

A Study of Rumor Detection based on Social Network Topic Models Relationship

Diogo Nolasco¹, Jonice Oliveira¹

¹Programa de Pós-Graduação em Informática (PPGI) – Universidade Federal do Rio de Janeiro (UFRJ) – Rio de Janeiro - RJ

diogo.sousa@ppgi.ufrj.br, jonice@dcc.ufrj.br

***Abstract.** The rumor detection problem on social networks has attracted considerable attention in recent years with the rise of concerns about fake news and disinformation. Most previous works focused on detecting rumors by individual messages, classifying whether a post or blog entry is considered a rumor or not. This paper proposes a method for rumor detection on topic-level that identifies whether a social topic related to a scientific topic is a rumor. We propose the use of a topic model method on social and scientific domains and correlate the topics found to detect the most prone to be rumors. Results applied in the Zika epidemic scenario show evidence that the least correlated topics contain a mix of rumors and local community discussions.*

1. Introduction

The traffic and discussions generated on social networks have been increasing with their development over time. In recent years, breaking news appears first on microblogs, before making it through to traditional media. A discovery, event, or any information can become viral almost instantly. While the quantity generates massive data for analysis, the quality of information does not become better. All kinds of false information, especially rumor information, have acquired an unprecedented range and permeates most social communities. Consequently, the means of automatically detecting the information credibility and monitor public subjects has been getting increased attention.

Rumor detection is one of the research topics critical to social networks. Rumor itself is often viewed as a tale of explanations of events circulating from person to person and pertaining to an object, event, or issue in public concern [Peterson and Gist 1951]. With the massive amount of data in social networks, it is hard to distinguish reliable information from false information. The rumor diffusion can happen inadvertently or maliciously, so it is common that the message appears to be truthful. The difficulty of the task is so that many news agencies and organizations have departments to assess and inform the public about false information vehiculated on social media. This spread can cause people to make wrong or misinformed decisions and could even harm social stability. Thus, its detection is a major concern for social networks and society.

There are a number of studies on rumor detection, with most of them basically consisting of the task of classifying a message into reliable or unreliable (binary classification). However, there are few studies that try to analyze rumor subjects and detect rumor topics at a coarse-grained level [Cao et al. 2018].

In this paper, we study the problem of automatically detect rumor topics spreading in social media. We propose a method for detecting rumors using topic models to find

unreliable topics given a main event or subject. Our methods use two datasets, the social network dataset, which possibly contains rumors, and a ground truth dataset, that shares the same main event or subject of the social network one but contains curated or technical reliable data. With these topics, we are going to use a topic correlation approach to establish relationships between the two datasets at a topic level. Finally, with the topic correlation, we draw some features that suggest which topics contain rumors and discuss the results to bring more evidence. The main contributions of this work are:

- The cross-topic model method for inferencing rumors.
- A topic correlation approach to detect rumor topics.
- An assessment of the nature of rumor topics included in the relationships between different topics.

The rest of the paper is organized as follows; Section 2 covers related works. Section 3 provides background knowledge on rumors and topic models. Section 4 presents the proposal and methods used. Section 5 presents the results followed by an in-depth discussion. Finally, Section 6 concludes this work.

2. Related Works

There are limited works that study the rumor detection at a topic level. Several survey papers exist on the subject but none of them presents relevant works on the matter [Cao et al. 2018; Zubiaga et al. 2018]. Some of them address the detection in the context of fake news [Ahsan and Kumari 2019; Sharma et al. 2019].

Although not at the topic level, the traditional methods for rumor detection are related to this work. Basically, related studies focus on extracting useful and efficient features for rumor detection. Generally speaking, features for rumor detection can be divided into three types: (1) content-based features; (2) user-based features; (3) propagation-based features.

For content-based features, [Ratkiewicz et al. 2011] identifies misleading political memes on Twitter using content-based features, including hashtags, links, and mention. [Takahashi et al. 2015] computes the ratio of the number of rumor and non-rumor messages vocabulary words as a feature to detect rumors. They found that vocabulary distributions are different between them.

