THE DENIERS ON TWITTER. THE NO MASK GROUPS AND THEIR COMMUNICATION[1]

Rosario D'Agata, Simona Gozzo

1. Introduction

The Sars-CoV-2 pandemic caused millions of deaths all over the world but, despite this, a lot of people still claim that it is only the result of a plan of "strong powers" to control people, or, worse, that the virus does not even exist but it is a media invention to scare and control people. A current problem is that, nowadays, deniers find support for their ideas through social networks, an echo chamber for their messages. This implies a huge problem of public health. In fact, disinformation can trigger rule braking: suffice to think about social distancing, wearing masks, distrust towards institutions and vaccines.

Paradoxically, considering that the specific purpose of the web is spreading news and information to a wide audience, it is showing messages of disinformation and/or incorrect information, producing as effect the moral panic (Thompson, 2006). This was a specific feature of social-network communication related to the issue of the pandemic crisis. An important aspect, currently at the heart of the academic debate, is the individual tendency to approve disinformation, avoiding factual control. Some studies point out that improving scientific knowledge could help reduce susceptibility to disinformation. These types of solicitations begin, today, to be incorporated into the design of social media platforms or messages from health organizations to make the public more attentive to the content they are reading and sharing. Other strategies identified to limit this phenomenon are:

- showing the techniques used to spread misinformation (Schmid and Betsch, 2019);
- engaging social media companies by asking to verify the accounts of credible experts and organizations and mark them as "authenticated" (Trivedi *et al.*, 2020);

[1] This work is the result of the joint efforts of both authors. However, Simona Gozzo wrote paragraphs 1, 2 and 5 while Rosario D'Agata wrote paragraphs 3 and 4. Researcb program "Programma di Ricerca di Ateneo Unict 2020-22 – linea 2".

- promoting wider environmental changes (Lewandowsky *et al.*, 2012) and changes in social norms (Paynter *et al.,* 2019)

Some studies show that, unfortunately, the attempts to change individual acts with rational arguments rarely produces large-scale behavioural changes, as we need. It emerges, in fact, that the cognitive mechanisms linked to mistrust of scientifically validated communications derive from non-rational mechanisms and, in particular:

- crisis, fear and uncertainty increase the likelihood of conspiratorial thoughts, and so sites and conspiratorial articles generate greater user engagement than more reliable sources (Stein *et al.*, 2021);

- the belief in conspiracy theories is not attributable to the desire to acquire information from multiple sources or more reliable ones but to states of anger and frustration, rather than detecting a close association with the justification of the use of violence (Jolley *et al.*, 2020);

- although, overall, social media hosts a larger volume of accurate information than fake information, misinformation seems to be more popular and spreads faster, farther, and deeper (Miller *et al.*, 2020);

- no-mask groups base their thesis precisely on a lack of confidence in information from official sources and not on rational arguments. As a result, people who reject vaccines and refuse the use of masks are more confident to obtain information from social networks than from verified health care professionals or health care websites (Ortiz-Sànchez *et al.*, 2020). So, the question that drove this research is: what information does the no-mask movement use in social networks? Are there any differences with respect to specific languages and contingencies? Given that Twitter seems to be the most used social network by the movement, we have chosen this social network to monitor communication, analysing then the available data through the triangulation of different techniques. Certainly, it has emerged that, although no-vax users are fewer than those in favour of vaccines, their on-line communication is strongly visible because they are very active.

2. Method

In this paper we used web tools to extract tweets containing the lemma ‘No-mask’. The choice to monitor the use of this lemma depends on several considerations. First of all, the protest movement was initially named as ‘no-mask’ and only later the term ‘no-vax’ emerged. Then, the lemma no-vax is not always used as hashtag but many of its variants are distinguished (no-vaccine, notest, novaccineforme, stopvaccine, vaccinkill etc.). Besides, on a theoretical plane, ‘nomask’ is useful for pointing out, in general, the position of the deniers. The allencompassing extension of “no-mask” as hashtag allows, in fact, to identify a huger amount of comments. Finally, a strong association is found between the two lemmas, overlapping in Italian co-occurrences.

