

Caracterização da reação de agências de fact-checking às publicações sobre a pandemia da COVID-19 em redes sociais

Cefas Garcia Pereira¹, Humberto Torres Marques-Neto¹

¹Programa de Pós-graduação em Informática (PPGInf)
Departamento de Ciência da Computação (DCC)
Instituto de Ciências Exatas e Informática (ICEI)
Pontifícia Universidade Católica de Minas Gerais (PUC Minas)
Caixa Postal 30535-901 – Belo Horizonte – MG – Brazil

cefas.pereira@sga.pucminas.br, humberto@pucminas.br

Abstract. *Since the beginning of the COVID-19 pandemic, it has been noticed, especially in social media, the generation of content about this subject. Social medias are an important tool of communication, however, they also create a space to misinformation spreading. This work aims to characterize how fact-checking agencies have been combating false information about COVID-19 that circulates on Twitter and Facebook. Fact-checking materials about COVID-19 written by specialized agencies from different countries were collected. Through the news, we searched for social media posts which the misinformation started to be spread. After collecting this material, it was verified how long it took for the agencies to analyze the veracity of the post and react to it. Besides, the news texts were also processed to detect whether the subjects being dealt with by the agencies are, in fact, the ones that have the greatest user engagement within the analyzed social networks. The results showed that the agencies' response time was, on average, 23 days in the case of false posts on Twitter and 6 days on Facebook.*

Resumo. *Desde o início da pandemia da COVID-19, percebe-se, especialmente em redes sociais online, a geração de muito conteúdo sobre esse assunto. As redes sociais são um meio importante de comunicação, porém, criam um espaço bem utilizado para disseminação de desinformação. Este trabalho se propõe a caracterizar como as agências de fact-checking têm reagido no combate às informações falsas sobre a COVID-19 que circulam no Twitter e no Facebook. Foram coletadas matérias de fact-checking sobre a COVID-19 escritas por agências especializadas de diferentes países. Através das notícias verificadas, buscou-se por postagens de mídias sociais em que a desinformação começou a ser disseminada. Após a coleta desse material, verificou-se em quanto tempo as agências analisam a veracidade da notícia e reagem a ela. Além disso, também realizou-se o processamento dos textos das notícias a fim de detectar se os assuntos que estão sendo tratados pelas agências são, de fato, os que possuem maior engajamento dos usuários dentro das redes sociais analisadas. Os resultados mostraram que o tempo de resposta das agências foi, em média, de 23 dias no caso de publicação da notícia falsa no Twitter e de 6 dias no Facebook.*

1. Introdução

O ano de 2020 foi marcado pela pandemia da COVID-19, a qual, em 2021, ainda abala o mundo inteiro. O vírus SARS-CoV-2, também chamado de Coronavírus, foi classificado como uma ameaça global grave pela OMS – Organização Mundial da Saúde [WHO 2020]. Após a declaração de março de 2020 da OMS, os países passaram a adotar medidas de prevenção ao vírus. Muitos deles começaram a implantar o isolamento social, através do bloqueio de suas fronteiras e restrição da circulação dos moradores. Os comércios fecharam em grande parte, permanecendo aberto apenas supermercados, farmácias, padarias e outros serviços essenciais. O mundo teve que se adaptar a esse novo formato de rotina e como consequência do isolamento social, as pessoas passaram a exercer suas atividades relacionadas à educação e trabalho de forma online, geralmente em suas residências.

A informação e relação social através do espaço virtual, que naturalmente já vem crescendo com os anos, se tornou quase que uma possibilidade exclusiva. As pessoas passaram a se informar e também a expressar suas opiniões, ainda mais através da internet, em especial em redes sociais online como Facebook, Twitter, Instagram e WhatsApp. De fato, não só as redes sociais, mas a internet como um todo, tem um poder enorme para contribuir em situações como essa, de isolamento social decorrente de uma pandemia. A velocidade na disseminação da informação, a simplicidade como ela é passada e também as diversas opções de forma e estilo textual faz com que as pessoas encontrem os meios em que ela se identifique mais e estejam muito mais engajadas e informadas. Diversos pesquisadores mostram que as redes sociais desempenham um papel importante como fonte de informação para as pessoas e que também auxiliam no entendimento de atitudes e comportamentos públicos durante eventos importantes [Shah et al. 2019, Jordan et al. 2019, Abd-Alrazaq et al. 2020].

Contudo, a presença da tecnologia na vida das pessoas traz consigo efeitos positivos e negativos, pois, ao mesmo tempo em que ela democratiza o acesso à informação, cria-se oportunidade para a manipulação política, comercial e a disseminação de notícias falsas [Sousa Jr et al. 2020]. Segundo a *Agência Lupa*, o Brasil é o líder mundial no ranking de ocorrências de desinformação [Moraes 2020]. Na segunda posição, os Estados Unidos. O combate não é exclusivo ao vírus, mas também à desinformação das pessoas. Teorias conspiratórias podem não só anular o efeito do trabalho sério que é realizado no enfrentamento ao coronavírus, mas também piorar. No Brasil, foram registrados golpes no auxílio emergencial liberado pelo governo, notícias de que pacientes assintomáticos não transmitem o vírus causador da COVID-19, desaconselhamento ao isolamento social defendido pela OMS e muitas outras notícias falsas circularam através das mídias sociais [Goeking 2020, Amapa 2020, Afonso 2020]. Por isso, é importante identificar rapidamente os temas que estão sendo tratados de forma equivocada para compreender o interesse e a interação que é feita pelos usuários por esses assuntos.

