Italian top actors during the Covid-19 infodemic on Twitter

Paola Zola

Institute of Informatics and Telematics (IIT),
National Research Council (CNR), Pisa, Italy

Guglielmo Cola*

Institute of Informatics and Telematics (IIT),
National Research Council (CNR), Pisa, Italy
*Corresponding author: guglielmo.cola@iit.cnr.it

Antonio Martella

Department of Sociology and Social Research,
University of Trento, Trento, Italy

Maurizio Tesconi

Institute of Informatics and Telematics (IIT),
National Research Council (CNR), Pisa, Italy

Abstract: The Covid-19 pandemic has led to a corresponding infodemic, emphasized by the use of social media as the primary communication channel during lockdowns. This study was aimed at finding the accounts that spread information in Italian on Covid-19, and how such information was propagated in the first Western country to face a lockdown. The presented analysis shows that, besides authoritative news media and institutional accounts, a relevant role was played by actors from the "Civil society", which included a popular virologist as well as a far-right activist and an unfamiliar account supporting anti-government and anti-immigration ideas. Quite surprisingly, this latter account achieved the highest number of retweets despite a relatively low number of followers. Also, it showed information propagation paths similar to health experts and institutions.

Keywords: coronavirus; covid-19; information disorder; network analysis; retweet cascades; social media; Twitter

1 Introduction

On the 11^{th} of March 2020, the World Health Organization (WHO) declared Covid-19 a pandemic. The virus probably originated in the Chinese region of Wuhan and then

quickly spread to other countries, leading governments to adopt unprecedented measures[1]. According to the WHO Director-General's speech, the Covid-19 pandemic has been accompanied by another relevant problem: the *infodemic* (Zarocostas, 2020). Measures like social distancing and self-quarantine have significantly limited real-life social interactions, thus encouraging online discussion and turning social media into the primary means of communication (Cronjé, 2021; Beldad, 2021). However, it is well known that alongside reliable and authoritative sources, social media are characterized by the presence of malicious accounts aimed at creating various forms of information disorder, which can expose individuals and society to health risks (Gallotti et al., 2020). Several studies already dealt with the use of social media as a communicative channel for health information and/or misinformation (Hagen et al., 2018). For instance, in the Republic of Iran, fake media messages encouraged people to fight Covid-19 by drinking whiskey and honey (Baines et al., 2020), while in Italy the suggestion was to wash the legs of pets with bleach[2]. Fear surrounding Covid-19 has been particularly intensified by social media, which have contributed to foster widespread uncertainty. This, in turn, has fueled people's anxiety and panic (Vaezi and Javanmard, 2020) to the extent that many Governments took legal actions to handle social media disinformation (Radu, 2020), while WHO established several guidelines to deal with information on the pandemic (Tangcharoensathien et al., 2020).

Initial studies on Covid-19 social media "contagion" have mainly focused on the concept of infodemic in terms of fake information diffusion. Cinelli et al. (2020) explored five different online platforms (i.e., Instagram, Reddit, Gab, YouTube, Twitter) to measure the infodemic dispersion and the incidence of fake news. It turned out that Twitter and YouTube exhibit the highest volume of posts related to Covid-19, and the first one appears more neutral in the amplification of unreliable sources' messages. Similarly, Pulido et al. (2020) have studied a sample of 18,000 tweets collected between the 6^{th} and the 7^{th} of February 2020. Then, the authors have checked the different information sources among false information, science-based evidence, fact-checking, etc. This analysis shows that, although false contents are very frequent on Twitter, they are less retweeted compared to scientific and reliable sources. Misinformation is also studied in the work by Singh et al. (2020), where the authors expand the analysis investigating temporal and spatial patterns in Covid tweets around the world, highlighting that Twitter conversations have followed the increase of Covid-19 cases and that misinformation sources URLs have been retweeted more than credible health sources. Differently, Medford et al. (2020) proposed the use of textual and sentiment analysis on a sample of tweets from the 14^{th} to the 28^{th} of January 2020 containing specific keywords related to the pandemic. The study shows that racial prejudices in tweets rose together with the increase of Covid-19 cases. In a recent work, Ferrara (2020) has studied about 43.4 million tweets in English to investigate bots activity related to Covid-19. This study shows that possible automated accounts predominantly posted political conspiracy content alongside Covid-related information. In previous studies, the link between conspiracy theories and the pandemic was studied for other health-related crises such as Zika (Kou et al., 2017; Hagen et al., 2018) and Ebola (Feuer, 2014). In particular, Kou et al. (2017) proposed a qualitative study of the content from conspiracy theories on Reddit during the Zika epidemic, finding that the social platform fostered the propagation of conspiratorial ideas.

Considering that the "escalation of the epidemics leads people to progressively pay attention to more reliable sources" (Gallotti et al., 2020), but also that highly digital societies are susceptible to infodemics, the present paper aims to give an overview of the main actors in the Italian Twittersphere that posted Covid-related contents. In particular, our analysis aims

to investigate if, leveraging on the inevitable fear and panic caused by the pandemic, some anomalous accounts have managed to achieve higher visibility. Italy was the first European country to adopt restrictive measures like private business closures, self-quarantines, and transportation limitations. Moreover, Italy was the first Western country to face the citizens' reaction to these imposed limitations. As such, we believe that an analysis of the social impact of Covid-19 in Italy could also help achieve a better understanding of the effects of the pandemic on the online public sphere of other Western democracies.

The main contributions of this study can be summarized as follows:

- A novel dataset about Covid-related tweets in Italian was collected, between the 30^{th} of January and the 20^{th} of March 2020;
- These tweets were analyzed to find the ten most-retweeted accounts;
- Further analyses examined the role of these accounts, studying the information diffusion paths to understand how and why these Twitter actors attracted a large audience.

The rest of the paper is organized as follows. The next Section **Background and research questions** describes the previous studies on information diffusion and dynamics in social media, especially focusing on the Twittersphere. This Section ends with a list of research questions that this paper aims to investigate. Section **Methodology** describes the methods adopted in our analyses, whereas **Covid-19 most retweeted accounts in Italy** reports an initial analysis of the dataset and the principal Italian actors involved in spreading Covid-related contents. Section **Accounts activity and reach** attempts to investigate the rationale behind the presence of unexpected accounts among the most retweeted ones, and describes the retweeters' behavior analysis. Finally, Section **Conclusions** summarizes the main findings and implications.

2 Background and Research questions

In the last decades, social media have gradually become an essential source of news and information throughout the world (Boulianne and Theocharis, 2020; Dubois et al., 2020). In general, the lack of subscription costs in social media led to the so-called "democratizing effect" (Hagen et al., 2020; Ida et al., 2020) on public life, as it has expanded access to information and created greater opportunities for civic engagement. Alongside reliable and trustful information, social media are characterized by the proliferation of fake and tendentious news (Bastos and Mercea, 2019). Promoting false information may have three additional effects: i) to drown out legitimate grassroots/minority interest movements; ii) to ascribe false legitimacy to fringe ideas by creating the impression of widespread support; and iii) to possibly overwhelm critical information channels with noise, spam, and propaganda during crisis events and public emergencies (Hagen et al., 2020).

A wide range of studies focused on detecting malicious users and their contents aimed at reinforcing conspiracy theories, fake news, and extremist movements on social media (Alvari et al., 2020; Mancosu and Vegetti, 2020). Indeed, some malicious automated accounts have been caught fostering online polarization by entangling users into "echo chambers", which are difficult to penetrate by reliable sources (Yuan et al., 2019). In this context, many studies (Hilbert et al., 2017) focused on social media users' behavior patterns in spreading online information and their engagement in political campaigns,

highlighting different users' categories based on their communication flow. Bracciale et al. (2018) focused on the online discussion in the Italian electoral Twittersphere from January to March 2018, to evaluate public participation on Twitter during an election campaign and find out whether the online communication space was dominated by a handful of actors. Identification of Twitter users was based on seven categories, namely *Media* (media outlets and journalists), *Politics* (including official accounts linked to governments), *Citizens* (individuals without any openly disclosed affiliation), *Political supporters* (users who openly support a political movement), *social media experts* (bloggers and social media managers), and *Civil society* (celebrities as well as cultural and research organizations, activists and "intellectuals"). The analysis highlighted that new actors are potentially able to gain significant attention in online debates, based on their success at being retweeted (super-echoed). The authors underlined that *Politics and Institutions* and *Media* were the most retweeted accounts, followed by *Civil Society*.

