

Association for Information Systems

AIS Electronic Library (AISeL)

ISLA 2020 Proceedings

Latin America (ISLA)

8-10-2020

Análise de Tweets e de Hashtags sobre COVID-19 na Cidade de São Paulo

Silvano Daniel Miloca

Universidade Tecnológica Federal do Paraná - UTFPR, danielmiloca@gmail.com

Follow this and additional works at: <https://aisel.aisnet.org/isla2020>

Recommended Citation

Miloca, Silvano Daniel, "Análise de Tweets e de Hashtags sobre COVID-19 na Cidade de São Paulo" (2020). *ISLA 2020 Proceedings*. 21.

<https://aisel.aisnet.org/isla2020/21>

This material is brought to you by the Latin America (ISLA) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ISLA 2020 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Análise de Tweets e de Hashtags sobre COVID-19 na Cidade de São Paulo

Artigo em Desenvolvimento

Silvano Daniel Miloca

Aluno de Mestrado em Computação Aplicada
Universidade Tecnológica Federal do Paraná - UTFPR
danielmiloca@gmail.com

Abstract

This paper analyzes *tweets* in the city of São Paulo related to COVID-19 and the official numbers of new cases and deaths in the city. The objective of this work was to establish a relationship between *tweets* and the spread of the pandemic in the city. Based on the most used *hashtags*, those who defend staying at home were separated, grouping by week of posting and comparing the number of *tweets* favorable to social isolation with the numbers of cases and deaths. A network diagram was also set up between users who *tweet* and those who *retweet* in order to check the layout of that network. It was observed that, over the weeks, tweets began to use less *hashtags* recommending people to stay at home, despite the growing number of new cases and deaths by COVID-19 in the city.

Key-words

COVID-19, Tweets, Hashtags.

Resumo

Este trabalho analisa os *tweets* na cidade de São Paulo relacionados à COVID-19 e os números oficiais de casos novos e de óbitos na cidade. O objetivo deste trabalho foi estabelecer relação entre os *tweets* e a difusão da pandemia na cidade. Com base nas *hashtags* mais utilizadas, foram separados os que defendem a permanência em casa, fazendo-se o agrupamento por semana de postagem e a comparação do número de *tweets* favoráveis ao isolamento social com os números de casos e de óbitos. Também foi montado um diagrama de rede entre os usuários que *twittam* e os que *retwittam* a fim de verificar a disposição dessa rede. Foi observado que, com o passar das semanas, os *tweets* passaram a utilizar menos as *hashtags* recomendando às pessoas para ficarem em casa, apesar do número crescente de novos casos e de óbitos por COVID-19 na cidade.

Palavras-chave

COVID-19, Tweets, Hashtags.

1. Introdução

A ocorrência de um surto de pneumonia que se iniciou em um mercado de animais vivos e atacadista de frutos do mar na cidade de Wuhan, China, chegou ao conhecimento da Organização Mundial da Saúde (OMS) em 31 de dezembro de 2019. (Zhu *et al.*, 2020). Tal surto foi identificado como um novo coronavírus: SARS-COV-2. Referenciando-se ao tipo do vírus e ao ano de surgimento, a OMS nomeou esta

doença COVID-19, um acrônimo de *CO*rona*VI*rus *D*isease-2019. (WHO, 2020; Croda e Garcia, 2020). Em 30 de janeiro de 2020, a OMS declarou-a como Emergência de Saúde Pública de Interesse Internacional. (Oliveira *et al.*, 2020). No Brasil, em 03 de fevereiro de 2020 foi declarada Emergência em Saúde Pública de Importância Nacional, por meio da Portaria 188 do Ministério da Saúde. (Brasil, 2020). Devido ao alto grau de transmissão desta doença entre os seres humanos as autoridades emitiram ordens para que as pessoas tivessem o mínimo de contato possível umas com as outras. (Brasil, 2020). Os locais de aglomeração de pessoas, como estádios, teatros, *shopping centers*, lojas e até escolas foram fechados, sendo estabelecidas também restrições de viagens e barreiras sanitárias. Os locais em que se desenvolvem atividades essenciais, como supermercados e farmácias, passaram a respeitar limites de número de clientes atendidos simultaneamente. Passou ainda a ser obrigatório o uso máscaras e de álcool gel para a limpeza e desinfecção das mãos. Há orientação para que as pessoas fiquem em casa e realizem teletrabalho (*home office*), frequentem aulas *online* e interajam por meio de videoconferências (Couto *et al.*, 2020; Gostin e Willey, 2020).