Esta pesquisa busca compreender como as agências de *fact-checking* têm se portado diante desse cenário, especialmente através da identificação do tempo em que a agência verifica a informação e como isto pode impactar o engajamento dos usuários. Entender essas questões é importante, pois, a desinformação tem o poder de agravar os efeitos de eventos como o surto da COVID-19. Para isso, foi realizada uma coleta de artigos de *fact-checking* na plataforma IFCN – *The International Fact-Checking Network*, desenvolvida

pelo instituto de jornalismo, sem fins lucrativos, *Poynter*. Posteriormente, foi realizado um mapeamento no corpo dos artigos do *Poynter* para encontrar todos os *hyperlinks* que fossem provenientes do Facebook ou do Twitter. Assim, foi possível chegar na origem das *fake news* nestas redes sociais, ou seja, em qual publicação do Facebook e/ou Twitter a desinformação começou a ser disseminada. Com estas publicações em mãos, foram extraídos seus metadados e então realizadas análises do número de publicações de *fake news* ao longo do tempo, quais publicações tiveram mais engajamento do público, quais tópicos foram mais discutidos e em quais países as notícias circularam.

Este trabalho, portanto, contribui na proposta de uma metodologia para entender o impacto do *fact-checking* sob as *fake news* que circulam em redes sociais para um determinado assunto. O trabalho também contribui com os resultados gerados através das análises, ao demonstrar que as agências de *fact-checking* levam, em média, 6 dias para verificar a informação no Facebook e 23 dias no Twitter. Geograficamente, a Índia, os Estados Unidos, Brasil e a Espanha foram detectadas como áreas quentes no que diz respeito à disseminação de desinformação sobre a pandemia. A pesquisa ainda mostrou que tópicos como cura, vacina e tratamento são comuns nas duas redes sociais analisadas e que, felizmente vêm sendo combatidos pelas agências de *fact-checking*, pois, esses mesmos assuntos foram identificados nas matérias escritas pelas agências.

O restante deste artigo está organizado como segue. A Seção 2 traz os trabalhos relacionados à desinformação, abordando os conceitos principais no qual essa pesquisa se baseia. Na Seção 3 é discutido o processo da coleta e pré-processamento dos dados utilizados para as análises. Na Seção 4 são apresentadas as análises e os resultados extraídos. As limitações são pontuadas na Seção 5; a conclusão é apresentada na Seção 6.

2. Trabalhos Relacionados

O termo “desinformação” é designado como informações que não refletem com precisão o verdadeiro estado da notícia [Lewandowsky et al. 2012]. Aplica-se o termo “desinformação” às informações que inicialmente são apresentadas como verdadeiras e que, posteriormente, são mostradas como falsas. As causas e efeitos da desinformação ainda são materiais de estudo de diversos pesquisadores. Embora a desinformação não seja um fenômeno recente [Tandoc Jr et al. 2018], os meios digitais de comunicação, por sua capacidade de transmitir uma grande quantidade de dados rapidamente, tem trazido essa questão à tona.

Os motivos da propagação de desinformação são diversos. Pesquisa [Larson 2018] alega que uma categoria de desinformação é originada da comunidade científica, nomeada “ciência ruim”. A autora exemplifica sua colocação através da publicação de 1998, realizada pelo então médico Andrew Wakefield, que pretendia mostrar uma ligação entre o autismo e a vacina contra sarampo, caxumba e rubéola. Apesar de ter sua licença revogada e seu trabalho retirado, Wakefield insiste nas suas campanhas antivacina. Larson defende que se Wakefield tivesse sido disciplinado e seu artigo retirado 12 meses após a publicação, ao invés de 12 anos, muitos problemas relacionados à saúde e desconfiança da vacina poderiam ter sido evitados.

Trabalho realizado [Broniatowski et al. 2018] revelou o uso de *bots* (robôs) para disseminar discurso de ódio associado a movimentos antivacina com o objetivo de segregar grupos políticos na Rússia, o que leva a outra categoria de desinformação, a ca-

tegoria movida por interesses políticos. Há também casos em que a desinformação é resultado de uma notícia que apresenta certa veracidade nos fatos, porém é distorcida em decorrência do exagero, sequência coincidente de fatos, mal entendido ou senso comum [Larson 2018].

Quando se trata de redes sociais, outro trabalho [Castillo et al. 2011] revelou que os usuários do Twitter são extremamente suscetíveis a serem enganados por desinformação compartilhada na rede, sobretudo os usuários que são inexperientes. As desinformações compartilhadas por figuras públicas, representam um grande problema, pois, elas geram um engajamento muito alto [Brennen et al. 2020].

É imprescindível, portanto, que formas de identificar e combater a desinformação sejam desenvolvidas. É difícil que uma solução permanente venha a existir, mas algumas medidas podem ser tomadas para frear o crescimento deste fenômeno. Algumas redes sociais, já passaram a adotar medidas de moderação baseadas em denúncia ou verificação de conteúdo por especialistas. Em março de 2020, o presidente Jair Bolsonaro teve algumas postagens retiradas de suas redes sociais por irem contra as políticas de conteúdo dessas empresas, já que continham desinformação sobre o coronavírus [Brennen et al. 2020].

Uma das ferramentas utilizadas para combater informações falsas é o *fact-checking*. A sua grande missão é elevar o custo da mentira, ou seja, não deixar que o processo de espalhar desinformação seja tão trivial. Ainda há uma dificuldade grande de se realizar a identificação automática de desinformação ou *fake news*. Muitos estudos estão sendo realizados a fim de resolver este desafio, mas, até o momento, não há um consenso de qual a melhor forma para realizar a detecção de desinformação. O *fact-checking*, geralmente, é realizado por jornalistas que buscam, através de novas fontes, pesquisas, entrevistas e outros recursos do jornalismo, verificar a veracidade da informação [Vlachos and Riedel 2014].

