Abstract

During the Covid-19 pandemic, many users consider Twitter as a reliable source of medical information, which is often distorted unintentionally while shared. Through the qualitative study of 10^6 tweets about controversial treatments, we show that despite the information quality in the initial tweet, the public figures’ feeds evoke cascades via their followers contributing to misinformation circulation. Medical information is often distorted accidentally and unconsciously due to misunderstanding of medical terminology, omission of essential details, insertion of erroneous background, overgeneralization of personal experience, and simplification of logical connections. Oversimplification of the medical issue allows lay users to conform controversial and unclear prescriptions of different doctors to the very clear political opposition. They apply simple rules and stereotypes to information processing on social media recollecting transparent slogans as ‘follow the money’ and ‘trust your eyes’.

1 Introduction

The misinformation is often disseminated in information cascades on social media (Jasser, 2019). During the Covid-19 pandemic, many users consider Twitter a reliable source of medical information (Bavel, 2020). They search public figures’ feeds for news and share description of ordinary person’s experience of dealing with the disease. The unverified information about Covid-19 is shared truthfully similar to the unverified rumours (Zubiaga et al., 2016). Mass information sharing generates information cascades (Jasser, 2019). The latter are developed on social media when a decision to quote or retweet an initial tweet is made by number of users one after another within the first few minutes after it was published (Galuba et al., 2010). In the cascades, medical information is at risk of distortion due to transformation of the initial tweet while quoting or retweeting (Boyd et al., 2010). Shortening and paraphrasing the initial tweet in comments, substituting medical terminology by lay people diminish reliability of medical information, which is sensitive to any modification of terminology and text structure made by incompetent people (Nye et al., 2018; Wilce, 2009).

Discussion of medical information requires professional competence. Social media provide inconsistent and ambiguous medical information since many users including public figures do not belong to healthcare (Bavel, 2020). Even professionals and researchers offer controversial opinions about the Covid-19 treatment due to the lack of verified data (Mehra et al., 2020). Sometimes, medical misinformation is spread on social media unconsciously, similar to an infection during epidemic (Kucharski, 2016). The influence of the public figures on misinformation distribution is one of the most important factors to study, as they attract many followers.

The objectives of our research are (1) to identify public figures the most involved in the discussion of medical information on Twitter, (2) examine the transformations of their initial tweet content bringing in information distortion, and (3) trace the dynamics of attitudes within information cascades. Contrarily to existing research about misinformation during disastrous events and the current pandemic (Chatfield and Brajavidaga, 2012; Pennycook et al., 2020; Pennycook and Rand, 2018), we explore the mechanism of the medical information distortion in line with socio-demographics of the initiators of the cascades, bringing special attention to politicians and healthcare workers. This allows us to disclose users’ political opposition and trust to the governmental decisions behind the discussion of Covid-19 treatment. To the best of our knowledge, we are the first to analyse the types of medical information distortion concerning Covid-19 treatment and to show the attitude shift within
cascades, even when the initial tweet is neutral. Our research is based on the original collection of tweets published between 30/03/20 and 13/07/20.

2 Methodology

Data Collection: We used Twitter API to collect $10^6$ tweets in English about controversy medical treatment of Covid-19 published between 30/03/20 and 13/07/20 in real time. The query contains the following terms ["chloroquine", "hydroxychloroquine", "Raoult", "remdesivir", "tocilizumab", "favipiravir", "Avigan", "azithromycin", "HCQ", "Plaquenil", "Axemal", "Dolquine", "Quensyl", "Hydroxychloroquinum", "Montagnier", "Hydroquin", "Quinoric"], the list of which was modified according to updates during the pandemic.

Most of these tweets were retweets (see Table 1 for examples of the most shared tweets). Note, that a retweet does not modify the original tweet content. Therefore, we selected only tweets containing original user generated content (141,866 tweets).

In order to retrieve the tweets forming the cascades, first, we selected 1,000 the most retweeted tweets and 1,000 the most quoted tweets, resulting in total in 1,356 tweet IDs with distinct text. These tweets are the initial tweets of the cascades. Second, we retrieved the hops of the corresponding cascades as follows. We selected all tweets if any of the fields retweeted_status.id, quoted_status.id or in_reply_to_status_id was in the set of initial 1,356 tweet IDs. They were added to the initial tweets. Thus, we obtained the cascades of the maximal depth of 9. For further analysis, we focused on the field text or extended_tweet.full_text if available of each of the tweets of the cascades.

We first identify and categorise the authors of tweets initiating information cascades, who are mainly represented by public figures. Then, we perform semantic analysis of texts of the discussions within the cascades. Finally, we examine the shift in the attitudes of cascades participants.

