Describing COVID-19 Pandemic by means of Tweets from Official Entities in Brazil

Matheus Henrique dos Santos[1] , **Fabiola S. F. Pereira**[1]

[1]Faculty of Computing - Federal University of Uberlândia

{matheus.santos,fabiola.pereira}@ufu.br

***Abstract.** In a world flooded with information, not always true, nor rarely biased, the communication of official entities assumes a key role. In this work we analyze a dataset composed of tweets from stakeholder entities regarding COVID-19 pandemic, to cite: National Agency of Sanitary Surveillance, Ministry of Health, World Health Organization and Brazilian Society of Infectious Diseases. We describe, by means of social, semantic and temporal patterns, the communication characteristics of above entities in social networks during COVID-19 pandemic. Further, we cross those patterns with key-facts occurred during pandemic. Results show that communication in social networks tend to be biased and not sufficient to comprehend the whole context. Furthermore, public entities are immature in their communication strategies in social networks.*

1. Introduction

In a world flooded with information, not always true, nor rarely biased, official entities communication assume a key role. Social networks have proven to be effective tools for communicating with society.

Understanding what attracts users to engage with social media content is important in domains such as advertising and market [Zafarani et al. 2014]. A new area that is emerging in this context is the communication strategy of official entities, in which the objective is not only to obtain engagement, but also to select the information to be disclosed, how it will be disclosed, as well as the communication timing. A good communication strategy in social networks is able to make the population allied to the conduct and decisions of an entity [Miletskiy et al. 2019].

Social patterns reflect engagement and how interactions with other users of the social network occur. Official entities can use networks to gain popularity [Dimitriadis et al. 2022]. Semantic patterns are good sources for understanding the subjects and the tone in which the communication is done. They can reveal, for example, if there is thematic diversity [Walter 2018]. Temporal patterns are good to describe scenarios in which you want to respond to key-facts or shift the focus away from negative marks associated with entities. Usually linked to posting frequency [Li et al. 2020].

In this work we investigate data from Twitter social network. Our focus is on a descriptive data analysis, contrasting communication strategies of official entities in social media. We combine social network analysis techniques with semantic networks and temporal descriptive statistics to perform a deep and innovative characterization of entities communication strategies.

In order to guide the achievement of this study, we define the following research question (RQ): *How was the communication strategy of key official entities on Twitter during COVID-19 pandemic?* To answer our RQ, we elaborate sub-questions regarding social (SO), semantic (SE) and temporal (TE) patterns as follows:

[SO-RQ1] *Were the subjects that most engaged during pandemic related with COVID-19?*
[SO-RQ2] *Were the profiles mentions during pandemic direct related with individuals?*
[SE-RQ3] *Was there a high thematic diversity in the posts themes during pandemic?*
[TE-RQ4] *Was there a correlation between posting frequency and key-facts during pandemic regarding official entities?*
[TE-RQ5] *Was there an increasing in the posting frequency during pandemic?*

2. Related Work

Much of the literature seeks to use social media to shape public opinion in relation to public events, such as emergency events [Dai et al. 2020], political events [Iyer et al. 2017] and the COVID-19 pandemic itself [Kwan and Lim 2021, Dimitrov et al. 2020]. From this perspective, social networks are seen as places for genuine expression of opinion by the population.

On the other hand, entities are increasingly cautious with the way in which communication is carried out on social networks. In particular, the communication strategy in social networks by public agencies has become increasingly relevant and has attracted the attention of the community [Miletskiy et al. 2019, Leyman et al. 2021].

In [Dejard et al. 2021] authors investigate the personal Twitter profiles of the heads of government of countries in South and North America and how they communicated with their audiences from November 2019 to November 2020. In the same line, [Li et al. 2021] perform a multi-country analysis of the discourse on Twitter investigating the reactions to government and public health agency social media accounts that share policy decisions and official messages. Results suggest that the most-discussed topic are evidence-based leadership and policy making. [Duque-Rengel et al. 2021] study Ecuador's government communication on Twitter with results showing an absence of engagement with digital audiences. In [Leyman et al. 2021] authors describe that the form of communication with the public chosen by the new acting Head of the Republic of Komi turned out to be the most in-demand and effective. On the other hand, in a local regional example, they observe leaders ignoring the rules and principles of crisis communication.