No Brasil, a *Agência Lupa*, é a primeira organização especializada em realizar o *Fact-Checking*. Nos Estados Unidos, há a *Politifact*, e na Argentina a agência *Chequeado*, que são outros exemplos de organizações que realizam a detecção de desinformação nos seus respectivos países. A metodologia adotada pela agência, é uma decisão individual, a qual os jornalistas priorizam os assuntos pelos critérios que julgarem mais relevantes, bem como os meios em que a informação circula e sua repercussão. Segundo a *Agência Lupa*, no *fact-checking* não se leva em consideração opiniões, não são feitas previsões e tampouco são apontadas tendências. A agência deve reservar seus esforços a validar a veracidade dos fatos. Após a conclusão do *fact-checking*, o jornalista escreve uma matéria esclarecendo os fatos e que justifique a classificação (falsa, verdadeira, exagerada etc.) da informação verificada.

O papel do *fact-checking* vem ganhando cada vez mais importância, visto que no contexto da comunicação digital, avaliar a qualidade do discurso público pode elevar a qualidade da informação que circula nas redes e oferecer dados mais precisos à sociedade. Este trabalho, optou por utilizar a base de dados do *Poynter* como ponto de partida para identificar as desinformações, já que o intuito do trabalho não é propor um sistema automático de predição de desinformação, mas, sim compreender a interação dos usuários com a desinformação nas redes sociais e como o *fact-checking* pode impactá-las. Diferentemente de estudos anteriores que analisaram o discurso público durante a pandemia,

este trabalho teve o foco em descobertas direcionadas à disseminação de informação falsa e às atividades das agências de *fact-checking*.

3. Coleta dos Dados

A coleta dos dados se dividiu, basicamente, em duas partes, ambas apresentadas a seguir. A primeira foi a coleta dos artigos de *fact-checking* e a segunda foi a coleta das postagens do Facebook e do Twitter relacionados aos artigos da primeira coleta.

3.1. Fact-Checking

O instituto *Poynter*, fundado por Nelson Poynter em 1975, têm tido um papel importante durante o período de pandemia. Em 2015, o instituto abriu sua unidade IFCN com a intenção de fomentar e disponibilizar o *fact-checking* no mundo. Em 2020, a unidade disponibilizou uma base de dados online com *fact-checking* sobre o Coronavírus proveniente de vários países [Poynter 2020]. Através de uma interface web, qualquer usuário pode explorar a base de dados e realizar a verificação de informações sobre a pandemia. O acesso a este tipo de informação é extremamente importante, pois, dá autonomia ao leitor para verificar os fatos de forma fácil e rápida.

Essa base de dados funciona como um centralizador do *fact-checking* que tem sido realizado ao redor do mundo, ela foi utilizada pelos autores para a coleta das matérias escritas pelas agências. Os autores desenvolveram um *crawler* utilizando a linguagem de programação *Python*¹. Os *crawlers* são programas que simulam o acesso realizado por humanos à páginas web. Inicialmente, todo o universo de artigos acessível foi coletado. A coleta na base do *Poynter* retornou 11.666 publicações de *fact-checking*. As verificações foram realizadas por 99 agências diferentes e as informações verificadas circularam por 130 países. 37 rótulos (*labels*) de classificação da informação foram identificados, entre eles: verdadeiro, falso, *misleading*, não comprovado, imprecisa etc. A Tabela 1 apresenta o formato em que o conjunto de dados foi coletado.

Poynter	
Dado	Descrição
id	Identificador exclusivo da matéria
title	Título do artigo no <i>Poynter</i>
checked_link	Link da matéria escrita pela agência original
countries	Países que a informação circulou até ser verificada
date	Data de publicação do <i>fact-checking</i>
poynter_link	Link da matéria na plataforma <i>Poynter</i>
source	Agências que verificaram a informação
label	Classificação dada à informação pelas agências (falsa, verdadeira, imprecisa etc.)
justify	Explicação curta do motivo pelo qual a informação recebeu tal classificação

Tabela 1. Estrutura de dados extraída do *Poynter*

3.2. Redes Sociais: Facebook e Twitter

Posteriormente à primeira coleta, foi realizada a coleta das postagens em redes sociais. Para isso, a estratégia adotada, foi de verificar no corpo da matéria do *fact-checking* a presença de postagem nas redes sociais Facebook e Twitter, escolhidas por causa da alta popularidade em diferentes países, especialmente no conteúdo do *Poynter*. O *crawler*

¹<https://github.com/cefasgarciapereira/fact-checking>

percorreu todos os *hyperlinks* da matéria e filtrou aqueles provenientes do Facebook e do Twitter, portanto uma matéria pode ter mais de uma postagem associado a ela.

Os *hiperlinks* provenientes das redes sociais podem vir em diversos formatos, por isso, *scripts* com manipulação de *strings* e expressões regulares foram executados para conseguir extrair os *IDs* das publicações. Ainda, foi identificado que algumas agências hospedam os *links* das redes sociais no *archive.today*, que é uma espécie de servidor que garante que o conteúdo continuará online mesmo que ele seja excluído na fonte original. A coleta também considerou os *links* contidos no *archive.today* e os recuperou para garantir o maior número possível de material.

Posteriormente, os dados passaram por uma filtragem. As *labels* dos artigos foram agrupadas para que apenas informações falsas fossem analisadas, visto que o foco aqui é a desinformação. Portanto, as *labels* colocadas à frente foram todas convertidas para “False”: *False, False and Without Context, false context, false headline, Falso, MISLEADING/FALSE, Incorrect, Mainly False, Mostly False, NOT LEGIT, misleading*.

