e a sobrecarga de informações nas pesquisas sobre a temática.

Apesar de a publicação durante a pandemia parecer repleta de riscos, é tentadora pelo imenso interesse do público e, em comparação a outros tópicos, os artigos sobre COVID-19 demonstraram gerar mais citações (mediana 45 *vs.* 2 citações)

(CORTEGIANI *et al.*, 2021). Um estudo que usou como fonte o *site* Retraction Watch relatou que houve aumento na velocidade das revisões em periódicos da área médica e um alarmante crescimento nas retrações de artigos relacionados à COVID-19 (15 pré-prints e 24 artigos de periódicos retirados ou retratados até dezembro de 2020) (ELSE, 2020).

2.2. RETRATAÇÕES E CITAÇÕES

Estudos recentes, feitos em novembro de 2020, apresentam análises preliminares sobre retratações de artigos sobre COVID-19 usando as bases de dados Retraction Watch e PubMed. De acordo com eles, naquela época já existiam 39 casos de artigos retratados, encontrados tanto em periódicos bem estabelecidos quanto naqueles com fator de impacto mais baixo, tendo seus autores um *b-index* moderadamente alto (CORTEGIANI *et al.*, 2021).

Entre os principais motivos de retratações estão desde duplicatas e plágio até questões metodológicas e de má interpretação de dados (BOSCHIEROA *et al.*, 2021).

Duplicação, questões éticas e plágio são mais frequentes em periódicos com baixo indicador de *SCImago Journal Rank* (SJR), enquanto periódicos com alto indicador de SJR têm, em sua maioria, questões metodológicas como motivo de retratação.

De acordo com os levantamentos realizados, pode-se observar que grande parte dos trabalhos relacionados se concentra em caracterizar publicações e retratações sobre a COVID-19. Em uma investigação dos 200 artigos acadêmicos mais recentes publicados em 2020, mais da metade, incluindo os publicados em periódicos conceituados, usou os artigos retratados para apoiar suas descobertas científicas e não notificou as retratações (PILLER, 2021). Além disso, não foram encontrados estudos mais aprofundados em relação às redes de citações desses artigos retratados. Dessa forma, o presente trabalho visa realizar este estudo com o uso de técnicas de análises de redes sociais.

3. ANÁLISE DE REDES SOCIAIS

Uma rede social é definida por um conjunto finito de atores e as relações definidas sobre eles (WASSERMAN; FAUST, 1994). A presença de informações relacionais é um aspecto crítico e definidor de uma rede social. Na análise de redes sociais, uma rede representa um grupo de atores (nós) que se relacionam. Essas relações, *links* ou vínculos (arestas) caracterizam-se por fluxos de informação. Dessa maneira,

fluxos, nós e vínculos constituem os elementos básicos de uma rede (RIBEIRO; BASTOS, 2011).

Entre as métricas para a análise de redes sociais estão incluídas as de centralidade. O papel da centralidade dentro de uma rede é parte do que geralmente tentamos entender durante a análise. Existem diversas maneiras de medir a centralidade, cada uma das quais serve para auxiliar em um tipo de entendimento específico (CHERVEN, 2015). A seguir será feita uma breve introdução sobre essas métricas, especialmente as que serão empregadas para as análises apresentadas na seção 5.

- Centralidade local: é calculada como o número total de *links* diretos com os demais nós da rede; consequentemente, um valor elevado de centralidade representa uma posição mais centralizada do nó. Esses nós podem ajudar a facilitar o fluxo de informações de um grupo para o outro em um contexto organizacional (HATALA, 2006);
- Centralidade de intermediação (*Betweenness centrality*): é calculada como o número de menores caminhos entre quaisquer dois nós da rede que passem pelo determinado nó. Nós com alto valor de centralidade de intermediação podem oferecer o caminho mais direto a *clusters* desconectados, sendo denominados de pontes. No entanto, ser uma ponte não é um pré-requisito para ter um alto valor de centralidade de intermediação, mas é comum que esses nós sejam classificados como criticamente importantes usando essa medida (CHERVEN, 2015).