Considering that the epidemic escalation fostered citizens' attention to reliable news sources, but at the same time the infodemic was sustained by several unverified sources (Gallotti et al., 2020), our first research question is:

RQ.1 Which are the most retweeted accounts in the Italian Twittersphere, in the communication flows related to Covid-19?

Several other studies showed that online messages are often deployed to the wider public by *influencers* or opinion leaders, who are positioned at the center of social networks and information flows (Katz, 1957; Choi, 2015). In practice, the presence of influencers can amplify the visibility of specific contents through the *secondary audiences* dynamics (followers of followers) (Vaccari and Valeriani, 2015), which enable reaching a wider community and may lead to a "double-viral chain". In fact, many social media are characterized by the presence of recommendation algorithms that suggest content based on its diffusion (viral messages) reinforcing content "virality" (Proferes, 2017) and fostering the well-known Saint Matthew effect in content diffusion. Therefore, another research question arises:

RQ.2 Which and how wide are the online audiences of the most retweeted accounts?

2.1 Twitter social interactions

Twitter is one of the most adopted social media worldwide, with more than 330 million monthly active users [3]. Several studies argued that it can be considered as an information network given its particular affordances that promote information propagation beyond the author's social network (Myers et al., 2014). Indeed, even if born as a social media, Twitter has assumed a growing influence in political campaigns (Shmargad and Sanchez, 2020), political communication (Bracciale and Martella, 2017), financial predictions (Zola et al., 2020b), news diffusion (Bastos and Mercea, 2019), and civic e-participation (Al-Aama, 2015).

In particular, differently from other social media, Twitter offers three main actions to interact and express interest in a specific content, which enables a more interesting analysis of such interactions and information propagation. These three actions are: *favorite*, *reply*, and *retweet*. Among these actions, retweets represent the most popular way of spreading contents (Firdaus et al., 2018), making the original tweet available to a larger audience. Unfortunately, the Twitter APIs do not provide complete information about the sharing action: given a retweet posted by a user, it is not always possible to derive with certainty

from whom the user has retweeted the content. In particular, if the user does not follow the author of the original content, he/she could have retweeted the content from any other user who also retweeted the content.

Moreover, a higher engagement achieved by a tweet leads to a higher probability that the message's hashtags are included in the trending topic list, becoming visible to many users. Indeed, as argued by Zhang et al. (2016b), there are factors affecting the Twitter trending topic algorithm that might be easily manipulated by malicious accounts, such as topic popularity and the potential reach of a content. Considering that content spread in Twitter networks may depend on account popularity, opinion leaders, or crowds dynamics (i.e., the joint activity of several ordinary users (Zhang et al., 2016a)), our last research question is:

RQ.3 How do the most retweeted accounts spread their messages and reach their audiences?

3 Methodology

We collected tweets by using the Twitter Streaming API from the 30^{th} of January 2020 to the 20^{th} of March 2020, selecting the tweets that used at least one of the following keywords: coronavirus, coronavirusitalia. These keywords were chosen considering the *trending topics* of Twitter during the monitored period. The resulting dataset includes 3,690,196 contents in Italian, consisting of original tweets, retweets, and replies.

Following other studies that have found that participation and attention on Twitter are highly skewed (Graham and Wright, 2014; Larsson and Moe, 2012), we identified the ten most retweeted Italian Twitter accounts and analyzed each of these actors by evaluating their main features and their ability to spread their messages through retweets. Indeed, confirming previous studies on attention inequality in the Italian Twittersphere (Bracciale et al., 2018), these actors received 290,431 retweets corresponding to 11.2% of the total number of retweets (2,581,680), which represents a great share of attention considering that 354,926 users adopted the selected keywords. To highlight content overlaps, we calculated Jaccard similarity based on used hashtags, ignoring the above-mentioned keywords exploited for tweets collection. We adopted the same procedure to identify similar audiences among the monitored accounts, by computing Jaccard similarity among retweeter IDs for each couple of accounts. To evaluate the audiences reached by the top ten actors, we measured their *potential reach* as the sum of the followers number of the retweeters for each tweet. We relied on basic statistical representations to highlight differences in feature distributions among the top retweeted accounts.

As mentioned in Section 2.1, the Twitter APIs do not provide full information on how the content of a tweet was spread through retweets: a user may have retweeted a content because it was previously retweeted by one of his/her friends, instead of retweeting directly from the original tweet author (source). Researchers have proposed several techniques to estimate the so-called retweet cascade graph, which describes the flow of information from the source to the retweeters (Zola et al., 2020a). In this work, we adopted the technique based on social relationships within the platform (follower/following) and retweeting timestamps to determine the most likely "influencer" that spread the content to a specific user (Taxidou and Fischer, 2014). Basically, when a retweet is posted by a user U, all U's friends (i.e. the users followed by U) that retweeted the same content before U are considered as potential influencers; the most likely influencer is the last one that posted before U. An example is

User	Time [dd-mm-yy hh:mm]	Action
A	16-04-20 16:30	Original tweet
B	16-04-20 16:32	Retweet
C	16-04-20 16:34	Retweet
D	16-04-20 16:35	Retweet
E	16-04-20 16:36	Retweet

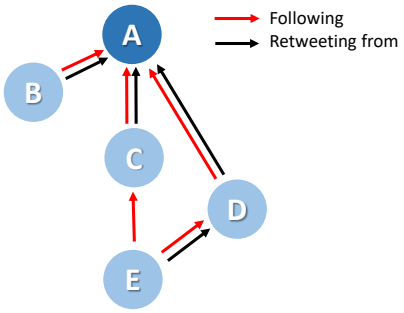


Figure 1: Retweet cascade graph estimation example.

shown in Figure 1. User A is the original source, whereas B, C, D, and E are retweeters. User B, C, and D do not have friends that retweeted before them, thus they retweeted from the source A. Conversely, user E has two friends who retweeted before: C and D. As D was the last to retweet before E, the latter is linked to D in the retweet cascade graph (black arrow).

This approach allows us to derive the *retweet cascade* information graph. The retweet cascade C is composed of a directed tree T and a set of sparse nodes S. The directed tree is defined by a set of nodes N which are linked by directed edges E. Each node $n \in N$ represents a user who retweeted the tweet and that is linked directly or indirectly to the tweet's author (root). Indeed, an edge $e = (n1, n2)$ means that user $n1$ retweeted from user $n2$. Hence, for each node belonging to the tree T there is a path to the root made of one or more edges. In addition to the users in N, there are also retweeters who are not linked (i.e., who do not follow) to any other user in the cascade: these users represent the sparse nodes S of the cascade.

Once the retweet cascade is found, it is possible to derive some metrics to compare the different cascades:

- *Cascade size (total node count)*, the number of nodes N in the directed tree plus the sparse nodes S, namely the total count of users that retweeted the root tweet;
- *Cascade levels (depth)*, the number of levels in the directed tree, which shows how much the original tweet was able to reach users far from the author's social network (not considering the sparse nodes);
- *Followers Ratio (FR)*, assessing the proportion of followers of the source (nodes in N connected to the root of the tree by an edge) with respect to the total number of retweeters;
- *Sparse nodes incidence*, measuring the incidence of sparse nodes in S with respect to the total number of nodes (i.e., node count).

In addition to these cascade metrics, we compute each nodes' *descendant count*, which denotes the number of retweets directly or indirectly caused by a specific node (i.e., all the nodes that are "below" the node in the tree graph). This metric is crucial in identifying influential nodes in the cascade propagation path.

