Covid or not Covid? Topic Shift in Information Cascades on Twitter

Liana Ermakova
HCTI - EA 4249
Université de Bretagne Occidentale
Brest, France
liana.ermakova@univ-brest.fr

Diana Nurbakova
LIRIS UMR 5205 CNRS
INSA Lyon
University of Lyon
Villeurbanne, France
diana.nurbakova@insa-lyon.fr

Irina Ovchinnikova
Sechenov First Moscow State
Medical University
Moscow, Russia
ovchinnikova.ig@1msmu.ru

Abstract

Social media have become a valuable source of information. However, its power to shape public opinion can be dangerous, especially in the case of misinformation. The existing studies on misinformation detection hypothesise that the initial message is fake. In contrast, we focus on information distortion occurring in cascades as the initial message is quoted or receives a reply. We show a significant topic shift in information cascades on Twitter during the Covid-19 pandemic providing valuable insights for the automatic analysis of information distortion.

1 Introduction

Social media is a valuable resource for all sorts of information. However, its power to shape public opinion can provoke serious societal issues such as misinformation. Words or actions by a *Public Figure (PF)* generate 69% of misinformation in discussions with ordinary users, while PFs themselves are responsible for 20% of the messages containing distorted information (Brennen et al., 2020). PFs post tweets that are likely to be shared by their followers (Romero et al., 2011), thus generating information cascades. The periodic repetitions provoke the mutability of information diffusion in the political domain (Shin et al., 2018). Recent studies show a similarity of distorted information dissemination during the pandemic to the distribution of political misinformation (Pennycook et al., 2020; Pennycook and Rand, 2018). The repetitions of rumours about conspiracy theories associated with Covid-19 led to the mutability of information; nevertheless, many users ridiculed these theories while repeating the rumors (Ahmed et al., 2020). During the Covid-19 pandemic, users look for medical information in PF feeds and follow personal stories of infected people who share unverified information because complicated medical texts deter lay readers (Ribeiro et al., 2019). Mass medical information sharing generates cascades where the probability to distort initial information increases due to omissions and paraphrases. As medical discourse is sensitive to any changes in terminology and text structure made by incompetent people (Nye et al., 2018), the impact of medical misinformation on social behaviour means that there is a pressing need to understand how it circulates on social media. To reveal crucial issues about Covid-19 that are of importance for lay people we need to understand topic shifts occurring within information cascades about the pandemic. Such understanding allows us to discover a particular lack of medical information and demand for clear explanation of the most important public problems of the current pandemic. In this paper, we present a preliminary study on medical information distortion occurring in cascades on Twitter due to topic shift. Several studies have focused on misinformation during the Covid-19 pandemic (Pennycook et al., 2020; Cuan-Baltazar et al., 2020; Nurbakova et al., 2020; Smith et al., 2020; Kouzy et al., 2020; Krause et al., 2020; Tasnim et al., 2020; Erku et al., 2020), but to the best of our knowledge, they assume that the initial message in a cascade is fake and do not study the mechanism of medical information distortion. We aim to answer two research questions: **RQ1**: *What are PF tweets on healthcare topics that generate information cascades?* **RQ2**: *How does a transformation of the initial tweet involve misinformation?*

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: http://creativecommons.org/licenses/by/4.0/.

Proceedings of the 3rd International Workshop on Rumours and Deception in Social Media (RDSM), pages 32–37
Barcelona, Spain (Online), December 12, 2020.

2 Materials and Methods for Detection of Topic Shift and Information Distortion

We collected 10M tweets in English about the controversy surrounding Covid-19 medical treatment published between 30/03/20-13/07/20 by querying Twitter API with the keywords, such as [*chloroquine*, *hydroxychloroquine*, *HCQ*, *Hydroxychloroquinum*, *azithromicyn*, *Raoult*, *remdesivir*, *tocilizumab*][1](Noel et al., 2020). The data contained 141,866 original tweets, the rest are retweets[2]. As we focused on the analysis of information cascades ($2 * 10^4$), we only considered a subset of the dataset. First, we determined the *initial tweets of the cascades* among the union of 10^3 the most retweeted and 10^3 the most quoted tweets (1,356 unique tweet IDs). Then we added *cascades hops*, i.e. tweets with fields quoted_status.id or in_reply_to_status_id containing initial tweet IDs. The maximal cascade depth with the initial tweet in the resulting dataset is 10 (see examples in Fig.1). For further analysis, we considered the field text.