Among the techniques used to analyze communication profiles, descriptive statistics of texts and wordclouds were used [Duque-Rengel et al. 2021], besides direct observations [Leyman et al. 2021, Li et al. 2021]. The literature lacks deeper and more complete characterizations of communication strategies. Our work stands out for considering temporal aspects, semantic patterns and social metrics.

3. Dataset Description

We selected four official entities involved with the COVID-19 pandemic from different perspectives. The entities are diverse in terms of political, health and non-governmental interests.

- *Ministry of Health (MH).* The official entity of the Federal Executive Branch responsible for the organization and elaboration of plans and public policies aimed at the promotion, prevention and health care of Brazilians[1].
- *Brazilian Health Regulatory Agency (ANVISA).* Entity that promotes the protection of the population's health by executing sanitary control of the production, marketing and use of products and services subject to health regulation, including related environments, processes, ingredients and technologies, as well as the control in ports, airports and borders[2].
- *Pan American Health Organization (PAHO).* The specialized international health agency for the Americas[3].
- *Brazilian Society of Infectology (BSI).* Entity that promotes the development, scientific exchange and defense of infectious disease doctors, with actions of collective interest, valuing professionals, benefiting the population and contributing to public health in Brazil[4].

For each account, we collected the maximum number of tweets allowed by Twitter API v2 Essential[5]. Table 1 summarizes the dataset statistics.

Table 1. Dataset statistics.

Entity	# tweets	Timespan
MH	3250	Apr/21 to Mar/22
ANVISA	3250	Feb/19 to Mar/22
PAHO	3225	Sep/20 to Mar/22
BSI	564	Sep/19 to Mar/22

Note that the entities were not considered within the same time interval, but in relation to the parameter of about 3200 tweets prior to March 2022. The exception is BSI, which joined Twitter in Sep 2019 and still does not have 3200 tweets posted on its account.

4. Patterns Discovery

We analyze three patterns o entities' communication: social, semantic and temporal patterns. For each pattern, we mainly considered a comparative analysis among entities in order to highlight how communication strategy in social networks is related with organizations interests [Li et al. 2020].

4.1. Social patterns

The goal is to investigate engagement and interaction of entities in the social network, taking into account user and social features. Such social patterns are part of the entity's communication strategy.

---

[1] https://www.gov.br/saude/
[2] https://www.gov.br/anvisa/
[3] https://www.paho.org/
[4] https://infectologia.org.br/
[5] https://developer.twitter.com/en/docs/twitter-api/

Engagement analysis. According to [Zafarani et al. 2014], social engagement corresponds to users' interactions with some content posted in a social network. We define the engagement $E(t)$ of a given tweet t as

$$E(t) = l(t) + rt(t) + m(t) \tag{1}$$

where $l(t)$ is the number of likes in t, $rt(t)$ is the number of times t has been retweeted and $m(t)$ corresponds to the number of mentions contained in t.

Table 2 describes top-3 most engaged tweets of each entity. 83.3% of those are directly related with Covid-19. Let us define the rate of engagement of a given entity o as

$$R(o) = E(t)/F(o) \tag{2}$$

where $F(o)$ is the number of followers of o and t the most engaged tweet of o. We have: $R(MH) = 0.014$, $R(ANVISA) = 0.057$, $R(PAHO) = 0.087$, $R(BSI) = 1.129$. Thus, BSI reached a high rate of engagement, what lead us to conjecture that the clear position of specialists in infectology against Brazil's main COVID-19 strategy advocated by Brazilian president [Freelon and S. 2021] fully impacted popular opinion.

Profiles mentions analysis. Observing the top-10 most mentioned profiles for each entity, only eight of them, 20%, correspond to natural persons. BSI has two doctors figuring among top-10 most mentioned profiles and MH has five natural persons in top-10 ranking, among which four correspond to president supporter profiles. From profiles mentions perspective, in general, official entities do not promote direct mentions of individual profiles in their communication strategy on Twitter. The exception among the entities investigated is MH, with 50% of its top-10 mentioned profiles made up of individuals.