Specifically, a samples of tweets were extracted every two weeks from November, 30th 2020 to the February, 6th 2021, by means of NodeXL Pro Twitter data importers (Smith *et al.*, 2009), an excel plug-in which binds the extraction of tweets to a time limit of approximately 2 weeks. So, we monitored the communication on twitter during these four months, every week. However, we couldn’t extract all comments because queries cannot return more than 18,000 tweets per extraction (twitter controls its API and throttles it based on unknowable parameters).

After collecting edges and comments in 5 waves, we observed different features of the nets representing links among users. Then we extracted the 10 most popular tweets for each wave. This was only the first step of our analysis. Technically, we propose a three steps procedure applied to obtain complementary information about data extracted from social networks. This approach captures and integrates information and it includes:

1) An in-depth qualitative analysis about the top 10 tweets extracted for each wave. The qualitative study carried out at this stage allowed us to exclude tweets in Portuguese, as they often refer to Brazilian issues and problems. Furthermore, during this step we noticed that the major part of tweets extracted are in English, Italian or French languages. This information led us to identify these three languages to build three corpus used for content analysis (step three).

2) A quantitative study about all comments, based on network analysis tools (Borgatti and Halgin, 2011). This step allowed us to select – for each week – the whole structure of web communications and the main groups as sub-networks obtained by extracting clusters mutually connected, with higher internal homogeneity and external heterogeneity in terms of links. Differences on web communities are also detected, both as quantitative structures and linkages.

3) An Automatic analysis of the content on tweets extracted during the five waves, selected distinguishing the main languages (see first point). This last step shows differences on topics and discussions both looking at languages (Italian, English and French) and time periods.

2.1 The Qualitative features

The first step’s aim is a reasoned in-depth study of the content, operated only on the main comments (10-top tweets automatically extracted for each monitored week). This is a “qualitative” phase, useful for identifying problems in attributing the language to a particular area and for the selection of categories, messages and users. The tweets are selected by the greatest weight in the page-rank index.

The communication on Twitter was monitored extracting tweets containing the word "no-mask" in the period November 30th - February 5th. Five extractions were carried out, repeating the operation every two weeks. This period goes through at least three specific moments of tension: a first distribution of the vaccine, the growing spread of the infection, the need to close shops, restaurants and attractions in many parts of the world.

Finally, it was built a matrix containing 50 main tweets (10 for each extraction) useful to identify information that requires an evaluation of the communicative content. The analysts have identified (when possible) information on the geographical origin of the tweets, senders (politicians, newspapers, private citizens, etc.), sentiment of tweets (i.e. the message is in favour, against or neutral with respect to the no-mask movement), main themes, tweets' addressees (political institutions, people or parties). Thanks to this qualitative analysis, all tweets that came from countries outside the European Union (excluding United Kingdom) were excluded.

2.2 The Quantitative features

After a preliminary qualitative step, in a second phase we focused on the analysis of the network obtained by selecting - for each extraction - the entire structure of the communication and the main sub-graphs with high internal homogeneity and external heterogeneity in terms of connections. Finally, the third step involved the automatic analysis of the content of all comments extracted during the reference period, in the main languages preliminarily identified.

While the first two steps required an analysis on the main tweets extracted in all languages (first step) or about the network structure of all the extracted tweets, the last phase - as analysis of content - needed the selection of specific languages (as corpus). So, we decided to build three corpus selecting all comments in the main languages used by users.

3. Findings

By means of traditional tools of network analysis applied in social media studies (Hansen *et al.,* 2010; Hansen *et al.*, 2012; Smith *et al.*, 2009) we examined in depth the links among twitter users. Our aim was to identifying the aggregative power of the comments, the specificity of the issue and the existence of clusters.

3.1 The network structures

Table 1 shows the values of mains network measures. As we can notice, the higher number of comments is observed in the second half of December. It needs to be kept in mind that on December 27th the whole Europe proclaimed the 'Vaccine day' that represented an important moment both for vaccine's supporters but, at the same time, for deniers and no vax groups too.

