

Análise da propagação de notícias falsas sobre COVID no Twitter

Lucas Freire
Giovani Valdrighi
Taisa Calvette

14 de janeiro de 2022

1 Introdução

O aumento de notícias falsas (*fake news*) está se tornando um problema mundial. Embora este não seja um problema recente, este vem se tornando um motivo de preocupação por causa da popularidade das redes sociais, que permitem a interação e difusão de notícias rapidamente e em grande escala. Por meio de compartilhamentos, curtidas ou retuítes, a informação falsa acaba se tornando incontável, altamente disseminada e de difícil localização da fonte. Desta forma, podemos dizer que as redes sociais se tornaram um lugar para divulgar informações e notícias falsas rapidamente. De fato o compartilhamento de notícias falsas tornou-se um galopante no mundo digital atual e se tornou também um incentivo para alguns indivíduos que se beneficiam com a proliferação de desinformação para um grande público.

Em dezembro de 2019, um novo vírus chamado COVID-19 foi relatado na China e, nos últimos meses, o vírus se espalhou para outras partes do mundo, levando muitas pessoas a óbito. Inicialmente, foi alegado que era transmitido de animais para humanos. No entanto, agora é transmitido de humanos para humanos por meio de contatos e gotículas. Em meio a incertezas e desconhecimento, muitos rumores e notícias falsas começaram a circular sobre o COVID-19. Consequentemente, a desinformação nas redes sociais alimentou o pânico entre o público em relação à pandemia COVID-19. O fato é que vem se tornando cada vez mais difícil distinguir notícias falsas de reportagens fidedignas. Neste contexto específico de notícias falsas de COVID-19, alguns autores vem fazendo a análise de redes sociais.

Apuke et. al (2021) usou a análise para entender a motivação por trás das *fake news* através do modelo *Use and Gratification Framework*. Os dados foram analisados com *Partial Least Squares* (PLS) para determinar os efeitos de seis variáveis no resultado do compartilhamento de notícias falsas. Os resultados encontrados pelos autores mostraram que o altruísmo foi o fator mais significativo que previu o compartilhamento de notícias falsas do COVID-19 [1].

Moscadelli et. al (2020) realizou a análise com o objetivo de medir quantas notícias falsas e as notícias verificadas correspondentes circularam na Itália no período entre 31 de dezembro de 2019 e 30 de abril de 2020. Os autores encontraram que os links que continham notícias falsas foram compartilhados 2.352.585 vezes, representando 23,1% do total de compartilhamentos de todos os artigos analisados. O que mostra como é preocupante a disseminação de *fake news* em relação ao COVID-19 [4].

Yang et. al (2020) estimaram a predominância de *links* para informações com baixa credibilidade no Twitter durante o surto e o papel dos *bots* na disseminação desses *links*. Os autores descobriram que o volume combinado de *tweets* com *links* para informações de baixa credibilidade é comparável ao volume de artigos do *New York Times* e *links* do CDC. Além disso, a análise de conteúdo revela uma politização da pandemia. A maior parte desse conteúdo se espalha por meio de retuítes. Os *bots* sociais estão envolvidos tanto na postagem quanto na amplificação de informações de baixa credibilidade, embora a maior parte do volume seja gerada por prováveis humanos [5].

Para o caso específico do Brasil, Garcia e Berton (2020) usaram a identificação de tópicos e a análise de sentimento para explorar um grande número de *tweets* no Brasil e nos EUA, que tiveram um alto número de propagação e mortes por COVID-19. Foram usados 3.332.565 *tweets* em inglês e 3.155.277 *tweets* em português para comparar e discutir a eficácia da identificação do tópico e da análise de sentimento em ambos os idiomas. O período analisado foi de abril a agosto de 2020 e os autores encontraram dez tópicos principais relacionados ao COVID-19 em ambos os idiomas, onde sete tópicos são equivalente como por exemplo: proliferação, tratamento e impacto econômico [3].

