



OPEN

# The COVID-19 social media infodemic

Matteo Cinelli<sup>1,2</sup>, Walter Quattrocchi<sup>1,2,3✉</sup>, Alessandro Galeazzi<sup>4</sup>, Carlo Michele Valensise<sup>5</sup>, Emanuele Brugnoli<sup>1</sup>, Ana Lucia Schmidt<sup>2</sup>, Paola Zola<sup>6</sup>, Fabiana Zollo<sup>1,2,7</sup> & Antonio Scala<sup>1,3</sup>

We address the diffusion of information about the COVID-19 with a massive data analysis on Twitter, Instagram, YouTube, Reddit and Gab. We analyze engagement and interest in the COVID-19 topic and provide a differential assessment on the evolution of the discourse on a global scale for each platform and their users. We fit information spreading with epidemic models characterizing the basic reproduction number  $R_0$  for each social media platform. Moreover, we identify information spreading from questionable sources, finding different volumes of misinformation in each platform. However, information from both reliable and questionable sources do not present different spreading patterns. Finally, we provide platform-dependent numerical estimates of rumors' amplification.

The World Health Organization (WHO) defined the SARS-CoV-2 virus outbreak as a severe global threat<sup>1</sup>. As foreseen in 2017 by the global risk report of the World Economic forum, global risks are interconnected. In particular, the case of the COVID-19 epidemic (the infectious disease caused by the most recently discovered human coronavirus) is showing the critical role of information diffusion in a disintermediated news cycle<sup>2</sup>.

The term *infodemic*<sup>3,4</sup> has been coined to outline the perils of misinformation phenomena during the management of disease outbreaks<sup>5–7</sup>, since it could even speed up the epidemic process by influencing and fragmenting social response<sup>8</sup>. As an example, CNN has recently anticipated a rumor about the possible lock-down of Lombardy (a region in northern Italy) to prevent pandemics<sup>9</sup>, publishing the news hours before the official communication from the Italian Prime Minister. As a result, people overcrowded trains and airports to escape from Lombardy toward the southern regions before the lock-down was put in place, disrupting the government initiative aimed to contain the epidemics and potentially increasing contagion. Thus, an important research challenge is to determine how people seek or avoid information and how those decisions affect their behavior<sup>10</sup>, particularly when the news cycle—dominated by the disintermediated diffusion of information—alters the way information is consumed and reported on.

The case of the COVID-19 epidemic shows the critical impact of this new information environment. The information spreading can strongly influence people's behavior and alter the effectiveness of the countermeasures deployed by governments. To this respect, models to forecast virus spreading are starting to account for the behavioral response of the population with respect to public health interventions and the communication dynamics behind content consumption<sup>8,11,12</sup>.

Social media platforms such as YouTube and Twitter provide direct access to an unprecedented amount of content and may amplify rumors and questionable information. Taking into account users' preferences and attitudes, algorithms mediate and facilitate content promotion and thus information spreading<sup>13</sup>. This shift from the traditional news paradigm profoundly impacts the construction of social perceptions<sup>14</sup> and the framing of narratives; it influences policy-making, political communication, as well as the evolution of public debate<sup>15,16</sup>, especially when issues are controversial<sup>17</sup>. Users online tend to acquire information adhering to their worldviews<sup>18,19</sup>, to ignore dissenting information<sup>20,21</sup> and to form polarized groups around shared narratives<sup>22,23</sup>. Furthermore, when polarization is high, misinformation might easily proliferate<sup>24,25</sup>. Some studies pointed out that fake news and inaccurate information may spread faster and wider than fact-based news<sup>26</sup>. However, this might be platform-specific effect. The definition of "Fake News" may indeed be inadequate since political debate often resorts to labelling opposite news as unreliable or fake<sup>27</sup>. Studying the effect of the social media environment on the perception of polarizing topics is being addressed also in the case of COVID-19. The issues related to the current infodemics are indeed being tackled by the scientific literature from multiple perspectives including the

<sup>1</sup>CNR-ISC, Rome, Italy. <sup>2</sup>Università Ca' Foscari di Venezia, Venice, Italy. <sup>3</sup>Big Data in Health Society, Rome, Italy. <sup>4</sup>Università di Brescia, Brescia, Italy. <sup>5</sup>Politecnico di Milano, Milan, Italy. <sup>6</sup>CNR-IIT, Pisa, Italy. <sup>7</sup>Center for the Humanities and Social Change, Venice, Italy. ✉email: w.quattrocchi@unive.it

dynamics of hatespeech and conspiracy theories<sup>28,29</sup>, the effect of bots and automated accounts<sup>30</sup>, and the threats of misinformation in terms of diffusion and opinions formation<sup>31,32</sup>.

In this work we provide an in-depth analysis of the social dynamics in a time window where narratives and moods in social media related to the COVID-19 have emerged and spread. While most of the studies on misinformation diffusion focus on a single platform<sup>17,26,33</sup>, the dynamics behind information consumption might be particular to the environment in which they spread on. Consequently, in this paper we perform a comparative analysis on five social media platforms (Twitter, Instagram, YouTube, Reddit and Gab) during the COVID-19 outbreak. The dataset includes more than 8 million comments and posts over a time span of 45 days. We analyze user engagement and interest about the COVID-19 topic, providing an assessment of the discourse evolution over time on a global scale for each platform. Furthermore, we model the spread of information with epidemic models, characterizing for each platform its basic reproduction number ( $R_0$ ), i.e. the average number of secondary cases (users that start posting about COVID-19) an “infectious” individual (an individual already posting on COVID-19) will create. In epidemiology,  $R_0 = 1$  is a threshold parameter. When  $R_0 < 1$  the disease will die out in a finite period of time, while the disease will spread for  $R_0 > 1$ . In social media,  $R_0 > 1$  will indicate the possibility of an infodemic.

Finally, coherently with the classification provided by the fact-checking organization Media Bias/Fact Check<sup>34</sup> that classifies news sources based on the truthfulness and bias of the information published, we split news outlets into two groups. These groups are either associated to the diffusion of (mostly) reliable or (mostly) questionable contents and we characterize the spreading of information regarding COVID-19 relying on this classification. We find that users in mainstream platforms are less susceptible to the diffusion of information from questionable sources and that information deriving from news outlets marked either as reliable or questionable do not present significant difference in the way it spreads.

