The diffusion of mainstream and disinformation news on Twitter: the case of Italy and France

Francesco Pierri
francesco.pierri@polimi.it
Dipartimento di Elettronica, Informazione e Bioingegneria
Politecnico di Milano

ABSTRACT

In this work we provide preliminary results from an ongoing investigation on the Twitter diffusion of news pertaining to two classes of sources, namely websites which notably produce *disinformation*, i.e. misleading and harmful information, opposed to more traditional and *mainstream* websites which instead publish credible information. We used the Twitter Streaming API to collect a large-scale dataset of thousands of tweets containing links to news articles in two different countries, Italy and France. We show that mainstream news outlets generate a much larger engagement in both settings, with a larger discrepancy between the two news domains in France. We also show that only a handful of Italian outlets actively engage with Twitter users, whereas in France there is a larger number of outlets sharing misleading information which exhibit a non-negligible volume of shares. We observe a strong tendency towards sharing mainstream news in those users who also share non-credible information in both countries. Analyzing the diffusion networks of distinct news domains and countries, we observed that disinformation networks are more clustered and connected, but much smaller than the mainstream ones (with the largest discrepancy in the French scenario).

CCS CONCEPTS

• **Information systems → Social networking sites**.

KEYWORDS

disinformation, Twitter, network science

ACM Reference Format:
Francesco Pierri. 2020. The diffusion of mainstream and disinformation news on Twitter: the case of Italy and France. In *Companion Proceedings of the Web Conference 2020 (WWW '20 Companion), April 20–24, 2020, Taipei, Taiwan.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3366424.3385776

1 INTRODUCTION

In recent times, a growing concern has raised over the spread and the impact of misleading and harmful information spreading on online social media in Europe.

Still, there is a huge debate in the research community for a single definition of malicious information [12]. In this work we use the term *disinformation* as a short-hand to indicate a broad spectrum of misleading and potentially harmful information, which ranges from false news intended to harm to hyper-partisan stories, click-bait, incorrect news reporting and unverified rumours [17].

Online misbehavior on Twitter social platform in Europe–such as the presence of deceptive information and/or malicious agents (e.g. trolls and bots)–has been reported in several political circumstances including 2016 Brexit referendum [4] and 2017 Catalan referendum [24], but also 2017 French Presidential elections [8, 11], 2018 Italian General elections [6, 9, 23] and 2019 European Parliament elections [16].

In this work we aim to compare the diffusion of news articles pertaining to two classes of outlets, namely websites which notably produce *disinformation*, i.e. misleading and harmful information, opposed to a set of more traditional and *mainstream* outlets which instead publish credible and trustworthy information. We perform experiments with Twitter data from two different countries: Italy and France.

According to 2019 digital news report by Reuters [15], trust in news is particularly low in Italy: 40% of people trust news most of the time whereas only 23% trust news present on social media most of the time; this is reported as a long-standing trend due to the political polarization of mainstream news organizations (and Italian journalism in general). Exposure to online disinformation in the context of 2018 Italian General elections, with a peak of interactions in correspondence of the day when the vote took place, has been previously reported by [9] and the Italian Authority for Communications Guarantees (AGCOM) [2]. Another contribution provides an estimate of the impact of false news on the electoral outcomes with a focus on the populist vote in Trentino-Alto-Adige region [7]. Finally, we refer the reader to our previous work [16], where we investigated the presence and the impact of online disinformation on Twitter in the run-up to 2019 European Parliament elections. We showed that malicious information was mainly shared by a few active outlets, which focused their agenda on controversial topics such immigration and national safety, and that there were interactions between active sharers of misleading news and the Italian far-right community (including Matteo Salvini and his "Lega" party). We draw this contribution from such analyses, extending our scope to mainstream news outlets and another country (France). This shows a worse scenario when it comes to consume online news [15]: overall trust in news in France is the lowest in Europe (24%); Yellow Vest protests have increased the consumption of online news but affected their reputability. In fact, it gets worse on social media which are generally blamed (with a 14% trust score) for spreading conspiracy theories, but also for reporting biases and algorithmic filtering.