4.2. Semantic patterns

In order to obtain visual insights from the textual content of tweets, we first built a hashtag network for each entity. Then we quantified the entities' variation of themes using semantic networks.

Hashtags network. Hashtags are powerful linguistic resources able to summarize posts' content [Budnik et al. 2019]. In all entities, more than 50% of the tweets contain at least one hashtag. Considering hashtags as summaries, we analyze the main themes covered by each entity during COVID-19 pandemic.

We map the hashtags in a network where edges link hashtags appearing in the same tweet [Wang et al. 2016, Praznik et al. 2019]. The more times two hashtags appear together, the greater the edge weight. Further, we filter nodes with a minimum degree of 5 and detect the communities by modularity metric [Zafarani et al. 2014]. Figure 1 illustrates the resultant hashtags networks.

In all hashtags networks covid19/coronavirus stands out as one of the recurring theme among the hashtags addressed by the entities. The modularity class to which the

Table 2. Top-3 engaged tweets for each entity

Entity	Created at	E	Tweet subject	Related with COVID-19?
MH	07/08/2021	19271	Call for 2nd dose of vaccine	yes
	07/29/2021	17553	Breastfeeding awareness	no
	05/21/2021	15598	Daily preventive care against COVID-19	yes
ANVISA	03/20/2020	10631	Clarification that there are no conclusive studies that prove the use of drugs containing hydroxychloroquine and chloroquine to treat Covid-19	yes
	07/28/2021	4670	Promoted action with a famous skater athlete	no
	01/25/2021	3920	List of vaccines authorized for emergency use	yes
PAHO	12/01/2021	3107	Omicron variant	yes
	05/03/2021	3049	Clarification about vaccines	yes
	05/01/2021	1292	Announcement of the receipt of vaccines through the COVAX Mechanism	yes
BSI	01/14/2021	20785	BSI does not recommend early treatment for COVID-19 with any drug (chloroquine, hydroxychloroquine, ivermectin, azithromycin et al.)	yes
	01/14/2021	5639	(cont.) BSI does not recommend early treatment for COVID-19 with any drug	yes
	01/14/2021	4145	(cont.) BSI does not recommend early treatment for COVID-19 with any drug	yes

respective covid19/coronavirus hashtags belong is strongly connected in relation to the others in all entities' networks.

