Diffusion of Fake News a.k.a Rumors

SI 608 - Networks | Project Report | April 2017
Team: Jasmit Kaur, Lia Bozarth, Omkar Sunkersett, Younghun Lee

Summary

Our study intended to understand the network characteristics fake news and real news. To gather our data, we first created a list of fake news and real news domain names. We then used Twitter API to gather 7 million plus tweets that included URL pointing to fake news over a period of ten days. In that same time period, we also gathered 5 million plus tweets containing URLs pointing to some fake news. To build a network, we downloaded friends and follower network for the top 100 users (those who tweeted the maximum number of times). Out of this dataset, we kept only those users that tweeted a particular type of news at least 100 times. Although, this may be a limitation of our study, containing our data with these cut-offs enabled us to compute network statistics and build graphs more effectively.

Our key takeaways included the following: First, Fake news network is larger than real news network. Just the number of nodes and edges alone in fake news were more. This might imply that there is more concentrated and organized effort towards the spread of fake news than there is for real news. Second, Fake news networks are more interconnected compared to real news networks. This is evident from the fact that 68% of nodes are in the largest SCC (compared to only 44% in real news), the diameter is 7 (compared to 9 in real news network) for the largest SCC, and the clustering coefficient for the fake news network is higher at 0.26 with a higher number of closed triangles (14,529) compared to real news network which has 0.01 network clustering coefficient and fewer number of closed triangles (27). Third, when we looked at the top users of the fake news and real news network, we learned that top users in fake news network are more connected with others in their network and are closer to many more nodes in their network compared to those in real news networks. In terms of their degree, closeness and betweenness centrality measures, we learned that degree centrality in case of fake news users is higher and so is closeness centrality. Betweenness centrality is similar in case of both the networks possibly because the fake news networks have more interconnected nodes to the prestiges associated with 'betweenness' is missing. Apart from these observations, we learned about the geographic location of spread of fake and real news users which largely mirrored each other except for higher concentration fake news in certain areas. We learned what the top hashtags were and build a network around the hashtags which further confirmed the interconnectivity or "sharedness"

of interests among the fake news users. From our analysis of the fake news sites, we gathered the behavioral patterns of users in terms of mentioning top websites.

Overall, we believe that the more compactness and interconnectivity of the fake news network is because of a higher degree of homophily. Fake news users perhaps connect as friends and followers with each other at a higher degree than those in the real news network.

Research Objectives

According to Pew Research Center (Link; July 2016) 38% of Americans get their news from online - social media, apps, etc. According to The Guardian (Link; Dec 2016), "Fake news is an insidious trend that is fast becoming a global problem". According to Jumpshot (Link, Nov 2016), over 70% of desktop-traffic from fake or hyper partisan news sites comes from Facebook. The role of social media in the diffusion of rumors and fake news cannot be denied. To understand how fake news spreads on social media and what could be possibly done to curb it at inception was our motivation.

With the above context we framed two key problem statements - First, we wanted to understand and compare network characteristics of fake news and "real" (non-fake) news. Second, we wanted to understand and compare other attributes of such networks like the hashtags used, the geographic origin of the users, the top websites referred. Based on our findings from these two key problem statements, we believed that we could get a deeper understanding of the spread of fake news and thereby identify mechanisms to challenge the spread of fake news.

We reviewed several existing research on these topics to develop our methodology and approach for the analysis. Our first big challenge was to get the right set of data with retweet chain related information to show diffusion or cascade of information. The three key papers which helped contribute to developing our methodology are discussed here. Kupavskii et al, in their paper "Predicting the Audience Size of a Tweet" collected all public tweets over two month period March 1 2012 - April 30 2012 using data from the Twitter \rehose API. The further explain their approach to building a dataset as follows -For the data for the first six weeks, they extracted all the ordered pairs of users who did a retweet via \retweet" button during these six weeks, and the time of each retweet. (750M pairs of users and 1.5B retweets.). Although, for their research study it was not relevant to get the exact chain of retweets, we examined the mechanics of retweets on twitter based on information in this paper. Jin et al, in their paper "Epidemiological Modeling of News and Rumors" studied news and rumors that were drawn from a variety of regions and across a diverse set of topics. Their data collection was aimed at gathering tweets highly related to the events under study. Jin et al described their approach to data collection as follows - They employed customized sets of keywords and hashtags pertaining to each incident and date range restrictions were used to define relevant tweets for each event. In their study, "they compared the basic properties of news and rumor propagation, by characterizing tweet volume over time, follower/follower distributions, the `response ratio' of a story, and the retweet cascades". Their retweet cascade was analysed in terms of time intervals during which bursts of information happened and they also used follower-followee information to build network. This paper provided us several clues about developing our approach. Lastly we reviewed the paper "Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter" by Romero et al. This research study aimed at predicting a user's probability to mention a particular hashtag after "successive exposure" to it. To build a network of users for their study, Romero et al connected "user X to user Y if X directed at least t @-messages to Y". Although this idea was very compelling, we did not use it for our study. The difference being that Romero et al used a "push" strategy for understanding how tweets cascaded and we were more interested in a person tweeting certain information regardless of whether it was a "push" or whether they did it without any explicit external influence.