Table 1 – *Network measures employed in the analysis.*

Network measures	30th-Nov	14th-Dec	30th-Dec	5th-Jan	05th-Feb
Vertices (number of tweets)	1,352	1,749	2,923	901	606
Unique Edges	1,336	1,674	2,938	1,466	545
Edges With Duplicates	190	161	117	189	630
Total Edges	1,526	1,835	3,055	1,655	1,175
Self-Loops	355	291	219	136	298
Connected Components	444	381	309	183	219
N. max of vertices in Connected Components	45	665	1,826	212	51
Maximum Geodesic Distance (Diameter)	8	10	6	13	6
Average Geodesic Distance	2.28	2.05	2.00	3.64	2.05
Modularity	0.801	0.728	0.583	0.793	0.469
Degree Centrality (average)	1.04	0.99	1.02	1.72	1.11
Closeness Centrality (average)	0.26	0.18	0.09	0.17	0.33
Betweenness Centrality (average)	14.05	278.28	1,155.3	183	11.81

The number of vertices goes along with the number of edges. It's noteworthy, however, the difference between the unique edges and the edges with duplicates. These latter, in fact, while in the third period show the lowest value, in the fifth one they reach the peak. In other words, we note a decrease of broad communication (probably aimed to acquire information) over time and the rise, in the last observed period, of a more 'closed' communication as the number of vertices (606) suggests. This assumption appears supported by the high number (298) of isolated cases (self-loops) on February and the maximum number of vertices in connected components observed on 30th of December. A connected component is a subgraph in which any two nodes are connected to each other by a path. In our study we can identify connected components as 'discussion groups'. The highest number of connected components is detected on the first period (444) but these components appear to be very small (max 45 vertices). The number of the nodes increases in the first half of December (max 665 vertices) and reaches the maximum (1.826) in the second half of December when the communication about Sars-CoV-2 on twitter became increasingly pervasive.

Considering the third period, if – on the one hand, the communication appears strong and pervasive, on the other - it is also true that it does not seem to be embracing many topics. In other words, we note the rise of components with a large number of nodes, close to each other (Average Geodesic Distance =2) but with a low Modularity (0.583). Modularity, whose range varies from -1 to 1, quantifies the quality of the division of the network into modules or communities[2]. So, the more the value approaches to 1, the more different groups are. In the first period (0.801) and in the fourth one (0.793) we observed the highest values. This means that, in these periods, we registered different and opposite subject groups (i.e. vaccine supporter vs. no-vax groups, etc…).

In order to analyse the links among twitter users we used network centrality measures (Hansen *et al.*, 2010; Hansen *et al.*, 2012; Smith *et al.*, 2009); specifically, *degree*, *closeness* and *betweenness* (Junlong and Yu, 2017). The standardised centrality degree (1) of node d - C_d'- is based on the number of links held by node d - $d(n_i)$ - on the total of all possible links $(g-1)$.

$$C_d' = \frac{d(n_i)}{g-1} \tag{1}$$

In our study, the *degree* measures the number of nodes linked with the each other through a reaction, i.e., visualizations, retweets, likes, answers, etc. The highest average value of Degree (1.72) is observed in the fourth period (Tab.1). In the first two weeks of January, therefore, the highest proportion (out of the total) of users communicating without intermediaries is found. This fact appears in line with the information about modularity previously descripted. In other words, the increase of subject groups (especially in small groups) increases the likelihood of direct contacts within it.

The *closeness* of *i*-th node - $C_c(n_i)$- can be defined as the reciprocal of the sum of the distances between the *i*-th node and other *g* nodes (2).

$$C_c(n_i) = \left[\sum_{j=1}^{g} d(n_i, n_j)\right]^{-1} \tag{2}$$

Closeness, whose range varies between 0 and 1, gives information about the relevance of peripheric nodes. The highest value of closeness is observed in the last wave. In this period, we noted a lowest number of vertices, the highest number of edges with duplicates and the lower value of modularity. The *closeness*

[2] Modularity in NodeXl is computed through the number of edges that move from a group to connect themselves to vertices in a different group. If modularity is high, the clusters or groups created may be of low quality. If modularity is low, the groups are well defined.

highlights the presence of small groups, with similar themes of discussion and, therefore, the average *closeness* among the nodes tends to be higher.