O objetivo principal do trabalho foi o de construir uma visualização para analisar um conjunto de *tweets* sobre a pandemia de coronavírus que foram classificados pelo Aos Fatos, que é um jornal verificador de informações falsas. Vale ressaltar que os jornalistas de Aos Fatos acompanham declarações de políticos e autoridades de expressão nacional, de diversas colorações partidárias, de modo a verificar a sua veracidade. As declarações são então classificadas com uma das sete categorias, dentre elas: Verdadeiro, Impreciso, Exagerado, Contraditório, Insustentável, Distorcido ou Falso.

A visualização foi feita pra junto com uma análise de como a narrativa dos perfis de desinformação se alterou ao longo do tempo, especificamente em relação ao COVID-19. O tema é de extrema relevância uma vez que foram disseminadas

diversas *fake news* sobre a pandemia. Para tal, foi utilizado o paper de [2] como referência na apresentação final dos dados.

2 Metodologia

Os dados do Twitter são amplamente usados para fins de pesquisa e são coletados por meio de uma variedade de métodos e ferramentas. Devido aos termos de uso dos dados do Twitter, os bancos de dados disponibilizados em geral são no formato de *tweets* desidratados. Em outras palavras, em vez de coletar o conteúdo do *tweet* que inclui geolocalização, tempo, imagens e outras informações anexadas aos *tweets*, o que é disponibilizado é um arquivo de texto simples que consiste em uma lista de IDs de *tweets* exclusivos.

Identificou-se que os dados encaminhados pelo AOS fatos continham um grande número de registros inválidos no campo de identificação do usuário. Utilizou-se o ID, exclusivo de cada *Tweet* e criando segundo o protocolo *Snow Flake*. De posse destes IDs foi possível buscar para cada *tweet* seus respectivos metadados, em especial os campos de informações relativos a sua autoria. Assim, os dados precisam passar um processamento para torná-los utilizáveis e então fazer a análise textual das redes sociais.

2.1 Hidratação dados do Twitter

Com o conjunto de IDs do *tweet*, o primeiro passo foi a 'hidratação' dos dados. Para hidratar o arquivo foi usado o *DocNow Hydrator*, um *software* de hidratação gratuito e de código aberto encontrado no *GitHub*. Uma vez feito o *upload* do conjunto de dados do *Tweet ID*, o resultado foi uma série de números de identificação de 18 dígitos, cada um correspondendo a um *tweet* específico no Twitter, e que o *Hydrator* usou para vincular o *tweet* junto com os metadados associados.

Os dados foram carregados para serem hidratados e os dados vistos foram da longa lista de IDs de *tweet* individuais para um conjunto de dados robusto e multidimensional no formato *.json*. Desta forma, cada *tweet* possui metadados úteis, incluindo o tempo em que foi criado, as *hashtags* incluídas, o número de retuítes e curtidas, além de outras informações específicas de localização.

2.2 Análise exploratória dos dados

Para conseguir fazer a modelagem visual dos dados o trabalho foi dividido em algumas etapas. Inicialmente foi feita uma análise exploratória dos dados para compreender o comportamento, as principais características presentes e ter *insights* iniciais, como por exemplo, a distribuição de palavras chaves ao longo do tempo, e a distribuição da quantidade de *retweets* entre os perfis observados. Os dados disponibilizados pela AOS Fatos já foram previamente organizados, entretanto, alterações pontuais o tratamento da base dados foram necessários para a utilização.

A primeira análise exploratória foi feita para analisar o comportamento dos usuários, verificando como é a distribuição de *tweets*, curtidas obtidas, compartilhamentos obtidos e número de seguidores dos usuários. Foi possível notar que os dados são formados por uma grande maioria de usuários que utilizam pouco a rede, com poucas postagens e curtidas obtidas, enquanto alguns perfis são os principais geradores de conteúdo.