Methodology and Dataset

To obtain online data in order to understand how fake news spreads, we evaluated data that can be obtained through various sources. We finalized on using Twitter API, because it is the most friendly social media API in terms of giving us vast amounts of data about the content and user profiles. To address our problem statements, we adopted the following methodology:

- **Clarify what is "fake" and "real" news:** Labelling a tweet as a fake news tweet or real news tweet can be very subjective. One person's fake news is another person's real news. Add to that mix, the tweets which are satirical. To address this complexity, we labelled tweets as "fake news" if they were propagating tweets with a URL to news sites and blogs which have been identified to be a "fake news" site. Appendix I contains the list of 1058 news sites that have been identified to have been a source of fake news by lists published by academicians and experts (like, Fake News Watch, NPR, Melissa Zimdar, etc.). Similarly, we labelled tweets as "real news", if they contained a URL pointing to a domain name of a site deemed non-fake or real news by a list prepared by forbes.com. Appendix II contains the list of 22 sites that have been identified to be source of "real" news.

- **Collect and refine sample data representing the spread of "real" and "fake news:** We downloaded tweets over a period of ten days - March 18 till March 28, 2017. For fake news, we downloaded We found 277 sites mentioned (out of original list of 1058 fake news sites) in the over 7 million (7,086,625) tweets we downloaded. For real news, we downloaded over 5 million (5,302,946) tweets.

- **Gather network data:** Since twitter does not allow us to get the chain of retweets, we adopted a proxy method wherein we identified the top 100 users from our dataset and then downloaded friends and followers network for these top 100 users. We ensured that only those friends and followers are included who have themselves tweeted the same "type" (real or fake) of news URL as the top 100 users. This proxy is based on the assumption that for followers, since the top 100 users' tweets appear on the twitter feed, it is likely that the users "retweeted" a particular tweet because they were influenced by the top 100 user they are following. Similarly, the friends' tweets show in the feed of the top 100 users and it is likely therefore that our top 100 users were influenced by the tweets posted by the friends they follow. This process allowed us to also build in directionality for our networks. To further contain the data, we limited the overall network to users who "frequently" tweeting a certain type of news - We used the limit for frequency to be 100.

  So, overall our final dataset contained friends and followers of top 100 tweeters of fake/real news over a period of ten days who frequently tweeted fake/real news (at least 100 times).

- **Analyse data to understand network characteristics:** We analyzed the network graphs and network measures like degree centrality, closeness centrality, clustering coefficient, and diameter. This allowed us to do visual and numerical comparison for the fake and real news networks.

- **Analyse data to understand other aspects of news related network (geographic location, hashtags, and top sites):** We also analysed a data to derive the qualitative differences between the networks in terms of users' geographic location, the hashtags they used and the top sites they mentioned in their tweets.

The above methodology led us to the findings presented in this paper.

Findings and Analysis

We did four different kinds of analysis on the networks - basic and centrality measures to understand network characteristics, geographic distributions of the users in the two networks, top websites referred, and hashtags used.

Network characteristics using basic and centrality measures

In our basic analysis of the network statistics (Table 1) we found differences between the fake and real news network.

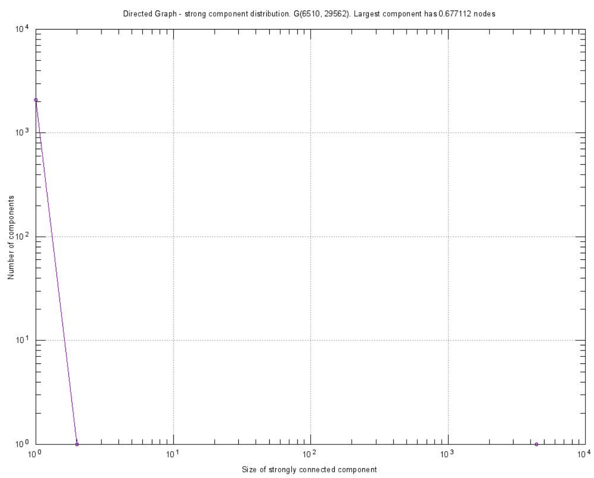
Table 1: Basic network measures
These network measures helped us get a basic understanding of the composition of the fake and real news networks on twitter.

No.	Measures	Fake News	Real News
1	Nodes	6510	591
2	Edges	29562	1321
3.	Bidirectional edges (%age of total edges)	17816 (60%)	608 (46%)
4.	Zero InDeg Nodes (%age of total nodes)	706 (11%)	92 (16%)
5.	Zero OutDeg Nodes (%age of total nodes)	1241 (19%)	106 (18%)
6.	Size of the largest SCC	4408	265
7.	Portion of the largest SCC among all nodes	68%	44%

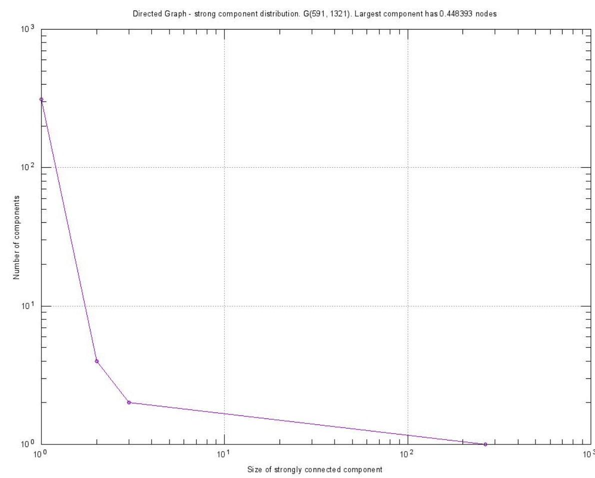
The first takeaway from these measures was that the **during the same time period, fake news network was much larger than the real news network** (terms of total number of *nodes and edges*). This observation is further bolstered by the size of the largest SCC (strongest connected component) size which has 4408 nodes (representing 68% of nodes in fake news network), in case of fake news and only 265 nodes in case of real news (representing 44% of nodes in real news network). Figure 1 represents the distribution of the strongest connected components and we observe that in the case of fake news, there one large SCC and then the size of SCC's becomes negligible. In case of real news, we see a higher number of significantly sized SCC's besides the largest component.