The *betweenness* centrality of i-th node - $C_b(n_i)$ - is the sum of all partial betweenness computed for each couple of nodes. So, in the (2), $g_{jk}(n_i)$ is the number of geodesics connecting two nodes that include i-th node.

$$C_b(n_i) = \sum_{j<k} g_{jk}(n_i)/g_{jk} \quad (3)$$

From a substantive point of view, the *betweenness* reports a high presence of users acting as intermediaries between other users or groups. The higher is the *betweenness*, the less redundant communication is. The highest value of *betweenness* (Table 1) is observed in the third wave (1,155.3). In the second half of December, so, communication is wide, homogeneous, as the low value of modularity suggests and seems, above all, as an attempting to search for information in a period of worrying uncertainty.

3.2 The Content analysis

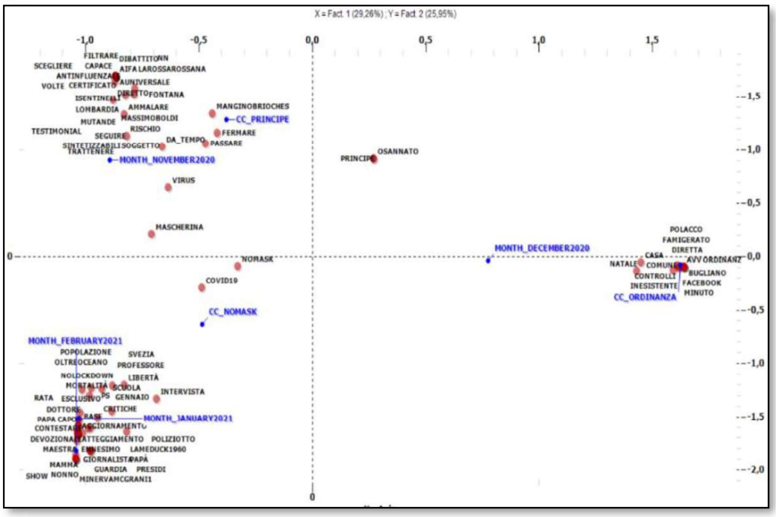
So far, we have considered the structural features of communication on twitter in the periods under analysis. In the next pages, we analysed the content of the tweets in order to identify the deniers and their messages. For this sake, we detected all the tweets extracted in the main languages (Table 2). Nevertheless, in this paper, we chose to only deal with Italian tweets.

Table 2 – *Descriptive information about the comments collected.*

	Italian	English	French
Types	4762	7369	1795
Tokens	47407	96470	29373
Hapax	1914	2849	662
Types/Token	0,1	0,07	0,06
Root TTR	21,876	23,725	10,473
Corrected TTR	15,465	16,776	7,406
Log TTR	0,787	0,776	0,728
Hapax/Types	0,402	0,387	0,369

The analysis was applied using T-lab. A segmentation of the *corpus* in elementary contests is carried on. In our case the elementary contest is a single tweet. Then, thematic analysis allowed to identify the main themes and to clusters them. Doing so, it was possible to identify three thematic clusters corresponding to three communication features: fake news and comments to these (cluster NOMASK, 47%), political debates and irony (cluster ORDINANZA, 27%), threats and insults (cluster PRINCIPE[3], 26%). In the next step, a multiple correspondence analysis was carried out selecting as variables both thematic clusters and periods of extraction, identifying two dimensions on which the lemmas have been projected. Figure 1 shows the thematic clusters obtained considering also the periods[4] of tweets' scraping and the main lemmas.

Figure 1. - *Correspondence analysis. Plot of clusters***.**



The first cluster is characterized by the lemmas related to juridical aspects, polemic against deniers and statistical information, summarised with the lemma 'no-mask'. This is the largest cluster and is observed in January and February. Many tweets within this cluster are ironic comments against deniers, but it is also possible to identify polemic tweets against restrictions. The second cluster (ORDINANZA) contains mainly ironic issues and it is observed in December when, as we saw before by means of Network Analysis, the communication is pervasive and widespread. Comments within the last cluster (PRINCIPE) are

[3] It needs to be specified that the Italian word "PRINCIPE" - "PRINCIPE DEL FORO" (prince of the forum) – is often referred to a great lawyer. In this case, however, the great lawyer Mr. Polacco did not revealed to be such because it was defending a cause which did not exist in the reality. Particularly, the cause was linked to a fake news announcing strict control measures in respect to COVID19 in the municipality of Bugliano, not existing in the reality but emerging in social network fake news.