Após este agrupamento, as matérias foram filtradas pelo rótulo “False”, restando 10.170 (87,17%) da coleta inicial. Foi possível identificar quantos artigos falsos do *Poynter* contêm os *links* para redes sociais. Das 10.170 matérias filtradas, identificou-se que 42% referenciam alguma postagem de rede social analisada neste trabalho (Twitter 21%, Facebook 16% e 5% em ambas). Assim, podemos dizer que algumas agências não inserem o link de redirecionamento para a postagem no Facebook e/ou no Twitter ou que realmente 58% das notícias não circularam nestas redes sociais, sendo veiculadas exclusivamente em outras redes sociais fechadas como o WhatsApp, na TV, mídia impressa, rádio, blogs ou até mesmo em outro meio de comunicação.

Também foi identificado que algumas notícias colocaram apenas imagens (*screenshots*) da publicação. Como uma tentativa de aproveitar esse conteúdo, os autores usaram um OCR – *Optical Character Recognition* para tentar reconhecer textos em imagens. Todavia, os resultados não foram promissores. O código escrito em *Python*, fez uso das bibliotecas *open-source Tesseract, Microsoft Computer Vision API* e *Google Vision*. Nenhuma das alternativas foi capaz de extrair com precisão o conteúdo da imagem devido à baixa qualidade e falta de padrão das imagens. Assim, deixamos essa abordagem para ser explorada em um trabalho futuro.

Foi decidido, portanto, seguir com a análise dos 42% de notícias coletados pelo *crawler* no *Poynter*. O *dataset* com o conteúdo mostrado na Tabela 1 foi acrescido dos *IDs* das postagens das redes sociais e submetidos a outra coleta, responsável por extrair os *metadados* das postagens. Nesta fase, foi utilizada a biblioteca *open source Tweepy* que é uma abstração da API oficial do Twitter para Python. Para o Facebook, a coleta foi implementada pelos próprios autores, também através de *crawlers*. A coleta resultou em um total de 2.908 postagens no Facebook e 4.096 no Twitter. É importante ressaltar que estas postagens são aquelas que ainda permanecem online, pois, aquelas que foram banidas ou excluídas, não podem ser mais capturadas. Inclusive, foi possível detectar que 931 postagens (32%) do Facebook estão marcadas com um alerta de informação falsa, ainda que o conteúdo esteja disponível. Esta medida já mostra que há um combate por parte da própria rede social a este tipo de conteúdo. No Twitter esta análise não foi possível, pois a biblioteca não retorna esta informação.

3.3. Pré-Processamento

No momento da coleta, algumas técnicas já haviam sido aplicadas para que os dados fossem extraídos e estruturados. Entretanto, para que a análise fosse possível, foi necessário realizar alguns tratamentos nos dados coletados. Começando pelas datas, já que cada plataforma utiliza uma forma de representação, os dados foram submetidos a um pré-processamento escrito pelos autores, para padronizar as datas. O conteúdo de texto também passou por uma padronização. Aqueles que não pertencem a língua inglesa, foram traduzidos de forma programática pela API *Google Translate*. Posteriormente, foi realizada a etapa de limpeza, a qual caracteres especiais e *stopwords* foram removidos. Ainda, os textos foram divididos em palavras, processo denominado de *tokenização* e por fim passaram pelo passo de *stemming*, que se trata do processo de agrupar as formas flexionadas de uma palavra para que possam ser analisadas como uma única palavra.

4. Análises e Resultados

Com as análises dos *datasets* gerados foi possível ver o comportamento das publicações ao longo do período de pandemia. As publicações vão de janeiro de 2020 a janeiro de 2021, como mostra a Figura 1. É possível perceber que a curva ao longo dos meses é extremamente parecida para todas as plataformas. Nota-se que o período com a maior quantidade de publicações foi entre março e maio de 2020, ou seja, no início da pandemia da COVID-19. Um indicativo positivo em relação à reação das agências de *fact-checking*, é que o número de matérias também foi mais intenso no início da pandemia.

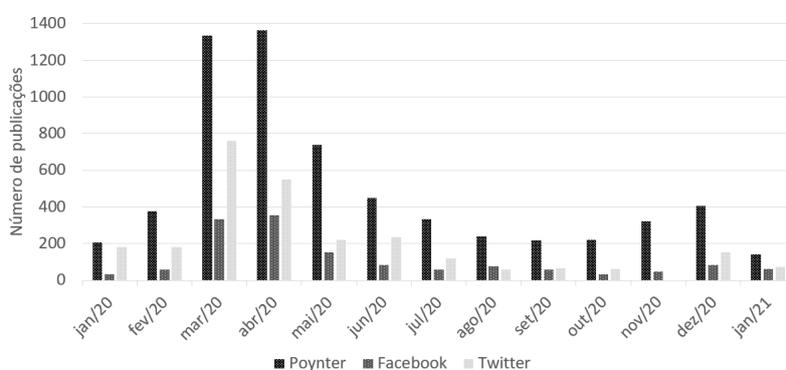


Figura 1. Número de publicações por mês em cada plataforma

A Figura 2 mostra que a média de compartilhamento, assim como o número de publicações da Figura 1, foi mais alto nos primeiros meses da pandemia mas segue um movimento de queda, como mostra a linha de tendência linear do gráfico. É importante ressaltar que no Facebook, a média de compartilhamento do mês de janeiro de 2021 foi o maior desde julho de 2020, o que pode indicar uma ascensão no número de publicações para o primeiro semestre de 2021.

