The Polarized Web of the Voter Fraud Claims in the 2020 US Presidential Election

Sreeja Nair and Adriana Iamnitchi
University of South Florida
sreejas@usf.edu, anda@cse.usf.edu

Abstract

We show that the polarization of the Twitter discussions around the fraud claims in the US presidential election is evident without considering tweet content other than URLs. In addition, we show that YouTube is connecting the two communities identified solely by structural community detection, thus being in the rare position of bringing together both right and left-leaning users. This enviable position could be used by YouTube video recommendation algorithms to alleviate some of the polarization in the society.

Introduction

Twitter plays a pivotal role in the US political discourse. Politicians, parties, media and the general public use Twitter as a medium to promote/demote their own or opposite parties, to conduct debates, and to interact with other voters. This de facto communication and engagement medium is thus the ideal place for spreading misinformation and political propaganda (Grinberg et al. 2019; Jasny and Rai 2019). Researchers have shown evidence of the correlation of major political/social events such as the 2016 U.S. presidential election (Allcott and Gentzkow 2017), Brexit (Bastos and Mercea 2019), and Covid-19 (Calvillo et al. 2020) with the diffusion of fake news.

Users' partisan behavior and confirmation biases are key factors in the spreading of misinformation. Selective exposure to contents from social media facilitate the formation of echo chambers. Many studies have shown the existence of polarization in Twitter. An early study on Twitter (Conover et al. 2011) shows that the retweet network exhibits a highly modular structure, separating users into two communities corresponding to the political left and right leaning. Most of the existing studies on Twitter polarization are centered on structural patterns of user activities such as tweet, retweet, mention, hashtags. Studies are also done on follower/followed network, tweeting behavior, and message content (Garimella and Weber 2017).

With all this previous evidence of the polarization on Twitter, we expect that another political situation, namely the claims of election fraud in the 2020 US presidential election, will show clear signs of polarization on Twitter. In this paper we investigate how polarization is reflected in the sharing of news media domains on Twitter. We use a publicly available Twitter dataset (Abilov et al. 2021) on 2020 U.S. presidential election voter fraud claims. We start by deriving a domain co-sharing network and label them based on the political bias of these websites, as reported by Media Bias Fact Check. We show that the polarization of user engagement is evident in the structure of the domain co-sharing network. We also investigate the dominant websites that drive conversation in each polarized communities. We then report the major domains that connect the polarized communities.

Related Work

Previous studies have presented evidence of the level of overall exposure to fake news in the 2016 US presidential election (Allcott and Gentzkow 2017; Bovet and Makse 2019; Grinberg et al. 2019). Many recent studies focused on misinformation and fake news in Twitter (Del Vicario et al. 2017; Vosoughi, Roy, and Aral 2018; Shao et al. 2018) as well as other online social media platform such as YouTube (Hussain et al. 2018; Lemos, Bitencourt, and dos Santos 2020), Facebook (Guess, Nagler, and Tucker 2019; Geeng, Yee, and Roesner 2020), or Reddit (Carman et al. 2018; Setty and Rekve 2020).

The spreading of misinformation and fake news is interconnected to the emergence of polarization and filter bubbles in social media. Zollo and Quattrociocchi (Zollo and Quattrociocchi 2018) show that online users connect to polarized communities by sharing common opinions and try to obtain information from sources that confirm their beliefs (confirmation bias) even if they contain false claims or fake news. The bias in online media and cable television news increases polarization (Martin and Yurukoglu 2017), which in turn affects the voting behavior and the taste of like minded news. Bail et al. (Bail et al. 2018) showed that polarization in social media can exacerbate with exposure to those with opposing political views.

Most studies on the polarization in Twitter analyzed topological structures of networks such as tweet, retweet, mentions network, and or analyzed the content of tweets. Adamic and Glance (Adamic and Glance 2005) studied the linking patterns among the discussion topics of political bloggers. Conover et al. (Conover et al. 2011) demonstrated that a network of political retweets exhibits a highly polarized structure, with extremely limited connectivity between left- and right-leaning Twitter users.