Our findings suggest that the interaction patterns of each social media combined with the peculiarity of the audience of each platform play a pivotal role in information and misinformation spreading. We conclude the paper by measuring rumor’s amplification parameters for COVID-19 on each social media platform.

## Results

We analyze mainstream platforms such as Twitter, Instagram and YouTube as well as less regulated social media platforms such as Gab and Reddit. Gab is a crowdfunded social media whose structure and features are Twitter-inspired. It performs very little control on content posted; in the political spectrum, its user base is considered to be far-right. Reddit is an American social news aggregation, web content rating, and discussion website based on collective filtering of information.

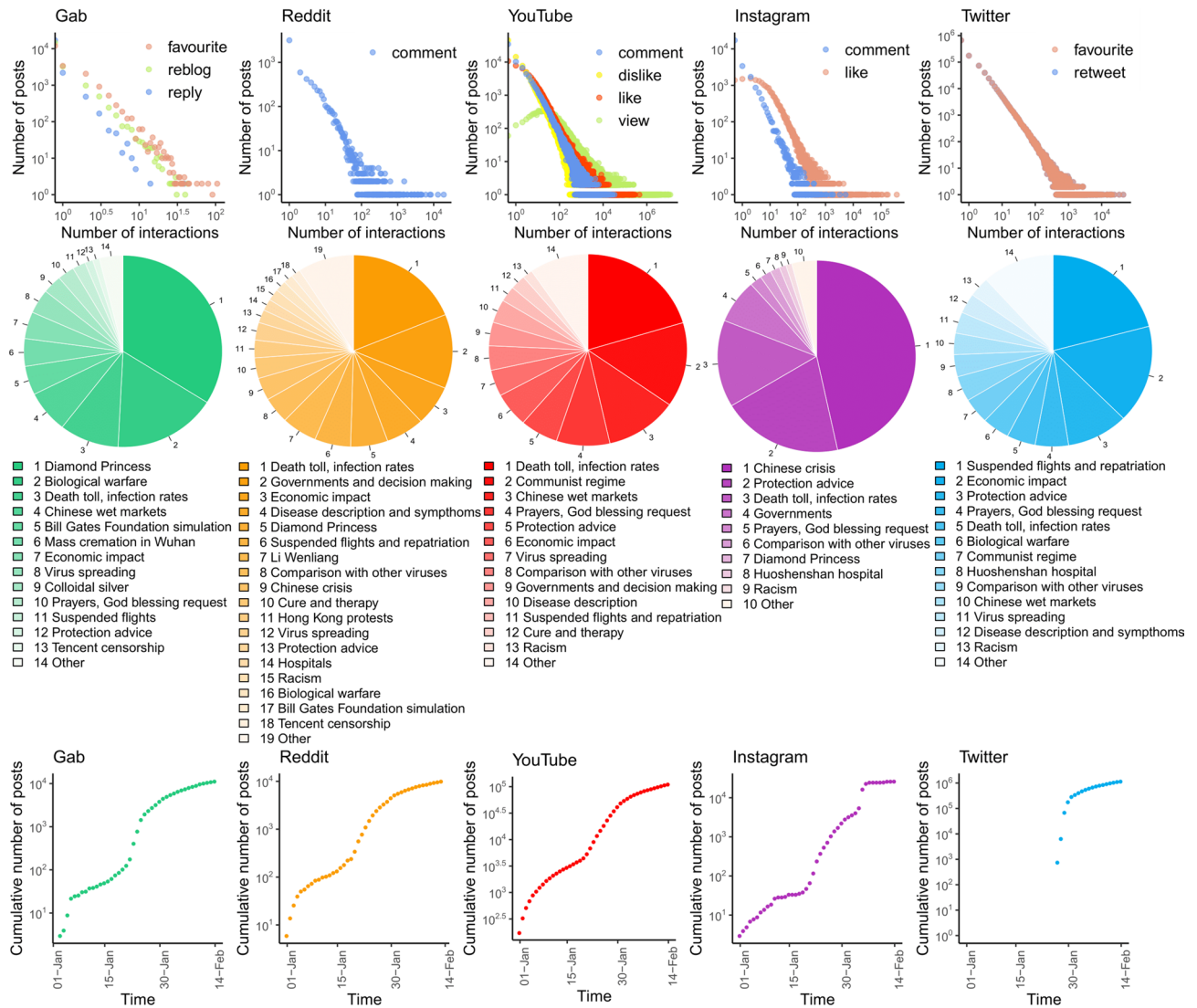
We perform a comparative analysis of information spreading dynamics around the same argument in different environments having different interaction settings and audiences. We collect all pieces of content related to COVID-19 from the 1st of January to the 14th of February. Data have been collected filtering contents accordingly to a selected sample of Google Trends’ COVID-19 related queries such as: *coronavirus*, *coronavirusoutbreak*, *imnotavirus*, *ncov*, *ncov-19*, *pandemic*, *wuhan*. The deriving dataset is then composed of 1,342,103 posts and 7,465,721 comments produced by 3,734,815 users. For more details regarding the data collection refer to Methods.

**Interaction patterns.** First, we analyze the interactions (i.e., the engagement) that users have with COVID-19 topics on each platform. The upper panel of Fig. 1 shows users’ engagement around the COVID-19 topic. Despite the differences among platforms, we observe that they all display a rather similar distribution of the users’ activity characterized by a long tail. This entails that users behave similarly for what concern the dynamics of reactions and content consumption. Indeed, users’ interactions with the COVID-19 content present attention patterns similar to any other topic<sup>35</sup>. The highest volume of interactions in terms of posting and commenting can be observed on mainstream platforms such as YouTube and Twitter.

Then, to provide an overview of the debate concerning the disease outbreak, we extract and analyze the topics related to the COVID-19 content by means of Natural Language Processing techniques. We build word embedding for the text corpus of each platform, i.e. a word vector representation in which words sharing common contexts are in close proximity. Moreover, by running clustering procedures on these vector representations, we separate groups of words and topics that are perceived as more relevant for the COVID-19 debate. For further details refer to Methods. The results (Fig. 1, middle panel) show that topics are quite similar across each social media platform. Debates range from comparisons to other viruses, requests for God blessing, up to racism, while the largest volume of interaction is related to the lock-down of flights.