Considering the number of communities and the volume of hashtags (nodes) in each community, we have MH as the entity with less diversity, having most of the tweets concentrated on the themes covid19, health news, masks, daily care and SUS (Brazil's publicly funded health care system). PAHO network indicates high diversity, addressing not only topics related to COVID-19, but also hepatitis, sifilis, poliomyelitis, cancer and hypertension.
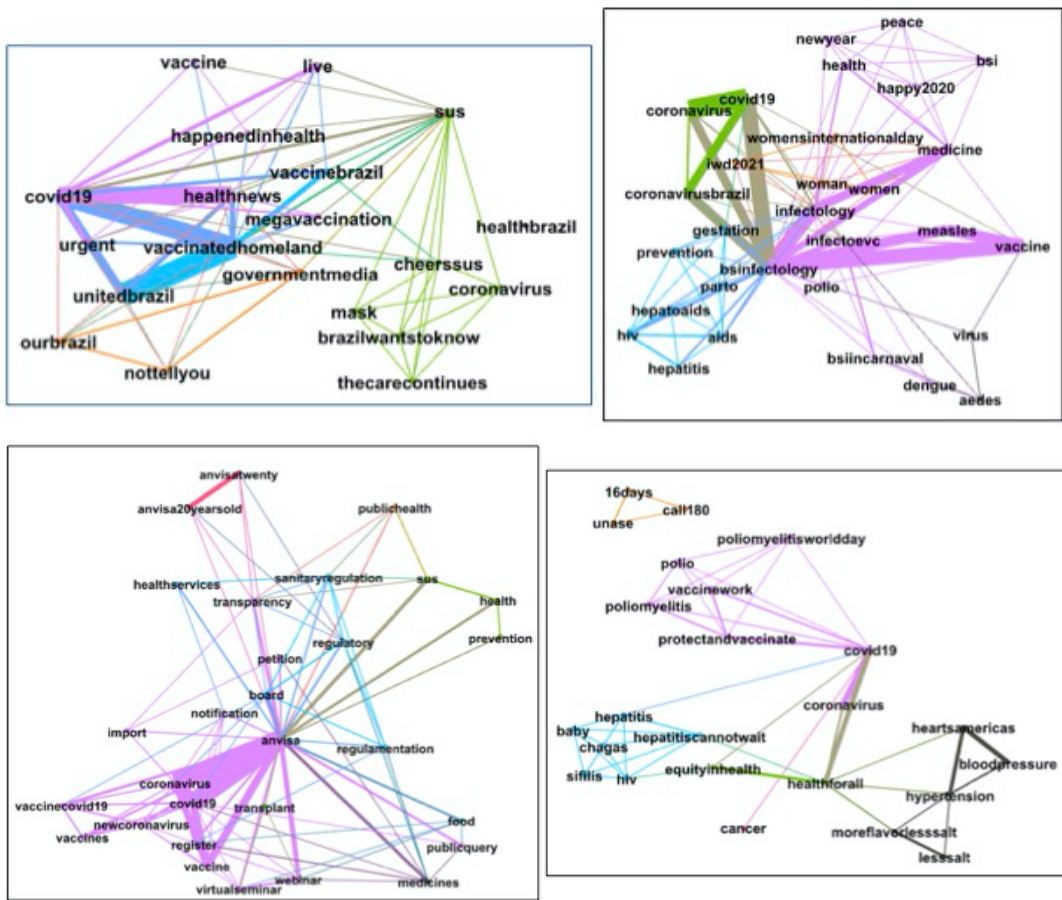

Figure 1. Hashtags network. From left up to right down: MH, BSI, ANVISA and PAHO. Edge strength is the number of times connected hashtags appear in the same tweet. Topology filter considers a minimum node degree of 5. Colors represent modularity class.

Thematic diversity. According to Walter [Walter 2018], to understand and measure diversity, we need to ask not only how many themes are in a system, or even how balanced their distributions are, but also whether and to what extent these themes differ from each other. The following formula, accounting for variety, balance, and disparity, was offered by Stirling [Stirling 2007] based on Rao's diversity coefficient [Rao 1982]:

$$D = \sum_{ij(i \neq j)} (\delta_{ij}).(p_i p_j) \tag{3}$$

where δ_{ij} indicates the distance between each two themes, with p_i indicating the prevalence of category i in the corpus. Thus, we call D_C the thematic diversity factor of a given corpus C. The higher D_C the higher the thematic diversity encompassed in C. The approach based on semantic network analysis was used to infer the thematic diversity factor of each entity.

Each dataset composes a corpus for which we build a semantic network. Nodes are words and a link between two nodes exists if they belong to the same tweet. Thus, the window size W for terms relationship considered the whole extension of a tweet. Then, we run modularity algorithm in the semantic networks in order to obtain the themes (modularity classes) [Zafarani et al. 2014]. From the semantic network we can calculate the thematic diversity of each agency using Eq. 3. Table 3 summarizes results.

Table 3. Thematic diversity of entities.

Entity	Thematic diversity (D)
MH	0.1850
ANVISA	0.1945
PAHO	0.2633
BSI	0.2221

The thematic diversity factor shows that PAHO and BSI have the higher diversity in themes posted on Twitter, confirming hashtags networks visual insights. PAHO is a comprehensive organization, which has not focused its communication strategy essentially on issues related to the COVID-19 pandemic. MH and ANVISA are entities that focused on the topic of COVID-19 and adopted as strategy a maximum communication on social networks related to the pandemic.

4.3. Temporal patterns

We look for temporal patterns by observing entities frequency of posting. From Table 1, clearly MH is the entity with higher number of tweets, reaching an average of 2.5 tweets/day against 0.2, 0.3 and 0.9 from BSI, ANVISA and PAHO, respectively.