Most of the existing studies showed the evidence of polarization on Twitter by mainly focusing on the structural patterns of networks built from tweets, retweets, reply or mentions, and analyzing the content of tweets (Yang et al. 2017). They often make use of vast metadata from the Twitter API such as user details, geography, or follower/followee networks. In this work, we only use the domains of the URLs shared by Twitter users related to the voter fraud claims in the 2020 US presidential election. We show that even by only considering the website domains from a Twitter dataset (Abilov et al. 2021), polarization is evident.

Dataset

Our dataset is derived from the publicly available VoterFraud2020 dataset (Abilov et al. 2021), a multi-modal Twitter dataset collected between between October 23rd and December 16th, 2020 and related to voter fraud claims about the US 2020 Presidential election. We used hydration to retrieve tweet data for the tweet identifiers provided in the dataset. From the hydrated data, we collected the URLs shared in tweets. We extracted the domains from the URLs. We removed the domain 'twitter.com' from the dataset, as such inclusions typically indicate quoted tweets.

We obtained 15,637 domains tweeted by 202,827 distinct users. We created an undirected domain network, where domains are nodes and an edge between two domains exists if a user tweeted both domains during our observation period (not necessarily in the same tweet). Edges are weighted with the number of users who tweeted the given pair of website domains. In this network, 8,352 nodes are isolates, that means the users who shared these domains have not shared any other urls/domains. We removed these isolated domains from the network. The final network contains 7,285 nodes and 207,194 edges. Figure 1 is a visualization of the domain network. This network of websites is structurally divided into many communities.

We label the domains based on their media bias and credibility using publicly available sources such as *Media Bias Fact Check (MBFC)* and *iffy.news*. We label 1,522 nodes as non-credible (nc), least bias (cr), left bias (lb), left center (lc), right bias (rb) and right center (rc). The remaining nodes are labeled as unknown (un). Around 78% domains are marked as unknown.

We removed the edges with weight 1 to focus only on the pairs of websites shared by two users or more. We consider an edge significant only if multiple users share the same domain pairs. The resulting network contains 2,607 nodes and 60,964 edges, of which 1,665 are unknown. We run the Louvain community detection algorithm (Blondel et al. 2008) on the reduced network and obtained 13 communities, of which the two largest contain 98.7% of the nodes (Figure 2), with 1,693 and 880 nodes, respectively.

Observations

Our first investigation is to understand if the structure of the website co-sharing network can reveal the polarization of the users engaged in the discussion of the alleged fraud, without any other information about the tweets. Intuitively, users who believe in the fraud election narratives would share URLs who support their beliefs; these websites will consistently share the same views via different articles they host. But how much of this intuition is supported by data? For this, we look at the distribution of labels as applied by Media Bias to the websites in the two clusters. Table 1 shows the node distribution based on labels. The percentage from the total labeled nodes is given within parenthesis.

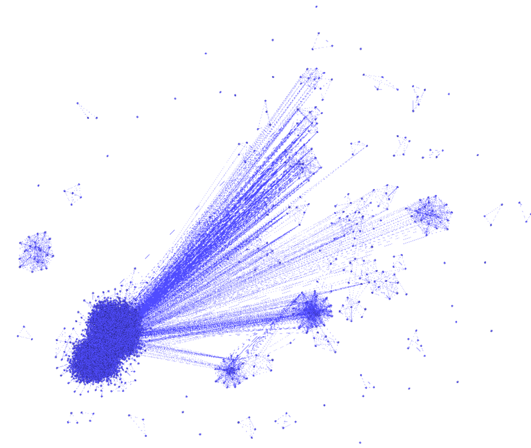


Figure 1: Visualization of domain network.

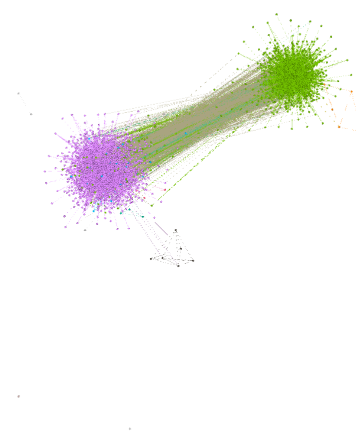


Figure 2: Visualization of communities in the domain network after running the Louvain community detection.