Sendo um assunto que envolve o mundo todo, e sendo as mídias sociais um importante meio de transmissão da informação, as notícias e os comentários sobre este assunto passaram a frequentar estes meios de comunicação. O Twitter, em particular, foi declarado como bom ou extremamente bom para compartilhar informações preventivas de saúde por 59% dos seus usuários. (Willford, Osann e Wenzel, 2018). Entretanto está repleto de informações erradas sobre saúde e a desinformação médica e conteúdo não verificável estão sendo propagados em um ritmo alarmante nas mídias sociais. (Singh, *et al.*, 2020, Kouzy, *et al.*, 2020).

O Twitter é uma rede social e *microblog* que permite aos usuários postarem mensagens com até 280 caracteres, as quais podem ser retransmitidas por *retweets* (CONOVER *et al.*, 2011). Dentro destes *tweets* são utilizadas *hashtags*, uma palavra-chave prefixada com um símbolo *hash* (#), que é uma forma utilizada pelo Twitter para facilitar a busca efetiva em um massivo volume de dados e de o usuário definir o contexto de compartilhamento e definir uma categoria de sua postagem sem ficar preso a categorias predefinidas pelo *blog*. (Otsuka, Wallace e Chiu, 2014). Os *tweets* podem ser retransmitidos por qualquer usuário, sendo denominados *retweets*. Neste contexto, esta pesquisa objetiva fazer análise das mensagens que possuem *hashtags* relacionadas à ideia de ficar em casa durante a pandemia, doravante denominados *tweets fiqueemcasa*, contabilizando o número de *tweets* por semana comparando a intensidade com que foram utilizados ao longo das semanas da pandemia à evolução da própria doença. Também são analisados os *retweets* e a ligação entre os usuários que escreveram o *tweet* original e os usuários que fizeram a retransmissão do *tweet*.

De acordo com o Ministério da Saúde do Brasil (Brasil, 2020), o número de novos casos e de óbitos do COVID-19 tem aumentado de forma rápida no país em geral, principalmente na cidade de São Paulo, que é a cidade em estudo. Este artigo considera apenas o número de casos novos e de óbitos novos por semana e não o número acumulado.

2. Dados Preliminares e Procedimentos Metodológicos

Esta seção explica a forma de obtenção dos dados do Ministério da Saúde do Brasil, como os *tweets* foram coletados, como foi feita a seleção dos *tweets* e de suas *hashtags* e como foi elaborada a representação da rede de *tweets* e de *retweets*.

2.1. Coleta dos Dados

Os dados sobre o número de casos novos e de óbitos relacionados à COVID-19 são divulgados diariamente pelo Ministério da Saúde do Brasil (Brasil, 2020) e são separados por dia, para o Brasil como um todo, para cada um dos 26 estados da federação e o Distrito Federal, e para os 5.570 municípios do país. Dentre estes dados foram selecionados somente os dados referentes à cidade de São Paulo e consolidados por semana de ocorrência.

A numeração das semanas foi seguida de acordo com o Ministério da Saúde que, conforme o número da semana constante nos dados disponibilizados, considerou a primeira semana do ano a que contém o dia

primeiro de janeiro. As semanas são todas consideradas de domingo até o sábado seguinte, inclusive. A primeira semana analisada neste artigo é a 13 por ser a primeira semana com dados disponibilizados pelo Ministério da Saúde. A semana 13 compreende os dias de 22 a 28 de março de 2020.