For the user-based features, [Castillo et al. 2011] used a number of user features like age, number of posts, followers, number of friends, and others to detect rumors. Other works like [Al-Khalifa and Al-Eidan 2011] also used several other user attributes.

For the propagation-based features, [Mendoza et al. 2010] analyzed the retweet network topology and the diffusion patterns of rumors and discovered that they are different from traditional news, they also found that rumors tend to be questioned more than news by the Twitter community. There is also the work of [Kwon et al. 2013], that discovered that rumor tweets had more cycle volatility, compared with non-rumor tweets.

3. Background

The proposal presented in this work is based on topic modeling techniques that are used to do rumor detection at a topic level, i. e., detecting rumor topics. Thus, an overview of rumors and topic models are presented in this Section.

3.1. Rumor

There are various definitions of rumor in different areas. There is the view that a rumor is a story or statement in general circulation without confirmation or certainty to facts [DiFonzo and Bordia 2007]. While [Allport and Postman 1947], define it as a story or a statement whose truth value is unverified or deliberately false.

The existence of different definitions makes it hard to compare the effectiveness of different methods for rumor detection. However, there are some typical definitions usually found in the literature (as in Figure 1) of this research area[Cao et al. 2018]:

- **General Rumors:** Concerns pieces of information where the truth value is unverified, both inadvertently or on purpose. Gossips, fake news, and unverified information are examples of this broad concept.
- **Objective Rumors:** Rumors that are verified fake information, i.e., where there exist reliable sources that show that the information is false. Messages that are spread or viral by people that do not know the sources or by people that deliberately spread the rumor.
- **Subjective Rumors:** Rumors where the truth value is determined by the subjective judgment of users. For example, where veracity is based on people's subjective feelings.

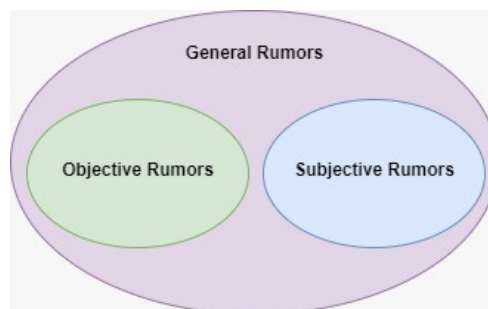


Figure 1. Rumor types

In this work, we study mainly the objective rumors, those that are proven to be false information. These include “fake news”, disinformation and misinformation. This kind of rumor is more easily comparable and can be assessed by using a ground source of truthful values for the information.

3.2. Topic Modeling

Topic models are a suite of algorithms used for discovering the abstract "topics" that occur in a large collection of documents through statistical models. Topic modeling is a frequently used text-mining tool for the discovery of hidden semantic structures in a text body [Blei et al. 2010]. It is considered an unsupervised learning technique that learns classes or topics from previously untrained data, effectively finding patterns on text and relations between terms.

As an unsupervised task, there is no need for training data and the classes called topics on this context to emerge from the application of the statistical model on the collection. In the context of social networks, it is usually unknown to the user which topics

exist in the data beforehand. The social discussions are fluid and dynamic, so the topics of discussion change over time. The very task of classifying those in categories needs an increased effort of domain specialists. New topics emerge every day in social media and the supervised learning techniques are often not feasible to be applied in this scenario.

In this work, we use topic models to find the topics present in a social network and scientific datasets respectively. It is suitable for the use in social networks as it is difficult to know the discussion topics beforehand, and in the scientific dataset, it provides an overview of the research topics that are researched in the academic community.

For this task, we use the Latent Dirichlet Allocation (LDA) model [Blei et al. 2003]. It is one of the most popular and is the source of many recently created models. The LDA was based on two other popular topic models: Latent Semantic Allocation (LSA) [Steyvers and Griffiths 2007] and Probabilistic Latent Semantic Indexing (pLSI) [Blei and Lafferty 2009].