WWW '20 Companion, April 20–24, 2020, Taipei, Taiwan

ACM ISBN 978-1-4503-7024-0/20/04.
https://doi.org/10.1145/3366424.3385776

We recognize that Twitter usage exhibits a decreasing trend in both countries, with only 8% of online users using it for consuming news in Italy and 9% in France [15]. It is clear that other platforms, such as Facebook and WhatsApp, are overtaking it, but at the same time they offer little opportunities for easily analyze the spread of credible vs non-credible information. Therefore, we follow a huge corpus on literature on the subject [5, 10, 16, 18, 21, 22] to collect relevant data on Twitter and shed light on the different consumption of misleading and more credible information.

In this work we present preliminary results of our ongoing investigation on the Twitter diffusion of URLs containing links to news articles pertaining to two classes of outlets: *disinformation* and *mainstream* websites. We collected a large-scale dataset of tweets–running Twitter Streaming API for over 3 weeks–relative to both Italian and French sources. In the following we first provide a descriptive statistics of data in terms of sources, URLs, users and tweets; then we analyze diffusion networks, built as in [16, 22], for each distinct news domain (and country), and we show a comparative analysis of resulting networks according to several indicators coming from the network science toolbox [3, 14]. Finally, we draw conclusions and future directions of our research.

2 DATA COLLECTION AND DESCRIPTION

Following a consolidated strategy [16, 18, 20–22], we employed Twitter Streaming API to collect tweets containing URLs of news articles pertaining to two classes of news outlets: **(1)** *disinformation* outlets, which notably produce and share misleading and harmful information, from false news to propaganda and conspiracy theories; **(2)** *mainstream* outlets, which convey reliable and credible information.

In particular, we assume that each article published on the former class is effectively a disinformation article, and likewise for mainstream news, although this might not be always true as deceptive websites do not share solely false stories, and reported case of misleading information on traditional outlets are not rare[12].

It is reported in [13] that Twitter API has an inherent limitation: whenever the matched number of tweets in a query exceeds the 1% of the global daily volume of tweets, the results are randomly filtered out. We never exceeded such quota in our collection process (which is approximately $5 \cdot 10^8$ tweets[1]), as we collected always less than $2 \cdot 10^6$ tweets each day, and thus we did not incur in missing data issues.

In the following we provide some statistics on data for both countries. Both datasets are available[2] at: https://doi.org/10.7910/DVN/CBHUMA.

2.1 Italy

For what concerns the Italian scenario, we referred to [16] to obtain a list of 60+ Italian disinformation websites, and to [25] to obtain a list of 25 Italian mainstream outlets; the former was compiled by resorting to blacklists on Italian fact-checking websites whereas the latter is provided by the association for the verification of newspaper circulation in Italy (Accertamenti Diffusione Stampa).

[1] https://www.internetlivestats.com/twitter-statistics/
[2] We can only provide id for tweets in agreement with Twitter terms of use.

Domain	No. URLs	No. Tweets	No. Users
Disinformation	6,732	99,569	11,293
Mainstream	103,214	592,463	85,146

Table 1: Breakdown of the Italian dataset in terms of URLs, tweets and users.

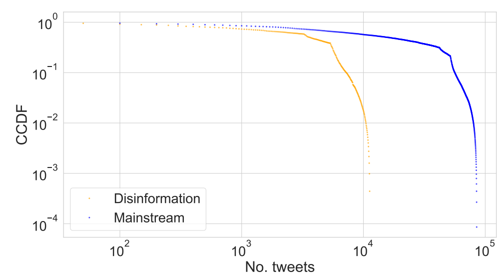

Figure 1: Complementary cumulative distribution of the number of shared tweets for mainstream (blue) and disinformation (orange) news articles in the Italian scenario.