We analysed **topic shift** within information cascades by comparing (1) two neighbouring hops within a cascade $\Delta^{(i-1)}$ (Fig. 2(a)) and (2) each hop within a cascade with the initial tweet $\Delta^{(0)}$ (Fig. 2(b)). To analyse topic shift, we encode tweets with the state-of-the-art sentence embedding model USE (Universal Sentence Encoder) (Cer et al., 2018). Then, we computed the cosine similarity between USE embeddings (Singhal, 2001) and transformed it into distance by subtracting the obtained values from 1.

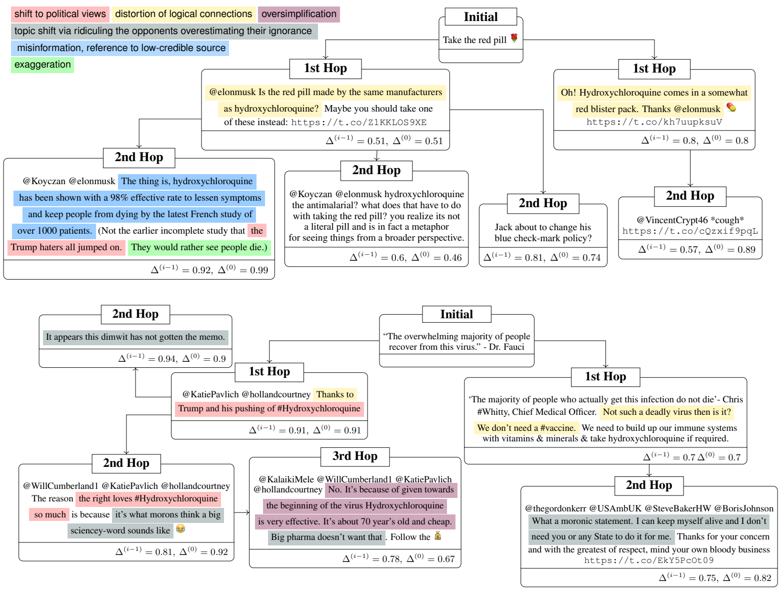


Figure 1: Examples of cascades and information distortion within them

(a)

(b)

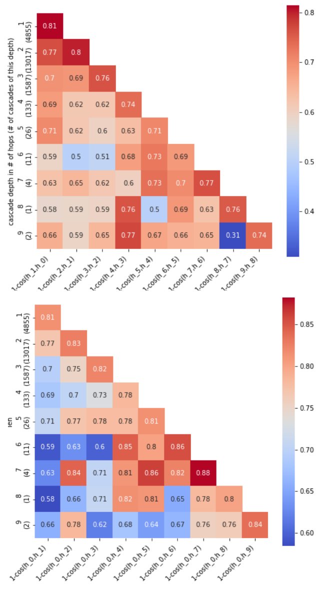


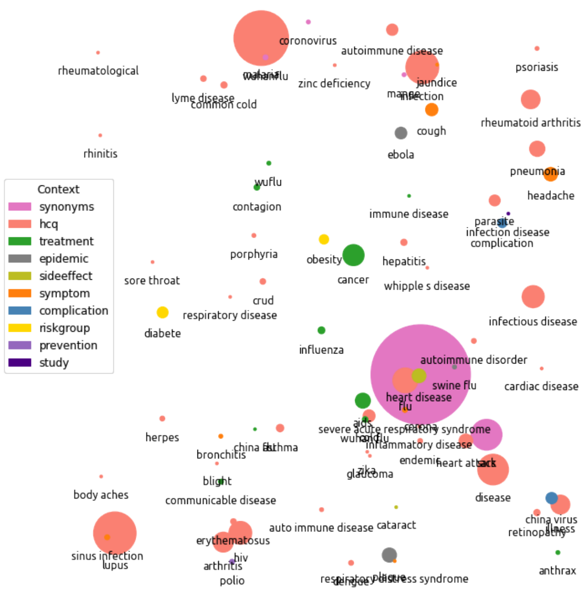
Figure 2: Distances ($1 - cos$) between: (a) neighbouring hops, (b) hops and initial tweets