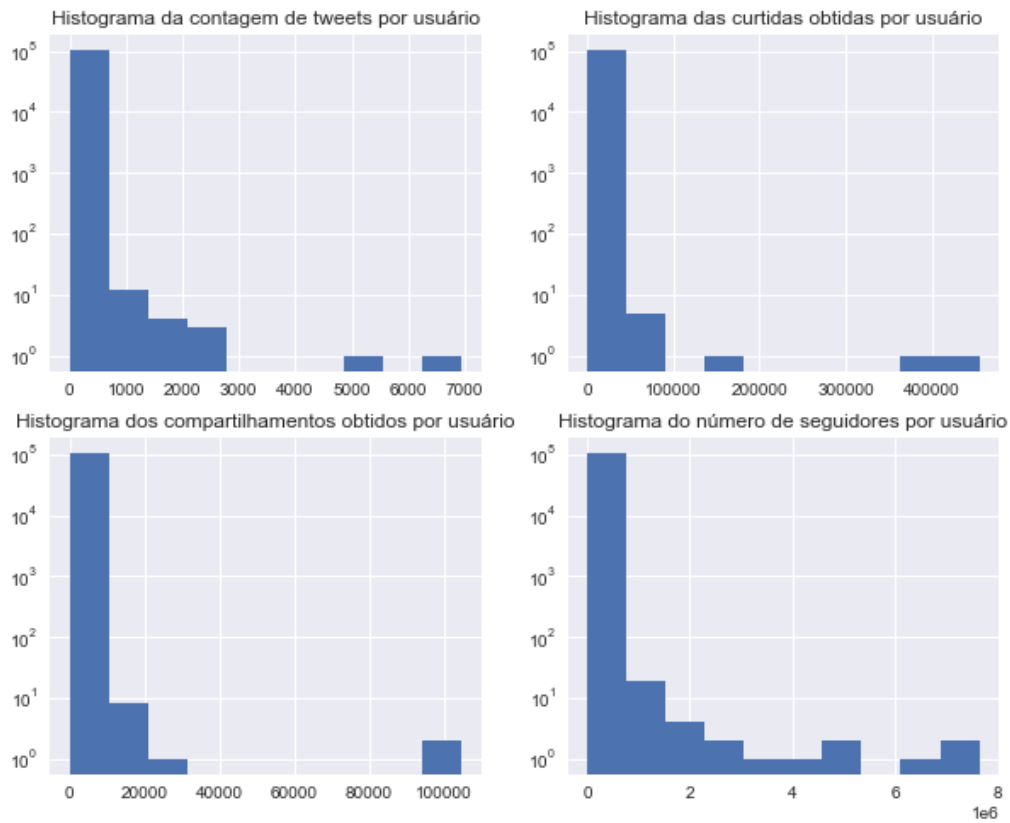


Figura 1: Análise inicial dos dados

Como pode ser visto na Figura 1, existe uma grande lacuna nos dados quando se observa usuários específicos. Diante disso, o trabalho teve um novo direcionamento focado então nos tópicos, ao invés dos usuários. Foi realizada também uma análise da existência de correlação entre essas variáveis, por exemplo, podemos verificar se os usuários com menores notas médias sem aqueles que possuem maior quantidade de *tweets* publicados. É possível

	Contagem	Soma curtidas	Soma compartilhamentos	Seguidores	Nota
Contagem	1.00	0.06	0.10	0.00	-0.04
Soma curtidas	0.06	1.00	0.99	0.05	-0.00
Soma compartilhamentos	0.10	0.99	1.00	0.05	-0.00
Seguidores	0.00	0.05	0.05	1.00	0.01
Nota	-0.04	-0.00	-0.00	0.01	1.00

Tabela 1: Correlação entre as variáveis

notar pela Tabela 1 que as correlações obtidas entre as variáveis são mínimas, a única significativa obtida é a correlação entre o número total de curtidas obtido e o número total de compartilhamentos obtido. No tangente às notas médias, não foi possível notar uma clara relação entre as variáveis, conforme os gráficos da Figura 2 .