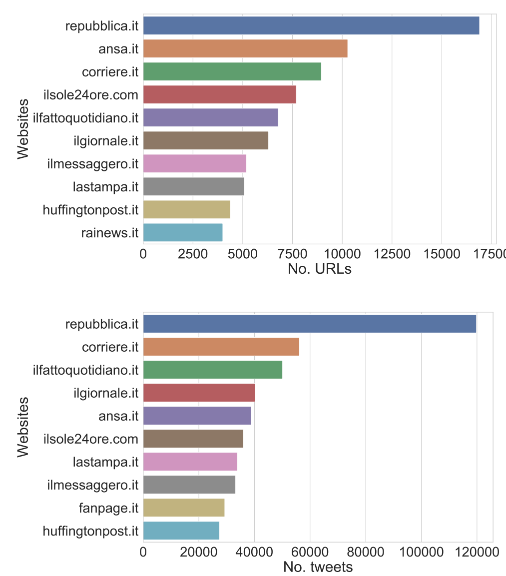

Figure 2: Distribution of the number of associated URLs (top) and tweets (bottom) shared for Top-10 Italian mainstream sources.

We collected data continuously with Twitter Streaming API in the period 16/11/2019-16/12/2019. A breakdown of the dataset in terms of articles, tweets and users (who authored tweets) is available in Table 1. We can notice a 6:1 ratio in the volume of shared tweets between mainstream and disinformation, and an 8:1 ratio in the number of users sharing mainstream news compared to disinformation.

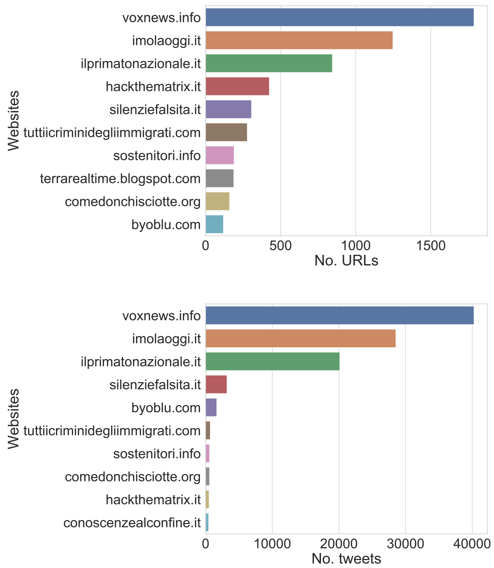

Figure 3: Distribution of the number of associated URLs (top) and tweets (bottom) shared for Top-10 Italian disinformation sources.

We show in Figure 2 (on a logarithmic scale) the complementary cumulative distribution of the number of tweets shared per number of users for both disinformation and mainstream news articles. We can observe that the two domains exhibit different distributions–which are both heavy-tailed–, and that users sharing mainstream news tend to be more active on average.

In Figure 3 we show the Top-10 mainstream sources w.r.t to the number of associated tweets and unique URLs; in Figure 4 we show same statistics for Top-10 disinformation outlets. We can observe that disinformation outlets have a very small engagement on Twitter, with the only exception of Top-3 outlets–namely "voxnews.info" (a repository of false stories), "imolaoggi.it" (a generic untrusthworthy news outlet) and "ilprimatonazionale.it" (the official newspaper of former far-right party "CasaPound")–which reach volumes which are comparable to a few mainstream outlets in the Top-10 ranking. Among the latter, "repubblica.it" clearly stands out on all other news outlets both in terms of URLs and tweets shared.

2.2 France

For what concerns the French context, we referred to [19] to obtain a list of both French disinformation and mainstream outlets. In this inquiry of 2018, journalists of LeMonde report their analysis on the engagement of different news sources on Facebook from January 2015 to September 2018; they used the Decodéx [1] to differentiate sources according to (a) satirical websites, (b) websites which notably share fake news, (c) websites which are untrustworthy and (d) websites which produce credible and reliable information. We excluded the set of (a) sources, and we considered as disinformation websites those (b) and (c) sources which generated at least 1 M engagements according to the results reported in the investigation; we then considered Top-50 (d) websites w.r.t to their engagement as our mainstream outlets.