Finally, to characterize user engagement with the COVID-19 on the five platforms, we compute the cumulative number of new posts each day (Fig. 1, lower panel). For all platforms, we find a change of behavior around the 20th of January, that is the day that the World Health Organization (WHO) issued its first situation report on the COVID-19<sup>36</sup>. The largest increase in the number of posts is on the 21st of January for Gab, the 24th January for Reddit, the 30th January for Twitter, the 31st January for YouTube and the 5th of February for Instagram. Thus, social media platforms seem to have specific timings for content consumption; such patterns may depend upon the difference in terms of audience and interaction mechanisms (both social and algorithmic) among platforms.

**Information spreading.** Efforts to simulate the spreading of information on social media by reproducing real data have mostly applied variants of standard epidemic models<sup>37–40</sup>. Coherently, we analyze the observed monotonic increasing trend in the way new users interact with information related to the COVID-19 by using epidemic models. Unlike previous works, we do not only focus on models that imply specific growth mechanisms, but also on phenomenological models that emphasize the reproducibility of empirical data<sup>41</sup>.

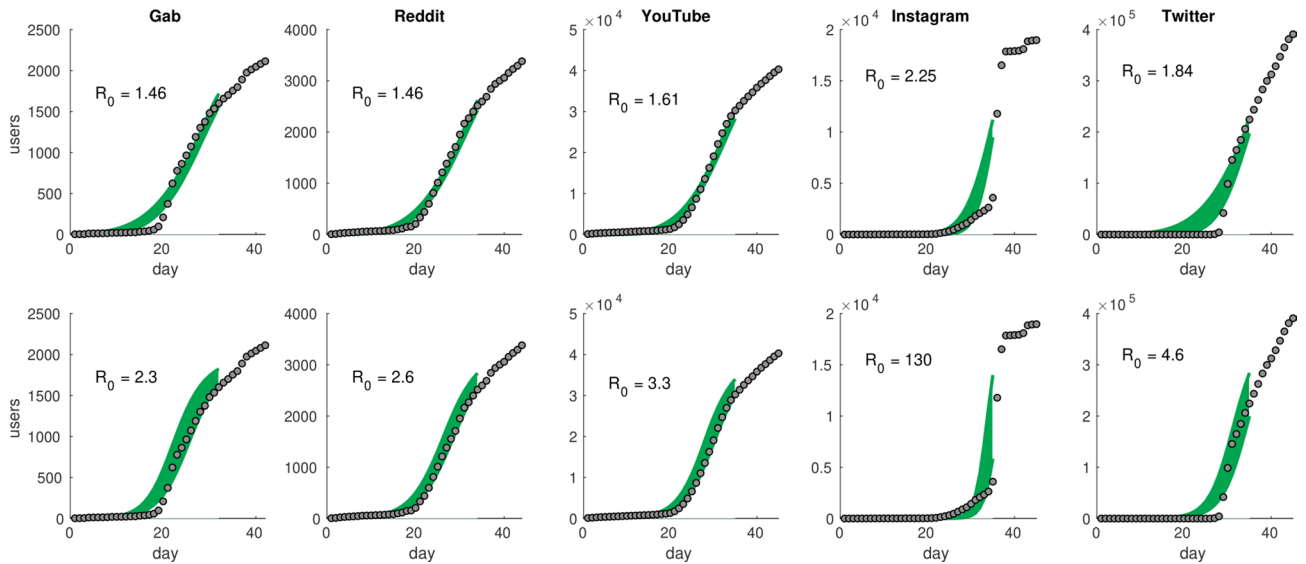


**Figure 1.** Upper panel: activity (likes, comments, reposts, etc.) distribution for each social media. Middle panel: most discussed topics about COVID-19 on each social media. Lower panel: cumulative number of content (posts, tweets, videos, etc.) produced from the 1st of January to the 14th of February. Due to the Twitter API limitations in gathering past data, the first data point for Twitter is dated January 27th.

Most of the epidemiological models focus on the basic reproduction number  $R_0$ , representing the expected number of new infectors directly generated by an infected individual for a given time period<sup>42</sup>. An epidemic occurs if  $R_0 > 1$ ,—i.e., if an exponential growth in the number of infections is expected at least in the initial phase. In our case, we try to model the growth in number of people publishing a post on a subject as an infective process, where people can start publishing after being exposed to the topic. While in real epidemics  $R_0 > 1$  highlights the possibility of a pandemic, in our approach  $R_0 > 1$  indicates the emergence of an infodemic. We model the dynamics both with the phenomenological model of<sup>43</sup> (from now on referred to as the EXP model) and with the standard SIR (Susceptible, Infected, Recovered) compartmental model<sup>44</sup>. Further details on the modeling approach can be found in Methods.

As shown in Fig. 2, each platform has its own basic reproduction number  $R_0$ . As expected, all the values of  $R_0$  are supercritical—even considering confidence intervals (Table 1)—signaling the possibility of an infodemic. This observation may facilitate the prediction task of information spreading during critical events. Indeed, according to this result we can consider information spreading patterns on each social media to predict social response when implementing crisis management plans.

While  $R_0$  is a good proxy for the engagement rate and a good predictor for epidemic-like information spreading, social contagion phenomena might be in general more complex<sup>45–47</sup>. For instance, in the case of Instagram, we observe an abrupt jump in the number of new users that cannot be explained with continuous models like the standard epidemic ones; accordingly, the SIR model estimates a value of  $R_0 \sim 10^2$  that is way beyond what has been observed in any real-world epidemic.



**Figure 2.** Growth of the number of authors versus time. Time is expressed in number of days since 1st January 2020 (day 1). Shaded areas represents [5%, 95%] estimates of the models obtained via bootstrapping least square estimates of the EXP model (upper panels) and of the SIR model (lower panels). For details the SIR and the EXP model, see SI.