Os *tweets* foram obtidos utilizando-se a *Application Programming Interface* (API) do Twitter, e foram coletados em uma região com raio de 30 quilômetros a partir do centro da cidade de São Paulo (Praça da Sé, com latitude -23.550552 e longitude -46.632917), entre os dias **22 de março e 19 de julho de 2020**. Desta forma, foram obtidos 1.824.469 *tweets* únicos. O guia de utilização da API informa que:

“ao realizar pesquisas geográficas, a API de pesquisa tentará primeiro encontrar tweets com latitude/longitude no código geográfico consultado e, no caso de não ter êxito, tentará encontrar tweets criados por usuários cuja localização do perfil pode ser reversamente geocodificada em um lat/long no código geográfico consultado, o que significa que é possível receber tweets que não incluem informações de lat/long.” (Twitter Guide, 2020).

Nem todos os *tweets* foram coletados, devido a restrições do equipamento utilizado.

2.2. Classificação dos Tweets

Os *tweets* foram agrupados por semana de acordo com o seu dia de criação e seguiu-se a numeração das semanas conforme os dados do Ministério da Saúde. Foram considerados apenas *tweets* únicos enviados até a semana 29 (**de 12 a 18 de julho de 2020**), pelo fato que, quando esta análise foi feita, os dados do Ministério da Saúde para a semana 30 ainda não estavam finalizados. Foram selecionados na análise apenas os *tweets* que continham alguma das seguintes expressões relacionadas ao tema: *coronavirus, covid, covid19, covid_19, virus, pandemia, fiqueemcasa, ficaemcasa, emcasa, lockdown, stayhome, quarantín, vaipassar, quarenten, saude, health, preven, paciente, hospital, obito, falec, hospital, emprego, intern, 40tena, economia, depress, fome, alimento, miseria, recessao, sustentabilidade, falencia, confinamento, isolamento ou cloroquina*. Também foi identificado um usuário com nome de ‘Bot_Corona_V’ que não possuía nenhum *tweet* apenas *reweets* e sua descrição de usuário é ‘Apenas um bot que *retweeta* assuntos com relação ao COVID-19’, continha 203 *retweets* e foram removidos seus *retweets* da análise. Depois da aplicação deste filtro restaram 86.235 *tweets*. Para a semana 15, por problemas técnicos enfrentados durante a coleta, não foram obtidos *tweets* para a cidade de São Paulo e, portanto, essa semana ficou fora da análise. Um *retweet* começa com as letras ‘RT’ seguidas de um espaço, depois o caracter @ e o nome do usuário que foi *retwittado*. Desta forma é possível identificar o usuário do *tweet* original e o que executou o *retweet*. Foram separados os *tweets* dos *retweets*, ficando 33.839 *tweets* e 52.396 *retweets*.

3. Análise

Foram realizadas análises sobre os *tweets*, conforme mostrado na seção 3.1, e sobre os *retweets*, conforme a seção 3.2.

3.1 Análise sobre os Tweets

Os dados do Ministério da Saúde mostram que houve aumento significativo nos casos novos e de óbitos por COVID-19 na cidade de São Paulo, ao longo das semanas de 2020, desde o início da pandemia, como mostra o gráfico da figura 1. Este gráfico apresenta duas escalas, uma à esquerda e outra à direita, devido à diferença de magnitude do número de casos e de óbitos. A primeira morte de COVID-19 no Brasil aconteceu no dia 17 de março de 2020 (semana 12) e foi na cidade de São Paulo (BRASIL, 2020).

Sendo São Paulo uma metrópole, o raio de 30 quilômetros a partir da Praça da Sé abrange várias cidades próximas, então, para fins de comparação com a cidade de São Paulo, foram selecionados os dados do Ministério da Saúde incluindo as cidades de Barueri, Caieiras, Carapicuíba, Cotia, Diadema, Embu das Artes, Ferraz de Vasconcelos, Franco da Rocha, Guarulhos, Itapeverica da Serra, Itapevi, Itaquaquecetuba, Jandira, Mairiporã, Mauá, Osasco, Poá, Ribeirão Pires, Rio Grande da Serra, Santo André, São Bernardo do Campo, São Caetano do Sul, Suzano e Taboão da Serra e verificada a evolução da

doença em todas essas cidades. Praticamente não houve mudança das linhas de evolução da doença somente na cidade de São Paulo e na cidade de São Paulo com a inclusão das cidades próximas.