Domain	No. URLs	No. Tweets	No. Users
Disinformation	28,499	247,173	38,938
Mainstream	225,304	2,125,781	425,960

Table 2: Breakdown of the French dataset in terms of URLs, tweets and users.

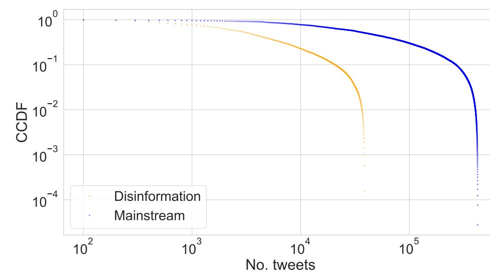

Figure 4: Complementary cumulative distribution of the number of shared tweets for mainstream (blue) and disinformation (orange) news articles in the French scenario.

We collected data continuously with Twitter Streaming API in the period 25/11/2019-25/12/2019. A breakdown of the dataset in terms of articles, tweets and users (who authored tweets) is available in Table 2. We can notice a larger discrepancy w.r.t the Italian scenario for what concerns the volumes of tweets shared across the two news domains (9:1 ratio between mainstream and disinformation); besides, we argue that a much larger volume of mainstream news is also due to the fact that we monitored a larger number of sources compared to the Italian setting (50 vs 25). Overall, disinformation exhibits double the number of users compared to Italy, but only a slightly larger number of associated tweets.

We show in Figure 4 the complementary cumulative distribution of the number of tweets shared by users for both domains. We can observe a similar behaviour as in the Italian scenario, i.e. both are heavy-tailed distributions and users sharing mainstream news are more active in the sharing.

In Figure 5 we show the Top-10 mainstream sources w.r.t to the number of associated tweets and unique URLs; in Figure 6 we show same statistics for Top-10 disinformation outlets. For what concerns mainstream sources, we notice that "leparisien.fr" stands out on other sources by far, and that these exhibit larger gaps in the ranking compared to the Italian scenario (where all sources, except the 1st one, have a similar number of associated tweets). For what concerns disinformation sources we can notice similar cardinalities in the number of associated tweets compared to the Italian scenario, but with a more homogeneous ranking, with a larger number of outlets with non negligible engagement; also, some of them have a number of associated tweets which is comparable to a few mainstream sources in the Top-10 ranking.

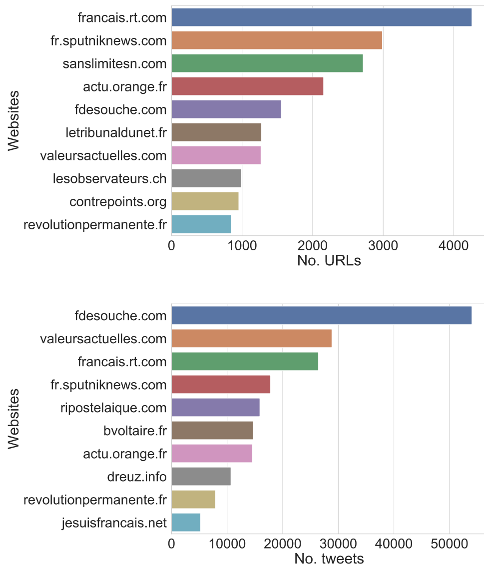

Figure 5: Distribution of the number of associated URLs (top) and tweets (bottom) shared for Top-10 French disinformation sources.

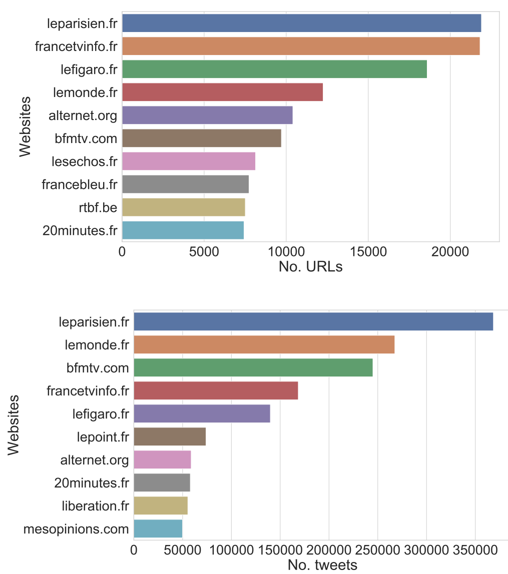

Figure 6: Distribution of the number of associated URLs (top) and tweets (bottom) shared for Top-10 French mainstream sources.