	Gab	Reddit	YouTube	Instagram	Twitter
$R_0^{EXP}$	[1.42, 1.52]	[1.44, 1.51]	[1.56, 1.70]	[2.02, 2.64]	[1.65, 2.06]
$R_0^{SIR}$	[2.2, 2.5]	[2.4, 2.8]	[3.2, 3.5]	$[1.1 \times 10^2, 1.6 \times 10^2]$	[4.0, 5.1]

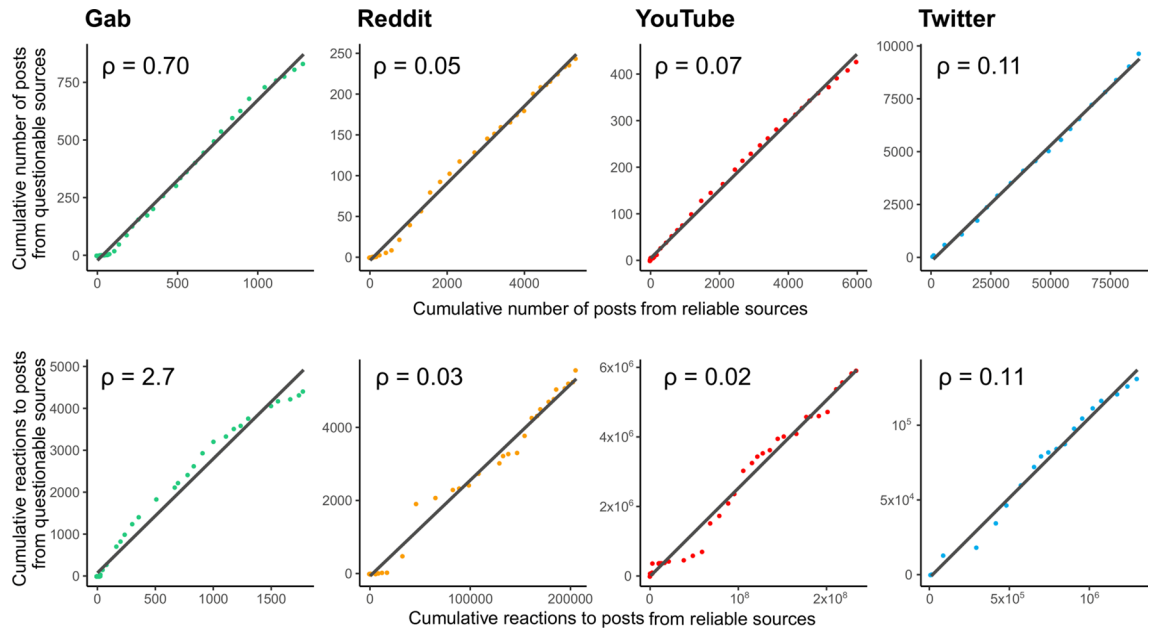
**Table 1.** [5%, 95%] interval of confidence  $R_0$  as estimated from bootstrapping the least square fits parameter of the EXP and of the SIR model. Notice that, due to the steepness of the growth of the number of new authors in Instagram,  $R_0$  assumes unrealistic values  $\sim 10^2$  for the SIR model.

**Questionable VS reliable information sources.** We conclude our analysis by comparing the diffusion of information from questionable and reliable sources on each platform. We tag links as reliable or questionable according to the data reported by the independent fact-checking organization Media Bias/Fact Check<sup>34</sup>. In order to clarify the limits of an approach that is based on labelling news outlets rather than single articles, as for instance performed in<sup>33,48</sup>, we report the definitions used in this paper for questionable and reliable information sources. In accordance with the criteria established by MBFC, by questionable information source we mean a news outlet systematically showing one or more of the following characteristics: extreme bias, consistent promotion of propaganda/conspiracies, poor or no sourcing to credible information, information not supported by evidence or unverifiable, a complete lack of transparency and/or fake news. By reliable information sources we mean news outlets that do not show any of the aforementioned characteristics. Such outlets can anyway produce contents potentially displaying a bias towards liberal/conservative opinion, but this does not compromise the overall reliability of the source.

Figure 3 shows, for each platform, the plots of the cumulative number of posts and reactions related to reliable sources versus the cumulative number of posts and interactions referring to questionable sources. By interactions we mean the overall reactions, e.g. likes or other form of endorsement and comments, that can be performed with respect to a post on a social platform. Surprisingly, all the posts show a strong linear correlation, i.e., the number of posts/reactions relying on questionable and reliable sources grows with the same pace inside the same social media platform. We observe the same phenomenon also for the engagement with reliable and questionable sources. Hence, the growth dynamics of posts/interactions related to questionable news outlets is just a re-scaled version of the growth dynamics of posts/reactions related to reliable news outlets; however, the re-scaling factor  $\rho$  (i.e., the fraction of questionable over reliable) is strongly dependent on the platform.

In particular, we observe that in mainstream social media the number of posts produced by questionable sources represents a small fraction of posts produced by reliable ones; the same thing happens in Reddit. Among less regulated social media, a peculiar effect is observed in Gab: while the volume of posts from questionable sources is just the  $\sim 70\%$  of the volume of posts from reliable ones, the volume of reactions for the former ones is  $\sim 3$  times bigger than the volume for the latter ones. Such results hint the possibility that different platform react differently to information produced by reliable and questionable news outlets.

To further investigate this issue, we define the amplification factor  $\mathcal{E}$  as the average number of reactions to a post; hence,  $\mathcal{E}$  is a measure that quantifies the extent to which a post is amplified in a social media. We observe that the amplification  $\mathcal{E}^U$  (for unreliable posts produced by questionable outlets) and  $\mathcal{E}^R$  (for reliable posts



**Figure 3.** Upper panels: plot of the cumulative number of posts referring to questionable sources versus the cumulative number of posts referring to reliable sources. Lower panel: plot of the cumulative number of engagements relative to questionable sources versus the cumulative number of engagements relative to reliable sources. Notice that a linear behavior indicates that the time evolution of questionable posts/engagements is just a re-scaled version of the time evolution of reliable posts/engagements. Each plot indicates the regression coefficients  $\rho$ , representing the ratio among the volumes of questionable and reliable posts ( $\rho^{post}$ ) and engagements ( $\rho^{eng}$ ). In more popular social media, the number of questionable posts represents a small fraction of the reliable ones; same thing happens in Reddit. Among less popular social media, a peculiar effect is observed in Gab: while the volume of questionable posts is just the  $\sim 70\%$  of the volume of reliable ones, the volume of engagements for questionable posts is  $\sim 3$  times bigger than the volume for reliable ones. Further details concerning the regression coefficients are reported in Methods.