Para saber quanto tempo uma agência levou para identificar a informação no Facebook e/ou no Twitter, verificá-la e publicar o *fact-checking*, foi calculado a diferença entre a data de publicação do *fact-checking* e a data de publicação das postagens nas redes sociais Facebook e/ou Twitter.

Os resultados apresentados na Figura 3 revelam que, em média, as agências levaram 23 (DP=33,6) dias para publicar a verificação da informação compartilhada no

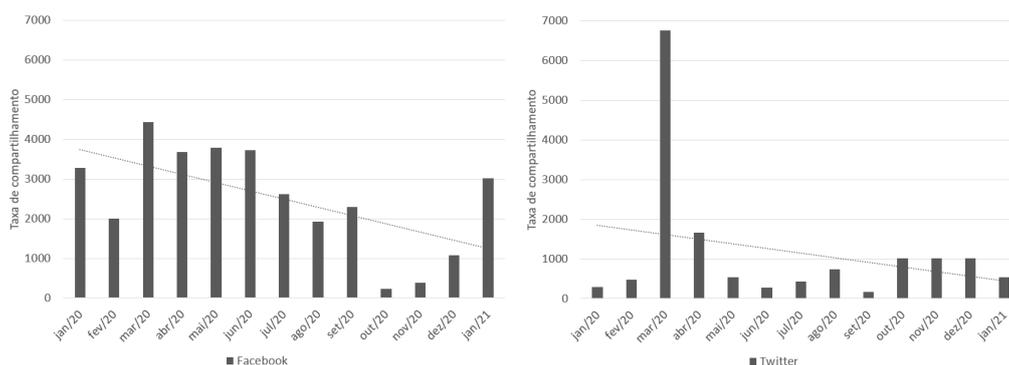


Figura 2. Média de compartilhamento por mês em cada rede social

Twitter e 6 (DP=7,04) dias para o Facebook. O alto valor do desvio padrão e a grande quantidade de *outliers* (pontos fora da curva) permite separar os casos em que a resposta da agência à postagem na rede social foi rápida e os casos em que a resposta demorou a ocorrer. Um tratamento nos valores foi aplicado para a remoção dos *outliers*. A técnica escolhida foi o método de Tukey, pois se mostrou recomendada em trabalhos passados [Bento and Santos 2018]. Entretanto, para entender o fenômeno dos *outliers* uma inspeção manual foi realizada pelos autores em uma parcela das postagens e foi identificado que alguns assuntos foram discutidos por um longo tempo, e portanto, as matérias de *fact-checking* revisitaram algumas postagens. Por exemplo, como a vacina é um assunto recorrente, postagens sobre o assunto “vacina” no início da pandemia foram referenciadas em *fact-checking* de 2021 e isso justifica o longo prazo, nesses casos.

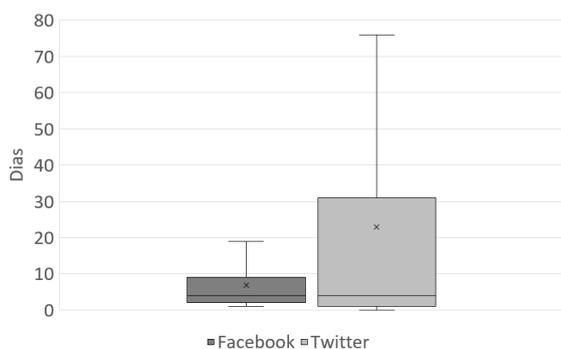


Figura 3. Diferença do número de dias entre a data de postagem na rede social e a data de publicação do fact-checking sem outliers

Outra situação que ocorreu, é quando algum usuário de rede social faz uma postagem de um conteúdo antigo, porém, manipulando algumas informações para trazê-lo ao contexto da pandemia. Para exemplificar, uma postagem de vídeo que foi compartilhada diversas vezes no Facebook e no Twitter, afirma mostrar uma procissão fúnebre simulada para “celebrar” o ministro do Interior indiano, Amit Shah, com teste positivo para o coronavírus. A afirmação, contudo, é falsa, pois o vídeo mostra, na realidade, uma procissão fúnebre simulada durante um protesto contra a lei de anticidadania no estado indiano de Bengala Ocidental em 2019 [AFP 2020]. Por este motivo a postagem de 2019 apareceu em um *fact-checking* de agosto de 2020, e resultou, portanto, em uma diferença grande de dias entre o *fact-checking* e a postagem nas mídias sociais.

Quanto à disposição geográfica, foi levado em conta quantas informações foram verificadas como falsas pelas agências de *fact-checking* em cada país. A Figura 4 mostra que a Índia, Brasil, Estados Unidos e Espanha são áreas quentes no número de circulação de informações falsas. É importante salientar que a Índia e os Estados Unidos são países que nativamente usam a língua inglesa para se comunicar e isso pode facilitar a verificação de informação pelas agências ao redor do mundo que podem não encontrar a mesma facilidade nos demais idiomas, e como consequência esses dois países se encontram no topo da lista, como mostra a Figura 4.



Figura 4. Áreas quentes na disseminação de desinformação

De qualquer forma, as agências costumam ter o cuidado de verificar a circulação da informação em outros países, como é o caso da informação verificada pelo Estadão, de um suposto vídeo em que uma mulher tremendo ao caminhar, seria em decorrência de efeitos colaterais da vacina da *Pfizer* [Lima 2021]. Entretanto, nenhuma evidência foi encontrada pelos especialistas que pudesse associar os dois fatos. Foi constatado que esse boato circulou não apenas no Brasil, mas também nos Estados Unidos; logo, o idioma não é um limitador, mas, indiretamente, pode favorecer um número maior de casos naqueles países que falam inglês.