To identify information distortion types in cascades, we **manually performed semantic analysis** of tweet content. We examined key term distribution in cascades, explored their context in tweets and verified logical relations among medical terminology (see Table 1). The context analysis helped to recognise term substitutions and the substitution analysis to detect information distortion w.r.t. the initial tweet.

Our analysis also leans on **topic modelling**. We used Latent Dirichlet Allocation, LDA (Hoffman et al., 2010) from the `scikit-learn` tool. As tweets are short, we considered only the first topic. (Abd-Alrazaq et al., 2020) distinguish four main discussion themes on Twitter during the current pandemic: origin of the virus; its sources; its impact on people, countries, and the economy; and ways of mitigating the risk of infection. This set lacks medical disease description and ways to treat Covid-19 (symptoms, diagnosis, drugs, etc). Thus, we were also interested in **references to other disease related terms (DRT)** within cascades, as they can indicate distortion. To examine them, we extracted a list of hyponyms of the word *disease* from WordNet corpus accessed via `NLTK` library, to which we added terms like *plague,*

[1]The query was updated throughout the collection period based on new information about Covid-19 and its possible treatment.
[2]Some tweets attained 422K retweets, e.g. `https://bit.ly/32TPeSt` or `https://bit.ly/3gYDcfE`

Table 1: Disease related terms (DRTs) in cascades and their prevailing context



Visualisation of mentions of DRTs	Context and terms
	hcq = {*hydroxychloroquine hcq hc azithromycin chloroquine zpack z.pac antimalarial zinc sulfate zithromax*}
	symptom = {*symptom congestion bood cough aches lungs fever antibody headache mucus signs asymptomatic respiratory shortness.of.breath symptom.free back.pain diarrhea nausea*}
	treatment = {*treatment cure curing treat pill medicament remedy therapy drug acetaminophen prescribe prescription breathlessness medications diagnos recovery*}
	prevention = {*vaccin mask hand.wash distanc prevention detection test cover.*mouth self.isol prophylaxis immunity stay-home staying.home stay.home prophylactic serum.test preventative*}
	study = {*study control.group randomi.ed research treatment.group trial expert scientific.evidence success.rate science protocol effective.rate placebo*}
	complication = {*ventilator complication transfusion coma hospitalization death severe.case critically critical.condition severe urgent.care.center emergency icu intensive.care.unit*}
	epidemic = {*epidemic pandemic plague zika ebola lockdown locked.down outbreak swine.flu*}
	side effect = {*side.effect heart.disease cardiac.problem hallucination psychiatric.symptom vision.loss vomiting loss.of.appetite dizziness slow.heartbeat heart.failure swelling.ankles*}
	risk group = {*elderly diabete obesity obese asthma comorbidity 60.plus 60.year*}
*radius is proportional to # of mentions of a term	synonyms = {*corona wuhan.virus wuhan.disease sars.cov.2 covid19 covid c19 coronovirus chinese.flu china.flu cv.19 sars.cov sars chinese.plague coronahoax wuhanflu*}

swine.flu, bird.flu, hiv, malaria, cough, wuhanflu, sars, cardiac.disease, china.flu, covid, coronovirus, cancer, obesity, diabete. We checked the appearance of these terms in the texts. In addition, we investigated the context in which these DRTs were mentioned such as: *hydroxychloroquine (HCQ), symptom, treatment, prevention, propagation, study, complication, epidemic, side effect, risk group, synonyms, plague reference, other issues*. Each context is defined by a set of terms (see Table 1). We then looked at the co-occurrence of DRTs and context terms in a tweet in order to predict the DRT context. This allowed us to gain a better understanding of topic shift related to references of other diseases.