Além disso, com o propósito de prover melhor visualização para grafos, comumente utilizados na representação de redes sociais, alguns algoritmos foram desenvolvidos. Neste artigo, optou-se por usar o algoritmo de distribuição Fruchterman-Reingold, que faz um bom trabalho em distribuir os vértices de maneira uniforme, tornando os comprimentos das bordas homogêneos e refletindo a simetria (FRUCHTERMAN; REINGOLD, 1991).

4. CONSTRUÇÃO DO CONJUNTO DE DADOS

Para realizar as análises propostas, algumas etapas foram necessárias, incluindo a construção de um conjunto de dados que tivesse todas as informações necessárias. Os dados foram coletados de três fontes distintas: Retraction Watch Database, para obter a

listagem dos artigos retratados, Semantic Scholar, para obter as citações e informações deles, e a DOI API, para obter sua data de publicação. A segunda fonte mencionada ainda foi utilizada para verificar se o trabalho que fez referência ao retratado foi fortemente influenciado por ele, dependendo do número de vezes que a referência aparece, assim como a sua localização e contexto.

Assim, foi utilizado o Retraction Watch Database filtrando o campo de título por: “COVID-19” or “coronavirus disease 2019” or “coronavirus 2019” or “SARSCOV-2” or “2019-nCov”. Por meio dessa consulta, realizada em 19 de março de 2021, foram obtidos 92 artigos. Destes, nove não tinham DOI disponível, restando 83 resultados, dos quais seis estavam aparecendo repetidamente na listagem, portanto resultando em 77 artigos retratados para análise.

Em seguida, a Semantic Scholar API foi utilizada para obter mais informações sobre cada artigo retratado e os trabalhos que os citavam. Foi verificado que quatro dos artigos retratados não estavam disponíveis na API, restando 73 artigos nessa situação. Por fim, o *script*, por meio de requisições para a DOI API, obteve a data de publicação dos artigos, um dado importante para a análise e que não estava presente na requisições iniciais.

O conjunto de dados foi criado em 18 de abril de 2021 no formato GraphML, contendo uma rede com 5.884 nós (artigos retratados e os que os referenciavam) e 6.623 arestas (citações). Entre elas, as que representavam citações altamente influentes receberam peso 2 e as demais, peso 1. Além disso, sobre os nós e arestas, são armazenadas as informações apresentadas no Quadro 1.

Com o objetivo de que o código para a construção do conjunto de dados, feito em linguagem Python, e o *dataset* gerado possam ser aperfeiçoados e utilizados em outras pesquisas, eles foram desenvolvidos no formato *open source* e encontram-se disponíveis no GitHub¹.

5. ANÁLISES E RESULTADOS

Após a geração do conjunto de dados, foi utilizada a ferramenta Gephi para realizar as análises referentes às perguntas de pesquisa previamente estabelecidas (ver seção 1).

Por meio delas e com as métricas e filtragens disponíveis na ferramenta, descobertas relevantes sobre os artigos retratados puderam ser vislumbradas.

¹ <https://github.com/ingridpacheco/RetractedArticlesDatasetGenerator>

Quadro 1. Informações sobre nós e arestas da rede.

Atributo	Para	Descrição
id	nó	DOI do artigo, ou identificador único
retractionDate	nó	Data de retratação, no formato yyyy-mm-dd
publicationDate	nó	Data de publicação do artigo, no formato yyyy-mm-dd
retracted	nó	Flag que identifica se um artigo é retratado
name	nó	Título do artigo
weight	aresta	Peso da aresta (se for artigo influente, aresta tem peso maior)
afterRetraction	aresta	Flag para identificar que foi uma citação feita após a retratação
influential	aresta	Flag para identificar uma citação influente

5.1. CARACTERIZAÇÃO DA REDE DE CITAÇÕES

O estudo da estrutura da rede é uma etapa importante para o entendimento do tipo de análise que pode ser realizada e, neste caso, possibilita uma visão geral da problemática da citação de artigos retratados. A rede possui algumas sub-redes egocentradas (*ego-centered network*), em que o nó central (*ego*) de cada sub-rede é um artigo retratado e o conjunto de nós que o cercam (*alter*) são artigos que fazem referência ao ego. Esse tipo de rede pode ser usada em diversas aplicações, como a análise de transmissão de doenças, estudos de suporte social e redes de discussão (WASSERMAN; FAUST, 1994).