Para extrair um conhecimento maior sobre os assuntos discutidos nas redes sociais e as informações verificadas pelas agências de *fact-checking*, os textos pré-processados foram submetidos a uma análise utilizando a frequência de palavras (unigrama) e a frequência de pares de palavras (bigramas). Para a visualização desta análise foram geradas nuvens de palavras como mostra a Figura 5. A nuvem de palavras consegue dar visibilidade às palavras que mais apareceram no conjunto de documentos e isso auxilia a identificação dos tópicos mais discutidos.

Para a identificação dos tópicos foi aplicado o LDA – *Latent Dirichlet Allocation* [Blei and Lafferty 2009], um algoritmo de modelagem de tópicos amplamente utilizado [Abd-Alrazaq et al. 2020, Cinelli et al. 2020]. A modelagem de tópicos é uma técnica de aprendizado de máquina não supervisionada que pode localizar grupos em uma coleção de documentos de texto. Foi utilizada a biblioteca *Gensim* para *Python* para a implementação do código responsável por processar os textos. O LDA recebe como parâmetro um conjunto fixo de tópicos. Cada tópico é representado por um conjunto de palavras. O objetivo do LDA é mapear os documentos fornecidos para o conjunto de tópicos de modo que as palavras em cada documento estejam associadas, principalmente, a esses tópicos definidos pelo algoritmo. Este trabalho utilizou a modelagem para encontrar grupos de tópicos no conteúdo de texto do Facebook, Twitter e matérias do *Poynter*. Para a modelagem é necessária a geração de um *corpus*, que se trata de uma representação das palavras e

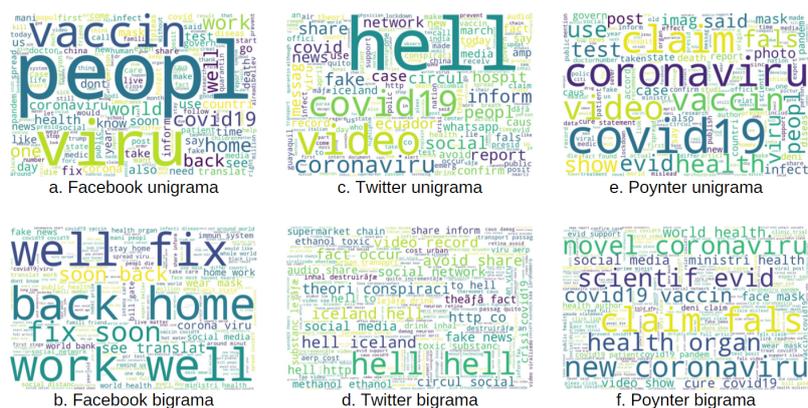


Figura 5. Nuvens de frequência de palavras e de pares de palavras por conjunto de documentos

suas frequências de ocorrência no conjunto de documentos. Um *corpus* foi gerado para o Facebook e outro para o Twitter, em ambos casos, o texto contido na publicação foi utilizado para este fim. No *Poynter*, o *corpus* é o texto de esclarecimento da agência, que se encontra no campo “*justify*” da Tabela 1.

Antes da aplicação da modelagem, foram executados testes que avaliam o nível de coerência entre as palavras e o seu respectivo tópico. Este teste foi aplicado para determinar qual número de tópicos é o mais coerente para cada conjunto de documentos. A medida de coerência do tópico pontua um único tópico medindo o grau de similaridade semântica entre as palavras com maior valor dentro do tópico. Um conjunto de afirmações ou fatos é considerado coerente, se apoiarem um ao outro. Assim, um conjunto de fatos coerente pode ser interpretado em um contexto em que ele cobre todos ou a maioria dos fatos [Röder et al. 2015].

Após os testes, o LDA foi executado com 6 tópicos para o Twitter, 7 para o Facebook e 6 para o *Poynter*. Esses números de tópicos como parâmetro foram escolhidos com base no teste de coerência e não levam os autores a uma decisão arbitrária. Posteriormente a análise manual das saídas do LDA associadas com as nuvens de palavras, chegou-se a um consenso sobre o conjunto de tópicos relevantes para cada coleção de documentos. Em seguida, já com a definição dos tópicos, foi executado um *script* de classificação baseado em regras escrito em *Python* para verificar a presença de qualquer uma das palavras que compõem determinado tópico pré-identificado em cada coleção de documentos. O *script* de classificação usou uma técnica de correspondência simples para ver se um determinado texto contém as palavras-chave extraídas dos tópicos. Um texto que contenha uma palavra-chave relacionada a um determinado tópico foi classificado como pertencente a esse tópico e, portanto, um texto pode ser classificado como pertencente a mais de um tópico simultaneamente.

Para determinar o nível de engajamento dos usuários para cada tópico, o número total de interações com a publicação (curtidas, comentários e compartilhamentos) foi somado e dividido pelo total de postagens pertencentes ao mesmo tópico. No caso das matérias do *Poynter*, como não há interação do usuário, foi considerado apenas o número de ocorrências, ou seja, quantas matérias pertencem a cada tópico. Assim, chegou-se nos valores contidos na Tabela 2, que mostram os tópicos identificados para cada contexto e o

nível de interação para cada tópico, que aqui foi denominado de “nível de engajamento”. É importante ressaltar que os valores contidos na coluna do nível de engajamento não são comparáveis entre conjunto de dados diferentes, ou seja, comparar o nível de engajamento de um tópico do Twitter com o Facebook, não faria sentido, já que se trata de outros tipos de interações, universo de usuários, tópicos etc. Este valor é importante para mostrar quais assuntos os usuários reagiram mais, dentro de cada rede social e no caso do *Poynter*, qual tópico teve maior número de publicações pelas agências.