It works by creating two types of multinomial probabilistic distributions. The first is a distribution of a term over the topics where each term has a probability associated with each topic (which represents the relevance of the term for each topic). The second one is a distribution of a document over the topics (which represents the relevance that each topic has in the document). This second distribution can be found by using the first one as a document is viewed as a collection of terms. Based on these distributions one can group the documents according to the probabilities associated with each topic.

Thus, given a vocabulary $V = \{w_1, w_2, \dots, w_{|V|}\}$ consisting of all terms that exist in the collection $C = \{d_1, d_2, d_3 \dots\}$, where w_n is the n th-term of the vocabulary and d_n is the n th-document of the dataset, we define a topic model from the algorithm as:

DEFINITION 1. A topic model θ in C is a probability distribution of terms such that:

$$\theta = \{p(w_1|\theta), p(w_2|\theta), \dots, p(w_{|V|}|\theta)\} \quad (1)$$

and

$$\sum_{w \in V} p(w|\theta) = 1 \quad (2)$$

A ‘‘Zika’’ topic, for example, would assign higher probabilities to words like ‘‘epidemic’’, ‘‘vaccine’’ and ‘‘victims’’ and lower probabilities to less relevant words such as ‘‘fun’’ and ‘‘music’’. Basically, a topic is a probabilistic distribution where words with high probability are more relevant to the assigned topic, while low probability words are irrelevant or stop words.

4. Proposal

The objective of the proposal is to find objective rumors from topic correlation obtained from social network and authoritative dataset topics respectively. In this work, we use a scientific dataset as our authoritative dataset, i.e., a dataset whose topics are reliable and verified. Our study is conducted mainly in the Zika epidemic scenario to find rumors that were spread through social media and thus, a scientific dataset provides the topics and labels that could be used to verify information credibility of what is discussed about the disease in the social networks.

The proposal technically consists of three different tasks performed in those datasets: (1) Topic Detection; (2) Topic Labeling, and (3) Topic Correlation. Topic detection consists in extracting topics from collections. Topic Labeling is used to analyze, correlate, and provide a comprehensive overview of the topics extracted. Topic Correlation is the task where the relationships between the two datasets are established and where we discuss the findings concerning rumor detection.

The techniques used for Topic Detection and Labeling are an implementation of the ones described by [Nolasco and Oliveira 2018], as they prove successful and efficient when applied in the dataset types used in our experiments, namely, in the academic and social network domains. The Topic Correlation approach is based on the Kullback-Leibler divergence [Kullback and Leibler 1951], a measure used to associate two probabilistic distributions and that we use on the distributions resulted from topics models.

As the results and findings depend on the methods used, we give a brief explanation of these three tasks in the next sections to provide the necessary technical background to understand our study.

4.1. Topic Detection

In this task, we are using the Latent Dirichlet Allocation (LDA) to detect topics from the textual collections. For this, we also use the topic detection methods described in [Nolasco and Oliveira 2018].

Usually, a topic model algorithm (including LDA) expects two main input parameters, a number of topics to find K and a collection of N documents. In the social network domain, the topics are not known by the user and thus the K value can vary according to the dataset. The topic detection method used here circumvents this by automatically estimating the best K value for a particular dataset based on Stability Analysis.

After execution, this algorithm results in a series of topics Θ where each topic is a set of terms with associated probabilities.

4.2. Topic Labeling

While the topics are just probabilistic distributions of terms over the topics, it is essential to the rumor analysis that these distributions could translate into textual themes or subjects for comparison and assessment. For this task, we use the topic labeling method presented in [Nolasco and Oliveira 2016]. It was made to work with topic models with successful results when labeling social and science topics. It consists of three steps illustrated by Figure 2, they are: 1) Candidate Label Selection, where keywords and keyphrases are extracted from the topic relevant documents; 2) Score and Ranking, where the keywords are ranked based on a function score and 3) Label Selection. Where the labels are attributed to each topic.

4.3 Topic Correlation

For comparing the potentially unreliable social topics and the scientific truth source topics, we are using the correlation between the two topics as a means to validate the topic's reliability. We calculate it using a distance function based on the KL-Divergence.