	$\mathcal{E}^U$	$\mathcal{E}^R$	$\alpha$
Gab	5.6	1.4	3.9
Reddit	22.7	40.1	0.55
Twitter	15.1	15.6	0.97
YouTube	$1.4 \times 10^4$	$3.9 \times 10^4$	0.35

**Table 2.** The average engagement of a post is the number of reactions expected for a post and is a measure of how much a post is amplified in each social media platform. The average engagement  $\mathcal{E}^U$  (for unreliable post) and  $\mathcal{E}^R$  (for reliable post) vary from platform to platform, and are the largest in Twitter and the lowest in Gab. The coefficient of relative amplification  $\alpha = \mathcal{E}^U/\mathcal{E}^R$  measures whether a social media amplifies more unreliable ( $\alpha > 1$ ) or reliable ( $\alpha < 1$ ) posts. Among more popular social media platforms, we notice that Twitter is the most neutral ( $\alpha \sim 1$  i.e.  $\mathcal{E}^U \sim \mathcal{E}^R$ ), while YouTube amplifies unreliable sources less ( $\alpha \sim 4/10$ ). Among less popular social media platforms, Reddit reduces the impact of unreliable sources ( $\alpha \sim 1/2$ ) while Gab strongly amplifies them ( $\alpha \sim 4$ ).

posts produced by reliable outlets) vary from social media platform to social media platform and that assumes the largest values in YouTube and the lowest in Gab. To measure the permeability of a platform to posts from questionable/reliable news outlets, we then define the coefficient of relative amplification  $\alpha = \mathcal{E}^U/\mathcal{E}^R$ . It is a measure of whether a social media amplifies questionable ( $\alpha > 1$ ) or reliable ( $\alpha < 1$ ) posts. Results are shown in Table 2. Among mainstream social media, we notice that Twitter is the most neutral ( $\alpha \sim 1$  i.e.  $\mathcal{E}^U \sim \mathcal{E}^R$ ), while YouTube amplifies questionable sources less ( $\alpha \sim 4/10$ ). Among less popular social media, Reddit reduces the impact of questionable sources ( $\alpha \sim 1/2$ ), while Gab strongly amplifies them ( $\alpha \sim 4$ ).

Therefore, we conclude that the main drivers of information spreading are related to specific peculiarities of each platform and depends upon the group dynamics of individuals engaged with the topic.

	Posts	Comments	Users	Period
Gab	6,252	4,364	2,629	01/01–14/02
Reddit	10,084	300,751	89,456	01/01–14/02
YouTube	111,709	7,051,595	3,199,525	01/01–14/02
Instagram	26,576	109,011	52,339	01/01–14/02
Twitter	1,187,482	–	390,866	27/01–14/02
Total	1,342,103	7,465,721	3,734,815	

**Table 3.** Data breakdown of the number of posts, comments and users for all platforms.

## Conclusions

In this work we perform a comparative analysis of users' activity on five different social media platforms during the COVID-19 health emergency. Such a timeframe is a good benchmark for studying content consumption dynamics around critical events in a times when the accuracy of information is threatened. We assess user engagement and interest about the COVID-19 topic and characterize the evolution of the discourse over time.

Furthermore, we model the spread of information using epidemic models and provide basic growth parameters for each social media platform. We then analyze the diffusion of questionable information for all channels, finding that Gab is the environment more susceptible to misinformation dissemination. However, information deriving from sources marked either as reliable or questionable do not present significant differences in their spreading patterns. Our analysis suggests that information spreading is driven by the interaction paradigm imposed by the specific social media or/and by the specific interaction patterns of groups of users engaged with the topic. We conclude the paper by computing rumor's amplification parameters for social media platforms.

We believe that the understanding of social dynamics between content consumption and social media platforms is an important research subject, since it may help to design more efficient epidemic models accounting for social behavior and to design more effective and tailored communication strategies in time of crisis.

## Methods

**Data collection.** Table 3 reports the data breakdown of the five social media platforms. Different data collection processes have been performed depending on the platform. In all cases we guided the data collection by a set of selected keywords based on Google Trends' COVID-19 related queries such as: coronavirus, pandemic, coronaoutbreak, china, wuhan, nCoV, IamNotAVirus, coronavirus\_update, coronavirus\_transmission, coronavirusnews, coronavirusoutbreak.

The Reddit dataset was downloaded from the [Pushift.io](#) archive, exploiting the related API. In order to filter contents linked to COVID-19, we used our set of keywords.

In Gab, although no official guides are available, there is an API service that given a certain keyword, returns a list of users, hashtags and groups related to it. We queried all the keywords we selected based on Google Trends and we downloaded all hashtags linked to them. We then manually browsed the results and selected a set of hashtags based on their meaning. For each hashtag in our list, we downloaded all the posts and comments linked to it.

For YouTube, we collected videos by using the YouTube Data API by searching for videos that matched our keywords. Then an in depth search was done by crawling the network of videos by searching for more related videos as established by the YouTube algorithm. From the gathered set, we filtered the videos that matched coronavirus, nCov, corona virus, corona-virus, corvid, covid or SARS-CoV in the title or description. We then collected all the comments received by those videos.

For Twitter, we collect tweets related to the topic coronavirus by using both the search and stream endpoint of the Twitter API. The data derived from the stream API represent only 1% of the total volume of tweets, further filtered by the selected keywords. The data derived from the search API represent a random sample of the tweets containing the selected keywords up to a maximum rate limit of 18000 tweets every 10 minutes.

Since no official API are available for Instagram data, we built our own process to collect public contents related to our keywords. We manually took notes of posts, comments and populated the Instagram Dataset.