The table shows that in Community 1 there is a larger percentage of non credible websites: 18% vs 1% of the labeled nodes are non credible. In the end, after many investigations and vote recounts, no proof of significant election fraud was discovered. The fact that the non-credible sources were shared more heavily in one community discovered based on classic community detection algorithm suggests the polarization of the Twitter user community engaged in this conversations. Moreover, Community 1 con-

Labels	Community 1	Community 2
Unknown	71.18	48.75
Non Credible	5.43 (18.85)	0.68 (1.33)
Least Bias	5.79 (20.08)	14.88 (29.04)
Right Bias	14.12 (48.86)	0.79 (1.55)
Right Center	1.65 (5.74)	4.89 (9.53)
Left Bias	0.59 (2.05)	11.13 (21.73)
Left Center	1.24 (4.3)	18.86 (36.81)

Table 1: Node distribution in Community 1 and Community 2 based on news site labels applied by Media Bias Fact Check and iffy.news. The numerical values represent the percentage of nodes for each label in each community. The numbers in parentheses show the percentage of labeled nodes in the community when unknown nodes are ignored.

tains a very low percentage of left-leaning websites (6.35% of the labeled nodes), while Community 2 contains a low percentage of right leaning websites (about 11% of the labeled nodes), which suggests that the first community is right leaning (about 54.6% of the nodes are labeled right bias or right-leaning) while community 2 is left-leaning (with 58.5% nodes).

For further clarity about these two communities we plot in Figure 3 the edge distribution between sites of different labels. Figures 3 (a) to (c) show the edge distribution based on the number of edges between domains (thus, ignoring the edge weights). Community 1 contains a large proportion of connections between right bias and unknown nodes (39%). 18% of edges connect right-bias to right-bias domains. Similarly, we see strong connections between left-bias, left-center, least-bias and unknown in Community 2. 16.45% of edges are between unknown and left center domains. The edge between left-center and left-center and left-bias account for 13% and 15%, respectively.

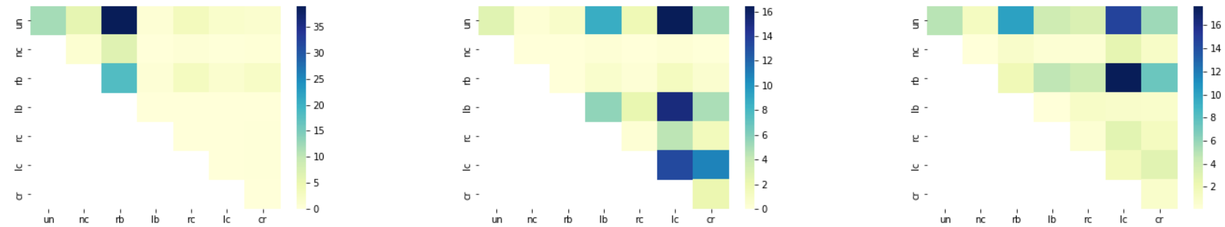
When we consider the edge weights as well, thus the number of times two websites appeared in users' history of tweeting about the election fraud claims, we see a similar story of right leaning pairs and left leaning pairs separated into two communities. Figure 3 (d) to (f) show the edge distribution based on edge weights between domains in each community. In Community 1, 31% of edge weights are distributed among right-bias – right-bias pairs. 29% edge weight are among right-bias and unknown pairs. On the other hand, 23% of edge weights in Community 2 are between left-bias and left-center domains. 20% edge weight are left-center domains.

Based on these observations, we conclude that the Louvain community detection algorithm applied to the website co-sharing network ended up clustering websites by political leaning despite the fact that it only used the network topology. This is due to the biased views of the Twitter users who chose to promote the websites that best reflect their own beliefs with little concern to source credibility or reputation.