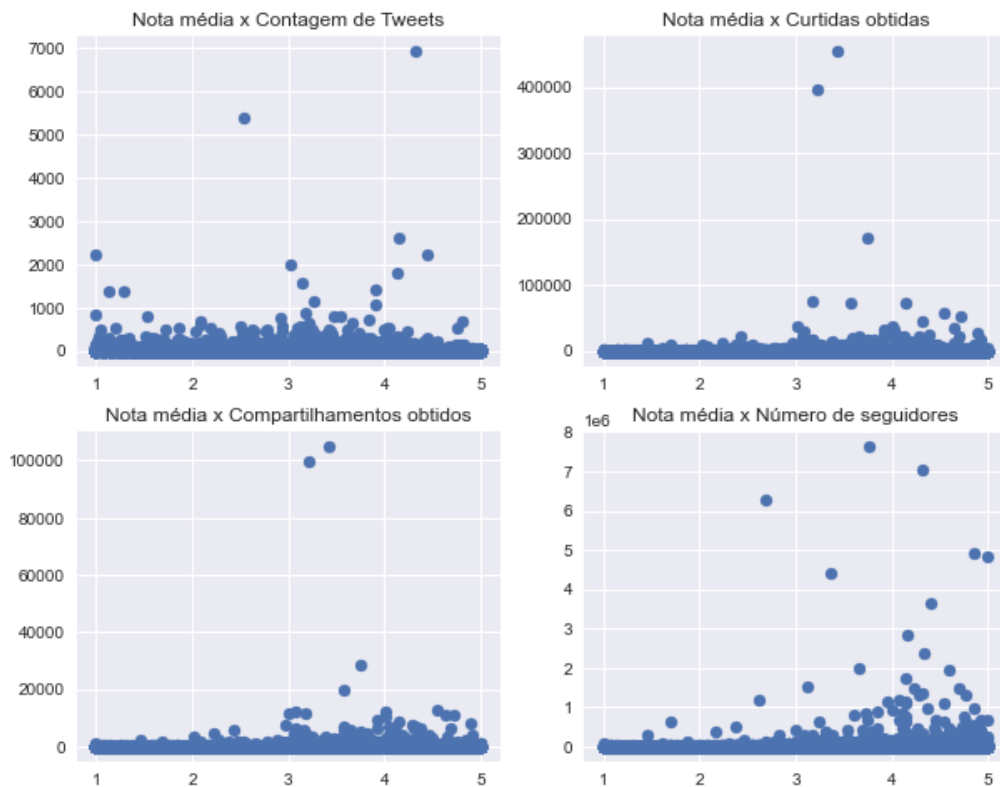


Figura 2: Análise da nota média

O arquivo em Python da análise exploratória de dados pode ser visto [neste link](#). Para mais detalhes da base de dados e demais arquivos basta acessar [este link](#).

2.3 Modelo

Para compreender o conteúdo textual que os *tweets* da base de dados possuíam, uma modelagem de tópicos permite encontrar quais foram os principais assuntos presentes através de uma combinação de palavras e classificar cada *tweet* em um desses assuntos. Foi utilizado o modelo *Latent Dirichlet Allocation* (LDA), um dos mais populares atualmente, em que cada tópico possui uma distribuição de palavras e cada texto possui uma distribuição de tópicos. Com a distribuição de palavras somos capazes de compreender o que cada tópico representa e com a distribuição de tópicos temos a informação se um texto é característico de um tópico específico, ou possui alta probabilidade para diferentes tópicos.