**Matching ability.** We consider all the posts in our dataset that contain at least one URL linking to a website outside the related social media platform (e.g., tweets pointing outside Twitter). We separate URLs in two main categories obtained using the classification provided by MediaBias/FactCheck (MBFC). MBFC provides a classification determined by ranking bias in four different categories, one of them being Factual/Sourcing. In that category, each news outlet is associated to a label that refers to its reliability as expressed in three labels, namely Conspiracy-Pseudoscience, Pro-Science or Questionable. Noticeably, also the Questionable set include a wide range of political bias, from Extreme Left to Extreme Right.

Using such a classification, we assign to each of these outlets a binary label that partially stems from the labelling provided by MBFC. We divide the news outlets into Questionable and Reliable. All the outlets already classified as Questionable or belonging to the category Conspiracy-Pseudoscience are labelled as Questionable, the rest is labelled as Reliable. Thus, by questionable information source we mean a news outlet systematically showing one or more of the following characteristics: extreme bias, consistent promotion of propaganda/conspiracies, poor or no sourcing to credible information, information not supported by evidence or unverifiable, a

	Gab	Reddit	YouTube	Instagram	Twitter
Posts containing a URL	3,778	10,084	351,786	1,328	356,448
Matched	0.47	0.55	0.035	0.09	0.27
Questionable	0.38	0.045	0.064	0.05	0.10
Reliable	0.62	0.955	0.936	0.95	0.90

**Table 4.** Number of posts containing a URL, matching ability and classification for each of the five platforms.

	Gab	Reddit	YouTube	Instagram	Twitter	Facebook
Gab	0.003	0.002	0.001	0.002	0.138	~ 0
Reddit	0.043	0.006	0.009	0.001	~ 0	0
YouTube	0	~ 0	0.292	~ 0	0.088	0.081
Instagram	0	0	0.003	0	0.001	0.001
Twitter	0.059	0.001	0.257	0.003	~ 0	~ 0

**Table 5.** Fraction of URLs pointing to social media. Table should be read as entries in each row link to entries in each column. For example, Gab links to Reddit 0.003.

complete lack of transparency and/or fake news. By reliable information sources we mean news outlets that do not show any of the aforementioned characteristics. Such outlets can anyway produce contents potentially displaying a bias towards liberal/conservative opinion, but this does not compromise the overall reliability of the source.

Considering all the 2637 news outlets that we retrieve from the list provided by MBFC we end up with 800 outlets classified as Questionable 1837 outlets classified as Reliable. Using such a classification we quantify our overall ability to match and label domains of posts containing URLs, as reported in Table 4. The matching ability that is low doesn't refer to the ability of identifying known domain but to the ability of finding the news outlets that belong to the list provided by MBFC. Indeed in all the social networks we find a tendency towards linking to other social media platforms, as shown in Table 5.

**Text analysis.** To provide an overview of the debate concerning the virus outbreak on the various platforms, we extract and analyze all topics related to COVID-19 by applying Natural Language Processing techniques to the written content of each social media platform. We first build word embedding for the text corpus of each platform, then, to assess the topics around which the perception of the COVID-19 debate is concentrated, we cluster words by running the Partitioning Around Medoids (PAM) algorithm on their vector representations.

Word embeddings, i.e., distributed representations of words learned by neural networks, represent words as vectors in  $\mathbf{R}^n$  bringing similar words closer to each other. They perform significantly better than the well-known Latent Semantic Analysis (LSA) and Latent Dirichlet Allocation (LDA) for preserving linear regularities among words and computational efficiency on large data sets<sup>49</sup>. In this paper we use the Skip-gram model<sup>50</sup> to construct word embedding of each social media corpus. More formally, given a content represented by the sequence of words  $w_1, w_2, \dots, w_T$ , we use stochastic gradient descent with gradient computed through backpropagation rule<sup>51</sup> for maximizing the average log probability

$$\frac{1}{T} \sum_{t=1}^T \left[ \sum_{j=-k}^k \log p(w_{t+j} | w_t) \right] \quad (1)$$

where  $k$  is the size of the training window. Therefore, during training the vector representations of closely related words are pushed to be close to each other.

In the Skip-gram model, every word  $w$  is associated with its input and output vectors,  $u_w$  and  $v_w$ , respectively. The probability of correctly predicting the word  $w_i$  given the word  $w_j$  is defined as

$$p(w_i | w_j) = \frac{\exp(u_{w_i}^T v_{w_j})}{\sum_{l=1}^V \exp(u_l^T v_{w_j})} \quad (2)$$

where  $V$  is the number of words in the corpus vocabulary. Two major parameters affect the training quality: the dimensionality of word vectors, and the size of the surrounding words window. We choose 200 as vector dimension—that is typical value for training large dataset—and 6 words for the window.

	Cleaned contents	Vocabulary size	Topics	Contents with max $\Theta > 0.5$
Instagram	21,189 posts	15,324	17	4,467
Twitter	638,214 posts	22,587	21	369,131
Gab	5,853 posts	3,024	19	2,986
Reddit	10,084 posts	1,968	34	6,686
YouTube	815,563 comments	35,381	30	679,261

**Table 6.** Results of text cleaning and analysis for all the corpora.

Before applying the tool, we reduced the contents to those written in English as detected with `cld3`. Then we cleaned the corpora by removing HTML code, URLs and email addresses, user mentions, hashtags, stop-words, and all the special characters including digits. Finally, we dropped words composed by less than three characters, words occurring less than five times in all the corpus, and contents with less than three words.

To analyze the topics related to COVID-19, we cluster words by PAM and using as proximity metric the cosine distance matrix of words in their vector representations. In order to select the number of clusters,  $k$ , we calculate the average silhouette width for each value of  $k$ . Moreover, for evaluating the cluster stability, we calculate the average pairwise Jaccard similarity between clusters based on 90% sub-samples of the data. Lastly, we produce word clouds to identify the topic of each cluster. To provide a view about the debate around the virus outbreak, we define the distribution over topics  $\Theta_c$  for a given content  $c$  as the distribution of its words among the word clusters. Thus, to quantify the relevance of each topic within a corpus, we restrict to contents  $c$  with  $\max \Theta_c > 0.5$  and consider them uniquely identified as a single topic each. Table 6 shows the results of the text cleaning and topic analysis for all the data.

**Epidemiological models.** Several mathematical models can be used to analyse potential mechanisms that underline epidemiological data. Generally, we can distinguish among phenomenological models that emphasize the reproducibility of empirical data without insights in the mechanisms of growth, and more insightful mechanistic models that try to incorporate such mechanisms<sup>41</sup>.