Outro aspecto interessante é que, dos 5.884 nós da rede, 73 representam artigos que sofreram retratação e os outros 5.811 nós restantes são os artigos que os citam. Com relação às arestas, a rede é formada por 6.623, das quais 423 têm o atributo *influential* igual a 1, o que significa que pelo menos 6,39% dos artigos dessa rede foram fortemente influenciados pelos artigos retratados. Além disso, o número médio de citações (arestas) por artigo retratado (nó) é de 90,72.

Por fim, as características principais da rede são: grau médio de 1,126 e coeficiente de clusterização igual a 0,18 (esses valores baixos são esperados, uma vez que essa rede inclui apenas as citações a artigos retratados). Os números de componentes fracamente e fortemente conectados são, respectivamente, 39 e 5.884. Na Figura 1, pode-se ver uma representação do grafo da rede completa usando o algoritmo de visualização Fruchterman-Reingold. Cada nó da rede representa um artigo, estando os retratados em vermelho e o

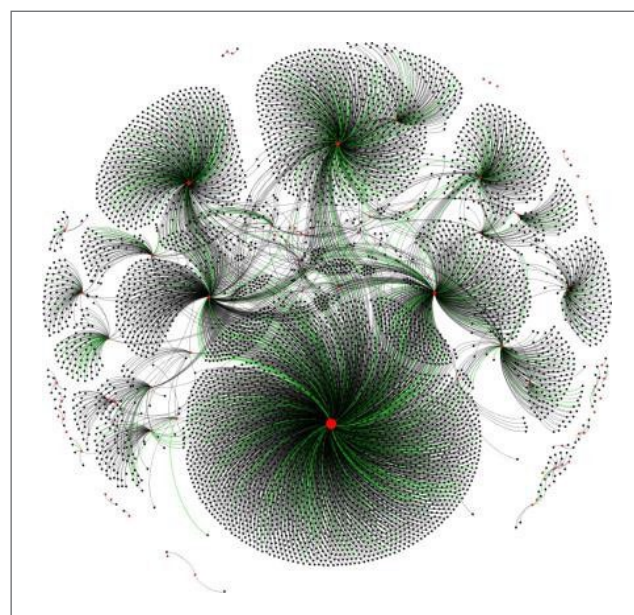


Figura 1. Rede de Citações de Artigos Retratados sobre COVID-19.

tamanho de cada nó sendo diretamente proporcional ao seu grau (centralidade local). Por fim, as arestas representam as referências aos artigos retratados, estando em verde as arestas que ligam os artigos que foram fortemente influenciados ao artigo retratado (ver seção 4).

5.2. ARTIGOS RETRATADOS MAIS REFERENCIADOS

Apesar de todos os artigos retratados terem grande impacto sobre a validade dos que os referenciam, compreender os mais citados serve para dar uma visão mais concreta e

estreita sobre os principais temas que causam retratação no escopo da temática da COVID-19, além de elucidar o seu impacto e a forma como a rede está disposta.

Levando em consideração esse contexto, e para responder à P1, uma métrica que se mostrou interessante para esta análise foi a de centralidade local, com a finalidade de encontrar os artigos retratados que fossem mais referenciados. Ao aplicar tal métrica sobre a rede social, o grafo da Figura 2a foi gerado. O nó em destaque, na cor vermelha, representa o artigo com maior valor de centralidade local, correspondente ao artigo retratado mais citado (referenciado por 2.745 artigos), de título “*Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial*” e DOI 10.1016/j.ijantimicag.2020.105949. Dado que o segundo artigo retratado com maior centralidade é citado por 799 artigos, esse valor passa a ser ainda mais significativo, expressando o quanto o artigo mais referenciado é influente dentro da rede.

Estendendo as análises, os cinco artigos retratados mais referenciados foram filtrados, ainda utilizando a métrica de centralidade local, a fim de tentar encontrar similaridades entre eles. A Figura 2b representa os cinco principais artigos retratados na cor amarela e suas citações influentes (que possuem o atributo *influential* na aresta de ligação igual a 1) em diferentes cores, dependendo do artigo que referenciam.