Para criar o modelo, os dados passaram por uma limpeza, com o objetivo de remover textos que sejam ruídos, *tweets* não relacionados a COVID. Inicialmente, foi-se considerado apenas os *tweets* publicados após agosto de 2020, pois o período anterior possuía diversas semanas de dados ausentes. Foram removidos *tweets* que possuíam algumas palavras chaves não relacionadas a COVID, termos relacionados a cultura pop. *Tweets* muito curtos (com menos de 12 caracteres) e com muitas menções (mais de 7) também não trariam informações pertinentes a modelagem de tópicos, pois textos muito curtos não representam tão bem um tópico e é comum textos de spam em que só possuem menções, sem um conteúdo de fato. Também foram removidos *tweets* que possuíam 0 curtidas e 0 compartilhamentos. Em sequência, o texto foi limpo removendo links para páginas, pontuações, números, *stopwords* (palavras que conectam ideias, mas não representam conceitos por si só). Para a modelagem foram consideradas apenas palavras que estavam presentes em menos de 50% dos textos pois palavras muito comuns não são capazes de representar as diferenças entre assuntos, além disso, foram desconsideradas as palavras que estavam em menos de 750 textos. Por fim, o treinamento do modelo foi feito com cerca de 130 mil *tweets*.

O modelo LDA possui como um parâmetro o número de tópicos, para escolher qual utilizaríamos foi levado em conta uma análise de perplexidade de diferentes modelos com diferentes visualizações, a característica de que os dados foram coletados com o objetivo de serem todas informações falsas sobre COVID, ou seja, era esperado de que tratassem de assuntos similares e que a visualização gerada representasse melhor modelos que não possuem muitos tópicos. Dessa forma, foi escolhido um modelo com 6 tópicos. As principais palavras para cada um dos tópicos obtidos encontram-se listadas abaixo junto com um nome representativo do tópico, tal nome foi utilizado na visualização para dar uma

breve noção do que se trata o tópico, ele foi escolhido através das análises principais dos tópicos e da diferença entre palavras dos tópicos.

- **Bolsonaro:** dinheiro, milhoes, bolsonaro, brasil, vacina, vacinas, governo, governadores, pais, pandemia, covid, stf, povo, presidente, contra.
- **gripezinha:** ne, vc, ser, doria, ninguem, vai, ainda, nada, covid, gripezinha, tomar, ai, pra, cloroquina, vacina.
- **tratamento precoce:** votar, voto, eleicoes, vidas, feitos, culpa, testes, tratamento, precoce, mil, ter, ivermectina, mortes, covid, casos.
- **mortes:** pessoas, deus, curados, mortalidade, aqui, vao, casa, vai, coronavirus, vacinha, pra, gente, covid, ta, pq.
- **china:** medico, pode, comunista, hidroxicloroquina, novo, cade, cu, saude, federal, estado, desde, jdoriajr, vamos, virus, chines.
- **lockdown:** lockdown, china, fez, existe, sempre, brasil, chinesa, bem, tudo, pessoas, sobre, mundo, pandemia, nunca, todo.

Palavras estão escritas na sua forma após a limpeza, a ordem não representa importância. O notebook que contém o processo de limpeza e treinamento do LDA se encontra no seguinte [link](#).

3 Visualização

Conforme citado na análise exploratória dos dados, o foco do trabalho teve um redirecionamento dos principais usuários para os tópicos mais populares, sendo assim, foi proposto um modelo mais adequado e que permita melhor visualização e análise de dados relacionados. Em linhas gerais, cada *tweet* contém a informação da data de publicação e assim, foi gerada a distribuição através da frequência semanal de cada tópico nos *tweets* ao longo de 39 semanas.

Os gráficos buscam dar duas perspectivas complementares ao desafio de representar a propagação de notícias falsas no *Tweeter*: uma análise micro, representando o movimento do *tweet* mais popular da semana, e uma análise macro que traz uma visão global do material repercutido no conjunto das 130 mil postagens. Portanto, a análise visual foi dividida em dois gráficos.