To fit our cumulative curves, we first use the adjusted exponential model of<sup>43</sup> since it naturally provides an estimate of the basic reproduction number  $R_0$ . This phenomenological model (from now on indicated as EXP) has been successfully employed in data-scarce settings and shown to be on-par with more traditional compartmental models for multiple emerging diseases like Zika, Ebola, and Middle East Respiratory Syndrome<sup>43</sup>.

The model is defined by the following single equation:

$$I = \left[ \frac{R_0}{(1+d)^t} \right]^t \quad (3)$$

Here,  $I$  is incidence,  $t$  is the number of days,  $R_0$  is the basic reproduction number and  $d$  is a damping factor accounting for the reduction in transmissibility over time. In our case, we interpret  $I$  as the number  $C_{auth}$  of authors that have published a post on the subject.

As a mechanistic model, we employ the classical SIR model<sup>44</sup>. In such a model, a susceptible population can be infected with a rate  $\beta$  by coming into contact with infected individuals; however, infected individuals can recover with a rate  $\gamma$ . The model is described by a set of differential equations:

$$\begin{aligned} \partial_t S &= -\beta S \cdot I/N \\ \partial_t I &= \beta S \cdot I/N - \gamma I \\ \partial_t R &= \gamma I \end{aligned} \quad (4)$$

where  $S$  is the number of susceptible,  $I$  is the number of infected and  $R$  is the number of recovered. In our case, we interpret the number  $I + R$  as the number  $C_{auth}$  of authors that have published a post on the subject.

In the SIR model, the basic reproduction number  $R_0 = \beta/\gamma$  corresponds to the ration among the rate of infection by contact  $\beta$  and the rate of recovery  $\gamma$ . Notice that for the SIR model, vaccination strategies correspond to bringing the system in a situation where  $S < N/R_0$ ; in such a way, both the number of infected will decrease.

To estimate the basic reproduction numbers  $R_0^{EXP}$  and  $R_0^{SIR}$  for the EXP and the SIR model, we use least square estimates of the models' parameters<sup>42</sup>. The range of parameters is estimated via bootstrapping<sup>41,52</sup>.

**Linear regression coefficients.** Table 7 reports the regression coefficient  $\rho$ , the intercept and the  $R^2$  values for the linear fit of Fig. 3. High values of  $R^2$  confirm the linear relationship between reliable and questionable sources in information diffusion.



Dataset	Type	Intercept	Coefficient ( $\rho$ )	$R^2$
Gab	Posts	- 22.321	0.695	0.996
Reddit	Posts	- 4.111	0.047	0.997
Youtube	Posts	4.529	0.073	0.998
Twitter	Posts	- 151.44	0.110	0.998
Gab	Reactions	74.577	2.721	0.981
Reddit	Reactions	- 70.677	0.026	0.990
Youtube	Reactions	- 8854.33	0.025	0.986
Twitter	Reactions	- 2136.978	0.107	0.987

**Table 7.** Coefficients and  $R^2$  of the linear regressions displayed in Fig. 3.

## Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 11 April 2020; Accepted: 15 September 2020

Published online: 06 October 2020

## References

- Organization, W. H. Naming the coronavirus disease (COVID-19) and the virus that causes it. [https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-\(covid-2019\)-and-the-virus-that-causes-it](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it) (2020 (accessed April 9, 2020)).
- Quattrocchi, W. Part 2-social and political challenges: 2.1 western democracy in crisis? In *Global Risk Report World Economic Forum* (2017).
- WHO Situation Report 13. [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf?sfvrsn=195f4010\\_6](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf?sfvrsn=195f4010_6). Accessed: 2020-09-30.
- Zarocostas, J. How to fight an infodemic. *Lancet* **395**, 676 (2020).
- Organization, W. H. Director-general's remarks at the media briefing on 2019 novel Coronavirus on 8 February 2020. <https://www.who.int/dg/speeches/detail/director-general-s-remarks-at-the-media-briefing-on-2019-novel-coronavirus---8-february-2020> (2020 (accessed April 9, 2020)).
- Mendoza, M., Poblete, B. & Castillo, C. Twitter under crisis: Can we trust what we RT?. *Proceedings of the first workshop on social media analytics* **71–79** (2010).
- Starbird, K., Maddock, J., Orand, M., Achterman, P. & Mason, R. M. Rumors, false flags, and digital vigilantes: Misinformation on twitter after the 2013 boston marathon bombing. *ICConference 2014 Proceedings* (2014).
- Kim, L., Fast, S. M. & Markuzon, N. Incorporating media data into a model of infectious disease transmission. *PLoS ONE* **14**, 1 (2019).
- John, T. & Ben Wedeman, C. Italy prohibits travel and cancels all public events in its northern region to contain Coronavirus. <https://edition.cnn.com/2020/03/08/europe/italy-coronavirus-lockdown-europe-intl/index.html> (2020 (accessed April 9, 2020)).
- Sharot, T. & Sunstein, C. R. How people decide what they want to know. *Nat. Hum. Behav.* **2020**, 1–6 (2020).
- Shaman, J., Karspeck, A., Yang, W., Tamerius, J. & Lipsitch, M. Real-time influenza forecasts during the 2012–2013 season. *Nat. Commun.* **4**, 1–10 (2013).
- Viboud, C. & Vespignani, A. The future of influenza forecasts. *Proc. Natl. Acad. Sci.* **116**, 2802–2804 (2019).
- Kulshrestha, J. *et al.* Quantifying search bias: Investigating sources of bias for political searches in social media. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 417–432 (2017).
- Schmidt, A. L. *et al.* Anatomy of news consumption on Facebook. *Proc. Natl. Acad. Sci.* **114**, 3035–3039 (2017).
- Starnini, M., Frasca, M. & Baronchelli, A. Emergence of metapopulations and echo chambers in mobile agents. *Sci. Rep.* **6**, 31834 (2016).
- Schmidt, A. L., Zollo, F., Scala, A., Betsch, C. & Quattrocchi, W. Polarization of the vaccination debate on Facebook. *Vaccine* **36**, 3606–3612 (2018).
- Del Vicario, M. *et al.* The spreading of misinformation online. *Proc. Natl. Acad. Sci.* **113**, 554–559 (2016).
- Bessi, A. *et al.* Science vs. conspiracy: collective narratives in the age of misinformation. *PLoS ONE* **10**, e0118093 (2015).
- Cinelli, M. *et al.* Selective exposure shapes the Facebook news diet. *PLoS ONE* **15**, e0229129 (2020).
- Zollo, F. *et al.* Debunking in a world of tribes. *PLoS ONE* **12**, 1 (2017).
- Baronchelli, A. The emergence of consensus: a primer. *R. Soc. Open Sci.* **5**, 172189 (2018).
- Del Vicario, M. *et al.* Echo chambers: emotional contagion and group polarization on Facebook. *Sci. Rep.* **6**, 37825 (2016).
- Bail, C. A. *et al.* Exposure to opposing views on social media can increase political polarization. *Proc. Natl. Acad. Sci.* **115**, 9216–9221 (2018).
- Vicario, M. D., Quattrocchi, W., Scala, A. & Zollo, F. Polarization and fake news: early warning of potential misinformation targets. *ACM Trans. Web (TWEB)* **13**, 1–22 (2019).
- Wardle, C. & Derakhshan, H. Information disorder: Toward an interdisciplinary framework for research and policy making. *Council of Europe report 27* (2017).
- Vosoughi, S., Roy, D. & Aral, S. The spread of true and false news online. *Science* **359**, 1146–1151 (2018).
- Ruths, D. The misinformation machine. *Science* **363**, 348–348 (2019).
- Schild, L. *et al.* “go eat a bat, chang!”: An early look on the emergence of sinophobic behavior on web communities in the face of covid-19. *arXiv preprint arXiv:2004.04046* (2020).
- Velásquez, N. *et al.* Hate multiverse spreads malicious COVID-19 content online beyond individual platform control. Preprint [arXiv:2004.00673](https://arxiv.org/abs/2004.00673) (2020).
- Ferrara, E. What types of COVID-19 conspiracies are populated by twitter bots? *First Monday* (2020).
- Alam, F. *et al.* Fighting the COVID-19 infodemic: modeling the perspective of journalists, fact-checkers, social media platforms, policy makers, and the society. Preprint [arXiv:2005.00033](https://arxiv.org/abs/2005.00033) (2020).