The KL-Divergence is one of the most used when comparing two probability distributions (in this case the topics) but it is not asymmetric measure, i.e., inverting the parameter order results in different metrics values. Since the topic correlation needs to be the same on both directions we use a distance based on that divergence defined by its creators [Kullback and Leibler 1951] and calculated as:

$$\text{KL-Distance}(\theta_1, \theta_2) = D_{KL}(\theta_2 || \theta_1) + D_{KL}(\theta_1 || \theta_2) \quad (3)$$

Which is symmetric and nonnegative and where D_{KL} is the original KL-Divergence between two topics. A low total value means that little information is lost when comparing the two distributions while a high value means that they are very dissimilar.

5. Results

We conduct an experiment using the Zika epidemic scenario to showcase how the topic correlation between social network and science topics can be used to detect unrelated or unreliable information about the event and the disease itself. We also use the topics to differentiate general rumors, local discussions, and fake news.

The evaluation is made using two datasets, a Twitter dataset of Zika related posts and a PubMed corpus of Zika related articles. The scenario covered by these datasets is relative to the context of the Zika epidemic from 2015 to 2016, which contains a variety of topics like reports, propagation to various countries, associated diseases, and influence on the 2016 Olympic Games organization.

A qualitative study is conducted to analyze the social topics and its correlation to the science topics as a means to detect rumors about the disease. Since we work with objective rumors, we are using some “ground truth” data to support our claims and findings. Specifically, we are going to use the timeline of the Zika epidemic communications report [Fundação Oswaldo Cruz. and Araujo 2007], the news reported by media for social topics, and the World Health Organization list of rumors about the Zika virus [WHO 2016]. The timeline contains the main events reported to the public such as the virus identification and the outbreak declaration by authorities, while news can be used to verify associated secondary events and subjects such as local efforts to combat the epidemic and ways of preventing the infection. The list of rumors contains a compilation of the all the false information spread or divulged during the outbreak and can be used to verify reliability.

5.1. Datasets

The database for this experiment was made extracting posts from around the world with the #zika “hashtag” and articles from the PubMed database containing the keyword Zika. The term is popular in both domains and has little ambiguity, so the addition of other terms could introduce more noise to the data. The time span of the datasets covers documents created from May. 2015 to Dec. 2016, to cover the lifespan of the epidemic. A total of 85,601 tweets and 1769 articles were retrieved. A preprocessing was made in these data by removing emotes, links, and accents from the text.

5.2. Execution

The proposal consists of three tasks, Topic Detection, Labeling, and Correlation. In this Section, we present the details of each task execution.

1	ZIKV, virus, infection	Zika virus prevention, travellers concern, emerging infectious diseases
2	Zika, emerging doorstep, outbreak	Congenital fetal malformations, pregnant women, congenital microcephaly
3	Brazil, Bahia, Americas	Zika virus infection, emergency department, ZIKV IgM
4	Following dengue, dengue spread, zika	Dengue Virus, human semen, pregnant
5	Zika virus infection, co-infection, new threat	Counter zika virus, diagnostic challenge, detecting
6		Neurologic inhibition, inflammatory, imported arbovirus
7		Mosquito-borne arboviruses, African, saliva

Next, we use a topic correlation on these topics to calculate a similarity between them to find the reliability of the information. Specifically, we use the symmetric KL-Distance to measure the distance between social and science topic distributions. For this analysis, we calculate the correlation between social and science topics from the same period.

Table 3 shows the results of the two periods, both as a heat map. Results are truncated with a precision of two. The cell colors of the heat map are such that the green indicates the lowest distances or most similar while the red indicates the most distant in comparison with the others and to facilitate interpretation.