Coleção de Documento	Tópico	Palavras	Nível de Engajamento
Twitter	Fake News	conspiraci, fake news, false, suppos	1.989.888
	Vacina Russa	russia, vaccin, president, putin	1.036.230
	Índia	vandana, gehlot, ashok, chouhan shivraj, delhi	1723
	Crise Humanitária	crise, migrant, hospital, water, crisi, infect, peopl, health, protest	2.251.289
Facebook	Vacina, Cura e Tratamento	vaccin, mask, test, viamin	7517,96
	Política	trump, govern, china, economy, politic, presid	5269,31
	China	china, wuhan	31539,90
	Quarentena	lockdown, outbreak, home	1804,30
	Mortes	death, died	378,46
Poynter	Fake News	false, mislead, fake, claim	3117
	Vacina, Cura e Tratamento	vaccin, cure, mask, vitamin	1981
	Bill & Melinda Gates	bill, melinda, gates	370
	Mortes	death, dead, die	1148
	Quarentena	outbreak, lockdown	471

Tabela 2. Tópicos identificados, por conjunto de documentos, e seu nível de engajamento

É possível perceber que discussões em volta de métodos para a redução da doença, vacina, cura e tratamento apareceu com bastante evidência nas coleções de documentos, bem como assuntos relacionados a própria *fake news* e disseminação de desinformação. O tópico envolvendo mortes também se mostrou relevante. No caso do Twitter, o tópico Índia emergiu em decorrência de um conjunto de nomes de personalidades indianas e também pelo nome da capital Delhi. Discussões envolvendo política também apareceram, no caso do Facebook, através das relações entre Estados Unidos, China e o ex-presidente Donald Trump; no Twitter este assunto apareceu na forma de crise humanitária, crise de água, protestos, problemas envolvendo imigração e a situação dos hospitais. No *Poynter*, a associação de termos em volta do sobrenome Gates, trouxe à tona boatos envolvendo a fundação Gates de Bill e Melinda Gates. Diversos boatos envolvendo a fundação Gates com recebimento de dinheiro do governo, desenvolvimento da vacina e acordos comerciais envolvendo a fundação foram identificados pelas agências de *fact-checking*. China também foi um dos tópicos emergentes nas redes sociais, principalmente no Facebook, e que está fortemente associado à origem do vírus e da pandemia.

Sete principais tópicos que estão sendo discutidos recorrentemente nas redes sociais foram identificados. São eles: a própria disseminação de *fake news* e teorias da

conspiração, vacina russa, política, quarentena, mortes, origem do vírus e por fim a cura e tratamento da doença. As nuvens de palavras revelaram uma forte sensação de pânico, é possível detectar termos e publicações que falam sobre a quarentena, bloqueios e pedidos para ficar em casa, no Twitter a palavra “*hell*” (inferno) apareceu com destaque e reflete bem este sentimento. Muitos usuários demonstram resistência em acreditar no fato da pandemia e foi possível identificar um uso sensacionalista e distorcido das redes sociais. Outras pessoas parecem buscar soluções alternativas para a pandemia, explicações sem embasamento científico de supostas curas ou formas de combater o vírus. Ainda, teorias conspiratórias disseminadas através de vídeos e imagens descontextualizadas defendem, de forma dissimulada, efeitos da vacina ou que o vírus não existe.

Este é um grande desafio para a saúde pública, que deve desprender esforços para combater não somente a pandemia, mas também a “infodemia” [Cinelli et al. 2020]. A eclosão disseminatória de informações falsas nas redes sociais pode alimentar o pânico generalizado e provocar efeitos nocivos à população, ofuscando as evidências e dificultando a resposta dos profissionais e dos sistemas públicos de saúde [Abd-Alrazaq et al. 2020, Vogel 2017].

5. Limitações

Para atingir o objetivo do trabalho, houve a necessidade de explorar informações confirmadamente falsas nas redes sociais. Por isso, a extração das postagens foi realizada em matérias de *fact-checking*, ao invés da extração direta por *hashtags* ou buscas no próprio Facebook ou Twitter. Isso resultou em uma amostra relativamente pequena, se comparada a outros trabalhos que exploram as redes sociais. Por outro lado, isso garantiu que os assuntos analisados neste trabalho estejam, necessariamente, associados à desinformação, como era desejado.

Outro fator limitante, é que nem todas informações verificadas continham em seu corpo o *post* original, onde ela foi compartilhada nas redes sociais. Por isso, é importante que as agências tenham o cuidado de colocar este tipo de referência, pois isso possibilita a automação de alguns tipos de análises e reduzir o trabalho manual realizado nas agências.

6. Conclusão

A pandemia da COVID-19 impactou o mundo todo, através da economia, sistemas de saúde, relações sociais e políticas. Lidar com uma situação tão nova como essa na era da informação é propício para que teorias conspiratórias e informações falsas circulem pela rede. Este trabalho se empenhou em detectar os tópicos que têm sido discutidos de forma a disseminar informações falsas ao redor do mundo, através das redes sociais. Para isso, os autores se apoiaram nas agências de *fact-checking* e mostraram que assuntos como vacina, cura, tratamento e mortes estão sendo amplamente utilizados para disseminar desinformação sobre a pandemia. A pesquisa ainda mostrou que estes assuntos vêm sendo tratados pelas agências, com a missão de reduzir o dano causado por eles, pois foi possível notar uma queda no número de postagens realizadas nas redes sociais, bem como o número de compartilhamento dessas postagens.