4 Covid-19 most retweeted accounts in Italy

In order to answer RQ.1, we identified the ten most retweeted accounts and classified them according to categories proposed by Bracciale et al. (2018). Table 1 shows these accounts and some information on their tweets and received retweets. The dataset with the retweet IDs is publicly available[4]. Four out of the ten sources, *Agenzia ANSA* (@Agenzia_Ansa), *Sky Tg24* (@SkyTG24), *Mediaset TgCom24* (@MediasetTgcom24), and *Repubblica* (@repubblica) are well-known Italian online news providers. In particular, Agenzia Ansa is the leading news agency in Italy, Sky Tg24 and Mediaset TgCom24 are the official profiles of the respective news television channels, Repubblica is a popular newspaper founded in 1976. *Ultime Notizie* (@ultimenotizie), instead, is an online news provider only active as a web page and in social media. Recalling the social media profile classification by (Hilbert et al., 2017; Bracciale et al., 2018), these five accounts fall in the *Media* class. The profile named *Ministero della Sanità* (@MinisteroSalute) is the official account of the Italian Ministry of Health. *Matteo Salvini* (@matteosalvinimi) instead is the leader of the right-wing party Lega. Both these accounts belong to the *Politics and Institutions* category.

Francesca Totolo (@francescatotolo) is a social media manager and a contributor of *Il primato nazionale*, a sovereignist/nationalist online magazine linked to *Casa Pound*, a far-right political movement. *Roberto Burioni* (@RobertoBurioni) is an Italian professor and doctor expert in virology, who played a relevant role during the spread of the pandemic, as he was widely considered as a valuable source and appeared in several television shows. *RadioSavana* (@RadioSavana) is a Twitter account not specifically linked to any individual, political party, or movement. It publishes various kinds of content ranging from economics, politics, and videos against immigration, the European Union, and the Italian Government (mainly composed by Movimento 5 Stelle and center-left Partito Democratico in the monitored period). RadioSavana has already been object of attention from online portals and news providers, which described its contents as unreliable and distorted.[5,6,7] These last three profiles can be all identified as members of the *Civil Society*.

Table 1 Overview of the Twitter Accounts

Account	Category	Tweet Count	Received Retweets	Followers (Apr 2020)	Friends (Apr 2020)	Unique Retweeters	Retweets / Tweets	Retweets / Retweeters	Retweeters / Followers
Agenzia Ansa	Media	2,122	36,722	1,218,612	221	11,031	17.3	3.3	0.9%
Francesca Totolo	Civil Society	503	29,091	24,999	706	5,344	57.8	5.4	21.4%
Matteo Salvini	Politics & Inst.	96	21,827	1,270,788	1,834	6,267	227.4	3.5	0.5%
Mediaset TgCom24	Media	1,597	21,750	1,037,676	977	5,854	13.6	3.7	0.6%
Ministry of Health	Politics & Inst.	121	18,436	144,109	215	10,544	152.4	1.7	7.3%
RadioSavana	Civil Society	435	65,890	17,403	2,000	13,230	151.5	5.0	76.0%
Repubblica	Media	3,899	25,076	2,996,260	372	10,296	6.4	2.4	0.3%
Roberto Burioni	Civil Society	168	29,583	233,967	1,106	15,492	176.1	1.9	6.6%
Sky Tg24	Media	1,825	17,571	3,114,534	344	7,212	9.6	2.4	0.2%
Ultime Notizie	Media	2,164	24,485	105,935	341	7,644	11.3	3.2	7.2%

As shown in Table 1, the Twitter source with the highest number of retweets per tweet is Matteo Salvini (about 227 retweets for each tweet on average), followed by Roberto Burioni, the Ministry of Health, and RadioSavana. Considering the social network of the ten accounts, Sky Tg24 has the highest number of followers (more than 3 million) and, in general, authoritative news media have a significantly higher number of followers than the other accounts. RadioSavana has the lowest number of followers, with only $17,403$ users. Nevertheless, it received the highest number of retweets about Covid-19 ($65,890$), followed at a distance by Agenzia Ansa ($36,722$), Roberto Burioni ($29,583$), and Francesca Totolo

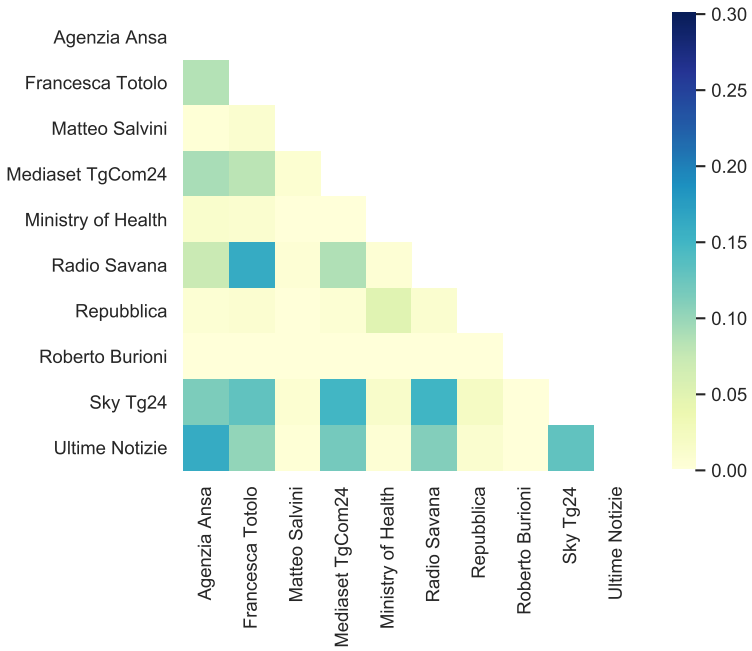


Figure 2: Hashtags similarities among the top ten Twitter Covid-19 actors

(29, 091). Thus, at first glance, the most retweeted accounts on the Covid-19 debate in Italy belong to the *Civil Society* category except for Agenzia Ansa. Moreover, from Table 1 it is apparent how RadioSavana's unique retweeters are 13, 230, about 76% of its total follower count, followed at a distance by Francesca Totolo (21.4%). In other words, RadioSavana seems to have extremely active followers, or it might have been able to attract a relatively high number of retweets from non-followers. These metrics may suggest the presence of an anomaly in the information diffusion path, which is further investigated in the following.

To better understand the public of these ten accounts and their potential audience, we analyzed tweeted contents and the similarities among retweeters. Observing the most adopted hashtags by each account, it is possible to identify the different topics (Small, 2011) associated with Covid-19 by the sources, and also the specific "ad-hoc publics" to which these hashtags were addressed (Bruns and Burgess, 2011). Jaccard similarity, reported in Figure 2, was used to compare the hashtags of each possible couple of sources. It turns out that the highest Jaccard value is between RadioSavana and Francesca Totolo, even if RadioSavana also shows similarities with Sky Tg24. The latter similarity is above all due to geo-localizing hashtags related to the spread of the pandemic in Italian cities (e.g., *Milano*, *Roma*, *Napoli*, etc). Naming cities and regions hit by Covid-19 emerged as a quite common pattern for all *Media* accounts, and this explains the high similarity among them. In contrast, Roberto Burioni, the Ministry of Health, and Matteo Salvini show a null similarity with the other sources. These three accounts used a very limited number of hashtags: the Ministry of Health did not cover other topics besides coronavirus and its consequences (the most frequent hashtag was *#iorestoacasa* [eng: *IStayAtHome*]), while Matteo Salvini is characterized only by self-referred hashtags (*#Salvini*) and hashtags against the Italian prime minister Giuseppe Conte (*#ConteDimettiti* [eng: *Conte resign*]).