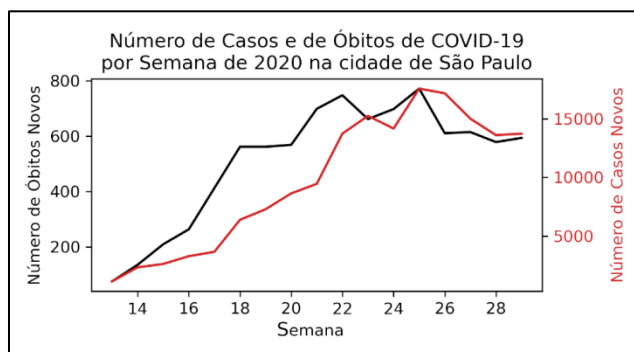


Figura 1. Número de casos e óbitos de COVID-19 na cidade de São Paulo.
Fonte: Brasil (2020)

Com a finalidade de verificar o número de *tweets* que utilizam *hashtags* defendendo a permanência em casa, dos *tweets* relacionados à COVID-19 nas semanas consideradas, foram listadas as 1.000 *hashtags* mais utilizadas e dentre estas foram selecionadas manualmente as que defendem a permanência em casa, que são as seguintes: #ficaemcasa, #Ficaemcasa, #FicaEmCasa, #ficaremcasa, #fiqueemcasa, #FiqueemCasa, #fiqueemcasacovid19, #fiqueemcasaecantecomigo, #fiqueemcasasepuder, #FiqueEmCasa, #FIQUEEMCASA, #fiqueemcasa, #FiquemEmCasa, #quedateencasa, #sepuderfiqueemcasa, #stayhome, #stayathome, #stayhomestaysafe, #StayHome. Para a continuidade das análises, foram selecionados os *tweets* que utilizaram qualquer destas *hashtags*, resultando em 2.050 *tweets*, que foram denominados *tweets fiqueemcasa*. Neste filtro não foram consideradas as *hashtags* #lockdown, #emcasa, #EmCasa, #confinamento, #quarentena, #quarantine, e #isolamentosocial por não ser possível definir se os usuários que utilizam estas *hashtags* estão a favor ou contra a permanência em suas residências e/ou ao isolamento social. Além disso, foram considerados apenas os *tweets* e não os *retweets*. A figura 2 mostra o número de *tweets* por tipo e por semana na cidade de São Paulo. **É possível perceber que, nas últimas semanas o número de *tweets* relacionados à COVID-19 é bem maior que nas semanas anteriores.**

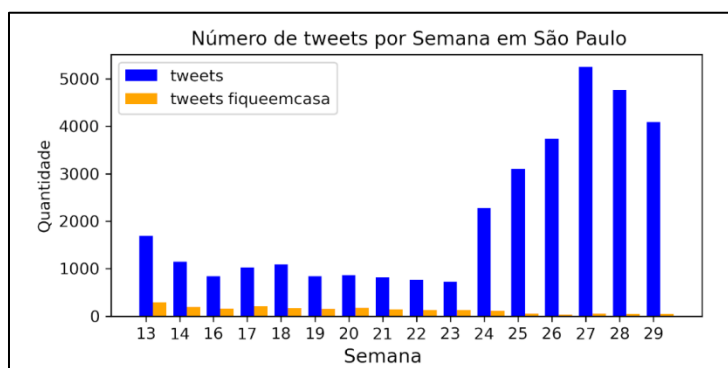


Figura 2. Número de *tweets* por tipo e por semana em São Paulo. (Em azul são *tweets* relacionados à COVID-19)
Fonte: dados da pesquisa

Como o número de *tweets* obtidos é diferente em cada semana, foi calculado o percentual dos *tweets* *fiqueemcasa* a cada semana em relação ao total de *tweets* relacionados à COVID-19. O número de casos

por semana e o percentual de *tweets fiqueemcasa* foram apresentados em um único gráfico, com duas escalas, conforme mostrado na figura 3, para permitir a análise dessas duas informações, conjuntamente.