O primeiro representa **evolução temporal da postagem mais popular**. Neste gráfico mostra-se a maneira como a postagem mais popular de determinada semana, entendida como a soma de curtidas e compartilhamentos, movem-se entre os tópicos de um conjunto de 6 categorias dos dados do *Twitter*. Como especificado no *project proposal*, trata-se de uma forma de visualização inspirada na proposta de visualização designada como : Speaker Movement Analysis: pelos criadores do ConToVi[2]. O painel em forma de círculo dividido entre os 6 tópicos modelados por LDA, permite que seja possível ver a postagem mais popular de cada semana posicionada no interior do círculo de maneira a refletir, por proximidade, os tópicos mais relevantes para aquele texto em particular. Ao passar do *mouse*, é possível inspecionar o texto ver também o número de *likes* e compartilhamentos, além de mostrar em qual categoria o *tweet* mais popular pertence. Este último se dá através da proximidade da circunferência central do tópico do painel o qual ele tem maior probabilidade de pertencer. Esta análise de probabilidade teve em vista uma maior liberdade para o *tweet* de pertencer a mais de uma categoria de tópico. Ainda no gráfico da **evolução temporal da postagem mais popular**, é possível examinar a lista de *tokens* mais frequentes de cada um dos tópicos. Os tópicos são identificados por cores diferentes (as mesmas que serão utilizadas no segundo gráfico). Por fim, destaca-se um último elemento visual associado à postagem mais popular da semana: as transições ao longo do tempo acumula um rastro em cada transição, permitindo portanto uma representação do histórico anterior até determinado momento.

Para abordar o tempo na visualização, foi feita uma linha do tempo semanal interativa onde é possível inclusive ver o traço da mudança temporal, conforme o exemplo da Figura 3. Por fim, é possível examinar o *tweets* mais popular da semana ao pairar o cursor do *mouse* sobre o círculo que representa o *tweets* mais popular de uma determinada semana.