Identification of the most involved public figures: First, we identify the most prominent public figures who actively contributed to medical information discussions, and whose tweets evoked cascades in our collection. We generated the frequency list of words from the most quoted and retweeted tweets. It contains approximately 2.5 mln words. We analysed high-frequency words with the occurrence above 100 per million words of text and checked the inclusion of our key terms and proper names (user_screen_name, user_name). We also considered the frequency of mentions of the public figures in hashtags (e.g. #Trump, #donaldtrump, #drfauci, #gates, and #billy gates) as an indicator of the importance of these people in discussing Covid-19 on Twitter. We did not include the names of those who did not publish information on Twitter due to the lack of his/her participation in the discussions (e.g. Dr. Vladimir Zelenko).

Thus, among people with more than 150 mentions, we identified 39 public figures (see Table 2). The majority of the most mentioned public figures belong to ‘Politics’. The politicians are in charge of informing people, arranging prevention and controlling Covid-19 outbreak, providing possibilities of treatment, and regulating everyday life routing during the current pandemic.

Expert analysis of information distortion within cascades: Given the text of the tweets forming the cascades, our expert performed semantic analysis of the content in order to identify the types of occurred information distortion. The analysis was based on the distribution of the key terms in the cascades, examination of their context in tweets and verification of logical relations among medical terminology. The context analysis allows us to recognise substitutions of the terms and words from the initial tweets. The substation analysis allows us to detect information distortion associated with the initial tweet of a prominent public figure.

Attitude shift within cascades: We perform sentiment analysis on every tweet of the cascades in order to determine the emotions and attitudes of their authors. Thus, we assign the polarity of the text to each of the tweets in our collection. We use TextBlob library to do so. Based on the polarity of tweets forming a cascade, we calculate the shift in the attitudes occurring at each of the hops of the cascade, denoted $\Delta$, as follows. Given the state at the $i$th and $(i-1)^{th}$ hops, where $state_i = \begin{cases} 1 & \text{if } polarity_i > 0 \\ 0 & \text{if } polarity_i = 0 \\ -1 & \text{if } polarity_i < 0 \end{cases}$, the attitude shift at hop $i$ is given by: $\Delta_i = state_i - state_{i-1}$.

---

1We do not claim the exhaustiveness of this list, but it includes the terms we came through during our research on the topic covering the most the discussions of Covid-19 treatment.

2https://textblob.readthedocs.io/
Table 1: Examples of the most shared tweets of the prominent public figures (April 6, 2020 and May 19, 2020)

<table>
<thead>
<tr>
<th>User</th>
<th>Text</th>
<th>#tweets</th>
<th>#quotes</th>
<th>#replies</th>
</tr>
</thead>
<tbody>
<tr>
<td>HillaryClinton</td>
<td>Please do not take medical advice from a man who looked directly at a solar eclipse. Be prepared, there is a small chance that our horrendous leadership could unknowingly lead us into World War III. HYDROXYCHLOROQUINE &amp; AZITHROMYCIN, taken together, have a real chance to be one of the biggest game changers in the history of medicine. The FDA has moved mountains. - Thank You! Hopefully they will BOTH! (H works better with A. International Journal of Antimicrobial Agents)....</td>
<td>422,447</td>
<td>57,326</td>
<td>78,621</td>
</tr>
<tr>
<td>realDonaldTrump</td>
<td>REOPEN OUR COUNTRY! A NY Doctor shared with Hannity his HydroxyChloroquin/Azithromycin results. 200mg 2x daily HydroxyChloroquine 500mg 1x daily Azithromycin 220mg 1x daily Zinc sulfate 350mg patients *Breathing restored 3-4 hours *Zero deaths *Zero hospitalizations *Zero intubations</td>
<td>103,609</td>
<td>30,058</td>
<td>70,802</td>
</tr>
<tr>
<td>elonmusk</td>
<td>Take the red pill</td>
<td>104,853</td>
<td>17,437</td>
<td>34,027</td>
</tr>
<tr>
<td>RossFairchild</td>
<td>REOPEN OUR COUNTRY! A NY Doctor shared with Hannity his HydroxyChloroquin/Azithromycin results. 200mg 2x daily HydroxyChloroquine 500mg 1x daily Azithromycin 220mg 1x daily Zinc sulfate 350mg patients *Breathing restored 3-4 hours *Zero deaths *Zero hospitalizations *Zero intubations</td>
<td>100,162</td>
<td>23,078</td>
<td>58,082</td>
</tr>
</tbody>
</table>