Temporal marks analysis. Given our datasets timespan, we selected the main temporal marks related with politics, vaccine and COVID-19 numbers [Saude 2022] (Table 4). Then, we correlate with posting frequency change of each entity. Figure 2 illustrates the percentage of change in the number of tweets over time considering the number of tweets in the first day as baseline.

Let us define direct correlation as a minimum of 200% of change in a maximum time interval of 3 days ahead the temporal mark. Three direct correlations can be identified in PAHO, which refer to *1stVac*, *BraDeadliest* and *Omicron* marks. In MH we can notice a variation in the number of tweets correlated with *Omicron* mark. ANVISA also just correlates with *Omicron* mark. BSI correlates with *3rdMH* mark. According to the results there is not a direct correlation in any entity with marks' categories. For instance, one could conjecture that MH has a sophisticated strategy of communication in the social network, by increasing the number of tweets in front of a politics mark, but our analysis did not confirm such strategy.

Table 4. Temporal marks.

| Date	Name	Mark	Category
2020-02-26	1stBraCase	First COVID-19 case in Brazil	covid numbers
2020-04-16	1stMH	1st Ministry of Health change	politics
2020-05-15	2ndMH	2nd Ministry of Health change	politics
2020-06-02	3rdMH	3rd Ministry of Health change	politics
2020-12-02	1stVac	United Kingdom's approval for the Pfizer–BioNTech vaccine. First country in the Western world to approve the use of any COVID-19 vaccine	vaccine
2021-01-06	GAMMA	Gamma variant first reported in Brazil	covid numbers
2021-01-17	1stBraVac	Brazil approves the first COVID-19 vaccine for emergency use in the country	vaccine
2021-03-23	4thMH	4th Ministry of Health change	politics
2021-03-29	BraDeadliest	Deadliest day of the pandemic in Brazil with 3541 deaths	covid numbers
2021-04-27	PCI	COVID Parliamentary commission of inquiry to investigate government conduct	politics
2021-11-26	Omicron	Omicron variant first reported in South Africa	covid numbers
2022-01-27	BraTopCases	Record of COVID-19 daily cases in Brazil - 228954 new infections	covid numbers
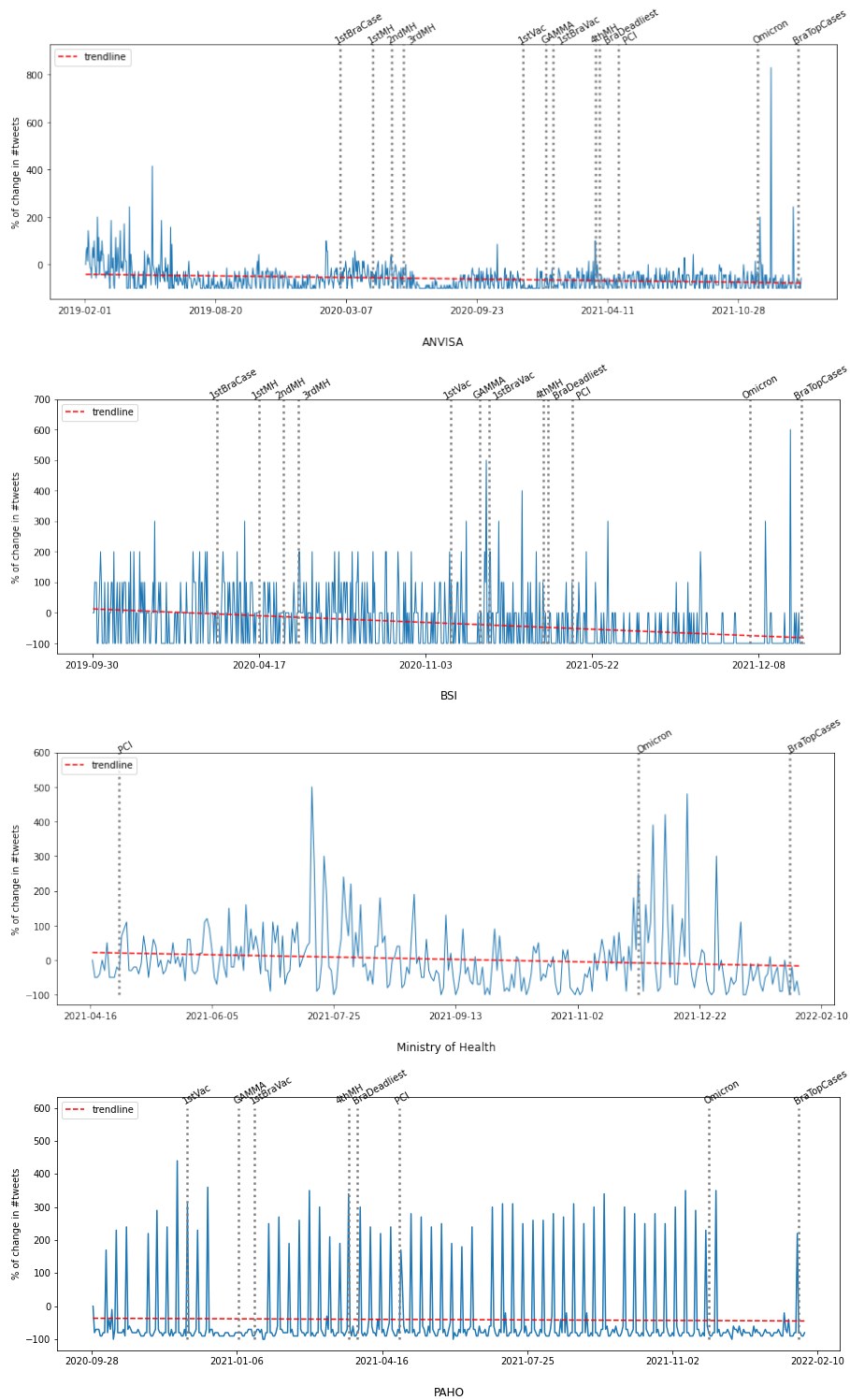
Figure 2. Posting frequency change over time and temporal marks of COVID-19 in Brazil.