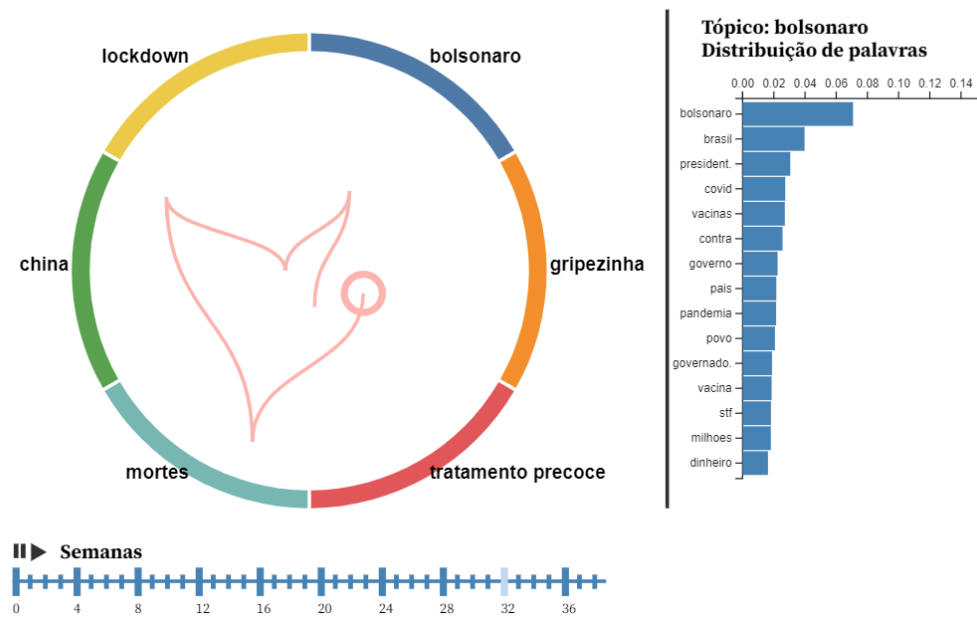


Figura 3: Painel principais tópicos

O segundo gráfico buscou apresentar uma **visão global da dinâmica entre os tópicos**. Os resultados da modelagem de tópicos aqui refletem o conjunto de *tweets*, não apenas o mais popular, no mesmo intervalo de tempo. O gráfico em linhas permite a seleção de e destaque de um tópico por vez, permitindo uma noção de importância relativa entre os tópicos. Nesse perspectiva global, optou-se por usar o ranqueamento dos tópicos e não os valores absolutos da contagem de postagens. O objetivo desta escolha de escala é dar ênfase às mudanças de discursos dominantes e evitar que as semanas mais ou menos ativas, sobre ou sub-representadas na base, criassem um componente adicional que ofuscaria a ênfase desejada. Elimina-se, assim, a variância do número de tweets por semana para que o usuário possa focar no elemento comparativo de perda ou ganho de importância do tópico relativamente aos demais tópicos. A série histórica dos 6 principais tópicos está representado na Figura 4 .

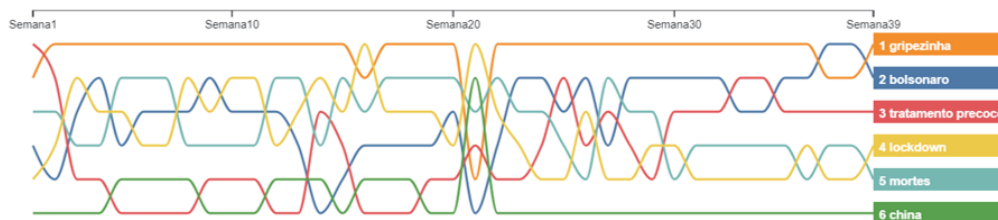


Figura 4: Painel séries históricas dos principais tópicos

Para a parte de interatividade, ao clicar em algum tópico podemos ver que a série histórica ganha destaque visual e na parte superior é possível ver as principais *hashtags* relacionadas a cada tópico, conforme a Figura 5.



Figura 5: Painel série histórica por tópico e *hashtags*

O notebook pode ser acessado no link: [este link](#).

4 Conclusão

Neste trabalho buscou-se produzir uma representação dos *tweets* mais populares identificados como potenciais propagadores de desinformação a respeito da Covid-19. A principal escolha estética do trabalho tem inspiração no ConToVi [2], com diversas adaptações. Assim como o ConToVi, os tópicos estão representados como um círculo dentro do qual nosso *corpus* é representado. Entretanto, no ConToVi representa-se a evolução do discurso de um conjunto pequeno de pessoas ao longo do tempo, ao passo que neste trabalho representamos os *tweets* de maior impacto em cada semana.

Desta forma, no círculo interior desenhamos um traçado que representa a trajetória do *tweet* mais popular da semana, mas não necessariamente abrangendo as mesmas pessoas. Aqui a representação busca refletir o tópico de maior repercussão independentemente do emissor da informação. Ao clicar em cada um dos tópicos, o painel à direita exibe os termos mais frequentes e, portanto, mais característicos daquele tópico. Os tópicos revelaram-se de interpretação pouco intuitiva em alguns casos, razão pela qual optou-se por representar cada tópico por uma dentre as palavras mais expressivas de cada tópico.

O traçado feito pelo círculo, além de representar o acúmulo na proximidade dos tópicos mais frequentes como um todo, permite identificar as mudanças de tópico ao passar das semanas. Nas primeiras 19 semanas os *tweets* giram entre os temas representados pelo terceiro e quarto quadrantes (a promoção do "tratamento precoce" e discursos de minimização ou relativização das estatísticas de mortes decorrentes da pandemia). Nas semanas que se seguem até a semana 39, verifica-se que o tema do "tratamento precoce" perde espaço para menções à China e a política de *lockdown*, com relativamente maior dispersão entre os temas.