Figure 1: Shortest connected component distribution for friends and followers network for frequent users

Fake News



Real News

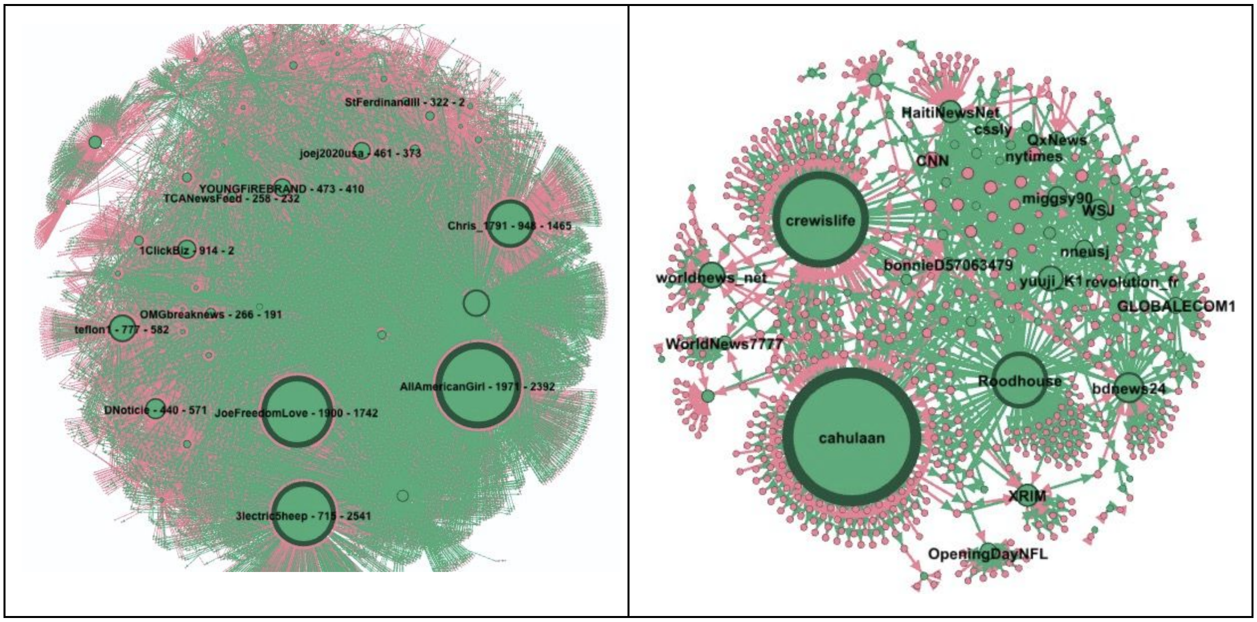


Although, we used a similar methodology for creating the sample data set for both fake and real news and given that, these differences in the size of nodes and edges along with size of single largest component indicates that **fake news circulates more proactively than real news.** Also, the size and number of SCCs indicates that compared to **fake news which has information concentrated in one large cluster, there are several 'clusters' of information groups in case of real news** (as seen in Figure 2). One possible reason for this difference could also be that the total number of sites we covered for fake news were way more than the total number of sites in case of real news. We discuss this further in the limitations of our study. Appendix III and Appendix IV represent the top 100 users in the real and fake news networks respectively.

Figure 2: Friends and Followers Network

The network representation below depicts the friends and follower network of top 100 users who tweet certain "type" (fake or real) news. The friends and followers in these networks also shared the same "type" of news as the top user and are frequent tweeters of the particular "type" of news (at least 100 tweets). Green color of nodes (and edges) represents that the users were the original top 100 users. Red represents that friends and followers of top 100 within the network who are themselves not among the top 100 users.

Fake News Network	Real News Network



We also saw that there is a higher percentage of *bidirectional edges* in fake news network than in real news networks. 60% of the edges are bidirectional indicating **users often are mutual followers in case of fake news network and less so in case of real news network** (46% edges were bidirectional). Following from this data, we also see that there are relatively lesser percentage of nodes that have no in-degree edges in case of fake news (11%), than in case of real news (16%).

As next step, we undertook more detailed analysis (Table 2) of the networks to be able to do a more equitable comparison across the networks.

Table 2: Advanced network measures for better comparison
These network measures helped us get a basic understanding of the composition of the fake and real news networks on twitter.

No.	Measures	Fake News	Real News
1	Average degree	9.1	4.5
2.	Diameter of the largest SCC	7	9
3.	90% effective diameter	3.8	4.94
4.	Closed triangles	14529	27

5.	Open triangles	12714264	38735
6.	Average/Global clustering coefficient	0.26	0.01
7.	Fraction of closed triads	0.001141	0.000697

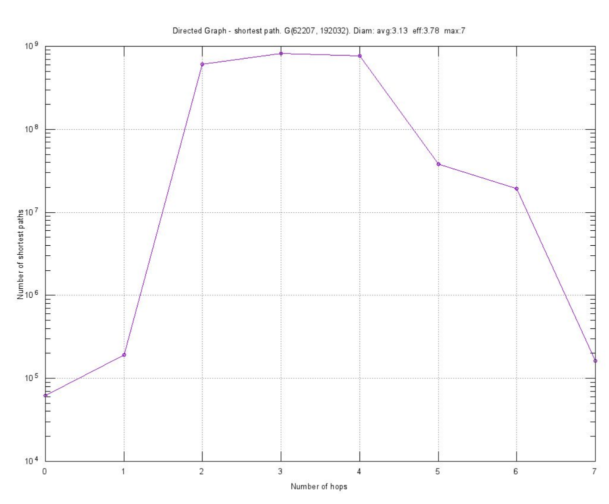
The average degree in case of fake news network (9.1) was almost twice the measure of ral news (4.5). This again points to our earlier observation that the network is more connected and larger in terms of nodes and edges.