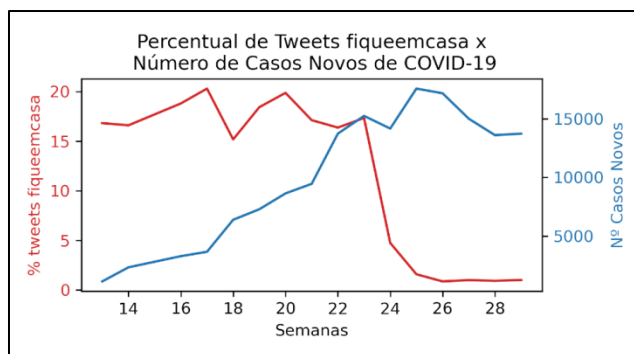


Figura 3. Percentual de *tweets fiqueemcasa* x número de casos novos de COVID-19.

Fonte: dados da pesquisa

Ao analisar a figura 3 é possível perceber que o número de casos de COVID-19 na cidade aumenta com o passar das semanas e o número de *tweets fiqueemcasa* diminui significativamente a partir da semana 24. Percebe-se que da semana 23 (31 mai a 06 jun) para a 24 (07 a 13 jun) houve uma queda bastante acentuada do percentual de *tweets fiqueemcasa*, caindo de 17,3% para 4,7% e este percentual baixo manteve-se com o passar das semanas. Uma das explicações a isto pode ser porque a cidade já estava há mais de noventa dias em quarentena (Zilberkan, 2020) e, no dia 30 de maio, o prefeito anunciou a continuidade da quarentena (Bocchini, 2020), ou que a quarentena não foi tão bem planejada e a pressão para arrefecer o isolamento estava sendo maior do que o desejo de mantê-lo (Mazza, 2020).

O fato de este percentual continuar abaixo dos 5% nas últimas semanas pode ser porque o isolamento tem efeitos profundos no corpo humano e no cérebro (Robinson, 2019). Além disso, os efeitos econômicos causados pelo isolamento social estão afetando a vida das pessoas mais do que a doença (Atalan, 2020).

Ainda com base na quantidade de *tweets* obtidos em cada semana, conforme mostrado nas figuras 2 e 3, é possível perceber que, embora nas semanas mais recentes o número de *tweets* obtidos tenha sido bem maior que nas semanas anteriores, o número de *tweets fiqueemcasa* está em nítida queda, ou seja, mesmo havendo mais *tweets* o número dos que defendem a permanência em casa é menor, confirmando uma tendência que já vinha ocorrendo desde a semana 24.

O porquê deste fenômeno precisará ser explorado em trabalhos futuros.

3.2 Análise dos retweets

Para a análise da rede de *retweets* foram criadas arestas entre os usuários que *twittam* e os que *retwittam*, a fim de verificar a disposição dessa rede. Nesta rede foram considerados todos os *retweets* relacionados à COVID-19.

Para cada *retweet* foram extraídos o usuário que escreveu o *post* e o usuário que teve seu *post retwittado*. Então, foi criada uma aresta entre estes dois. O usuário *retwittado* foi chamado de 'origem', considerando que foi a origem do texto, e o usuário que *retwittou* foi chamado de 'destino', considerando que o texto foi recebido por ele, antes de ser retransmitido. A rede mostrada na figura 4 foi gerada através do *software* Gephi (Bastian, Reymann e Jacomy, 2009). Cada nó representa um usuário e cada aresta representa a ligação entre o usuário que escreveu o *tweet* e o que encaminhou o *tweet*. O tamanho do nó é proporcional ao número de *tweets* originados a partir dele e a cor do nó é proporcional ao número de *retweets* que o usuário realizou. Não é possível distinguir as cores dos nós, porque os usuários que mais *retwittaram* tiveram poucos *tweets* seus *retwittados* e, portanto, o tamanho dos nós que os representam ficou pequeno em relação aos demais, mostrando, desta forma, que os usuários mais *retwittados* *retwittam* pouco, ao passo que aqueles que *retwittam* muito tem poucos textos seus enviados adiante.