O segundo gráfico representa um universo maior de informações. No primeiro gráfico, apenas o *tweet* de maior popularidade é representado. Já no segundo gráfico, utilizou-se toda a base de *tweets* para identificar as *hashtags* mais representativas de cada tópico. Ao longo do eixo horizontal podemos identificar o número total de *tweets* relacionados a cada um dos tópicos e as respectivas *hashtags* acima do gráfico. As duas visões se complementam e apresentam duas visões distintas sobre a potencial propagação de desinformação: uma focada no individual e outra mais geral.

No primeiro caso, o *tweet* de maior popularidade dá uma visão sobre o conteúdo que está sendo curtido e compartilhado mais intensamente, enquanto a segunda representação gráfica dá a dimensão do todo, da produção e repercussão de conteúdo pelos perfis menos influentes.

O tópico identificado como "gripezinha", que abrando discursos de desqualificação da gravidade da pandemia em geral, permanece como o mais importante por todo o período, mas podemos notar uma troca de importância relativa entre os temas do "tratamento precoce" e "lockdown". Este tópico apresenta grande importância entre agosto e início de janeiro, momento em que uma inflexão coloca "tratamento precoce" em maior evidência.

Como maior obstáculo ao trabalho, identifica-se uma ausência de controle sobre o processo de aquisição de dados. Se por um lado buscou-se utilizar uma base de dados produzida por uma agência jornalística especializada em checagem de informações foi um passo essencial para o desenvolvimento do projeto, a ausência de um registro detalhado do processo de coleta de dados coloca em questão o risco de vieses derivados do próprio método de modelagem do risco de desinformação associado a cada *tweets*.

Portanto, a visualização cumpre o propósito de comunicar uma modelagem de tópicos (LDA) como forma de compreensão do potencial de desinformação no debate público expresso no *Twitter*. Para eventuais desenvolvimentos futuros, um controle mais refinado da coleta de dados poderia permitir uma identificação mais fidedigna da propagação de desinformação na rede social em estudo. Além disso, sugere-se uma análise mais robusta juntamente com o número de mortes, média móvel e número de infectados.

5 Divisão das tarefas

- Lucas foi responsável pela preparação dos dados para visualização, em especial o modelo LDA e o cômputo das coordenadas cartesianas de cada *tweet* na visualização circular.
- Giovanni foi responsável pela implementação da visualização circular baseada na ferramenta ConToVi [2] que contém a animação do trajetório do *tweet* mais popular ao longo das semanas observadas, além disso, um painel para a visualização das palavras mais importantes entre cada um dos tópicos.
- Taisa foi responsável pela implementação da visualização da série histórica dos tópicos mais importantes e *hashtags* mais populares relacionados aos tópicos. Além disso, foi também responsável pela maior parte dos relatórios escritos.

Referências

- [1] Oberiri Destiny Apuke and Bahiyah Omar. Fake news and covid-19: modelling the predictors of fake news sharing among social media users. *Telematics and Informatics*, 56:101475, 2021.
- [2] Mennatallah El-Assady, Valentin Gold, Carmela Acevedo, Christopher Collins, and Daniel Keim. ConToVi: Multi-Party Conversation Exploration using Topic-Space Views. *Computer Graphics Forum*, 35(3):431–440, 2016.
- [3] Klaifer Garcia and Lilian Berton. Topic detection and sentiment analysis in twitter content related to covid-19 from brazil and the usa. *Applied Soft Computing*, 101:107057, 2021.
- [4] Andrea Moscadelli, Giuseppe Albora, Massimiliano Alberto Biamonte, Duccio Giorgetti, Michele Innocenzio, Sonia Paoli, Chiara Lorini, Paolo Bonanni, and Guglielmo Bonaccorsi. Fake news and covid-19 in italy: Results of a quantitative observational study. *International Journal of Environmental Research and Public Health*, 17(16), 2020.
- [5] Kai-Cheng Yang, Christopher Torres-Lugo, and Filippo Menczer. Prevalence of low-credibility information on twitter during the COVID-19 outbreak. *CoRR*, abs/2004.14484, 2020.