3 NETWORK ANALYSIS

3.1 Twitter diffusion networks

For each country, we built several Twitter diffusion networks according to the procedure described in [16, 21, 22]. In particular, we represent each of them as a weighted directed graph where nodes are users and edges are built according to the type of Twitter interaction. More precisely, given two nodes a and b we build an edge $a \longrightarrow b$ whenever

- a is retweeted by b
- a is quoted by b
- b mentions a
- b replies to a

In case of pure tweets, i.e. tweets which do not contain interactions with other users, we simply add the node to the graph (if not present).

For each country, we built three different diffusion networks: (a) the *disinformation* diffusion network, which is derived by tweets containing links to disinformation websites, (b) the *mainstream* diffusion network, which is obtained processing tweets containing links to mainstream outlets, and an *intersection* network (c) which corresponds to the union of edges and the intersection of nodes in (a) and (b); thus, (c) represents the set of users (and their interactions) who shared at least one disinformation and one mainstream news article.

In the following we provide some statistics on these networks for both Italy and France.

3.2 Network statistics

For each network we computed the following metrics, taken from the network science toolbox [3, 14]:

(1) Number of weak connected components (**WCC**)
(2) Size of the giant connected component (**S-GCC**)
(3) Number of disconnected nodes (**Disc.**)
(4) Average degree (which is equal to the mean in-degree and mean out-degree in a directed network) ($< k >$)
(5) Max in-degree (**max**(k_{in}))
(6) Max out-degree **max**(k_{out}))
(7) Density (d)
(8) Average clustering coefficient (**CC**)
(9) Main K-core number (**KC**)

We show in Table 3 values for all networks and countries.

For what concerns Italy, we first notice a larger network for mainstream news, but this is less clustered and connected compared to the disinformation network; mean degree and max out-degrees are comparable, whereas the max in-degree for mainstream news is much larger. We also observe a smaller giant connected component (in proportion) in the mainstream network w.r.t the disinformation network, but a higher K-core number and a larger number of disconnected nodes (i.e. isolated nodes). The intersection network is slightly smaller than the disinformation network, but it is more clustered and denser, and it also has a much higher K-core number (comparable to the mainstream network); also, the mean degree is much higher than the other networks, and there is a smaller number of disconnected nodes compared to the disinformation network.

For what concerns France, we observe similar differences in the three networks, with a less accentuated discrepancy in the connectedness (cf. CC) of the mainstream network compared to the disinformation. The mainstream network is by far bigger than the disinformation network–which is bigger than the Italian counterpart–and also than the Italian mainstream network, with a smaller number of disconnected nodes compared to the Italian scenario and a bigger

Network	Nodes	Edges	WCC	GCC	Disc.	$< k >$	max(k_{in})	max(k_{out})	d	CC	KC
IT Disinformation	11,632	40,252	1,843	9,614 (82.6%)	1,735	6.92	282	3,454	2.97e-04	0.773	27
IT Mainstream	88,607	224,111	20,590	65,555 (73.98%)	19,004	5.05	6,970	3,342	2.85e-05	0.028	39
IT Intersection	9,334	98,572	995	8,334 (89.28%)	988	21.12	725	3,130	1.12e-03	0.108	45
FR Disinformation	40,496	105,263	6,386	32,628 (80.57%)	5,494	5.19	984	5,344	6.41e-05	0.038	34
FR Mainstream	437,446	1,248,682	30,013	400,741 (91.60%)	25,365	5.70	6,752	35,338	6.52e-06	0.027	61
FR Intersection	31,092	366,776	2,580	28,434 (91.45%)	2,526	23.59	1,847	7,584	3.79e-04	0.086	68

Table 3: Indicators for diffusion networks of both news domains and countries.