O Quadro 2 apresenta informações sobre os cinco artigos retratados mais citados.

Salienta-se que a retratação do primeiro artigo ocorreu por problemas em seus resultados, o que possivelmente prejudicou os 150 artigos que foram fortemente influenciados por ele, conseqüentemente tendo suas validações questionadas. O segundo foi publicado no *National Science Review* e teve como principal motivo de retratação erro no texto, o que também pode acabar afetando a qualidade dos artigos que o referenciam. O terceiro foi publicado no *The New England Journal of Medicine*, e os motivos de sua retratação se expandem, não só havendo incertezas nos resultados como também nos dados utilizados, mostrando os problemas que referenciá-lo pode causar. O quarto, publicado no *Engineering*, não teve o seu motivo de retratação muito claro, havendo apenas a indicação de que o artigo foi retirado, sem maiores informações sobre o que o levou a isso. Por fim, o quinto artigo, publicado no *The Lancet*, é citado por outro retratado, 10.1016/S0140-6736(20)31174-0, e ainda é um dos sete retratados que cita outros na mesma condição. Aliás, entre eles, é o único que cita mais de um retratado, quais sejam, os artigos 10.1016/j.ijantimicag.2020.105949 e 10.1056/NEJMoa2007621. Diante desse fato, o artigo cria uma ponte entre dois trabalhos retratados muito relevantes na rede e fica com o maior valor de intermediação (711.833), como visto

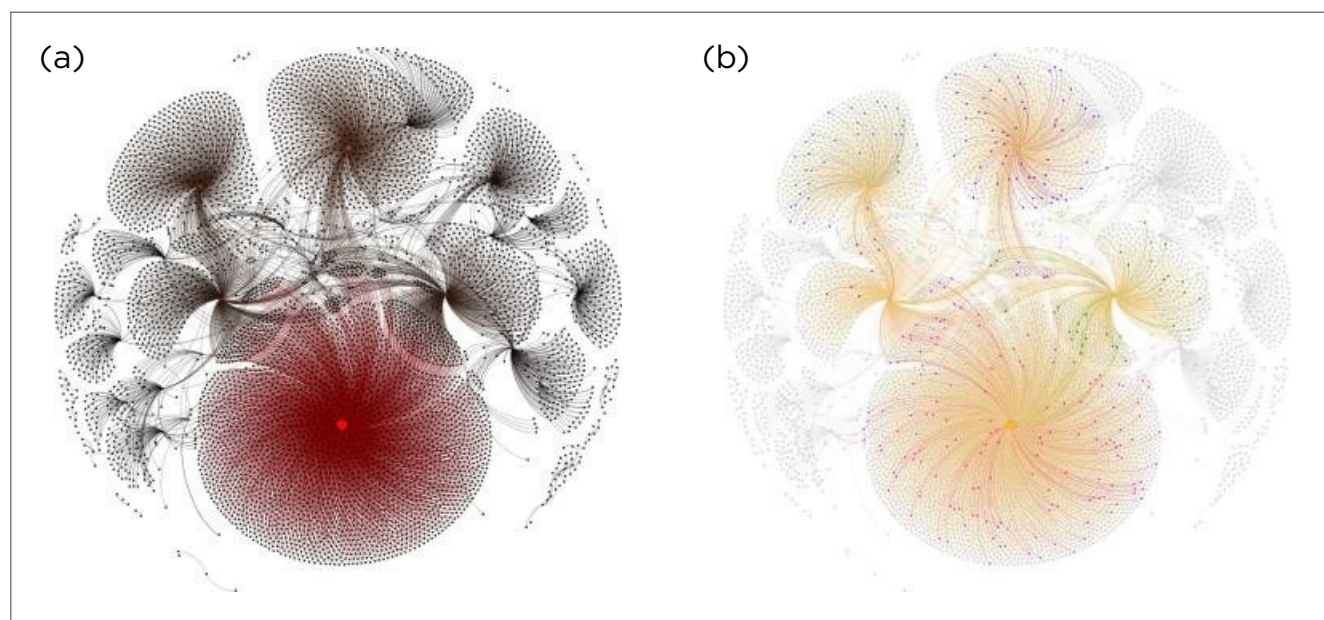


Figura 2. Rede: (a) com a métrica de centralidade; (b) de retratados com seus nós influentes.

na Figura 3a. Ademais, esse artigo e suas principais conexões são destacados na Figura 3b.