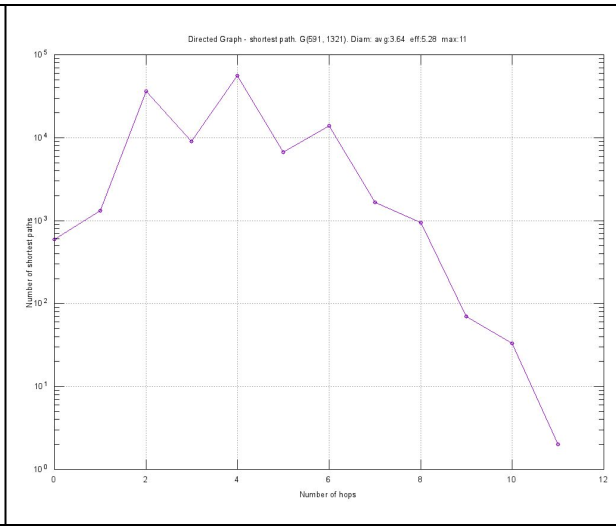
The diameter represents the largest geodesic distance between two nodes in a network and helps us assess the compactness of the network. We see that the diameter in case of the SCC for fake news is 7 and in case of real news is 9. Figure 3 shows the distribution of the shortest path between nodes for fake and real news networks. The two graphs are similar in the sense that majority of distance between the nodes can be covered between two to five hops. But even the distance between nodes in fake news network is smaller than in real new networks. A related metric is 90-percentile effective diameter, "which equals the number of edges needed on average to reach 90% of all other nodes[1]". Again, we see that even for the broader network, beyond SCC, the diameter is smaller for fake news than for real news implying that **there is compactness in the fake news network and more interconnectivity**.

Figure 3: Shortest path for friends and followers network for frequent users

Fake News	Real News

[1] Source: http://konect.uni-koblenz.de/statistics/diameter





The *open and closed triangle measures* gives us information about the clustering coefficient across the two networks, which an indication of the probability that two randomly selected nodes are connected with one another. Standalone, the measures suggests that there are far more closed triangles in the fake news network than in the real news network. This is also evident when we see the network clustering coefficient. These network characteristics (diameter and clustering coefficients) are indicative of the **overall network being more connected in case of fake news versus rela news**. Overall the number of closed triangles seem much smaller that what one would expect from a small world network (which we assume that twitter networks would look like) and this is likely because of the way we sampled our data.

We did further analysis to understand the centrality measures of nodes in the two networks (Table 3). In Table 3.1, we present the nodes with the top degree centrality measures. In case of fake news, this measure varies from 0.208 to 0.670 and in case of real news the range is between 0.08 and 0.529. Simply speaking degree centrality tells us who has more friends. Since the average degree in fake news is higher, there is no surprise that the degree centrality for the top five nodes is higher in case of fake news. We do not however know if this difference is significant or not.

Table 3.1: Centrality Measure - Degree Centrality (Top 5 nodes)
This table contains the information about the nodes/users which the highest degree centrality measure for the two networks.

No.		

	Fake News			Real News		
	Top 5 Users	Degree Centrality	Degree	Top 5 Users	Degree Centrality	Degree
1	AllAmericanGirl	0.670	4363	cahulaan	0.529	312
2	JoeFreedomLove	0.559	3642	crewislife	0.359	212
3	3lectric5heep	0.500	3256	Roodhouse	0.202	119
4	Chris_1791	0.371	2413	bdnews24	0.099	59
5	'teflon1	0.208	1359	worldnews_net	0.08	49

Closeness centrality is the "average length of the shortest paths between a vertex and all vertices in the graph"[2]. In other words, this measure tells us how far a node is from all the other nodes in the network. We see that in the fake news network the range of this centrality measure is from 0.391 tp 0.495. We see the range for the real news network to be between 0.244 and 0.293. For this particular centrality measure, the ranges don't even overlap and closeness centrality for the top five nodes in fake news network is clearly higher than those in the real news network. The is explainable because as we noted in case of other measures, the fake news networks are more compact. The nodes there are mode interconnected and therefore a given node is close to other nodes in the network much more than in real news network which is more spread out with fewer interconnections.

Table 3.2: Centrality Measure - Closeness Centrality (Top 5 nodes)
This table contains the information about the nodes/users which the highest closeness centrality measure for the two networks.

No.	Fake News			Real News		
	Top 5 Users	Closeness Centrality	Degree	Top 5 Users	Closeness Centrality	Degree
1	3lectric5heep	0.495	3256	cahulaan	0.293	312
2	AllAmericanGirl	0.484	4363	Roodhouse	0.273	119

[2] SI608 - Class lecture notes

3	JoeFreedomLove	0.453	3642	crewislife	0.262	212
4	Chris_1791	0.439	2413	JumpTheRework	0.251	7
5	PrisonPlanet	0.391	32	Honey17011	0.244	8

Betweenness centrality helps us understand how many pairs of nodes would have to go through you in order to reach one another through their shortest path. The range for this centrality measure in fake news is 0.067 and 0.241 and in real news is 0.0366 and 0.229. We don't see this difference to as large as closeness centrality perhaps because there is already a lot of interconnectivity in the fake news network so the nodes do not get the extra weightage or points to be the ones "through" which other nodes have to pass to connect with each other.