Table 3. KL-Distance between social and science topics over two periods

Period 1 – From May. 2015 to Feb. 2016									
		Science Topics							
Social Topics		1	2	3	4	5	Avg		
	1	4.02	2.28	4.71	1.03	2.48	2,90		
	2	2.44	1.99	1.33	1.47	1.98	1,84		
	3	1.21	1.34	3.17	4.93	1.61	2,45		
	4	2.86	3.96	4.41	2.87	1.09	3,04		
	5	4.58	3.98	2.71	4.13	2.75	3,63		
	6	1.46	4.30	3.14	1.47	2.16	2,51		
	7	0.74	2.34	0.67	2.71	1.72	1,64		
	8	3.44	4.26	4.92	3.74	2.80	3,83		
Period 2 – From Mar. 2016 to Dec. 2016									
		Science Topics							
Social Topics		1	2	3	4	5	6	7	Avg
	1	4.90	4.46	2.68	0.52	2.11	4.71	1.38	2,97
	2	4.61	1.12	3.55	0.70	3.75	4.22	2.14	2,87
	3	2.44	4.60	4.87	0.92	3.01	0.54	2.49	2,70
	4	2.60	2.18	2.65	2.50	2.97	2.89	2.61	2,63
	5	0.51	3.43	1.37	3.98	4.49	1.30	2.65	2,53
	6	3.13	4.50	3.96	1.93	3.92	1.41	3.27	3,16
	7	1.18	1.11	1.11	4.25	3.03	3.43	3.59	2,53

	8	3.54	3.59	4.70	2.45	2.84	3.84	4.01	3,57
--	---	------	------	------	------	------	------	------	------

5.3. Discussion

For an overview of the main topics detected, we can cite at the first period of social topics, topic 1 that is related to a rumor of a possible relationship between a company (Monsanto) and microcephaly. Topic 2 is about the epidemic affecting the Olympic Games preparation to be held in Brazil. Topics 3, 4, and 7 which are related to the WHO declaration of the epidemic as a Public Health Emergency of International Concern, the news about travel warnings for pregnant women and the cases related in various countries as the disease spread respectively. Topics 5 and 8 are related to case discussions in other languages (Spanish and Portuguese specifically) on the most affected countries.

In the second period of the social topics, we have topic 5, which contains posts reporting the spread of the virus to other countries not initially affected. Topics 3 and 7 are related to measures to prevent the contamination and dissemination of the disease. Topic 9 is a local discussion of Zika in Brazil and topic 1 has the main information about the disease. Finally, topic 4 shows population concerns about the Matthew Hurricane that hit Central and North America in that period.

For the science topics, topics 1 and 5 suggests that the academic community was aware of the initial stages of the epidemic as the labels are associated with infection vectors, how it is transmitted, and the virus threat. Topic 2 refers to the time when the researchers already considered the disease an outbreak. Topics 3 and 4 refer to studies about the first cases in Brazil (Particularly Bahia, a state where local researchers identified the Zika virus for the first time in the region) and the relationship between Zika virus (ZIKV) and Dengue virus (DENV), another virus that shares some of the same transmission vectors.

In the second period of the scientific topics, we have discussions focused on prevention in topic 1, new forms of transmission like saliva in topic 7, and human semen in topic 4. Topic 2 shows the discovery of the association between the occurrence of microcephaly and Zika virus infection in pregnant women. Topic 6 suggests research between Zika and neurological problems caused by it like the Guillain–Barré syndrome.

Since the similar topics suggest a correlation between the social topics and the science topics, which is a reliable source, we will focus our discussion of rumors evidence in the most distant topics. Based on the average values of the heat map, we analyze the ones with the highest average distance over all topics (values above 3) while still discussing particular cases when necessary.

Starting at the first period, the topic 1 shows a rumor that says that a Brazilian company (Monsanto) pesticides were the true cause of microcephaly in children, a disease associated with the Zika virus in many kinds of research and with clinical proof. This topic presented one of the highest distances overall (in relation to science topic 3). The topic 3 has a high distance related to science topic 4 too, but it seems like a dissimilarity issue because the first is related to the outbreak declaration and the latter with the relationship between the Dengue virus and the Zika virus.