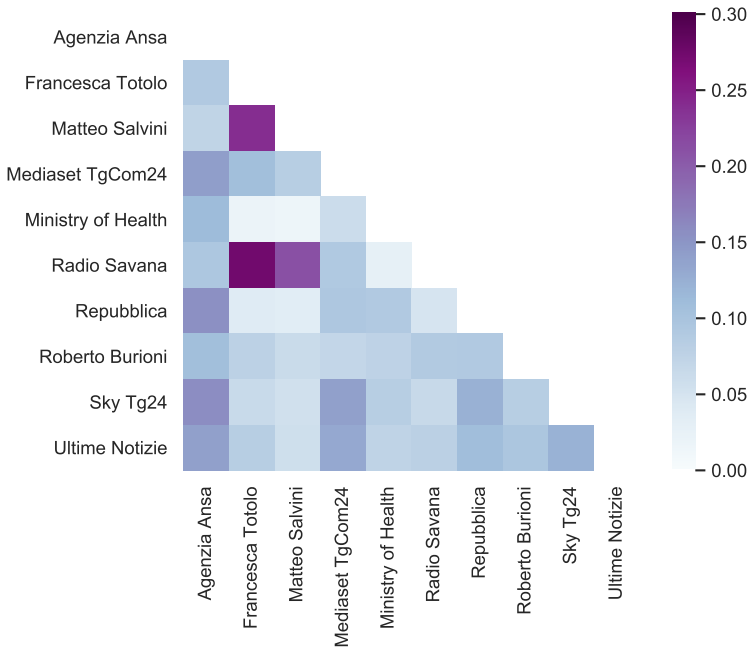


Figure 3: Audience similarities among the top ten Twitter Covid-19 actors

Roberto Burioni used only one hashtag, namely the abbreviation of a popular Italian talk show to which he participated as a guest during the pandemic (*#ctcf - Che Tempo Che Fa*). The common topics characterizing the high similarity between RadioSavana and Francesca Totolo mainly regard the Italian PM (*#Conte*), foreign politics related to the pandemic (e.g., *#Cina*, *#Wuhan*, *#Macron*, *#Erdogan*, etc) and immigration-related issues (e.g., *#oceanviking*, *#ong*, etc). In fact, among the most frequent hashtags adopted by RadioSavana, there are anti-immigration (*#risorseinps/#risorsainps*[8]) and anti-government phrases (*#governodelcontagio [eng:* Government of contagion*]*). It is worth noting that RadioSavana also adopted the hashtag *#Salvini*. Francesca Totolo was mainly devoted to anti-immigration ideas, for instance, *#migranti, #oceanviking, #ong, #torniamoitalia, #africa* [eng: *migrants, oceanviking, non-governmental organizations, let us be Italy again, Africa*] were the most used hashtags by the activist/social media manager.

The similarities pointed out from Figure 2 are also confirmed by the audience overlap among the ten profiles: Figure 3 reports the Jaccard similarity among retweeters. This metric measures the overlap between two sources in terms of retweeters. A high similarity indicates that there is a high number of shared listeners between two actors. From Figure 3 it follows that, among considered actors, *Media* audiences tend to retweet other *Media* sources, whereas Francesca Totolo, Matteo Salvini, and RadioSavana show a similar population of retweeters, meaning that Twitter accounts retweeting one of these sources are also likely to share contents from the other two.

Indeed, Francesca Totolo, Matteo Salvini, and RadioSavana audiences show the lowest level of similarity with users retweeting the Ministry of Health and Repubblica, followed by Sky Tg24. Therefore, both the analysis on contents' and audiences' overlap seem to show a political/ideological connection between ad-hoc publics and users who retweeted these

accounts. This leads to the conclusion that the unexpected online visibility of Francesca Totolo and RadioSavana (Table 1) might have been mainly driven by political beliefs and aims. On the other hand, *Media* audiences seem quite similar, indicating a low level of polarization.

These results lead to partially answer to RQ.2 about which are the audiences of the most influential accounts. In fact, according to our analysis, ad-hoc publics identified by hashtags are quite similar among *Media* category except for Repubblica. On the other hand, four accounts show high similarity based on adopted hashtags: Mediaset TgCom24, Sky Tg24, RadioSavana, and Francesca Totolo. The last two accounts are linked by similar topics (against the government and immigrants) but they also used geo-localizing hashtags similar to *Media* accounts. Indeed, similarities among retweeters showed strong connections beyond the topics and publics addressed through hashtags. There is a connected cluster of retweeters shared by Matteo Salvini, RadioSavana, and Francesca Totolo. This means that three of the most retweeted accounts during the pandemic in Italy were ideologically linked to each other and addressed similar audiences. Thus, our findings suggest that while *Media* audiences were not so polarized in the studied period (quite similar hashtags and retweeters), two accounts from *Civil Society* (Francesca Totolo and RadioSavana) and one account from *Politics and Institutions* (Matteo Salvini) had strongly interconnected audiences.