Table 3.3: Centrality Measure - Betweenness Centrality (Top 5 nodes)
This table contains the information about the nodes/users which the highest betweenness centrality measure for the two networks.

No.	Fake News			Real News		
	Top 5 Users	Betweenness Centrality	Degree	Top 5 Users	Betweenness Centrality	Degree
1	AllAmericanGirl	0.241	4363	cahulaan	0.229	312
2	JoeFreedomLove	0.207	3642	crewislife	0.181	212
3	3lectric5heep	0.174	3256	worldnews_net	0.046	49
4	Chris_1791	0.115	2413	bdnews24	0.044	59
5	JSavoly	0.067	607	XRIM	0.0366	40

Geographic Comparison - User locations for networks

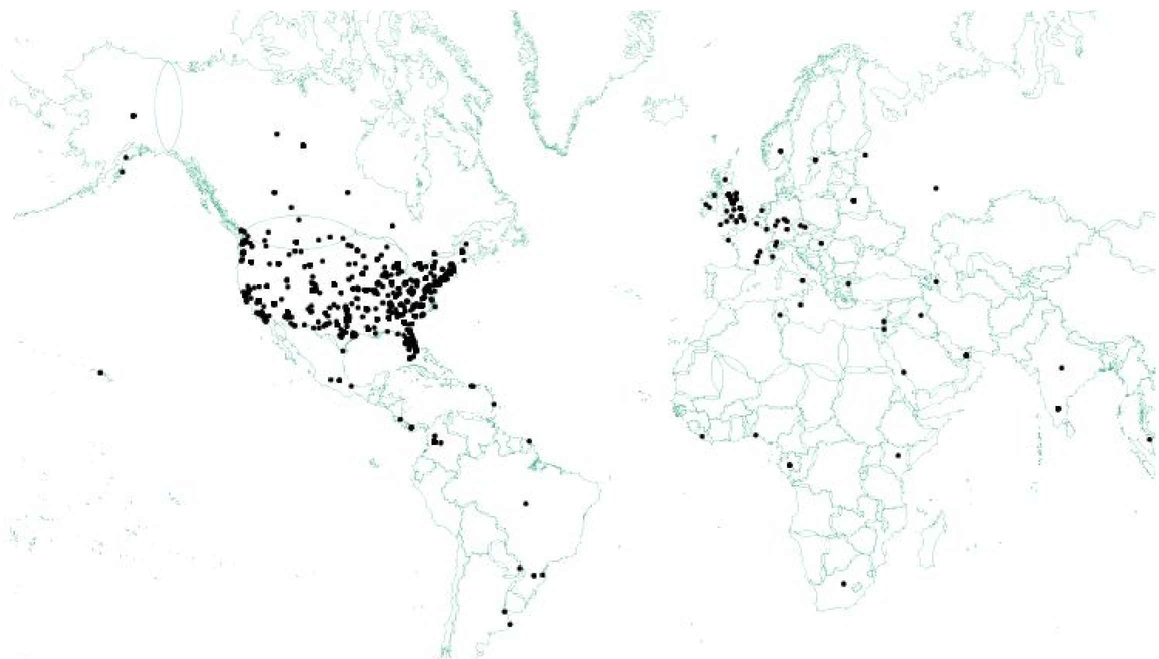
We looked at the user location for each of the networks. Not surprisingly, the concentration of users for both and fake and real news was highest in the USA (51% of the world's twitter users are in the USA, followed by UK which has 17% of Twitter's user base[3]). So, apart from the overall concentration of fake news being

[3] Source: http://www.beevolve.com/twitter-statistics/

higher, we did not see much of difference in the user location of fake and real news. In some ways, we may be seeing that from a geographical standpoint, fake news is being appropriately countered by real news.

Figure 4: User Location (Global)

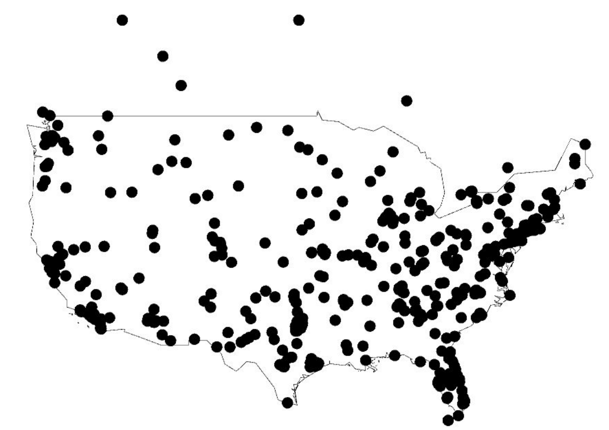
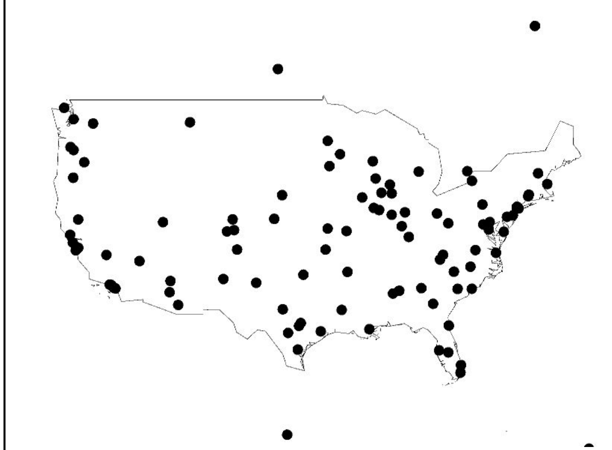
Fake News



Real News

Figure 5 provides a closer look at locations within the US. We provide this visual because even though our data set was location agnostic, a large population of users were located in the US. A visual glance suggests that fake news network is denser and there are visible locations like Florida, certain parts of midwest and east coast, California, Washington, New Mexico, and Arizona where the concentration of fake news is higher than other regions.