Topic 4 has a high average and its related to local concerns about the virus spread reaching Miami and concerns about the Mathew hurricane, but it seems that this high average is caused by the topic being a local discussion of the Miami city and thus,

unrelated to any science topic. This seems to be the case with the topic 5 and 8 too. Both have high averages and are linked to local discussions (Spanish and Portuguese speaking communities). Also, a curious case can be seen in topic 7 which has a low average but it is related to a car brand named “Zica” from the Tata company in India. The high average could be explained by the hashtags and the promoted terms used like “#zika”, “#zika virus”.

Finally, in the second period, we have Topic 1 that is dissimilar with the scientific topics concerning travels and neurologic diseases, but it has a relation with a rumor of vaccines causing babies to be infected with the disease.

Topics 6 and 8 have the highest average distances. Topic 6 is related to the rumor that a negative mutation in mosquitoes was causing the zika spread in the Americas. Topic 8 is a local community discussion in Brazil made in Portuguese.

The lowest averages are seen in topics 5 and 7. The first contains a mix or reliable information of the virus spreading to various countries along with some rumors of cases in countries that were not proven. The latter is associated with repellents marketing and relate to rumors of specific brands being more effective in combating the Aedes mosquito.

We present a summary of rumors found in the topics and rumors found in the reference rumor list [WHO 2016] to facilitate the discussion comprehension in Table 4.

Table 4. Comparison of WHO rumors and rumor topics.

Rumor Topic	WHO Rumor
Monsanto pesticides related to microcephaly (Period 1, Topic 1)	Evidence that pyriproxyfen insecticide causes microcephaly
Tata car named “Zica” that generated confusion with the virus name (Period 1, Topic 7)	Vaccines cause microcephaly in babies
Rumors of first cases in various countries (Period 2, Topic 5)	Most symptoms of Zika virus disease are equal from those of seasonal flu
Negative mutation in mosquitoes causing the spread of the virus (Period 2, Topic 6)	Bacteria used to control the male mosquito population are spreading Zika further
Repellent information and propaganda (Period 2, Topic 6)	Some repellents work better against the Aedes mosquito
	Evidence that Zika virus and its complications are linked to releases of genetically modified mosquitoes in Brazil
	Evidence that sterilized male mosquitoes contribute to the spread of Zika

The results of the analysis show that there were 5 topics related to rumors, with 3 of them related to rumors in the reference list and just one rumor in the list is not related with any topic (“Most symptoms of Zika virus disease are equal from those of seasonal flu”). Most of the rumor topics found had the highest averages (above 3,00 in both scenarios) or lowest averages (1,64 in the first scenario and 2.53 in the second) overall. Although one can conclude that the two extremes of the averages are evidence of false information there were also local discussion in other languages or limited to certain areas that had high averages too. Possibly this was caused by the different languages and thus different terms causing an increased distance over the topics.

Some other studies are necessary to verify that the patterns found here can be generalized in other cases and to study the types of rumors or fake information found in

the data. It seems that the topics more unrelated to authoritative sources correspond to local discussions or rumors, while topics that are related to anything could be a deliberate fake or propaganda when used in marketing like repellent sellers or cars.

6. Conclusions

The detection of false information had always been a topic that attracted much attention over the years. The modern threats of social networks, “fake news”, and the spread of viral false or unverified information demand different methods to detect rumor and analyze them to cope with the dynamic nature of the online communities.

Most of the works in the area have been trying to detect false messages but few of them were concerned with a topic level detection, i. e., which subjects are related to rumors. This work proposes a topic model approach to detect these subjects and uses topic correlation via KL-Distance between two domains: A social network domain and a “ground truth” domain where the topics are verified (in our case a scientific domain).

We conduct a study of how the relationship between topics of these different domains relate to each other in the Zika epidemic scenario and how they relate to verified rumors from the World Health Organization. We compare the results found and our findings suggest evidence that topics that have very low or high correlation could indicate rumors or limited local discussions (most in other languages). Our main contributions are thus: 1) The cross-topic model method for inferencing rumors; 2) The topic correlation approach to detect rumor topics; and 3) The analysis of the metric differences between localized discussions, rumors, and the behavior of rumor topics.