32. Shahi, G. K., Dirkson, A. & Majchrzak, T. A. *An exploratory study of COVID-19 misinformation on twitter*. Preprint [arXiv :2005.05710](https://arxiv.org/abs/2005.05710) (2020).
33. Bovet, A. & Makse, H. A. Influence of fake news in twitter during the 2016 us presidential election. *Nat. Commun.* **10**, 1–14 (2019).
34. (MBFC), M. B. C. Media bias/fact check, the most comprehensive Meida bias check resource. <https://mediabiasfactcheck.com/> (2020 (accessed April 9, 2020)).
35. Romero, D. M., Meeder, B. & Kleinberg, J. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In *Proceedings of the 20th international conference on World wide web*, 695–704 (2011).
36. Organization, W. H. Novel Coronavirus (2019-NCOV) situation report-1 21 January 2020. [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10\\_4](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4) (2020 (accessed April 9, 2020)).
37. Pellis, L. *et al.* Eight challenges for network epidemic models. *Epidemics* **10**, 58–62 (2015).
38. Liu, Y. *et al.* Characterizing super-spreading in microblog: an epidemic-based information propagation model. *Physica A* **463**, 202–218 (2016).
39. Skaza, J. & Blais, B. Modeling the infectiousness of twitter hashtags. *Physica A* **465**, 289–296 (2017).
40. Davis, J. T., Perra, N., Zhang, Q., Moreno, Y. & Vespignani, A. Phase transitions in information spreading on structured populations. *Nat. Phys.* **2020**, 1–7 (2020).
41. Chowell, G. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: a primer for parameter uncertainty, identifiability, and forecasts. *Infect. Dis. Model.* **2**, 379–398 (2017).
42. Ma, J. Estimating epidemic exponential growth rate and basic reproduction number. *Infectious Disease Modelling* (2020).
43. Fisman, D. N., Hauck, T. S., Tuite, A. R. & Greer, A. L. An idea for short term outbreak projection: nearcasting using the basic reproduction number. *PLoS ONE* **8**, 1 (2013).
44. Bailey, N. T. *et al.* *The mathematical theory of infectious diseases and its applications* (Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE, 1975).
45. Centola, D. The spread of behavior in an online social network experiment. *Science* **329**, 1194–1197 (2010).
46. Del Vicario, M., Scala, A., Caldarelli, G., Stanley, H. E. & Quattrociocchi, W. Modeling confirmation bias and polarization. *Sci. Rep.* **7**, 40391 (2017).
47. Baumann, F., Lorenz-Spreen, P., Sokolov, I. M. & Starnini, M. Modeling echo chambers and polarization dynamics in social networks. *Phys. Rev. Lett.* **124**, 048301 (2020).
48. Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B. & Lazer, D. Fake news on Twitter during the 2016 us presidential election. *Science* **363**, 374–378 (2019).
49. Mikolov, T., Yih, W.-T. & Zweig, G. Linguistic regularities in continuous space word representations. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 746–751 (Association for Computational Linguistics 2013 (Georgia, Atlanta, 2013)).
50. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. & Dean, J. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems—Volume 2*, NIPS'13, 3111–3119 (Curran Associates Inc., Red Hook, NY, USA, 2013).
51. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536. <https://doi.org/10.1038/323533a0> (1986).
52. Efron, B. & Tibshirani, R. J. *An Introduction to the Bootstrap* (CRC Press, London, 1994).

### Author contributions

M.C., A.G., C.M.V., A.L.S., P.Z. collected and prepared the data. All authors conceived the experiments. M.C., A.G., C.M.V., A.L.S., E.B., and A.S. conducted the experiments. All authors analysed the results, wrote and reviewed the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to W.Q.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

## Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).

Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.

These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.

We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.

While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval, sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.

These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.

To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.

Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.

If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

[onlineservice@springernature.com](mailto:onlineservice@springernature.com)