5 Accounts activity and reach

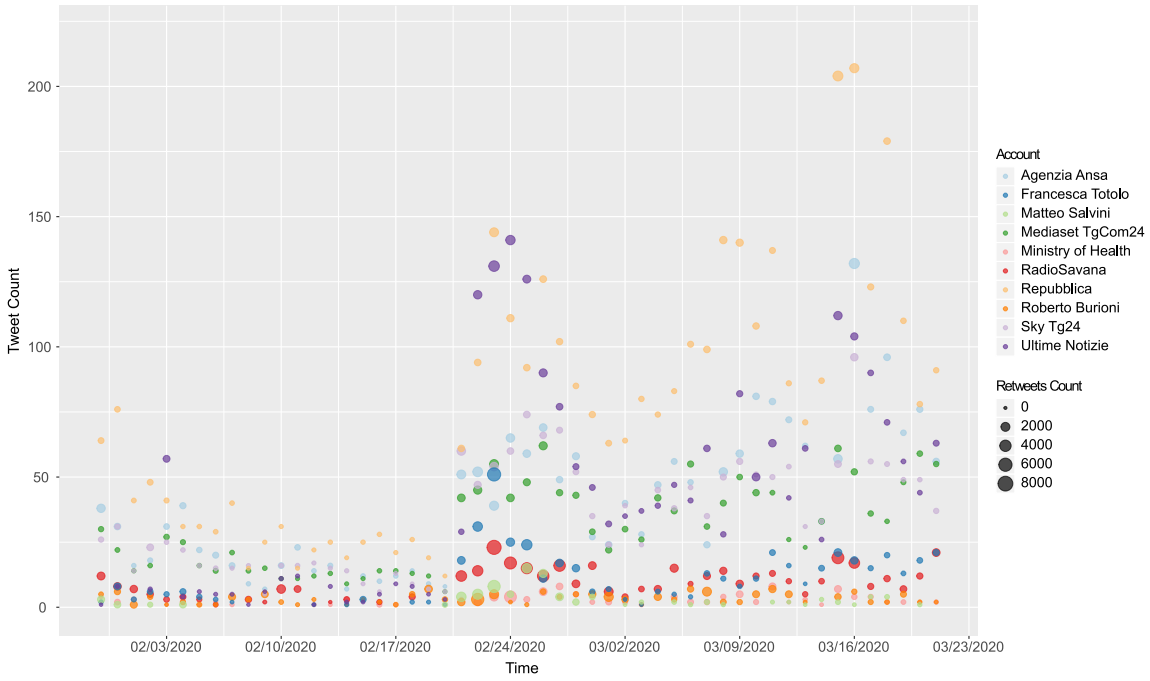


Figure 4: Tweets and Retweets temporal distribution

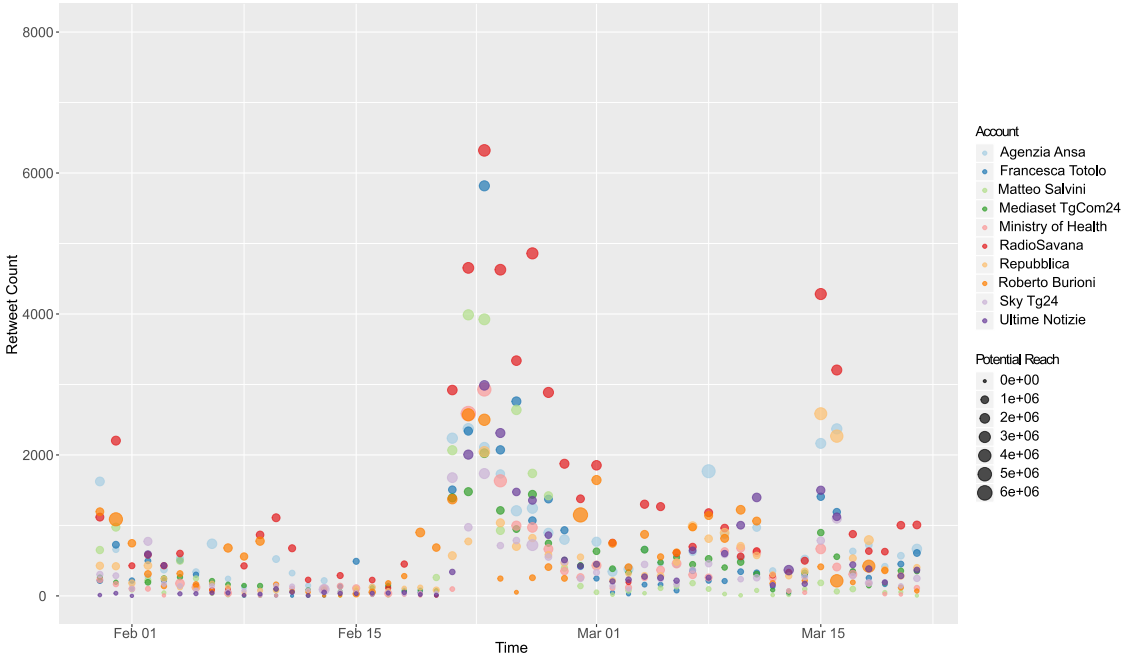


Figure 5: Tweets potential reach

In this section we explore the information spread linked to the Covid-related tweets of the top ten Italian actors to answer RQ.3. The first analysis involves the study of the accounts' activity related to Covid-19 on Twitter; then we show the information diffusion graph as described in section **Methodology**, highlighting patterns among the accounts and deriving possible anomalies. Starting from the tweet and retweet activity, Figure 4 reports the daily number of posted tweets by the top ten accounts and the respective received retweets, which are depicted in the dot sizes. The different dot colors denote the ten accounts. This plot shows the increasing dynamics of the tweets related to the Covid-19 topic, which reached its maximum at the beginning of March as the lockdown was confirmed all over the country. Another important spike is visible on the 20^{th} of February, when the first Italian citizen positive to Covid-19 was found. *Media* sources were, in general, the most active, posting more than 100 contents per day, while *Politics and Institutions* produced just a few tweets per day. Despite the lower activity of *Politics and Institutions* and *Civil society*, the accounts' dot size shows a vast amount of received retweets. In particular, the biggest dots are tweets belonging to RadioSavana, followed by Francesca Totolo and Roberto Burioni.

Analyzing the ten sources' *potential reach*, Figure 5 reports the daily accounts' activity and shows the number or received retweets for each produced content, which is represented by a dot. Dot color denotes the accounts as in Figure 4, while dot size depicts the total number of unique followers of the retweeters, i.e., the potential reach. We labeled it as *potential* because it represents the maximum number of users that might have been reached at the second step of the retweet chain (secondary audience). Figure 5 shows that, notwithstanding the high level of activity, *Media* accounts are characterized by small dots, denoting that people retweeting their content were usually not classifiable as influencers. Similarly, Francesca Totolo and Matteo Salvini show smaller dots with respect to RadioSavana, which

on certain days is comparable to the Ministry of Health. However, Roberto Burioni appears as the account with the highest *potential reach* and, despite the low activity of the account, he was able to attract influencers' interest. This result is not surprising considering the high popularity of the professor in TV shows during the pandemic, as hypothesized in Gallotti et al. (2020).

These results allow us to definitively answer RQ.2 about how wide were the audiences of the top retweeted accounts. The most engaging accounts, who were able to spread their messages far beyond their followers, were Roberto Burioni, the Ministry of Health, and RadioSavana. This pattern is quite expected for Burioni and the Ministry of Health, who are both authoritative sources, but it is quite surprising for a relatively unknown Twitter account like RadioSavana. Our analyses allow us to hypothesize that both the very active audiences of RadioSavana (Table 1) and the negative emotional content that emerged in its tweets (criticism against the government, immigrants, the EU, etc.) played a role in boosting visibility, as emerged in other studies (Hansen et al., 2011). This allowed RadioSavana to compete with the most reliable sources in the Italian Twittersphere. Indeed, this result is even more surprising if we consider that the evolution of the pandemic has led people to pay more attention to reliable sources (Gallotti et al., 2020).

5.1 Information propagation paths among Twitter users

Here we investigate the retweet cascade graphs adopting the state-of-the-art method previously described in Section 3. Figure 6 reports the distribution of the cascade size (i.e., the number of nodes involved in the propagation by retweeting the original content) in logarithmic scale. Interestingly, the cascade size is lower for *Media* compared to the other two groups. Moreover, the *Media* class is characterized by numerous outliers, meaning that while news typically received a low number of retweets, some contents were still able to get high visibility. Relatively high median values were achieved by *Civil Society* and *Politcs and Institutions* accounts. In particular, RadioSavana shows a median value in line with the Ministry of Health and Roberto Burioni, notwithstanding the huge gap in terms of followers ($17,403$ for RadioSavana vs $144,109$ and $233,967$ for the Ministry of Health and Roberto Burioni, respectively).

Overall, these results show that the monitored keywords have been better exploited by *Civil Society* and *Politcs and Institutions* accounts. It is worth noting that these categories are composed of very different kinds of actors: on the one hand, there are experts (Roberto Burioni) and institutions (Ministry of Health), on the other, a recognized populist leader (Matteo Salvini) opposed to the Government (Bobba, 2019), and an extremely active user (RadioSavana) not linked to any known personality, and which has been accused of spreading misinformation (mainly on immigration-related issues).

Another relevant metric is the number of levels in retweet cascades, also known as cascade depth, which is shown in Figure 7. *Media* accounts tend to have all the nodes of the retweet cascade graph in the first level, indicating that the majority of retweets come directly from followers. Similarly, the cascades of Matteo Salvini and Francesca Totolo are characterized by low depth, whereas Roberto Burioni, the Ministry of Health, and RadioSavana show a higher number of levels. Such high depth denotes that even accounts not directly linked to the source read the contents and contributed to spread their tweets. It follows that the incidence of sparse nodes (Figure 8), namely the proportion of nodes that are not connected to the source, is higher for Roberto Burioni, the Ministry of Health, and RadioSavana. However, it should be noted a substantial difference in terms of the distribution

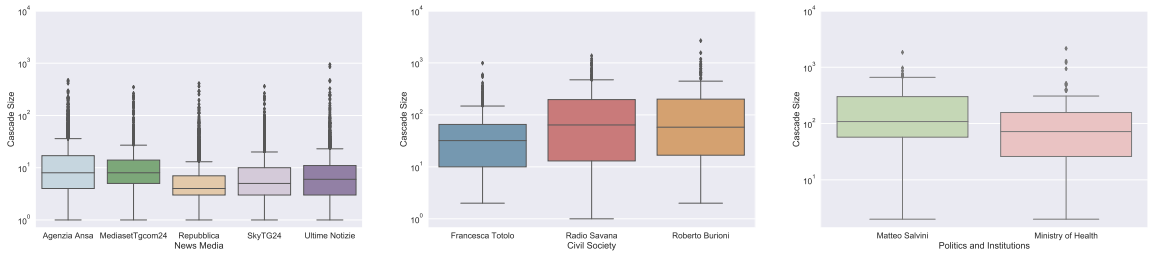


Figure 6: Retweets cascades size (Log scale)