Trending analysis. Using least squares polynomial fit of degree 2 we generate a trendline. The goal is to verify if the frequency of posting was affected during pandemic. Clearly, BSI and MH show such behavior as both have a decreasing trendline.

Contrary to what might be thought, two entities remained stable and two entities have decreased the number of posts in Twitter. Such a phenomenon may indicate lack of maturity in the social network communication.

5. Conclusions and Future Work

We perform an innovative analysis of communication strategies in social media of Brazilian official entities. Through the combination of social network analysis techniques, semantic networks and temporal descriptive statistics, we show a deep characterization encompassing social, semantic and temporal patterns of communication.

As results, all entities have most of the posts with higher engagement related with COVID-19 (SO-RQ1). MH is the unique entity with high interaction with individuals, being such individuals president's supporters (SO-RQ2). MH and ANVISA – government entities – have most of the tweets related with pandemic resulting in a low thematic diversity (SE-RQ3). The entities' strategy is to fully focus on COVID-19 themes. No entity showed a significant correlation between the number of posts and key-facts occurred during pandemic (TE-RQ4). BSI and MH showed a downward trend in the number of posts as the pandemic eases (TE-RQ5).

The official entities still have a long way to go to use communication through social networks in a strategic way to bring the population as an ally in their conduct and decisions. All of them proved to be immature, with MH being the entity that stood out the most in adopting a strategy.

Future work encompasses to characterize official entities communication in their websites, by means of their press releases. It is interesting to compare those releases with social media posts, evidencing eventual biases.

References

Budnik, E., Gaputina, V., and Boguslavskaya, V. (2019). Dynamic of hashtag functions development in new media: Hashtag as an identificational mark of digital communication in social networks. In *Proceedings of the XI International Scientific Conference Communicative Strategies of the Information Society*, CSIS'2019, New York, NY, USA. Association for Computing Machinery.

Dai, L., Wang, H., and Liu, X. (2020). St-etm: A spatial-temporal emergency topic model for public opinion identifying in social networks. *IEEE Access*, 8:125659–125670.