Figure 5 : User Location (Within USA)

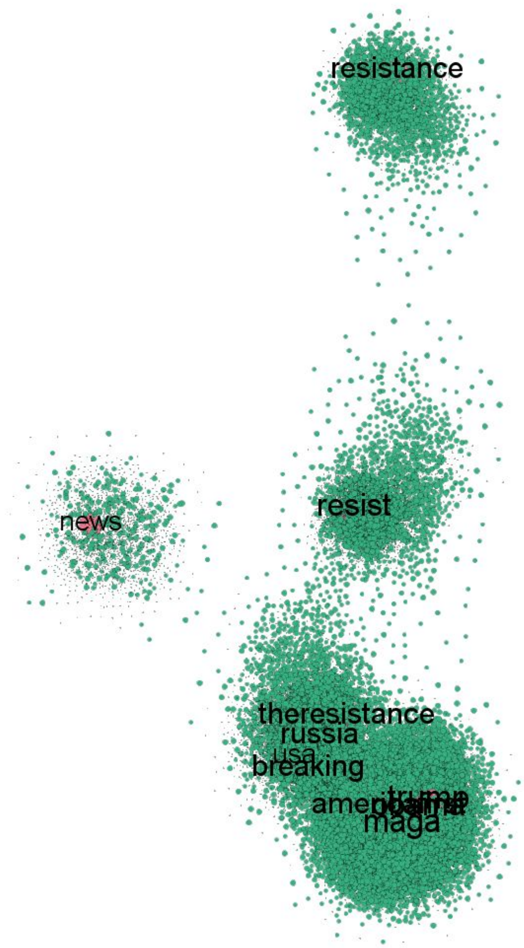
Fake News	Real News
	

Hashtag-based network comparison

To explore the real and fake news network further, we did some linguistic processing to extract the hashtags from the tweets sent by users. Table 6 shows the broad clusters of hashtags.

Figure 6 : Broad hashtags clusters

Fake News	Real News



Top 10 hashtags include:

Hashtags	**Number of tweets**
maga	164986
trump	66526
obama	26240
resist	19988
theresistance	17611
breaking	16453
russia	12658

Top hashtags include:

Hashtags	**Number of tweets**
trumprussia	67547
trump	35526
resist	32362
russiagate	27670
news	23769
aca	22926
theresistance	21069

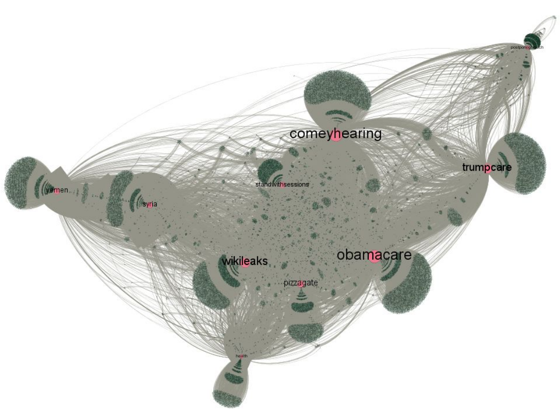
americafirst	16492
resistance	14554
usa	8247

breaking	19495
trumpcare	14117
cnn	13422

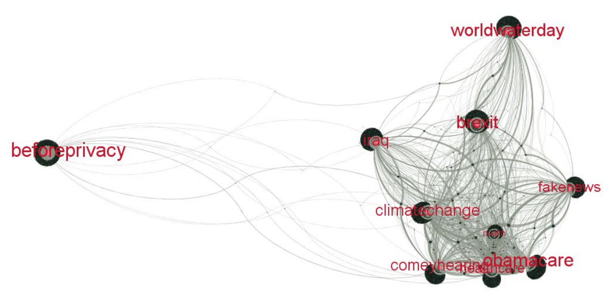
Our initial observation was that these hashtags did not provide any meaningful input. So instead, we did manual processing to create “issue-based” hashtags. By “issue-based” we mean hashtags which point to a particular action being taken. We also built a network of users using these issue based hashtags as seen in Figure 7. The node size is based on degree centrality and each red/black node represents a specific hashtag, whereas the edges represent the users. The network statistics revealed that the fake news network is more interconnected or has higher clustering coefficient (closed triangles). In this specific hashtag based network, this means that it is perhaps more likely for fake news users to tweet more common and shared hashtags than in case of real news networks.