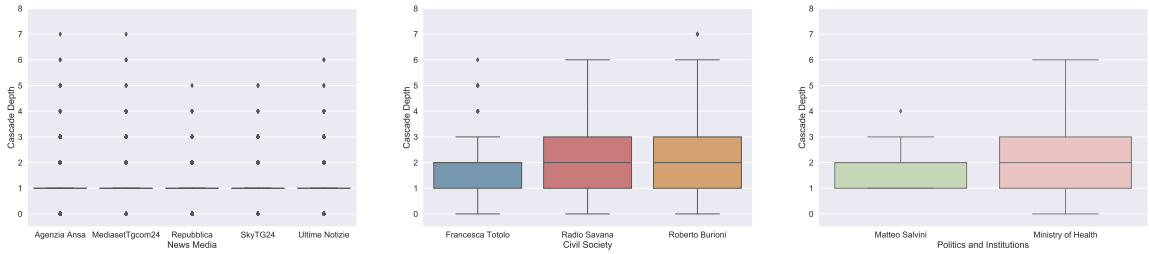


Figure 7: Cascades Depth distribution

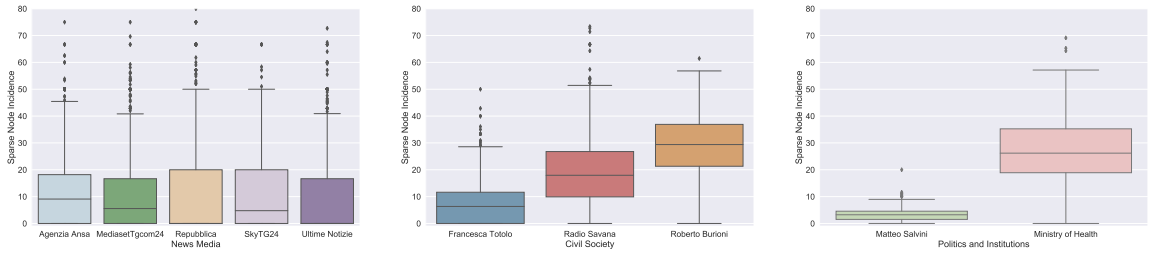


Figure 8: Sparse nodes incidence

of sparse nodes among these accounts. Indeed, Roberto Burioni and the Ministry of Health show an almost normal distribution, with a median incidence of sparse nodes around 30% and 25%, respectively, whereas for RadioSavana this distribution is right-skewed, with a median value below 20%. This suggests that the majority of RadioSavana's cascades have less sparse nodes than Burioni and the Ministry of Health. Finally, Matteo Salvini's sparse node incidence plot confirms the finding in Figure 7, as almost all his retweets came from followers.

The fact that *Media* retweets come mainly from followers is not surprising, considering the role of these accounts as news sources and their wide audiences ranging from 1 to 3 million followers (except for Ultime Notizie). The outliers in cascade depth distribution represent a minority of "catchy" contents that became viral and were thus spread even by non-followers.

On the other hand, the cascade depth of the other accounts shows some interesting cross-category patterns, notwithstanding the substantial difference in terms of the number

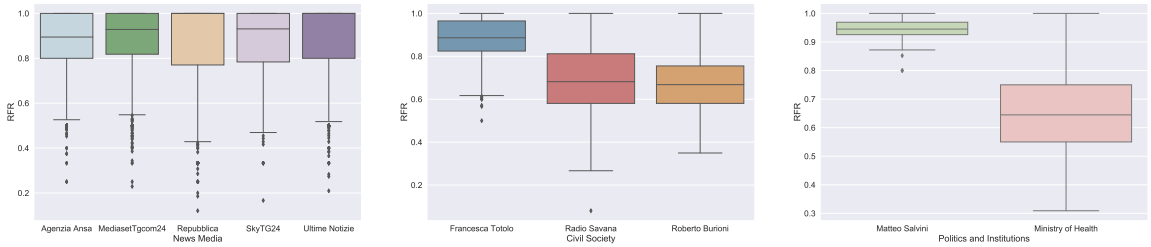


Figure 9: Followers Ratio (FR) distribution

of followers among the selected actors. First, *Civil Society* and *Politics and Institutions* cascades show a relatively low number of outliers and high median values in the depth distribution compared to *Media*. Hence, these accounts were able to trigger retweets chains more frequently than *Media* in the monitored period. In other words, their secondary audiences (Vaccari and Valeriani, 2015) were more engaged in the diffusion of their contents.

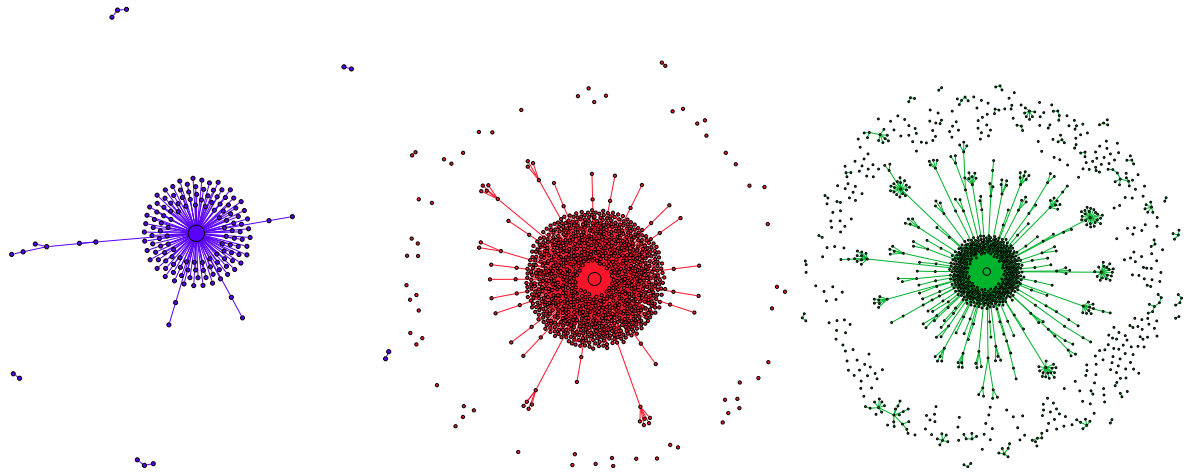
Regarding cross-category patterns, it is worth noting that the limited depth of Matteo Salvini's and Francesca Totolo's cascades means that their audiences were strongly connected to their accounts: from a network point of view, the majority of their retweeters are separated by no more than two levels from the source. This characteristic of Salvini and Totolo is also confirmed by FR analysis, shown in Figure 9. The higher the value of FR, the higher the percentage of accounts in the first level of the retweet cascade, namely users who directly follow the source. In comparison, Ministry of Health's, Roberto Burioni's, and RadioSavana's retweeters are substantially more distant from their respective source. This is apparent by looking at the median depth in Figure 7 and even more by considering Figure 9: these three accounts have the lowest FR median values (about 70%), which highlight a relatively higher incidence of accounts that were not direct followers.

Based on cascade depth (small, medium, and large), it is possible to identify three main groups for the studied accounts:

- *Two Step* (small depth), represented by *Media* accounts, whose contents are shared directly by their first-level followers recalling the two-step flow of communication (Katz, 1957) (see Figure 10a). Retweet cascades in this group are characterized by the so-called star shape, with the central node directly connected to the majority of retweeters (Zhou et al., 2017);

- *Small Words* (medium depth) represented by Matteo Salvini and Francesca Totolo, who are mostly retweeted by users belonging to their network, but are also able to attract audiences not too "far" from them and probably tightly knitted (Figure 10b). This group's cascades show a visible first level but also some "chains", indicating a relatively higher diffusion of information beyond first-level followers (Zhou et al., 2017);

- *Multi-center stars* (large depth), represented by Roberto Burioni, the Ministry of Health and RadioSavana, who are retweeted by a high number of users separated by more than two levels from the source and reaching a depth up to six levels (Figure 10c). These cascades are characterized by the presence of multi-stars, indicating the presence of multiple nodes able to spread the content to their respective communities (Zhou et al., 2017).



(a) Repubblica (b) Matteo Salvini (c) RadioSavana

Figure 10: Examples of retweet cascades.

In conclusion, it should be highlighted that the offline popularity and the recognized authority of Roberto Burioni and the Ministry of Health can explain a high incidence of retweets from non-followers. However, this does not apply to RadioSavana, which had a relatively small number of followers and had no access to mainstream media. Thus, we can hypothesize the presence of "hidden influencers"(Borge-Holthoefer et al., 2013), which helped spread information to non-followers, or some other anomaly among the population of retweeters.

5.2 Covid-19 Italian retweeters behavior

Given the results obtained in the previous sections, we here investigate the presence of *opinion leader* accounts in the tweet information graphs, which might have pushed the contents to new communities. The analysis is based on measuring the descendant count for the nodes belonging to each cascade. The descendants of a node are the nodes directly or indirectly linked to the node, namely direct or indirect followers. Figure 11 shows the descendants distribution for the ten accounts: the y-axis indicates a specific account, the x-axis shows bins of descendant count values (e.g., nodes with descendant count ranging from 10 to 20), while the gradient indicates the number of nodes (a darker color indicates a higher number of nodes).