[4] We aggregated all the periods in 4 mouths, namely, November (TEXT 001), December (TEXT002), January (TEXT 003) and February (TEXT 004).

widespread in November, when the communication on twitter about COVID 19 appears sectorial and mainly subject of discussion groups (see Modularity in Tab. 1). These comments are often filled with hate, implying disputes between individuals or groups and often use vulgar and offensive language.

4. Discussion

Our results show that the weight of deniers' communication increased sharply during December 2020 while it becomes sparser and specific in January and February. Evaluating the different structure of the communicative flows, we notice that the most representative groups in November were rather unstructured, with self-referential messages, while the first part of December shows a network with greater aggregating power and, during Christmas, the highest mobilization emerges, with very limited distance between vertexes (which implies a redundancy in information). Heterogeneity characterizes the communication flow in January, with an increasing number of groups, despite the reduction in the number of vertexes. If these considerations concern the structure of the network, further observations concern the meaning of the comments that convey the debate in the different languages. Although the work focuses on communication in Italian, it is worth noting that the topics are similar even in the tweets in English and French and, in particular, a connotation linked to the critique of the rules always emerges (the English cluster 'norms', Italian 'principe', French 'souffle') and one related to the generic theme (no mask, coronavirus, liberté). This last cluster can be split into two topics: strongly critical comments due to the economic situation and others that refer to anti-vaccination rhetoric (Miller *et al.*, 2020; Stein *et al.*, 2021).

The no-vax issues are surprisingly similar, although the most intense and numerous protests emerge at different times, depending on the languages. The most common topics are distrust of doctors or governments, repulsion at the idea of introducing unknown substances into the body and the suspicion that the real motivations behind vaccines are to make people sick or to control the population (Jolley *et al.*, 2020; Stein *et al.*, 2021). However, we can detect differences too. The Italian comments are often ironic. One of the two ironic clusters is characterized by vulgar and offensive language. The largest cluster is 'No-mask' (which also use ironic but even critical or aggressive-offensive language) and prevails in January.

English comments don't show irony at all. The main topic is the contestation of the norms, concerning the two minor clusters, which characterize the communication of January (cluster 'exemption') and the first weeks of February (cluster 'woman-shop'). The largest cluster is (such as in Italian) 'no-mask' but this term prevails in December as English comments and in January/February for the Italian ones.

Finally, French tweets often refer to the emotional communication, especially in December (cluster 'souffle') but here the most consistent cluster ('no-mask') characterizes the communication in November while February and it is mainly associated with protests (also with offensive language and vulgarity) and to the defense of freedom as value ('liberté').

5. Conclusions

The procedure we proposed is particularly useful when you need to extract main information and you are analyzing the on line communication. The extraction of the Top-10-tweets allows an in-depth analysis on a low number of relevant comments. In so doing, it is possible to quickly identify errors, problems and new topics. Besides, the Network Analysis tools allow to identify the structure of communities and how it changes over time, while the content analysis permits to evaluate (with attention) the content of comments and the importance of each cluster (also in terms of debate and construction of parallel communities). Finally, it is possible to compare communication through main topics, languages and time periods.

References

BORGATTI S.P., HALGIN D.S. 2011. On Network Theory. *Organization Science*, 22, 5, pp.1168-1181.

HANSEN D.L., SHNEIDERMAN B. AND SMITH M.A. 2010. Analyzing Social Media Networks With NodeXL: Insights From a Connected World, Elsevier Science.

HANSEN, D. L., ROTMAN D., BONSIGNORE E., MILIC-FRAYLING N., RODRIGUES E. M., SMITH M., SHNEIDERMAN B. 2012. Do You Know the Way to SNA?: A Process Model for Analyzing and Visualizing Social Media Network Data, *International Conference on Social Informatics*, pp. 304-313.

JUNLONG Z., YU L. 2017. Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network. *Advances in Intelligent Systems Research*, volume 132, pp. 300-303.