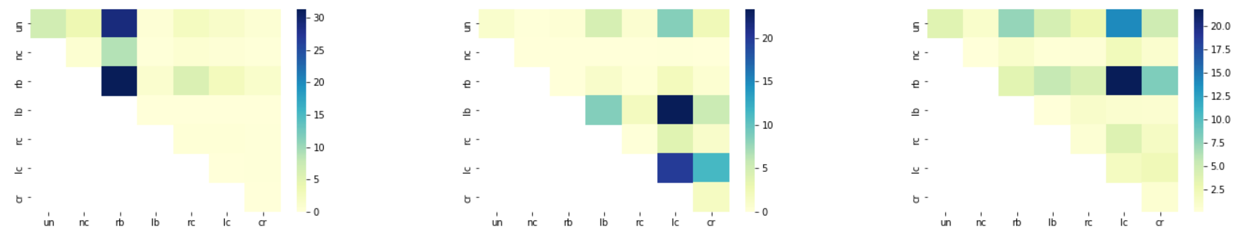
The next question we investigate is which websites are the drivers of conversation in each community. Table 2 shows the top five websites shared on each side. On the right side, `thegatewaypundit.com` is the node with the highest weighted degree in the website network, followed by the known drivers of the right bias media, such as `breitbart.com` and `newsmax.com`. YouTube also appears prominently in this group but because it is a platform for content creators, it does not have an assigned bias category. In Community 1 YouTube mostly connects to right bias domains (52% of its connections). We also note that the highest connection of YouTube is towards other hubs in the community. Note that four out of the five websites have a "right-bias" label applied. In contrast, in the left-leaning community, the top shared domains are not all towards the left: two labels are "left-bias" (for CNN and Raw Story), two "left-center" (Washington Post and New York Times) and one labeled as "least bias" (Associated Press).

We next ask: how well connected are the two communities and which are the websites that connect the two sides? Intuitively, we would expect web sites classified as towards the center (left-center, right-center, least bias) to be the connectors of the two clusters. 11% of edge weights are distributed in cross-community. Table 2 shows that YouTube is the top connector of the two clusters. As discussed earlier, YouTube is also one of the top hub in the Community 1. Figure 5 shows the edge weight distribution of YouTube in Community 1 and cross community. 43% of edge weight from YouTube to Community 2 goes towards left-center domains and 17% towards left-bias. In Community 1, 52% of Youtube's edges are towards right-bias domains and 27% towards unknown. An interesting finding about YouTube in this network is that, in its community, YouTube is mostly connected to low credible right leaning websites such as `thegatewaypundit.com`, `theepochtimes.com`, `breitbart.com` or `rumble.com`. On the other hand, it connects to mainstream least/left-bias media sources such as CNN, Washington Post, New York Times, or The Guardian.

Figure 4 shows how the highest degree nodes from the cross-community group of websites connect to the other community. The second highest hub that connects the two communities is Fox News, listed as right-bias by Media Bias, and part of the Community 1 in our analysis. 45% of the Community 2 domains co-shared with Fox News are left-center domains and 17% are left-bias. For comparison, `thehill.com`, which is labeled as a least-bias domain, has 53% of its connections to the opposite community labeled as right-bias domains and 25% as unknown domains. Similarly, `newsweek.com`, a left-center domain, has 50% of its cross community edges connecting to right-bias domains and 27% towards unknown. Figure 4 shows that left leaning cross community hubs `thehill.com` and `newsweek.com` are mostly connected to the right-bias domains, whereas right-leaning hubs `foxnews.com` and `thegatewaypundit.com` mostly connect to left-center websites in the other community. Figure 3 (f) shows that 22% of cross community edge weights are between right-bias and left-center. 14% weights are distributed among left-center and unknown domains. In short, right-bias nodes from Community 1 are the main connectors of the two clusters while the left-center nodes are the main connectors from Community 2. Even if a left-center or least-bias domain is



(a) Community 1 (b) Community 2 (c) Cross community

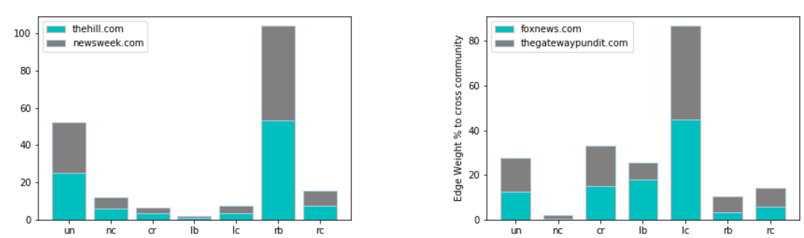


(d) Community 1 (e) Community 2 (f) Cross community

Figure 3: Heatmap showing the edge distribution of labels in each community. (a) to (c) show edge distribution based on unique connection between domains. (d) to (f) show edge distribution based on edge weight.