As expected, the majority of nodes are leaves having no descendants. A higher number of influential nodes is associated with the Ministry of Health, Roberto Burioni, and RadioSavana, which all belong to the *Multi-center stars* diffusion pattern. These three actors have some retweeters that were able to spread the information to more than 99 additional nodes. In contrast, for the other accounts there were no retweeters able to spread their messages to more than 40 nodes. More precisely, the descendant count is always below 40 for Agenzia Ansa, below 30 for Ultime Notizie, below 20 for Francesca Totolo, Mediaset TGCom24, Repubblica, Sky Tg24, and below 10 for Matteo Salvini. If the findings of Figure 11 are read in conjunction with Figure 5, it emerges that the higher descendant

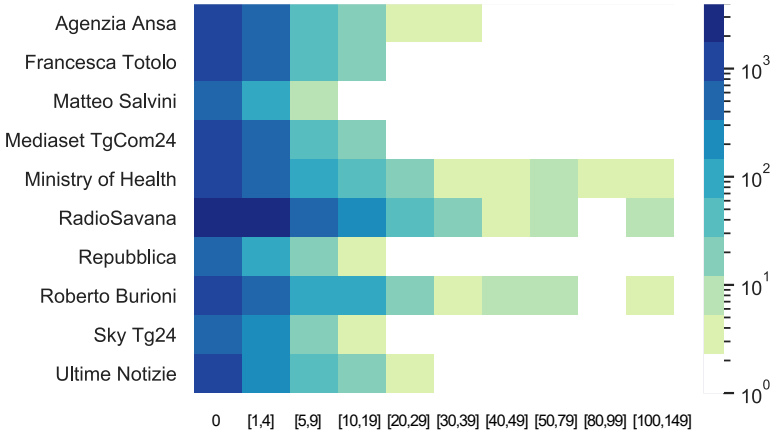


Figure 11: Descendants count distribution

count might be linked to the influential accounts that contributed to the propagation of contents. Again, this is quite expected for reliable accounts like the Ministry of Health and Roberto Burioni, whereas it is surprising for RadioSavana. In line with other studies (Zhang et al., 2016a), our findings show that content spread patterns on Twitter do not necessarily depend on account popularity (e.g., the number of followers), but that followers' activity can trigger molecular opinion leaders (Katz, 1957), able to reach *niche audiences* boosting content visibility despite the general trends emerged during the pandemic (Gallotti et al., 2020).

6 Conclusions

In this work we have identified and analyzed the ten most-retweeted Italian "actors" in the Twittersphere, considering contents related to the Covid-19 pandemic. In the period between the 30^{th} of January and the 20^{th} of March 2020, we collected 3,690,196 contents in Italian including Covid-related keywords. Considering the skewness of participation and attention on Twitter (Graham and Wright, 2014; Larsson and Moe, 2012; Bracciale et al., 2018), we focused our analysis on the ten most-retweeted accounts, which received 11.2% of all the retweets. As expected, *Media* and *Politics and Institutions* played an important role in Covid-related information diffusion (Gallotti et al., 2020). The first category (*Media*) is represented by well-known media sources (Agenzia Ansa, Repubblica, etc.), while the second is represented by Matteo Salvini (leader of the Lega party) and by the account of the Ministry of Health. However, also some actors from the *Civil Society* managed to assume a central role in the Covid-19 debate on Twitter. In our top ten selection, this category is represented by Roberto Burioni, a popular professor and doctor expert in virology, and also by two emerging players: Francesca Totolo and RadioSavana. Francesca Totolo is an activist/social media manager, close to far-right movements, while RadioSavana is a recent account (created in 2019) predominantly posting anti-government and anti-immigration videos. Both accounts achieved a high number of retweets during the considered period, despite a relatively low number of followers compared to the other sources in our list.

In particular, among the top ten Covid-19 retweeted accounts, RadioSavana achieved the highest number of retweets.

To explain the reasons behind the huge visibility achieved by these ten accounts, we investigated contents, communities, and behaviors, also studying the profiles' abilities in engaging audiences beyond their follower base. We derived the presence of three main different attitudes. The public information scope characterized the *Media* accounts, which posted several tweets per day mainly engaging their social network. Then, a relevant role was played by Roberto Burioni and the Ministry of Health, which achieved huge visibility during the pandemic given their health-related reputability. These two profiles showed a high potential reach, indicating their ability in being the object of interest from other influential accounts to reach secondary audiences. Finally, a populist point of view was offered by Matteo Salvini, Francesca Totolo, and RadioSavana. While Salvini is a well-known politician, Francesca Totolo and RadioSavana had a limited number of followers with respect to the other sources. Nevertheless, these two accounts achieved high visibility during the pandemic, with RadioSavana, which is not associated with any known movement or personality, being the most-retweeted Italian account on Covid-related contents. Moreover, it is worth noting the polarization of these three accounts and, in particular, the far-right anti-immigration, anti-government and anti-Europe ideas supported by RadioSavana and Francesca Totolo. In addition, it should be mentioned that several contents published by RadioSavana have been classified as unreliable by the Italian fact-checking website "Bufale.net" (`www.bufale.net`), confirming previous studies about epidemics and the related spread of fake news and conspiracy theories (Ferrara, 2020; Kou et al., 2017; Feuer, 2014).

In conclusion, the presented results highlight, on the one hand, the dynamism and democratization of the Twittersphere (Ida et al., 2020), and on the other hand the well-known "million follower fallacy" dynamic (Cha et al., 2010; Borge-Holthoefer et al., 2013), as a huge number of followers does not necessarily correspond to a wider audience. These audiences are sometimes characterized by the presence of molecular opinion leaders (Katz, 1957) and/or active crowds (Zhang et al., 2016a), who allowed less popular accounts to compete with more reliable and popular sources. Regarding this, we can make two final considerations: first, contrary to general trends (Gallotti et al., 2020), less reliable sources were able to gain substantial users' attention in the Italian Twittersphere on Covid-related debates; second, it would be important to understand whether the most retweeted accounts took specific actions to increase their retweet count, for instance by buying automated retweets. The possible presence of anomalies in the spread of information deserves further investigation in future work. Indeed, even if the results do not show clear evidence of social bots (i.e., automated retweeters), future analyses are necessary to evaluate the authenticity of retweeters. This will be a key to fully understand the strange case of RadioSavana, which despite its limited follower base turned out to be the most-retweeted account on Covid-19 in Italy.

Acknowledgment

This work is supported by the European Union – Horizon 2020 Program under the scheme "INFRAIA-01-2018-2019 – Integrating Activities for Advanced Communities", Grant Agreement n.871042, "SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics" (http://www.sobigdata.eu).

Note

[1] www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline
[2] www.agcom.it/documents/10179/4514383/Allegato+14-4-2020/080f9ab1-df4d-481a-ad10-d4a89237bd01
[3] www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/
[4] https://data.d4science.net/GjJz.
[5] www.butac.it/radio-savana-e-lavvelenamento-del-pozzo/
[6] www.repubblica.it/economia/rapporti/osserva-italia/stili-di-vita/2020/03/28/news/la_vita_in_quarantena_ecco_cosa_twittano_gli_italiani_chiusi_in_casa_per_il_coronavirus-252536689/
[7] www.bufale.net/tag/radio-savana/
[8] Sarcastic way of addressing immigrants as useful resources for Italy, as they may contribute to the national social welfare institution INPS

References

Al-Aama AY (2015) The use of twitter to promote e–participation: connecting government and people. *International Journal of Web Based Communities* 11(1): 73–96.

Alvari H, Beigi G, Sarkar S, Ruston SW, Corman SR, Davulcu H and Shakarian P (2020) A feature-driven approach for identifying pathogenic social media accounts. *CoRR* abs/2001.04624. URL https://arxiv.org/abs/2001.04624.

Baines D, Elliott RJ et al. (2020) Defining misinformation, disinformation and malinformation: An urgent need for clarity during the covid-19 infodemic. Technical report.