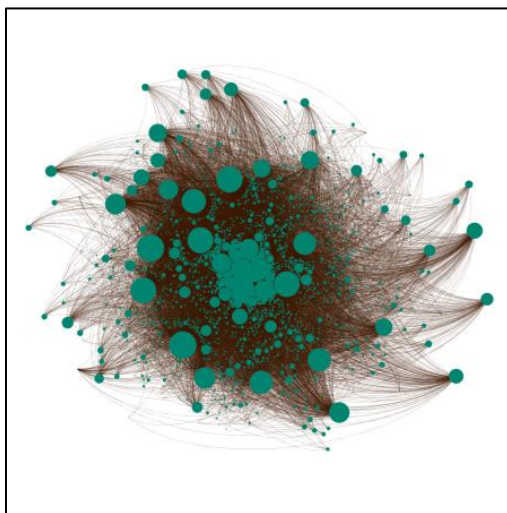


Figura 4. Rede de *retweets* entre os usuários.
Fonte: dados da pesquisa

4. Conclusão

Baseado nos dados coletados é possível perceber que no período analisado, com o passar da quarentena, as pessoas passaram a defender menos o isolamento social, mesmo com o número de casos e de óbitos de COVID-19 aumentando na cidade. Os motivos para isso precisam ser melhor investigados, mas, possivelmente, podem ser explicados pelo fato que a pandemia não foi tão bem planejada ou que o isolamento foi realizado de forma precoce. Com isso, a pressão para arrefecer o isolamento, ou por problemas associados ao corpo e à mente das pessoas ou pelos efeitos econômicos relacionados ao isolamento social, pode ter passado a contribuir para o relaxamento da posição das pessoas quanto à continuarem recomendando o isolamento social. Tais problemas podem ter contribuído para que, ao se chegar a um momento mais crítico da pandemia, as pessoas já estivessem menos imbuídas de continuar com os esforços de contenção da doença. Também é possível concluir que usuários que mais têm seus *tweets* encaminhados são usuários que *retwittam* pouco e os que mais *retwittam* têm poucos *tweets* seus encaminhados. Ou seja, que os formadores de opinião são muito menos numerosos que os difusores dessas opiniões.

REFERÊNCIAS

- Atalan, A. (2020) *Is the lockdown important to prevent the COVID-19 pandemic? Effects on psychology, environment and economy-perspective*. Annals of Medicine and Surgery, Volume 56, August 2020, Pages 217. <https://doi.org/10.1016/j.amsu.2020.06.010>.
- Bastian, M., Heymann, S., Jacomy, M. (2009). *Gephi: An Open Source Software for Exploring and Manipulating Networks*. In *Proc. of International AAAI Conference on Weblogs and Social Media*, <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154>.
- Bocchini, B. (2020). *Capital paulista prorroga quarentena até 15 de junho*. Agência Brasil – São Paulo. <https://agenciabrasil.ebc.com.br/saude/noticia/2020-05/capital-paulista-prorroga-quarentena-ate-15-de-junho>. Acesso em 05 Abr 20.
- Brasil, (2020). Ministério da Saúde. *Números da COVID-19*. <http://www.covid.saude.gov.br>, acesso em 05 Abr 20.
- Brasil, (2020). Ministério da Saúde. *Portaria MS/GM nº 188, de 3 de fevereiro de 2020. Declara Emergência em Saúde Pública de importância Nacional (ESPIN) em decorrência da Infecção Humana pelo novo Coronavírus (2019-nCoV)*. Diário Oficial da União, Brasília (DF), 04 Feb 2020 <http://www.in.gov.br/web/dou/-/portaria-n-188-de-3-de-fevereiro-de-2020-241408388>.