Figure 7 : Issue-based hashtag network

Fake News



Real News



Row Labels	**Number of Tweets**

Hashtags	***Number of Tweets***

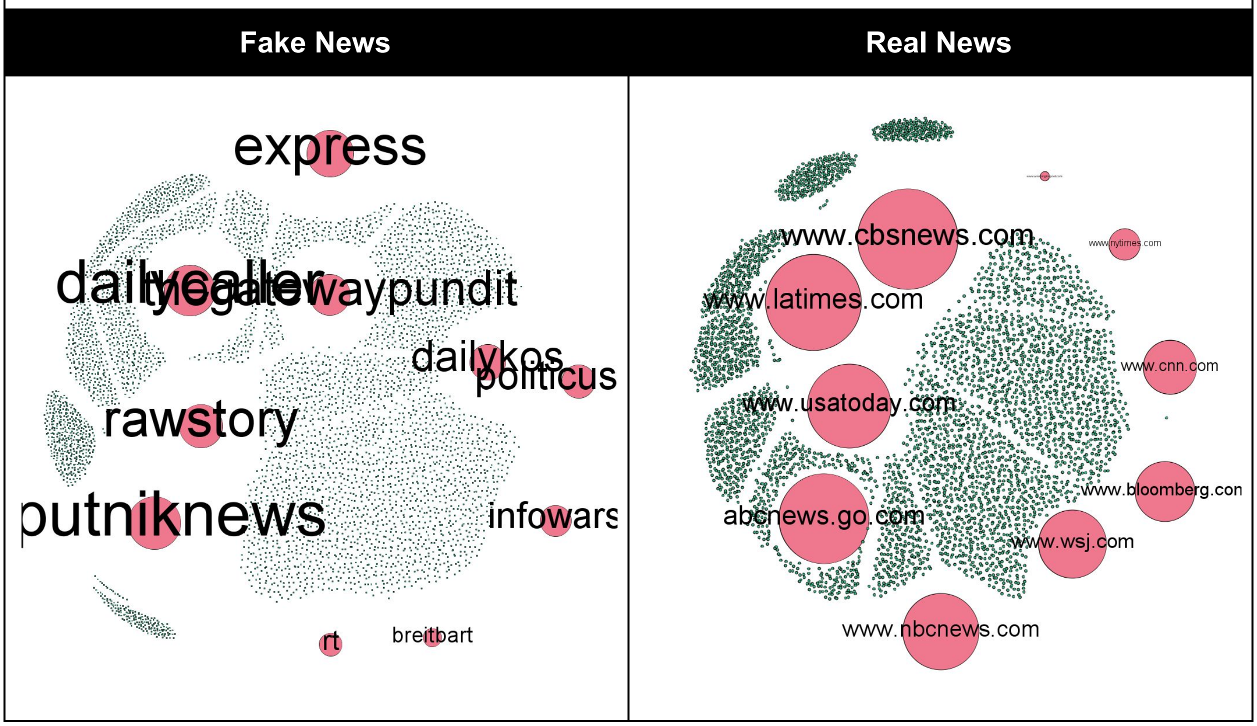
obamacare	15201	brexit	5484
postponegorsuch	13573	obamacare	5050
wikileaks	12489	climatechange	4666
comeyhearing	11557	profitbeforeprivacy	4640
standwithsessions	9684	worldwaterday	4580
pizzagate	9603	healthcare	4187
trumpcare	8730	health	4186
yemen	8700	comeyhearing	3952
health	8528	iraq	3836
syria	8454	fakenews	3567

Network Statistics	**Network Statistics**
Nodes: 34210	Nodes: 40634
Edges: 36682	Edges: 53003
Zero Deg Nodes: 0	Zero Deg Nodes: 0
Zero InDeg Nodes: 12819	Zero InDeg Nodes: 12632
Zero OutDeg Nodes: 20657	Zero OutDeg Nodes: 25439
NonZero In-Out Deg Nodes: 734	NonZero In-Out Deg Nodes: 2563
Unique directed edges: 36682	Unique directed edges: 53003
Unique undirected edges: 36682	Unique undirected edges: 53003
Self Edges: 0	Self Edges: 0
BiDir Edges: 0	BiDir Edges: 0
Closed triangles: 400	Closed triangles: 0
Open triangles: 69915181	Open triangles: 166833366
Frac. of closed triads: 0.000006	Frac. of closed triads: 0.000000
Connected component size: 1.000000	Connected component size: 1.000000
Strong conn. comp. size: 0.000029	Strong conn. comp. size: 0.000025
Approx. full diameter: 6	Approx. full diameter: 6
90% effective diameter: 3.938799	90% effective diameter: 3.881292

Network of websites mentioned

To understand the role of the websites in the fake news network, we built the network around top ten websites (based on the number of tweets) that were most frequently referred to by users in the tweets. Each red node represents a particular website and the size of the node represents degree centrality. We observe that in case of fake news the size of the nodes is roughly similar, except until we see the bottom few nodes which include sites like "Breitbart", which are arguably more hyper-partisan than fake news. This might indicate the affinity of the fake news users to tweet more news emanating from 'purely' fake news sites and not just partisan news sites.

Figure 8 : Website based network



Conclusion and Discussion

The vast amount of information we collated helped us understand fake news and real news from different dimensions - friends and followers network, geographic origin of users, hashtags, and top websites. For future work, we would try to address some of the limitations of our study. First, we needed a more scientific method of arriving at the list of sites considered to be fake or real. The imbalance of number of sites in the fake news versus real news perhaps contributed to use gathering 2 million less tweets in real news versus fake news. One method to do that would be to identify fake news sites which have been vetted by more than one expert or academician to avoid any bias. Another method would be to lengthen the list of real news sites to make the total number of sites in each group comparable. Second, for future work we suggest collecting data for different time intervals and conduct significance testing for various metrics to determine whether our conclusions about fake and real news sites is indeed meaningful or not. Third, we limited the dataset to enable easier computing of graphs. In this process, we perhaps missed some nuanced of the network. For future work we suggest looking at the data in totality.