As contribuições trazidas por este trabalho podem auxiliar outras pesquisas em volta da disseminação de desinformação, inclusive, replicando a mesma metodologia para outros assuntos, pois o LDA pode extrair tópicos de alta concordância, tornando-se

uma ferramenta adequada para estudar fenômenos de infodemia no contexto de políticas públicas. É importante que este tipo de colaboração seja feita para que não somente a curva de casos de COVID-19 diminua, mas também a curva da mentira.

Os dados das redes sociais e as abordagens de modelagem de tópicos facilitam a compreensão das discussões e preocupações públicas sobre a pandemia. Analisar e perceber as reais preocupações do público pode aumentar eficiência do trabalho realizado para combater a disseminação de informação falsa.

Como trabalhos futuros, pretende-se aprimorar as técnicas de OCR e outras abordagens a fim de permitir a exploração de uma gama de maior de conteúdo contido nas matérias de *fact-checking*, como aqueles que circularam em WhatsApp, jornais impressos, blogs e ainda os que circularam em redes sociais, porém não possuem o *link* de redirecionamento. Pretende-se realizar um estudo sobre o perfil de quem compartilha esse tipo de informação para compreender melhor o fenômeno da desinformação.

Agradecimentos

Este trabalho foi desenvolvido com o apoio da Fapemig (PPM-XII).

Referências

- Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., and Shah, Z. (2020). Top concerns of tweeters during the covid-19 pandemic: Infoveillance study. *J Med Internet Res*.
- Afonso, N. (2020). É falso que oms desaconselhou a aplicação de medidas de isolamento social. *Agência Lupa*.
- AFP (2020). This video shows a mock funeral procession during an anti-citizenship law protest in india in 2019. Acessado em 26/01/2021: <https://factcheck.afp.com/video-shows-mock-funeral-procession-during-anti-citizenship-law-protest-india-2019>.
- Amapa, O. (2020). Oms: transmissão de covid-19 a partir de assintomáticos é “muito rara. *Amapá Online*.
- Bento, G. M. and Santos, R. T. d. (2018). Avaliação de métodos de remoção de outliers e seus impactos na precisão dos métodos de interpolação. *1º Simpósio Mato-Grossense de Mecanização Agrícola e Agricultura de Precisão - SIMAP*, 1.
- Blei, D. M. and Lafferty, J. D. (2009). Topic models. *Chapman and Hall/CRC*, pages 101–124.
- Brennen, D. J. S., Simon, F., Howard, D. P. N., and Nielsen, P. R. K. (2020). Types, sources, and claims of covid-19 misinformation.
- Broniatowski, D. A., Jamison, A. M., Qi, S., AlKulaib, L., Chen, T., Benton, A., Quinn, S. C., and Dredze, M. (2018). Weaponized health communication: Twitter bots and russian trolls amplify the vaccine debate. *American Journal of Public Health*, 108(10):1378–1384. PMID: 30138075.
- Castillo, C., Mendoza, M., and Poblete, B. (2011). Information credibility on twitter. In *Proceedings of the 20th International Conference on World Wide Web, WWW '11*, page 675–684, New York, NY, USA. Association for Computing Machinery.

- Cinelli, M., Quattrocioni, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., Zola, P., Zollo, F., and Scala, A. (2020). The covid-19 social media infodemic. *Scientific Reports*, 10(1):16598.
- Goeking, W. (2020). Auxílio emergencial é alvo de golpes na internet; veja como evitar ciladas. *Valor Investe*.
- Jordan, S. E., Hovet, S. E., Fung, I. C.-H., Liang, H., Fu, K.-W., and Zion Tsz Ho Tse (2019). Using twitter for public health surveillance from monitoring and prediction to public response. *Data*.
- Larson, H. J. (2018). The biggest pandemic risk? viral misinformation. (562).
- Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., and Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3):106–131. PMID: 26173286.
- Lima, S. (2021). Não há provas de que vídeo de mulher tremendo ao caminhar tenha relação com vacina da pfizer. Acessado em 06/03/2021: <https://politica.estadao.com.br/blogs/estadao-verifica/nao-provas-de-que-video-de-mulher-tremendo-ao-caminhar-tenha-relacao-com-vacina-da-pfizer/>.
- Moraes, M. (2020). Brasil lidera desinformação sobre número de casos e mortes por covid-19 no mundo. Acessado em 12/06/2020.
- Poynter (2020). The coronavirusfacts/datoscoronavirus alliance database.
- Röder, M., Both, A., and Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, WSDM '15*, page 399–408, New York, NY, USA. Association for Computing Machinery.
- Shah, Z., Surian, D., Dyda, A., Coiera, E., Mandl, K. D., and Dunn, A. G. (2019). Automatically appraising the credibility of vaccine-related web pages shared on social media: A twitter surveillance study. *J Med Internet Res*.
- Sousa Jr, J. H., Raasch, M., Soares, J. C., Virgínia, L., and de Sousa Ribeiro, H. A. (2020). Da desinformação ao caos: uma análise das fake news frente à pandemia do coronavírus (covid-19) no brasil. *Cadernos de Prospecção*.
- Tandoc Jr, E. C., Lim, Z. W., and Ling, R. (2018). Defining “fake news”. *Digital Journalism*, 6(2):137–153.
- Vlachos, A. and Riedel, S. (2014). Fact checking: Task definition and dataset construction. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pages 18–22, Baltimore, MD, USA. Association for Computational Linguistics.
- Vogel, L. (2017). Viral misinformation threatens public health. *CMAJ*, 189(50):E1567–E1567.
- WHO (2020). Coronavirus disease (covid-19) situation report – 139. *World Health Organization*. Acessado em 07/06/2020: <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200607-covid-19-sitrep-139.pdf?sfvrs=79dc6d082>.