- Conover, M., Ratkiewicz, J., Francisco, M. (2011). *Political Polarization on Twitter*. In *Proc. of International Conference on Weblogs and Social Media*. Barcelona, Spain, 89-96. <https://doi.org/10.1021/ja202932e>.
- Couto, E. S., Couto, E. S., Cruz, I. M. P. (2020). #FIQUEEMCASA: EDUCAÇÃO NA PANDEMIA DA COVID-19. <http://dx.doi.org/10.17564/2316-3828.2020v8n3p200-217>.
- Croda, J. H. R., Garcia, L. P. (2020). *Resposta imediata da Vigilância em Saúde à epidemia da COVID-19*. <https://doi.org/10.5123/s1679-49742020000100021>.
- Gostin, L. O., Willey, L. F. (2020). *Governmental Public Health Powers During the COVID-19 Pandemic*. JAMA. 2020; 323 (21): 2137-2138. <http://doi.org/10.1001/jama.2020.5460>.
- Kouzy, R., Jaoude, J. A., Kraitem, A., El Alam, M. B., Karam, B., Adib, E., Zarka, J., Traboulsi, C., Akl, E. W., Baddour, K., (2020). *Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter*. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7152572>.
- Mazza, L. (2020). *Uma Pandemia, Muitas Quarentenas: Epidemia se aproxima do pico enquanto Brasil vai deixando isolamento iniciado sem planejamento prévio; cientistas defendem modelo inteligente e em ritmos distintos*. <https://piaui.folha.uol.com.br/uma-pandemia-muitas-quarentenas>. Acesso em 05 Jul 20.
- Oliveira, W. K., Duarte, E., França, G. V. A., Garcia, L. P., (2020). *How Brazil Can Hold Back COVID-19*. <https://www.scielo.org/article/ress/2020.v29n2/e2020044/en>.
- Otsuka, E., Wallace, S. A., Chiu, D., (2014). *Design and evaluation of a Twitter hashtag recommendation system*. <https://dl.acm.org/doi/abs/10.1145/2628194.2628238>.
- Robinson, S., (2019). *Isolation Has Profound Effects on The Human Body And Brain. Here's What Happens*. <https://www.sciencealert.com/isolation-has-profound-effects-on-the-human-body-and-brain-here-s-what-happens>. Acesso em 22 Jul 20.
- Singh, L., Bansal, S., Bode, L., Budak, C., Chi, G., Kawintiranon, K., Padden, C., Vanarsdall, R., Vraga, E., Wang, Y., (2020). *A First Look at COVID-19 information and misinformation sharing on Twitter*. <https://arxiv.org/abs/2003.13907>.
- Tsur, O., Rappoport, A., (2012). *What's in a hashtag?: content based prediction of the spread of ideas in microblogging communities*. <https://doi.org/10.1145/2124295.2124320>.
- Twitter Guide, <https://developer.twitter.com/en/docs/tweets/search/guides/standard-operators>, acesso em 01 Abr 20.
- Wilford, J.; Osann, K.; Wenzel, L., (2018). *Social Media Use Among Parents of Young Childhood Cancer Survivors*. *Journal of Oncology Navigation & Survivorship*. Jan2018, Vol. 9 Issue 1, p10-17. 8p.
- World Health Organization. (2020). *Novel Coronavirus (2019-nCoV): Situation Report – 22*. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200211-sitrep-22-ncov.pdf?sfvrsn=fb6d49b1_2.
- Zhu N, Zhang D, Wang W, Li X, Yang B, et al. *A Novel Coronavirus from Patients with Pneumonia in China, 2019*. (2020). <http://doi.org/10.1056/NEJMoa2001017>.
- Zilberkan, M., (09 Jun 2020). *Com restrições, comércio de rua vai reabrir amanhã na cidade de São Paulo*. <https://veja.abril.com.br/brasil/com-restricoes-comercio-de-rua-vai-reabrir-amanha-na-cidade-de-sao-paulo>. Acesso em 17 Jul 20.