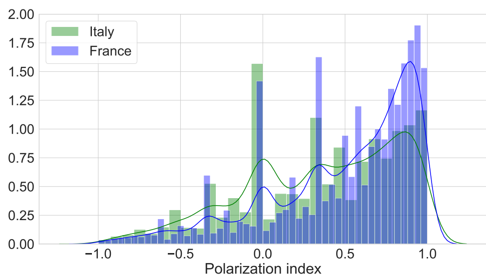

Figure 7: Normalized histogram and kernel density estimation (solid curve) for the polarization index of users in the intersection networks of Italy (green) and France (blue).

giant connected component (in proportion). The max out-degree is much larger than all other networks, including Italian ones, whereas the max in-degree is slightly bigger than its Italian counterpart; also, the mean degree of both mainstream and disinformation networks is smaller than their Italian counterpart.

We further computed for each user a polarization index ρ defined as:

$$\rho = \frac{t_m - t_d}{t_m + t_d}$$

where t_m is the number of tweets containing a link to mainstream outlets and t_d is the number of tweets containing a link to disinformation websites. We show in Figure 7 the normalized histogram and kernel density estimation of this index for users in the intersection network of both countries. We can see that users tend to be polarized towards mainstream news, with a peak in the number of users who share news from both domains and a negligible fraction of users sharing only disinformation articles. We observed similar results also when considering only users who actually authored tweets (excluding those who are just involved via Twitter actions).

4 CONCLUSIONS AND FUTURE DIRECTIONS

In this work we provided preliminary results from an ongoing investigation of the diffusion of news published on *disinformation* and *mainstream* news outlets.

By leveraging Twitter Streaming API, we collected thousands of tweets containing links to news articles published on sources in two different countries, Italy and France. Overall, the amount of misleading and potentially harmful information is small yet not negligible compared to mainstream news, with approximately 15% of the total share of news in Italy and 10% in France.

We showed that mainstream news outlets generate a much larger engagement in both countries: only a handful of Italian disinformation outlets are actively shared by Twitter users, whereas in France there is a larger number of disinformation outlets which exhibit a non-negligible volume of shares.

We observed a strong polarization towards mainstream news when considering users who also shared non-credible information, in both countries.

Finally, analyzing the diffusion networks pertaining to disinformation and mainstream news, we observed that in both countries disinformation networks are generally more clustered and connected, and that they are much smaller than the mainstream ones.

In the future we plan to analyze the news coverage of outlets belonging to distinct news domains, estimate the presence of any agenda-setting effects and uncover coordinated deceptive strategies between mainstream and disinformation sources. We also intend to better understand the political affiliations (if any) of most active users in the diffusion networks of both domains, to further explore the presence of echo chambers and bots. Finally, we aim to apply a multi-layer representation of Twitter diffusion networks in order to better understand differences and similarities in the news sharing behavior of users who engage more actively with disinformation rather than mainstream news.

5 ACKNOWDLEGMENTS

The author is partially supported by the PRIN grant HOPE (FP6, Italian Ministry of Education). The author is thankful to Stefano Ceri for constructive feedback, and to Alexios Mantzarlis for redirecting to the inquiry by Le Monde.