Passando para a análise da totalidade da rede, um dos primeiros pontos que se podem notar é um tema em comum que aparece nos títulos de 75% deles: *tratamento para a COVID-19*. A fim de responder à P2, foram feitas análises

sobre a incidência de termos nos títulos dos artigos usando a ferramenta *WordArt.com* e considerando: (a) todos os artigos da rede; e (b) apenas do subgrupo dos 73 artigos retratados. Foram executadas as seguintes etapas de pré-processamento: remoção de *stopwords*, remoção de números e *stemming* (BAEZA-YATES; RIBEIRO-NETO, 2013).

Quadro 2. Informações sobre citações dos primeiros 5 artigos retratados.

Título/DOI	Citações	Citações de Influentes
Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial 10.1016/j.ijantimicag.2020.105949	2.745	150
On the origin and continuing evolution of SARS-CoV-2 10.1093/nsr/nwaa036	799	68
Cardiovascular Disease, Drug Therapy, and Mortality in COVID-19 10.1056/NEJMoa2007621	627	36
Experimental Treatment with Favipiravir for COVID-19: An Open-Label Control Study 10.1016/j.eng.2020.03.007	508	33
RETRACTED: Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis 10.1016/S0140-6736(20)31180-6	471	8

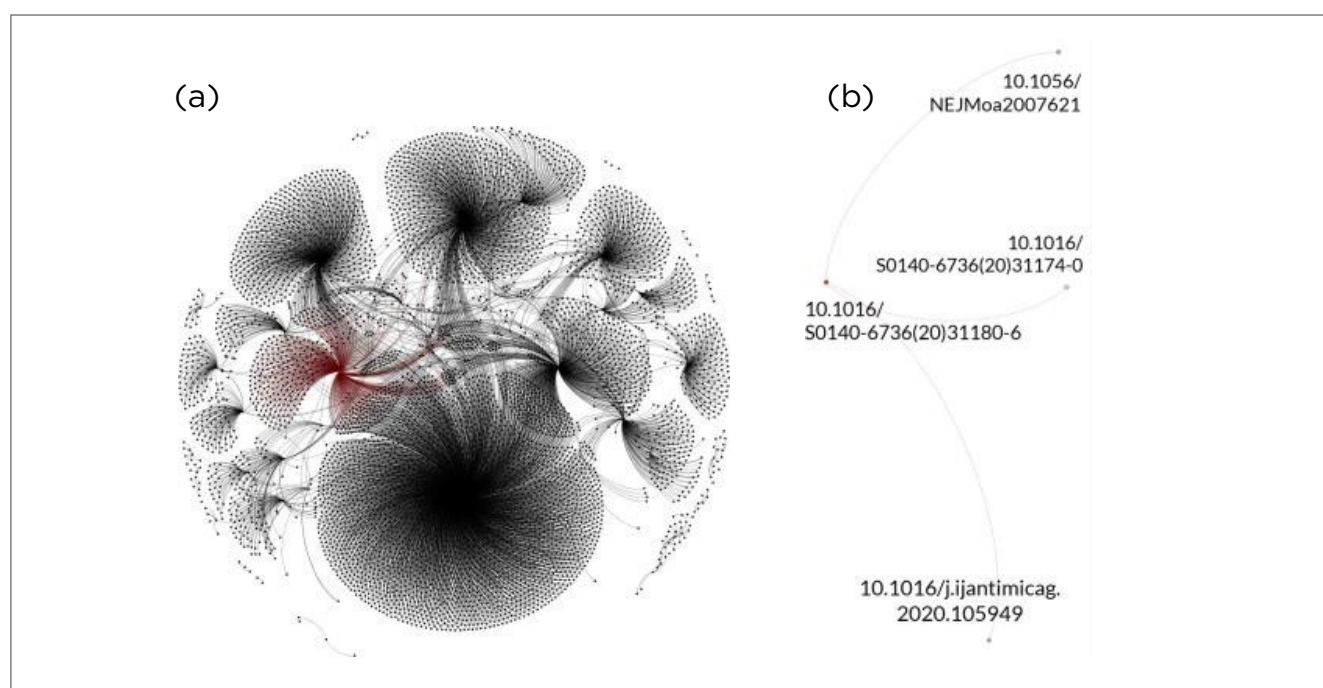


Figura 3. Rede: (a) destacando métrica de intermediação; (b) ego centrada no nó com maior valor de intermediação.