Table 2: The most frequently mentioned public figures in the quoted tweets

<table>
<thead>
<tr>
<th>Category</th>
<th>#members</th>
<th>Frequency</th>
<th>Top mentioned person in the category</th>
<th>Mentions</th>
<th>Hashtags</th>
</tr>
</thead>
<tbody>
<tr>
<td>Politics</td>
<td>17</td>
<td>41,974</td>
<td>Donald Trump, president of the U.S. - Thank You! Hopefully they will BOTH (H works better with A, International Journal of Antimicrobial Agents)...</td>
<td>33,821</td>
<td>1,218</td>
</tr>
<tr>
<td>Healthcare</td>
<td>7</td>
<td>12,327</td>
<td>Dr. A. Fanci, chief of the U.S. National Institute of Allergy and Infectious Diseases</td>
<td>7,196</td>
<td>448</td>
</tr>
<tr>
<td>Journalism</td>
<td>9</td>
<td>2,565</td>
<td>Nail Cavuto, business journalist for Fox News</td>
<td>1,051</td>
<td>26</td>
</tr>
<tr>
<td>Business</td>
<td>3</td>
<td>2,201</td>
<td>Bill Gates, business magnate</td>
<td>1,896</td>
<td>100</td>
</tr>
<tr>
<td>Arts</td>
<td>2</td>
<td>334</td>
<td>Scott Adams, creator of Dilbert comic strip</td>
<td>206</td>
<td>0</td>
</tr>
<tr>
<td>Academia</td>
<td>1</td>
<td>162</td>
<td>Bill Mitchell, Australian economist</td>
<td>162</td>
<td>0</td>
</tr>
</tbody>
</table>

3 Results

In this Section, we discuss medical information distortion observed within the cascades initiated by the tweets of the prominent public figures, and report the results of the analysis of attitude shift.

Information distortion in cascades evoked by prominent public figures: The list of high-frequency words in the quoted tweets is headed by Hydroxychloroquine (62,521 occurrences), while #Hydroxychloroquine is the most frequent hashtag (7,144 occurrences). Thus, the effect of Hydroxychloroquine (HCQ) in treating Covid-19 attracted attention of Twitter users and became an important topic in information cascades evoked by comments on public figures’ tweets.

A deep cascade (4 hops) about the effect of HQC was initiated in April, 2020 by a tweet of Donald Trump provoking 103,637 retweets and evoking a cascade of supporting messages with shared experience of preventing and treating Covid-19, on the one hand, and healthcare professionals clarifying the effect of the drug and trying to repair the damage made by the initial tweet and by supporters who exaggerated HCQ efficacy, on the other hand. Meanwhile, many politicians and media persons criticized HCQ efficacy and ridiculed Trump’s advice (total of 2,003,007 retweets in April-May).

The most retweeted message belongs to the congresswoman Hillary Clinton. It initiated an information cascade, shifting the topic of the debate from treatment to politics. The discussion contained medical misinformation and revealed the opposition Republicans vs. Democrats.

The effect of HCQ seems to be discussed in every cascade when a politician tweeted about Trump’s policy during the current pandemic. The cascades evoked by politicians contain medical misinformation regardless of the topic of the discussion; their participants used the effect of HCQ (positive or negative) as an argument in the political struggle between liberals and conservatives.

The medical information was distorted in the cascades evoked by comments on the politicians’ tweets due to the simplification of treatment procedures when a HCQ prescription is considered a prominent step in the treatment, while postponing the prescription due to bureaucratic rules or delay of drug delivery are shown as a violation of human rights and/or a manifestation of the Big Pharma conspiracy theory. Exaggeration of the drug impact on the Covid-19 cure leads to misinformation since the complicated interrelation of drugs and consequences of the treatment for different cases are often omitted. Figure 1 provides examples of the cascades with highlighted information distortion, polarity and attitude shift values.

Thus, in cascades evoked by the politicians’ tweets, the misinformation was revealed in simplifications, omissions, additions, insertion of false background or prerequisites and overgeneralization of a personal experience. It is connected to the Republicans and Democrats confrontation.

Contrarily to politicians, healthcare professionals rarely evoked information cascades in our dataset, being instead more often mentioned in
were able to terminate an information cascade.

cascades of the same depth. It can be seen that the first hop usually differs little from the initial tweet, while starting from the second hop we note the polarisation of the attitudes (blue-red column alternation). This could be explained by attitude polarisation, i.e., group’s attitude toward a situation changes in the sense that the individuals’ initial attitudes have intensified after discussion (Myers and Lamm, 1975). It is consistent with discussed confrontation mainly based on user political beliefs.

Figure 1: Examples of cascades and information distortion within them

![Image](Image 95x435 to 267x583)

Figure 2: Average attitude shift within cascades

Attitude shift: According to our sentiment analysis, initial tweets are neutral in average, yet attitude shift occurs between the hops of the cascades. Fig. 2 depicts attitude shifts averaged over the cascades of the same depth. It can be seen that the first hop usually differs little from the initial tweet, while starting from the second hop we note
References