Community 1		Community 2		Cross community	
Domain	Label	Domain	Label	Domain	Label
thegatewaypundit.com	rb	washingtonpost.com	lc	youtube.com	un
youtube.com	un	cnn.com	lb	foxnews.com	rb
breitbart.com	rb	nytimes.com	lc	thehill.com	cr
theepochtimes.com	rb	rawstory.com	lb	newsweek.com	lc
newsmax.com	rb	apnews.com	cr	thegatewaypundit.com	rb

Table 2: Top 5 hubs in the communities based on popularity. The table shows the hubs and its label in each community. Cross community means the nodes and edges that connect the Community 1 and the Community 2.



(a) Left-leaning cross-community hubs (b) Right-leaning cross community hubs

Figure 4: Stacked bar plot showing the percentage of connections to the opposite community of the two top hubs from each community. (a) Percentage of connections of the top two left leaning (Community 2) hubs towards the Community 1. (b) Percentage of connections of the top two right leaning (Community 1) hubs towards the Community 2.

a cross community hub, more than half of its connections go towards right-bias domains.

To conclude, our expectation was that the center bias websites (left-center, right-center or least bias) would form the majority of cross-community edges. Our analysis shows that, surprisingly, right bias websites are the main connectors to the other group. This is not only true for the most popular websites (top hubs), but over all websites engaged in cross edges, as shown by the correlation between right bias and left-center nodes in Figure 3 (f).

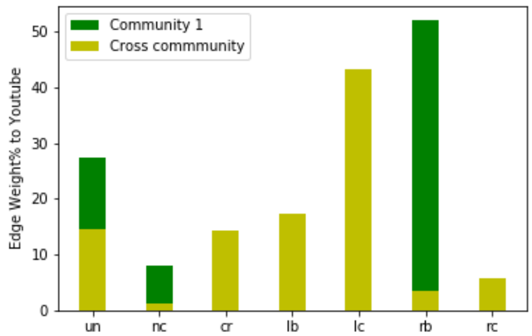


Figure 5: Edge weight distribution of Youtube to other labeled nodes in Community 1 and cross-community

Summary and Discussions

This paper investigates the possibility of finding polarized communities in Twitter from the domains of URLs shared in the tweet, without any other information about the tweet. We create the domain network from the publicly available Twitter data on 2020 U.S. presidential election voter fraud claims. Our main focus is to find whether the structure of the website co-sharing network shows sign of polarization among the users who engaged in the voter fraud claims discussions. We have confirmed that there is polarization in the Twitter dataset by only considering URLs from the original tweets. The Louvain community detection algorithm applied to the network creates two large communities. These communities show a dominant political leaning based on the websites' political bias.

The results show that in the right-leaning community most hubs (such as `thegatewaypundit.com`, `theepochtimes.com`, and `breitbart.com`) are low credible, as classified by Media Bias Fact Check. These websites were found in promotion of conspiracies and numerous instances of publishing false (fake) news. YouTube also appears predominantly in this group. The other community forms around left-leaning mainstream media as hubs.

We also investigated how well the two communities are connected and which are the websites that most often connects the two sides, that is, appear most often in the tweets of users who also tweet the other side. Around 11% of the edges in our domain network are between the two communities. YouTube is the top connector, and while it is assigned to the right-leaning community, it mostly connects to left-leaning mainstream media. We see that the most connections between the two political leaning communities are via right-bias websites and left-center websites.

This work opens up new research questions related to reliable user assignment to political leanings. Is it possible that the domain network as constructed in this work can help assign more reliable labels to Twitter users, especially when the users are hard to classify as right- or left-leaning based on the content of their Twitter activity? Further research is also needed to find the role of YouTube in polarization. Considering the important position of YouTube in the network we study and in the social media landscape, can YouTube contribute to a more civil and less polarized online discourse?

References

Abilov, A.; Hua, Y.; Matatov, H.; Amir, O.; and Naaman, M. 2021. VoterFraud2020: a Multi-modal Dataset of Election Fraud Claims on Twitter.

Adamic, L. A.; and Glance, N. 2005. The Political Blogosphere and the 2004 U.S. Election: Divided They Blog. In *Proceedings of the 3rd International Workshop on Link Discovery*, LinkKDD '05, 36–43. New York, NY, USA: Association for Computing Machinery.

Allcott, H.; and Gentzkow, M. 2017. Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives* 31: 211–236.