3 Results

We identified the top-10 words characterising the first topic of each hop using LDA. We represented each hop as a binary vector built over the words of all hops. For visualisation, we applied Principle Component Analysis (PCA) (Tipping and Bishop, 1999) with two variables (see Fig. 3). Note that the first three hops are rather distant from the initial tweet, while the fourth hop is quite close to the initial tweet (its role is not clear yet). Based on our analysis, 3,939 out of 21,585 unique tweets of cascades contain DRTs. Table 1 summarises the frequency of the terms and the contexts in which they were primarily used[3]. HCQ is the most used context. It brings up DRTs such as *lupus*, *rheumatoid arthritis*, *malaria*, *heart attack*, *respiratory disease*, etc. Chloroquine often substituted its derivative HCQ, as in the E. Musk's cascade about the research of French microbiologist D. Raoult. The terms *corona* and *sars* are often used to refer to Covid-19. As for the *treatment* context, the most typical DRTs are *cancer*, *aids*, *influeza*. Note that a given term is often used in multiple contexts but here, we report the dominant one. Thus, HCQ was discussed in the cascades regardless of their initial tweet and the PF who initiated them.

[3]We intentionally excluded the term '*covid*' from the plot, as it is the main topic of the cascades (mentioned in 2,200 texts).

Information cascades are rarely evoked by *healthcare professionals* (HS) (identified by their profile information) since they are less active on Twitter than politicians. HS quotes are often misinterpreted and their research results are misrepresented. Though most of the cascades in which HS took part were initiated by *journalists*, HS tweets were able to **terminate an information cascade**[4] by providing relevant information and ending discussion. Cascade terminators often come from the professional community (Ziegelmeyer et al., 2010). In Fig. 2b, distances between initial tweets and the last hops of cascades show essential semantic differences revealing topic shifts. Hops in 'heads' of deep cascades are closer to their initial tweets than those in 'tails'. As distances between neighbouring hops show more similarity to each other than to their initial tweet (Fig. 2a), hops accumulate information mutability. The last hop is able to exhaust the cascade new topic.

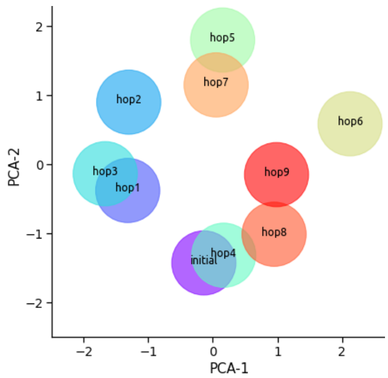


Figure 3: PCA on the first LDA topic

Medical information is distorted via **erroneous logical conclusions** and mental operations of **oversimplification**, **overgeneralisation**, **exaggeration**, **substitution**, **omission of facts**, **insertion of erroneous conclusions**, **misuse of medical concepts** and **distortion of their connections**. The instances of the distortions occur in comments on the initial tweet and hops of cascades generated by the PF tweets. In fragments of cascades generated by comments on PF tweets in Fig. 1, distortion and **misinformation** appear due to **oversimplification** and **distortion of logical links**. **Omission of facts** is connected to **oversimplification**: HCQ efficacy in the Covid-19 treatment depends on patient anamnesis. **Misinformation** appears when a user did not provide a link to results of a French study he referred to while reacting on a red pill. The red pill meme reveals an unpleasant truth and is derived from a scene in the film *The Matrix*; an insertion of an erroneous conclusion occurred in comments where red pill is associated with HCQ. Ordinary users **exaggerate** consequences of government decisions. They **politicise** and **criminalise** these actions shifting the topic to political and business disputes (*RT @TribeforFreedom: Cuomo is dedicated to a vaccine (Bill Gates) So he does not allow the use of hydroxychloroquine*). **Overgeneralisation** often appears in references to a personal experience when a single fact was considered as a trend[5].