REFERENCES

[1] Le decodex, un outil de verification de l'information. available at https://www.lemonde.fr/verification/.
[2] AGCOM. News vs fake nel sistema dell'informazione. *Report available at: https://www.agcom.it/documents/10179/12791486/Pubblicazione+23-11-2018/93869b4f-0a8d-4380-aad2-c10a0e426d83?version=1.0*, 2018.
[3] A.-L. Barabási et al. *Network science*. Cambridge university press, 2016.
[4] M. T. Bastos and D. Mercea. The brexit botnet and user-generated hyperpartisan news. *Social Science Computer Review*, 37(1):38–54, 2019.
[5] A. Bovet and H. A. Makse. Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications*, 10(1):7, 2019.
[6] G. Caldarelli, R. De Nicola, F. Del Vigna, M. Petrocchi, and F. Saracco. The role of bot squads in the political propaganda on twitter. *arXiv preprint arXiv:1905.12687*, 2019.
[7] M. Cantarella, N. Fraccaroli, and R. Volpe. Does fake news affect voting behaviour? *Available at SSRN: https://ssrn.com/abstract=3402913*, 2019.
[8] E. Ferrara. Disinformation and social bot operations in the run up to the 2017 french presidential election. *First Monday*, 22(8), 2017.
[9] F. Giglietto, L. Iannelli, L. Rossi, A. Valeriani, N. Righetti, F. Carabini, G. Marino, S. Usai, and E. Zurovac. Mapping italian news media political coverage in the lead-up to 2018 general election. *Available at SSRN: https://ssrn.com/abstract=3179930*,

2018.
[10] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer. Fake news on twitter during the 2016 u.s. presidential election. *Science*, 363(6425):374–378, 2019.
[11] P. N. Howard, S. Bradshaw, B. Kollanyi, and G. Bolsolver. Junk news and bots during the french presidential election: What are french voters sharing over twitter in round two?
[12] D. M. J. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, M. Schudson, S. A. Sloman, C. R. Sunstein, E. A. Thorson, D. J. Watts, and J. L. Zittrain. The science of fake news. *Science*, 359(6380):1094–1096, 2018.
[13] F. Morstatter, J. Pfeffer, H. Liu, and K. M. Carley. Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose. In *Seventh international AAAI conference on weblogs and social media*, 2013.
[14] M. Newman. *Networks*. Oxford university press, 2018.
[15] R. K. Nielsen, N. Newman, R. Fletcher, and A. Kalogeropoulos. Reuters institute digital news report 2019. *Report of the Reuters Institute for the Study of Journalism*, 2019.
[16] F. Pierri, A. Artoni, and S. Ceri. Investigating italian disinformation spreading on twitter in the context of 2019 european elections. *PloS one*, 15(1):e0227821, 2020.
[17] F. Pierri and S. Ceri. False news on social media: a data-driven survey. *ACM Sigmod Record*, 48(2), 2019.
[18] F. Pierri, C. Piccardi, and S. Ceri. Topology comparison of twitter diffusion networks effectively reveals misleading news. *Scientific reports*, 10:1372, 2020.
[19] A. Senecat. Les fausses informations circulent de moins en moins sur facebook. *LeMonde*, 2018.
[20] C. Shao, G. L. Ciampaglia, A. Flammini, and F. Menczer. Hoaxy: A platform for tracking online misinformation. In *Proceedings of the 25th International Conference Companion on World Wide Web*, WWW '16 Companion, pages 745–750, Republic and Canton of Geneva, Switzerland, 2016. International World Wide Web Conferences Steering Committee.
[21] C. Shao, G. L. Ciampaglia, O. Varol, K.-C. Yang, A. Flammini, and F. Menczer. The spread of low-credibility content by social bots. *Nature Communications*, 9(1):4787, 2018.
[22] C. Shao, P.-M. Hui, L. Wang, X. Jiang, A. Flammini, F. Menczer, and G. L. Ciampaglia. Anatomy of an online misinformation network. *PLOS ONE*, 13(4):1–23, 04 2018.
[23] M. Stella, M. Cristoforetti, and M. De Domenico. Influence of augmented humans in online interactions during voting events. *PloS one*, 14(5), 2019.
[24] M. Stella, E. Ferrara, and M. De Domenico. Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences*, 115(49):12435–12440, 2018.
[25] M. D. Vicario, W. Quattrociocchi, A. Scala, and F. Zollo. Polarization and fake news: Early warning of potential misinformation targets. *ACM Transactions on the Web (TWEB)*, 13(2):10, 2019.