Além disso, foram removidos os termos utilizados para a coleta dos dados no *Retraction Watch Database* (ver seção 4). Os resultados dessas análises estão sumarizados no Quadro 3. Entre as palavras-chave mais frequentes nos dois grupos, citam-se *hydroxychloroquine*, *treatment* e *drug*. Esses termos destacam-se em razão do contexto atual da pandemia da COVID-19, momento em que são divulgados estudos relacionados ao tratamento da doença.

Outra análise sobre a rede de citações de artigos retratados foi conduzida com o intuito de considerar o aspecto temporal das citações, verificando a quantidade de referências feitas antes e depois dos avisos de retratação. O Quadro 4 apresenta os quantitativos de citações antes e depois das retratações para

os cinco artigos mais citados na rede. Já a Figura 4 ilustra a rede, destacando em amarelo as citações efetuadas (a) antes e (b) depois das retratações. Como demonstrado, para os cinco artigos mais citados, o número de citações após a retratação foi expressivamente maior do que as citações anteriores ao aviso de retratação. Desta forma, 82,11% de todas as citações da rede aconteceram depois, e apenas 17,89% ocorreram antes, respondendo assim à P3.

5.3. ARTIGOS QUE MAIS REFERENCIAM RETRATADOS

Em resposta à P4, há sete retratados que referenciam artigos na mesma situação. Além disso, foi feita uma análise

Quadro 3. Comparativo entre palavras-chave de todos os artigos da rede versus palavras-chave dos artigos retratados.

Todos os Artigos			Artigos Retratados	
Termo	Peso	x	Termo	Peso
Patient	893	1°	Patient	20
Review	717	2°	Pandemic	9
Pandemic	663	3°	Infect	8
Treatment	629	4°	Analysis	7
Hydroxychloroquine	601	5°	Clinic	6
Infect	576	6°	Effect	6
Clinic	524	7°	Treatment	6
Drug	438	8°	Hospital	5
Analysis	416	9°	Hydroxychloroquine	5
Study	380	10°	Mortality	5

Quadro 4. Citações antes e depois da retratação para os primeiros 5 artigos retratados.

Título	Data Retratação	Antes	Após
Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial	11/04/2020	260	2485
On the origin and continuing evolution of SARS-CoV-2	19/06/2020	49	750
Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19	04/06/2020	117	510
Experimental Treatment with Favipiravir for COVID-19: An Open-Label Control Study	01/04/2020	30	478
RETRACTED: Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis	04/06/2020	36	435