4 Conclusions

A topic shift is like the broken telephone effect (Boyd et al., 2010) when the message is altered during transmission, a typical cascade feature (Ribeiro et al., 2019). Thus, we showed that through this effect, PFs influence misinformation distribution on Twitter regardless of the quality of the information in their initial tweet. Users consider HSs as sources of 'raw information', which needs PF evaluation and approval. An interesting finding is that medical experts were able to stop the development of cascades by providing their factual and knowledgeable opinion. Intellectuals have the most influence on ordinary users' evaluation of the drugs efficacy research that is similar to the results of the cascades study in (Cha et al., 2010). We see the effect in the cascade evoked by comments on Musk's tweet. In the contexts of DRTs, we discovered the instances of medical information distortion. In contrast to previous works mainly focused on the initial spreading of fake news (Ahmed et al., 2020; Brennen et al., 2020), here we clarified the mechanism of the medical information distortion during the Covid-19 pandemic by analysing topic shifts within cascades. Usually, the medical topic is shifted to political and business disputes. We showed that cascade hops accumulate mutability of information. We found that after a noticeable topic shift occurring in the first 3 hops, there is a return to the original topic. Through context analysis, we improved the list of topics of (Abd-Alrazaq et al., 2020) adding those that are sensitive to medical information distortion. Our analysis provides valuable insights for the automatic detection and classification of medical information distortion.

[4]Example: *@eugenegu Hydroxychloroquine has known side effects including prolonging the heart QT interval (time between the Q wave and the T wave on an EKG), which is the time it takes for the ventricles to contract and relax. QT prolongation can cause Torsades de Pointes, a deadly heart rhythm*

[5]Example: *HYDROXYCHLOROQUINE cured my cousin and his wife, after 10 days of insurmountable suffering, in a matter of 24 hours...and he had existing heart issues. Did great*

References

A. Abd-Alrazaq, D. Alhuwail, M. Househ, M. Hamdi, and Z. Shah. 2020. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. *J Med Internet Res*, 22(4):e19016, April.

W. Ahmed, J. Vidal-Alaball, J. Downing, and F.L. Seguí. 2020. COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. *Journal of Medical Internet Research*, 22(5):e19458.

D. Boyd, S. Golder, and G. Lotan. 2010. Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. In *2010 43rd Hawaii International Conference on System Sciences*, pages 1–10, January. ISSN: 1530-1605.

J.S. Brennen, F.M. Simon, P.N. Howard, and R.K. Nielsen. 2020. Types, Sources, and Claims of COVID-19 Misinformation. page 13.

Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St. John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, Brian Strope, and Ray Kurzweil. 2018. Universal Sentence Encoder for English. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 169–174, Brussels, Belgium, November. Association for Computational Linguistics.

Meeyoung Cha, Hamed Haddadi, Fabrício Benevenuto, and P. Krishna Gummadi. 2010. Measuring user influence in twitter: The million follower fallacy.

Jose Yunam Cuan-Baltazar, Maria José Muñoz-Perez, Carolina Robledo-Vega, Maria Fernanda Pérez-Zepeda, and Elena Soto-Vega. 2020. Misinformation of COVID-19 on the Internet: Infodemiology Study. *JMIR Public Health and Surveillance*, 6(2):e18444. Company: JMIR Public Health and Surveillance Distributor: JMIR Public Health and Surveillance Institution: JMIR Public Health and Surveillance Label: JMIR Public Health and Surveillance Publisher: JMIR Publications Inc., Toronto, Canada.

Daniel A. Erku, Sewunet A. Belachew, Solomon Abrha, Mahipal Sinnollareddy, Jackson Thomas, Kathryn J. Steadman, and Wubshet H. Tesfaye. 2020. When fear and misinformation go viral: Pharmacists' role in deterring medication misinformation during the 'infodemic' surrounding COVID-19. *Research in Social and Administrative Pharmacy*.