LEWANDOWSKY S., ECKER U. K., SEIFERT C. M., SCHWARZ N., COOK, J. 2012. Misinformation and its correction: Continued influence and successful debiasing, *Psychological Science in the Public Interest*, 13, 3, pp. 106-131.

MILLER J. 2020. Do COVID-19 Conspiracy Theory Beliefs Form a Monological Belief System?, *Canadian Journal of Political Science*, 53, 2, pp. 319-326.

ORTIZ-SÁNCHEZ E., VELANDO-SORIANO A., PRADAS-HERNÁNDEZ L., VARGAS-ROMÁN K., GÓMEZ-URQUIZA J.L., CAÑADAS-DE LA FUENTE G.A., ALBENDÍN-GARCÍA L. 2020. Analysis of the Anti-Vaccine Movement in Social Networks: A Systematic Review, *Int J Environ Res Public Health*, Jul 27, 17, 15, pp.53-94.

PAYNTER J., LUSKIN-SAXBY S., KEEN D., FORDYCE K., FROST G., IMMS C., MILLER S., TREMBATH D., TUCKER M., ECKER U. 2019. Evaluation of a template for countering misinformation Real-world autism treatment myth debunking. *PLoS One*, 14, 1, pp. 1-13.

SCHMID P., BETSCH C. 2019. Effective strategies for rebutting science denialism in public discussions, *Nature Human Behaviour*, 3, pp. 931–939.

SMITH M., SHNEIDERMAN B., MILIC-FRAYLING N., RODRIGUES, E. M., BARASH V., DUNNE C., CAPONE T., PERER A. AND GLEAVE E. 2009. Analyzing (Social Media) Networks with NodeXL. in *C&T '09: Proc. Fourth International Conference on Communities and Technologies*, pp. 255-264.

STEIN R.A., OMETA O., SHETTY S. P., KATZ A., POPITIU M.I., BORTHERTON R. 2021. Conspiracy theories in the era of COVID-19: A tale of two pandemics, *The International Journal of Clinical Practice*. Vol. 75, 2.

THOMPSON K. 2006. The History and Meaning of the Concept in Critcher Chas, Critical Readings: Moral Panics and the Media, *Open University Press*, pp. 60-66.

TRIVEDI N., KRAKOW M., HYATT HAWKINS K., PETERSON E.B., CHOU W.-Y.S., 2020. Well, the message is from the institute of something: Exploring source trust of cancer-related messages on simulated Facebook posts, *Frontiers in Communication*, 5, p. 12.

JOLLEY D, PATERSON J.L. 2020. Pylons ablaze: Examining the role of 5G COVID-19 conspiracy beliefs and support for violence, *Br. J. Soc. Psychol.,* 59, 3, pp. 628-640.

SUMMARY

The deniers on Twitter.
The No Vax groups and their communication title

The Sars-CoV-2 pandemic caused millions of deaths all over the world but, despite this, a lot of people still claim that the virus does not exist. This implies a huge problem public health is at stake. In this paper we aim to analyse the way deniers communicate on twitter. Using web scraping tools, we extracted the tweets containing the lemma “Nomask”. Tweet’s collection was run every two weeks from November, 30th 2020 to the February, 6th 2021. After collecting tweets in 5 waves, through Social Network Analysis measures, we observed the features of networks. Then, taking into account the weight of the "retweet count" index, we selected the first 10 tweets for each extraction. Once the tweets were selected, they have been analysed by means of both cluster and textual analysis. Analysis of the most representative groups of the various networks showed that communication in November was rather unstructured. It did not generate a network but rather self-referential messages. The second wave shows a network with greater aggregating power. In the groups in the third wave, we may notice a mobilization with very limited distance between vertexes which implies a redundancy of information. The fourth wave, instead, highlights an increase in the number of groups, despite the reduction in the number of vertexes. Finally, the fifth wave has four groups with a limited number of nodes.

Rosario D’AGATA, Dipartimento di Scienze Politiche e Sociali. Università di Catania, rodagata@unict.it

Simona GOZZO, Dipartimento di Scienze Politiche e Sociali. Università di Catania, simonagozzo@.yahoo.it