Bastos MT and Mercea D (2019) The brexit botnet and user-generated hyperpartisan news. *Social Science Computer Review* 37(1): 38–54.

Beldad AD (2021) Sustaining social capital online amidst social distancing during the covid-19 pandemic: web-based communities, their mitigating effects, and associated issues. *International journal of web based communities* 17(1): 35–52.

Bobba G (2019) Social media populism: Features and 'likeability'of lega nord communication on facebook. *European Political Science* 18(1): 11–23.

Borge-Holthoefer J, Banos RA, González-Bailón S and Moreno Y (2013) Cascading behaviour in complex socio-technical networks. *Journal of Complex Networks* 1(1): 3–24.

Boulianne S and Theocharis Y (2020) Young people, digital media, and engagement: A meta-analysis of research. *Social Science Computer Review* 38(2): 111–127.

Bracciale R and Martella A (2017) Define the populist political communication style: the case of italian political leaders on twitter. *Information, Communication & Society* 20(9): 1310–1329.

Bracciale R, Martella A and Visentin C (2018) From super-participants to super-echoed, participation in the 2018 italian electoral twittersphere .

Bruns A and Burgess JE (2011) The use of twitter hashtags in the formation of ad hoc publics. In: *Proceedings of the 6th European consortium for political research (ECPR) general conference 2011*.

Cha M, Haddadi H, Benevenuto F, Gummadi PK et al. (2010) Measuring user influence in twitter: The million follower fallacy. *Icwsm* 10(10-17): 30.

Choi S (2015) The two-step flow of communication in twitter-based public forums. *Social Science Computer Review* 33(6): 696–711.

Cinelli M, Quattrociocchi W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, Zola P, Zollo F and Scala A (2020) The covid-19 social media infodemic. *arXiv preprint arXiv:2003.05004* .

Cronjé JC (2021) Social distancing close together: the rhizomatic role of whatsapp in communities-a proposed research framework. *International Journal of Web Based Communities* 17(1): 65–76.

Dubois E, Gruzd A and Jacobson J (2020) Journalists' use of social media to infer public opinion: The citizens' perspective. *Social Science Computer Review* 38(1): 57–74.

Ferrara E (2020) Covid-19 on twitter: Bots, conspiracies, and social media activism. *arXiv preprint arXiv:2004.09531* .

Feuer A (2014) The ebola conspiracy theories. *The New York Times* 18.

Firdaus SN, Ding C and Sadeghian A (2018) Retweet: A popular information diffusion mechanism–a survey paper. *Online Social Networks and Media* 6: 26–40.

Gallotti R, Valle F, Castaldo N, Sacco P and De Domenico M (2020) Assessing the risks of 'infodemics' in response to covid-19 epidemics. *Nature Human Behaviour* 4(12): 1285–1293.

Graham T and Wright S (2014) Discursive Equality and Everyday Talk Online: The Impact of "Superparticipants"*. *Journal of Computer-Mediated Communication* 19(3): 625–642.

Hagen L, Keller T, Neely S, DePaula N and Robert-Cooperman C (2018) Crisis communications in the age of social media: A network analysis of zika-related tweets. *Social Science Computer Review* 36(5): 523–541.

Hagen L, Neely S, Keller TE, Scharf R and Vasquez FE (2020) Rise of the machines? examining the influence of social bots on a political discussion network. *Social Science Computer Review* : 0894439320908190.

Hansen LK, Arvidsson A, Nielsen FÅ, Colleoni E and Etter M (2011) Good friends, bad news-affect and virality in twitter. In: *Future information technology*. Springer, pp. 34–43.

Hilbert M, Vásquez J, Halpern D, Valenzuela S and Arriagada E (2017) One step, two step, network step? complementary perspectives on communication flows in twittered citizen protests. *Social Science Computer Review* 35(4): 444–461.

Ida R, Saud M and Mashud M (2020) Persistence of social media on political activism and engagement among indonesian and pakistani youths. *International Journal of Web Based Communities* 16(4): 378–395.

Katz E (1957) The two-step flow of communication: An up-to-date report on an hypothesis. *Public opinion quarterly* 21(1): 61–78.

Kou Y, Gui X, Chen Y and Pine K (2017) Conspiracy talk on social media: collective sensemaking during a public health crisis. *Proceedings of the ACM on Human-Computer Interaction* 1(CSCW): 1–21.

Larsson AO and Moe H (2012) Studying political microblogging: Twitter users in the 2010 swedish election campaign. *New media & society* 14(5): 729–747.

Mancosu M and Vegetti F (2020) "is it the message or the messenger?": Conspiracy endorsement and media sources. *Social Science Computer Review* DOI:10.1177/0894439320965107.

Medford RJ, Saleh SN, Sumarsono A, Perl TM and Lehmann CU (2020) An" infodemic": Leveraging high-volume twitter data to understand public sentiment for the covid-19 outbreak. *medRxiv* .

Myers SA, Sharma A, Gupta P and Lin J (2014) Information network or social network? the structure of the twitter follow graph. In: *Proceedings of the 23rd International Conference on World Wide Web*. pp. 493–498.

Proferes N (2017) Information flow solipsism in an exploratory study of beliefs about twitter. *Social Media+ Society* 3(1): 2056305117698493.

Pulido CM, Villarejo-Carballido B, Redondo-Sama G and Gómez A (2020) Covid-19 infodemic: More retweets for science-based information on coronavirus than for false information. *International Sociology* : 0268580920914755.

Radu R (2020) Fighting the 'infodemic': Legal responses to covid-19 disinformation. *Social Media+ Society* 6(3): 2056305120948190.

Shmargad Y and Sanchez L (2020) Social media influence and electoral competition. *Social Science Computer Review* : 0894439320906803.

Singh L, Bansal S, Bode L, Budak C, Chi G, Kawintiranon K, Padden C, Vanarsdall R, Vraga E and Wang Y (2020) A first look at covid-19 information and misinformation sharing on twitter. *arXiv preprint arXiv:2003.13907* .

Small TA (2011) What the hashtag? a content analysis of canadian politics on twitter. *Information, communication & society* 14(6): 872–895.

Tangcharoensathien V, Calleja N, Nguyen T, Purnat T, D'Agostino M, Garcia-Saiso S, Landry M, Rashidian A, Hamilton C, AbdAllah A et al. (2020) Framework for managing the covid-19 infodemic: methods and results of an online, crowdsourced who technical consultation. *Journal of medical Internet research* 22(6): e19659.

Taxidou I and Fischer PM (2014) Online analysis of information diffusion in twitter. In: *Proceedings of the 23rd International Conference on World Wide Web*. pp. 1313–1318.

Vaccari C and Valeriani A (2015) Follow the leader! direct and indirect flows of political communication during the 2013 italian general election campaign. *New media & society* 17(7): 1025–1042.

Vaezi A and Javanmard SH (2020) Infodemic and risk communication in the era of cov-19. *Advanced Biomedical Research* 9.

Yuan X, Schuchard RJ and Crooks AT (2019) Examining emergent communities and social bots within the polarized online vaccination debate in twitter. *Social Media+ Society* 5(3): 2056305119865465.

Zarocostas J (2020) How to fight an infodemic. *The Lancet* 395(10225): 676.

Zhang L, Zhao J and Xu K (2016a) Who creates trends in online social media: The crowd or opinion leaders? *Journal of Computer-Mediated Communication* 21(1): 1–16.

Zhang Y, Ruan X, Wang H, Wang H and He S (2016b) Twitter trends manipulation: a first look inside the security of twitter trending. *IEEE Transactions on Information Forensics and Security* 12(1): 144–156.

Zhou Y, Zhang B, Sun X, Zheng Q and Liu T (2017) Analyzing and modeling dynamics of information diffusion in microblogging social network. *Journal of Network and Computer Applications* 86: 92–102.

Zola P, Cola G, Mazza M and Tesconi M (2020a) Interaction strength analysis to model retweet cascade graphs. *Applied Sciences* 10(23): 8394.

Zola P, Cortez P and Brentari E (2020b) Twitter alloy steel disambiguation and user relevance via one-class and two-class news titles classifiers. *Neural Computing and Applications* .