Matthew D. Hoffman, David M. Blei, and Francis R. Bach. 2010. Online learning for latent dirichlet allocation. In John D. Lafferty, Christopher K. I. Williams, John Shawe-Taylor, Richard S. Zemel, and Aron Culotta, editors, *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010. Proceedings of a meeting held 6-9 December 2010, Vancouver, British Columbia, Canada*, pages 856–864. Curran Associates, Inc.

Ramez Kouzy, Joseph Abi Jaoude, Afif Kraitem, Molly B. El Alam, Basil Karam, Elio Adib, Jabra Zarka, Cindy Traboulsi, Elie W. Akl, and Khalil Baddour. 2020. Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter. *Cureus*, 12(3), March. Publisher: Cureus Inc.

Nicole M. Krause, Isabelle Freiling, Becca Beets, and Dominique Brossard. 2020. Fact-checking as risk communication: the multi-layered risk of misinformation in times of COVID-19. *Journal of Risk Research*, April. Publisher: Routledge.

Marianne Noel, Liana Ermakova, Pedro Rammaciotti, Alexis Perrier, and Bilel Benbouzid. 2020. Controverse scientifique.

Diana Nurbakova, Liana Ermakova, and Irina Ovchinnikova. 2020. Understanding the Personality of Contributors to Information Cascades in Social Media in response to COVID-19 Pandemic. In *2020 International Conference on Data Mining Workshops, ICDM Workshops 2020, Sorrento, Italy, November 17-20, 2020*, page 8. IEEE.

B. Nye, J.J. Li, R. Patel, Y. Yang, I. Marshall, A. Nenkova, and B. Wallace. 2018. A Corpus with Multi-Level Annotations of Patients, Interventions and Outcomes to Support Language Processing for Medical Literature. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 197–207, Melbourne, Australia, July. Association for Computational Linguistics.

G. Pennycook and D.G. Rand. 2018. Lazy, Not Biased: Susceptibility to Partisan Fake News Is Better Explained by Lack of Reasoning Than by Motivated Reasoning.

G. Pennycook, J. McPhetres, Y. Zhang, J.G. Lu, and D.G. Rand. 2020. Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychological Science*, 31(7):770–780, July. Publisher: SAGE Publications Inc.

Horta Manoel Ribeiro, Kristina Gligoric, and Robert West. 2019. Message distortion in information cascades. page 681–692.

D.M. Romero, W. Galuba, S. Asur, and B.A. Huberman. 2011. Influence and Passivity in Social Media. In Dimitrios Gunopulos, Thomas Hofmann, Donato Malerba, and Michalis Vazirgiannis, editors, *Machine Learning and Knowledge Discovery in Databases*, pages 18–33, Berlin, Heidelberg. Springer Berlin Heidelberg.

J. Shin, L. Jian, K. Driscoll, and F. Bar. 2018. The diffusion of misinformation on social media: Temporal pattern, message, and source. *Computers in Human Behavior*, 83:278–287, June.

Amit Singhal. 2001. Modern information retrieval: A brief overview. *IEEE Data Eng. Bull.*, 24:35–43.

Graeme D. Smith, Fowie Ng, and William Ho Cheung Li. 2020. COVID-19: Emerging compassion, courage and resilience in the face of misinformation and adversity. *Journal of Clinical Nursing*, 29(9-10):1425, May. Publisher: Wiley-Blackwell.

Samia Tasnim, Md Mahbub Hossain, and Hoimonty Mazumder. 2020. Impact of Rumors and Misinformation on COVID-19 in Social Media. *Journal of Preventive Medicine and Public Health*, 53(3):171–174. Publisher: The Korean Society for Preventive Medicine.

Michael E. Tipping and Christopher M. Bishop. 1999. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611–622.

A. Ziegelmeyer, F. Koessler, J. Bracht, and E. Winter. 2010. Fragility of information cascades: an experimental study using elicited beliefs. *Experimental Economics*, 13(2):121–145, June.