Dejard, L., Santos, A., Junior, H., Paulino, R., Figueiredo, K., Costa, F., Vidal, D., Pires, Y., and Seruffo, M. (2021). Social and institutional presence of the presidents of the americas on social media: An analysis of the communication on twitter about covid-19. *International Journal of Advanced Engineering Research and Science*, 8(8):512–526.

Dimitriadis, I., Poiitis, M., Faloutsos, C., and Vakali, A. (2022). Tg-out: temporal outlier patterns detection in twitter attribute induced graphs. *World Wide Web*, pages 1–25.

Dimitrov, D., Baran, E., Fafalios, P., Yu, R., Zhu, X., Zloch, M., and Dietze, S. (2020). Tweetscov19 - a knowledge base of semantically annotated tweets about the covid-19 pandemic. In *Proceedings of the 29th ACM International Conference on Information Knowledge Management*, CIKM '20, page 2991–2998, New York, NY, USA. Association for Computing Machinery.

Duque-Rengel, V. K., Calva-Cabrera, K. D., and Márquez-Domínguez, C. (2021). Government communication on twitter during the covid-19 pandemic in ecuador. In *2021 16th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–6.

Freelon, A. and S., H. (2021). Brazil's main covid strategy is a cocktail of unproven drugs.

Iyer, R., Wong, J., Tavanapong, W., and Peterson, D. A. M. (2017). Identifying policy agenda sub-topics in political tweets based on community detection. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, ASONAM '17, page 698–705, New York, NY, USA. Association for Computing Machinery.

Kwan, J. S.-L. and Lim, K. H. (2021). Tweetcovid: A system for analyzing public sentiments and discussions about covid-19 via twitter activities. In *26th International Conference on Intelligent User Interfaces - Companion*, IUI '21 Companion, page 58–60, New York, NY, USA. Association for Computing Machinery.

Leyman, I. I., Filimonov, V. V., and Ivanov, F. I. (2021). Social media and the strategy of crisis communication in covid-19 pandemic: a case study of komi republic (russia). In *2021 Communication Strategies in Digital Society Seminar (ComSDS)*, pages 191–195.

Li, L., Aldosery, A., Vitiugin, F., Nathan, N., Novillo-Ortiz, D., Castillo, C., and Kostkova, P. (2021). The response of governments and public health agencies to covid-19 pandemics on social media: A multi-country analysis of twitter discourse. *Frontiers in Public Health*, 9.

Li, L., Zhang, Q., Wang, X., Zhang, J., Wang, T., Gao, T.-L., Duan, W., Tsoi, K. K.-f., and Wang, F.-Y. (2020). Characterizing the propagation of situational information in social media during covid-19 epidemic: A case study on weibo. *IEEE Transactions on Computational Social Systems*, 7(2):556–562.

Miletskiy, V. P., Cherezov, D. N., and Strogetskaya, E. V. (2019). Transformations of professional political communications in the digital society (by the example of the fake news communication strategy). In *2019 Communication Strategies in Digital Society Workshop (ComSDS)*, pages 121–124.

Praznik, L., Srivastava, G., Mendhe, C., and Mago, V. (2019). Vertex-weighted measures for link prediction in hashtag graphs. In *2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 1034–1041.

Rao, C. (1982). Diversity and dissimilarity coefficients: A unified approach. *Theoretical Population Biology*, 21(1):24–43.

Saude, S. (2022). Linha do tempo do Coronavírus no Brasil. `https://www.sanarmed.com/linha-do-tempo-do-coronavirus-no-brasil`. [Online; accessed May-2022].

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface*, 4(15):707–719.

Walter, D. (2018). Exploring thematic diversity in news coverage and social media activity of political candidates using unsupervised machine learning.

Wang, Y., Liu, J., Huang, Y., and Feng, X. (2016). Using hashtag graph-based topic model to connect semantically-related words without co-occurrence in microblogs. *IEEE Transactions on Knowledge and Data Engineering*, 28(7):1919–1933.

Zafarani, R., Abbasi, M. A., and Liu, H. (2014). *Social Media Mining: An Introduction*. Cambridge University Press, New York, NY, USA.