This work open opportunities in the detection of fake news, disinformation or topics and labels that are indicative of unreliable information that could be used to assess message reliability, the impact of a rumor in the most important subjects of the social network and the very detection of rumor topics that are widespread in the social networks.

References

- Ahsan, M. and Kumari, M. (2019). Rumors and their controlling mechanisms in online social networks: A survey. *Online Social Networks and Media*
- Al-Khalifa, H. S. and Al-Eidan, R. M. (2011). An experimental system for measuring the credibility of news content in Twitter. *International Journal of Web Information Systems*, v. 7, n. 2, p. 130–151.
- Allport, G. and Postman, L. (1947). The psychology of rumor.
- Blei, D., Carin, L. and Dunson, D. (2010). Probabilistic topic models. *IEEE Signal Processing Magazine*, v. 27, p. 55–65.
- Blei, D. and Lafferty, J. (2009). Topic models. : *classification, clustering, and applications*,
- Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, v. 3, n. 4–5, p. 993–1022.
- Cao, J., Guo, J., Li, X., et al. (10 jul 2018). Automatic Rumor Detection on Microblogs: A Survey.
- Castillo, C., Mendoza, M. and Poblete, B. (2011). Information credibility on Twitter. In

Proceedings of the 20th International Conference Companion on World Wide Web, WWW 2011.

DiFonzo, N. and Bordia, P. (2007). *Rumor psychology: Social and organizational approaches.*

Fundação Oswaldo Cruz., R. and Araujo, I. S. (2007). *A mídia em meio às 'emergências' do vírus Zika: questões para o campo da comunicação e saúde.* Fundação Oswaldo Cruz. v. 10

Kullback, S. and Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, v. 22, n. 1, p. 79–86.

Kwon, S., Cha, M., Jung, K., Chen, W. and Wang, Y. (2013). Prominent features of rumor propagation in online social media. In *Proceedings - IEEE International Conference on Data Mining, ICDM.*

Mendoza, M., Poblete, B. and Castillo, C. (2010). Twitter under crisis: Can we trust what we RT? In *SOMA 2010 - Proceedings of the 1st Workshop on Social Media Analytics.*

Nolasco, D. and Oliveira, J. (2016). Detecting knowledge innovation through automatic topic labeling on scholar data. In *Proceedings of the Annual Hawaii International Conference on System Sciences.*

Nolasco, D. and Oliveira, J. (2018). Subevents detection through topic modeling in social media posts. *Future Generation Computer Systems,*

Nolasco, D. and Oliveira, J. (2020). Mining social influence in science and vice-versa: A topic correlation approach. *International Journal of Information Management*, v. 51.

Peterson, W. A. and Gist, N. P. (sep 1951). Rumor and Public Opinion. *American Journal of Sociology*, v. 57, n. 2, p. 159–167.

Ratkiewicz, J., Meiss, M., Conover, M., et al. (2011). Detecting and Tracking Political Abuse in Social Media. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media.*

Sharma, K., Qian, F., Jiang, H., et al. (2019). Combating fake news: A survey on identification and mitigation techniques. *ACM Transactions on Intelligent Systems and Technology*

Steyvers, M. and Griffiths, T. (2007). Probabilistic topic models. *Handbook of latent semantic analysis,*

Takahashi, B., Tandoc, E. C. and Carmichael, C. (2015). Communicating on Twitter during a disaster: An analysis of tweets during Typhoon Haiyan in the Philippines. *Computers in Human Behavior*, v. 50, n. 2015, p. 392–398.

WHO (2016). WHO - Dispelling rumours around Zika and complications. <http://www.who.int/emergencies/zika-virus/articles/rumours/en/>, [accessed on Apr 29].

Zubiaga, A., Aker, A., Bontcheva, K., Liakata, M. and Procter, R. (1 feb 2018). Detection and resolution of rumours in social media: A survey. *ACM Computing Surveys.* Association for Computing Machinery.