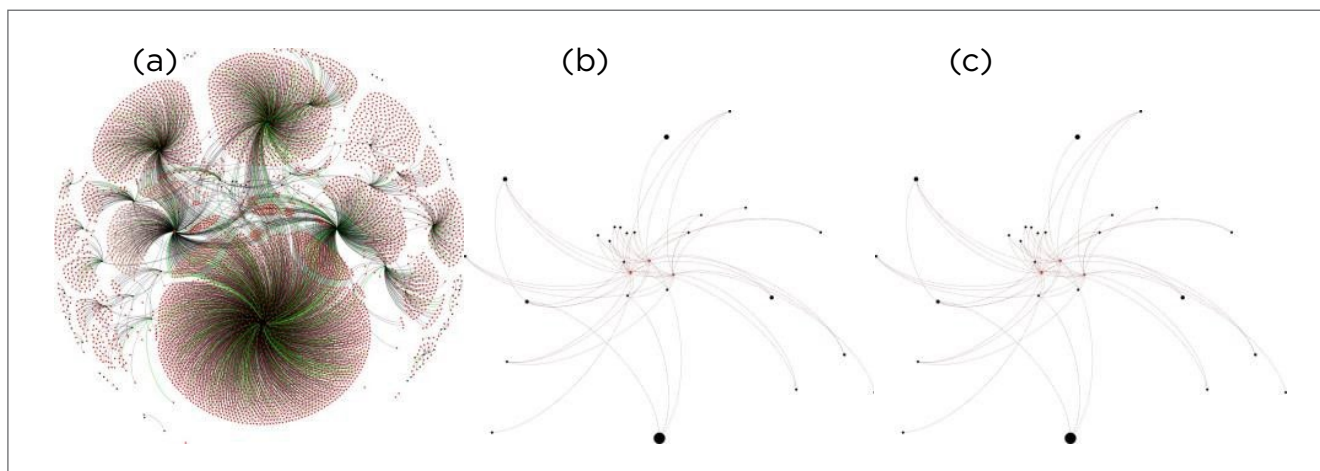


Figura 4. Rede destacando: citações feitas (a) antes e (b) depois das retratações; (c) artigos que mais referenciam retratados

para encontrar quais eram os artigos que mais citavam retratados. Ao analisar a rede procurando por esses nós, filtrando os que referenciavam mais de um retratado, foi possível encontrar 648 artigos (já considerando o quinto artigo retratado mais citado mencionado na seção anterior), dessa maneira respondendo à P5. Todavia, os três nós que mais tiveram conexões, destacados em vermelho, mostrados na Figura 4c, possuem um diferencial em relação aos outros por conta da quantidade de retratados que citam. Enquanto o quarto lugar referencia apenas cinco artigos, os três principais (0.1136/medethics-2020-106494, 10.1007/s11192-020-03661-9 e 10.1016/j.pulmo.2020.10.011) referenciam uma quantidade significativa de artigos retratados, respectivamente, 19, 17 e 16.

O grande número chama a atenção para o tema em comum de que eles tratam, pois todos abordam o assunto de artigos retratados sobre COVID-19 como forma de crítica à velocidade com que as publicações são realizadas, em vez de priorizar a qualidade e a busca por informações corretas, portanto fazendo as referências propositalmente para trazer exemplos que elucidem o problema.

Entre eles, entretanto, vale mencionar que o artigo de DOI 10.1007/s11192-02003661-9, que possui o título de “*Retracted COVID-19 articles: a side-effect of the hot race to publication*” (SOLTANI; PATINI, 2020), é o único que cita todos os cinco retratados mencionados antes, demonstrando o quão relevante ele é para o tema e para a rede social utilizada.

6. CONCLUSÕES

Com o avanço das descobertas científicas, o compartilhamento de todos os resultados encontrados no menor tempo possível é essencial. Entretanto, essa pressa por publicações pode implicar diversos problemas com relação a qualidade do artigo, o levando-o à retratação. Assim, mecanismos que analisem os artigos retratados ou auxiliem na hora de referenciá-los mostram-se muito importantes, principalmente diante do conhecimento adquirido por este trabalho.

Mediante o exposto, é possível afirmar que por meio da ARS foi possível descobrir os artigos retratados mais relevantes no contexto da COVID-19. Usando as métricas de centralidade, este trabalho mensurou a influência dos trabalhos retratados em artigos válidos ou não. Ademais, foi possível identificar pesquisas que sofreram retratações e faziam referência a artigos na mesma situação. Por fim, também foi verificado que, nesta rede, a maior parte das citações foi feita após o aviso de retratação.

Considerando o fato, este trabalho demonstra a importância de maior discussão sobre o tema, que hoje não é tão difundido como deveria. Além disso, foi detectado que as palavras mais utilizadas nos títulos destes trabalhos se referiam a possíveis tratamentos e medicamentos relacionados à doença, deixando em aberto a possibilidade de que essas publicações influenciem de alguma forma a disseminação das notícias falsas.

No entanto, durante o desenvolvimento deste trabalho, alguns desafios foram encontrados. O principal foi a falta de um

conjunto de dados único com informações sobre as retratações e citações ou referências feitas por esses artigos. Para mitigar esse desafio, foi necessário o desenvolvimento de um *script* em Python que, coletando dados de três bases distintas, criasse a rede de citações de artigos retratados a ser utilizada nas análises.