Bail, C. A.; Argyle, L. P.; Brown, T. W.; Bumpus, J. P.; Chen, H.; Hunzaker, M. B. F.; Lee, J.; Mann, M.; Merhout, F.; and Volfovsky, A. 2018. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences* 115(37): 9216–9221.

Bastos, M. T.; and Mercea, D. 2019. The Brexit Botnet and User-Generated Hyperpartisan News. *Social Science Computer Review* 37(1): 38–54.

Blondel, V. D.; Guillaume, J.-L.; Lambiotte, R.; and Lefebvre, E. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008(10): 10008. doi:10.1088/1742-5468/2008/10/P10008.

Bovet, A.; and Makse, H. 2019. Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications* 10.

Calvillo, D. P.; Ross, B. J.; Garcia, R. J. B.; Smelter, T. J.; and Rutchick, A. M. 2020. Political Ideology Predicts Perceptions of the Threat of COVID-19 (and Susceptibility to Fake News About It). *Social Psychological and Personality Science* 11(8): 1119–1128.

Carman, M.; Koerber, M.; Li, J.; Choo, K. R.; and Ashman, H. 2018. Manipulating Visibility of Political and Apolitical Threads on Reddit via Score Boosting. In *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, 184–190.

Conover, M. D.; Ratkiewicz, J.; Francisco, M.; Gonçalves, B.; Menczer, F.; and Flammini, A. 2011. Political Polarization on Twitter. In *ICWSM*.

Del Vicario, M.; Gaito, S.; Quattrociocchi, W.; Zignani, M.; and Zollo, F. 2017. Public discourse and news consumption on online social media: A quantitative, cross-platform analysis of the Italian Referendum .

Garimella, V.; and Weber, I. 2017. A Long-Term Analysis of Polarization on Twitter. In *ICWSM*.

Geeng, C.; Yee, S.; and Roesner, F. 2020. Fake News on Facebook and Twitter: Investigating How People (Don't) Investigate. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 1–14. New York, NY, USA: Association for Computing Machinery.

Grinberg, N.; Joseph, K.; Friedland, L.; Swire-Thompson, B.; and Lazer, D. 2019. Fake news on Twitter during the 2016 U.S. presidential election. *Science* 363(6425): 374–378. ISSN 0036-8075.

Guess, A.; Nagler, J.; and Tucker, J. 2019. Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances* 5(1).

Hussain, M. N.; Tokdemir, S.; Agarwal, N.; and Al-Khateeb, S. 2018. Analyzing Disinformation and Crowd Manipulation Tactics on YouTube. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 1092–1095.

Jasny, B. R.; and Rai, T. S. 2019. Finding facts about fake news. *Science* 363(6425): 360–360.

Lemos, A. L. M.; Bitencourt, E. C.; and dos Santos, J. G. B. 2020. Fake news as fake politics: the digital materialities of YouTube misinformation videos about Brazilian oil spill catastrophe. *Media, Culture & Society* .

Martin, G.; and Yurukoglu, A. 2017. Bias in Cable News: Persuasion and Polarization. *American Economic Review* 107: 2565–2599.

Setty, V.; and Rekve, E. 2020. Truth Be Told: Fake News Detection Using User Reactions on Reddit. In *Proceedings of the 29th ACM International Conference on Information amp; Knowledge Management*, CIKM '20, 3325–3328. New York, NY, USA: Association for Computing Machinery.

Shao, C.; Hui, P.-M.; Wang, L.; Jiang, X.; Flammini, A.; Menczer, F.; and Ciampaglia, G. L. 2018. Anatomy of an online misinformation network. *PLOS ONE* 13(4): 1–23.

Vosoughi, S.; Roy, D.; and Aral, S. 2018. The spread of true and false news online. *Science* 359(6380): 1146–1151.

Yang, M.; Wen, X.; Lin, Y.; and Deng, L. 2017. Quantifying Content Polarization on Twitter. In *2017 IEEE 3rd International Conference on Collaboration and Internet Computing (CIC)*, 299–308.

Zollo, F.; and Quattrociocchi, W. 2018. *Misinformation Spreading on Facebook*, 177–196. ISBN 978-3-319-77331-5. doi:10.1007/978-3-319-77332-2_10.