Como trabalhos futuros, pretende-se ampliar essas análises em uma rede completa sobre a COVID-19, também com as conexões entre artigos não retratados, diferentemente da usada neste trabalho (*ego-centered*), que foi gerada com base

nos retratados. Dessa forma, seria possível verificar mais amplamente quão influentes realmente são os artigos retratados. Além disso, pretendemos realizar a automatização de parte das análises a fim de evitar possível subjetividade nos resultados. Finalmente, outra abordagem interessante seria o cruzamento da rede de retratações com uma rede de *fake news* associadas à pandemia da COVID-19, a fim de encontrar possíveis conexões entre as notícias falsas e as publicações retratadas (como prováveis fontes delas).

REFERÊNCIAS

BAEZA-YATES, R.; RIBEIRO-NETO, B. *Recuperação de informação: conceitos e tecnologia das máquinas de busca*. 2ª ed. Porto Alegre: Bookman, 2013.

BOSCHIEROA, M. N.; CARVALHO, T. A.; LIMA MARSON, F. A. Retraction in the era of COVID-19 and its influence on evidence-based medicine: is science in jeopardy? *Pulmonology*, v. 27, n. 2, p. 97-106, 2021. <https://doi.org/10.1016/j.pulmoe.2020.10.011>

BROWMAN, H.; ALEXANDER, J.; FENNELL, C.; HODGKINSON, M.; TIERNEY, H. *Guidelines for retracting articles*. Technical report. Committee on Publication Ethics, 2019.

CHERVEN, K. *Mastering Gephi network visualization: produce advanced network graphs in Gephi and gain valuable insights into your network datasets*. Community experience distilled. Packt Publishing, 2015.

CORTEGIANI, A.; CATALISANO, G.; IPPOLITO, M.; GIARRATANO, A.; ABSALOM, A. R.; EINAV, S. Retracted papers on sars-cov-2 and covid-19. *British Journal of Anaesthesia*, v. 126, n. 4, p. e155-e156, 2021. <https://doi.org/10.1016/j.bja.2021.01.008>

ELSE, H. How a torrent of covid science changed research publishing in seven charts. *Nature*, v. 588, n. 7839, p. 553, 2020. <https://doi.org/10.1038/d41586-020-03564-y>

FRUCHTERMAN, T. M. J.; REINGOLD, E. M. Graph drawing by force-directed placement. *Software: Practice and Experience*, v. 21, n. 11, p. 1129-1164, 1991. <https://doi.org/10.1002/spe.4380211102>

HATALA, J.-P. Social network analysis in human resource development: A new methodology. *Human Resource Development Review*, v. 5, n. 1, p. 45-71, 2006. <https://doi.org/10.1177/1534484305284318>

PILLER, C. Disgraced COVID-19 studies are still routinely cited. *Science*, v. 371, n. 6527, p. 331-332, 2021. <https://doi.org/10.1126/science.371.6527.331>

RIBEIRO, E. M. B. A.; BASTOS, A. A. V. A.-L. B. Redes sociais interorganizacionais na efetivação de projetos sociais. *Psicologia & Sociedade*, v. 23, n. 2, p. 282-292, 2011. <https://doi.org/10.1590/S0102-71822011000200009>

SHETH, B. P.; THAKER, V. S. Scientific retraction: a synonym for pseudoscience? *Acta Bioethica*, v. 20, n. 1, 2014.

SILVA, J. A. T. D.; BORNEMANN-CIMENTI, H. Why do some retracted papers continue to be cited? *Scientometrics*, v. 110, p. 365-370, 2017. <https://doi.org/10.1007/s11192-016-2178-9>

SOLTANI, P.; PATINI, R. Retracted COVID-19 articles: a side-effect of the hot race to publication. *Scientometrics*, v. 125, n. 1, p. 819-822, 2020. <https://doi.org/10.1007/s11192-020-03661-9>

WASSERMAN, S.; FAUST, K. *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press, 1994. v. 8.

ZDRAVKOVIC, M.; BERGER-ESTILITA, J.; ZDRAVKOVIC, B.; BERGER, D. Scientific quality of COVID-19 and sars cov-2 publications in the highest impact medical journals during the early phase of the pandemic: A case control study. *PLoS One*, v. 15, n. 11, 0241826, 2020. <https://doi.org/10.1371/journal.pone.0241826>