References

1. Et Al, J. (2013). Epidemiological modeling of news and rumors on Twitter. Retrieved from http://people.cs.vt.edu/naren/papers/news-rumor-epi-snakdd13.pdf
2. Romero, Meeder, & Kleinberg. (2011). Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter. Retrieved from http://dromero.org.s3-website-us-east-1.amazonaws.com/info_diffusion_topics.pdf
3. Taxidou, & Fischer. (2014). Online Analysis of Information Diffusion in Twitter. Retrieved from https://websci.informatik.uni-freiburg.de/publications/LSNA2014-diffusion.pdf
4. Steinkirch. (2012). Information Diffusion in Twitter. Retrieved from http://astro.sunysb.edu/steinkirch/reviews/twitter.pdf
5. Rogers-Pettie, & Herrmann. (2015). Information Diffusion: A Study of Twitter during Large Scale Events. Retrieved from http://www.isr.umd.edu/~jwh2/papers/Pettie-Herrmann-ISERC-2015.pdf

Appendices

Appendix I - List of Real News Sites

SOURCE (DOMAIN),OWNERSHIP

1. www.nytimes.com
2. www.wsj.com
3. www.washingtonpost.com
4. www.bbc.com
5. www.economist.com
6. www.newyorker.com
7. www.ap.org
8. www.reuters.com
9. www.foreignaffairs.com
10. www.bloomberg.com
11. time.com
12. www.csmonitor.com
13. www.latimes.com
14. www.usatoday.com
15. www.cnn.com
16. www.nbcnews.com
17. www.cbsnews.com
18. abcnews.go.com
19. www.nationalreview.com
20. www.weeklystandard.com
21. newrepublic.com
22. www.thenation.com

Appendix II - List of Fake news sites (Partial)

Source	Listed by Fake News Watch	Listed by Melissa Zimdars	Listed by The Daily Dot	Listed by U.S. News and World Report	Listed by New Republic	Listed by CBS News	Listed by about.com	Listed by NPR	Listed by Snopes Field Guide
21stcenturywire.com	Yes	Yes	Yes	No	No	No	No	No	No
70news.wordpress.com	No	Yes	Yes	No	No	Yes	No	No	No
abcnews.com.co	No	No	Yes	No	No	Yes	No	No	No
activistpost.com	Yes	Yes	Yes	Yes	No	No	No	No	No
addictinginfo.org	No	Yes	Yes	No	No	No	No	No	No
americannews.com	Yes	Yes	Yes	Yes	No	No	No	No	No
americannewsx.com	No	Yes	No	No	No	No	No	No	No
amplifyingglass.com	Yes	No	No	No	No	No	No	No	No
anonews.co	No	No	Yes	No	No	No	No	No	No
beforeitsnews.com	Yes	Yes	No	Yes	No	No	No	No	No
bigamericannews.com	Yes	Yes	No	No	No	No	No	No	No
bipartisanreport.com	No	Yes	Yes	No	No	No	No	No	No
bluenationreview.com	No	Yes	Yes	No	No	No	No	No	No
breitbart.com	No	Yes	Yes	No	No	No	No	No	No
burrardstreetjournal.com	No	No	No	No	No	Yes	No	No	No
callthecops.net	No	No	Yes	No	No	No	Yes	No	No
christiantimes.com	No	No	No	No	No	Yes	No	No	No
christwire.org	Yes	Yes	Yes	No	No	No	No	No	No
chronicle.su	Yes	Yes	No	No	No	No	No	No	No
civictribune.com	Yes	Yes	Yes	No	No	Yes	No	No	No
clickhole.com	Yes	Yes	Yes	Yes	No	No	No	No	No
coasttocoastam.com	Yes	Yes	Yes	No	No	No	No	No	No
collective-evolution.com	No	No	Yes	No	No	No	No	No	No
consciouslifenews.com	Yes	Yes	Yes	No	No	No	No	No	No
conservativeoutfitters.com	Yes	Yes	Yes	No	No	No	No	No	No
countdowntozerotime.com	Yes	Yes	Yes	No	No	No	No	No	No
counterpsyops.com	Yes	Yes	No	No	No	No	No	No	No
creambmp.com	Yes	Yes	Yes	No	No	No	No	No	No
dailybuzzlive.com	Yes	Yes	No	Yes	No	No	No	No	No
dailycurrant.com	Yes	Yes	No	No	No	No	Yes	No	No
dailynewsbin.com	No	Yes	No	No	No	No	No	No	No
dcclothesline.com	Yes	Yes	No	No	No	No	No	No	No

demyx.com	No	No	No	No	Yes	No	No	No	No
denverguardian.com	No	No	No	No	No	No	No	Yes	No
derfmagazine.com	Yes	Yes	No	No	No	No	No	No	No
disclose.tv	Yes	Yes	No	Yes	No	No	No	No	No
duffelblog.com	Yes	Yes	Yes	Yes	No	No	No	No	No
duhprogressive.com	Yes	Yes	No	No	No	No	No	No	No
empireherald.com	No	Yes	No	No	No	Yes	No	No	No
empirenews.net	Yes	Yes	Yes	No	No	Yes	Yes	No	Yes
empiresports.co	Yes	No	No	No	Yes	No	Yes	No	Yes
en.mediamass.net	Yes	Yes	Yes	No	Yes	No	Yes	No	No
endingthefed.com	No	Yes	No	No	No	No	No	No	No
enduringvision.com	Yes	Yes	Yes	No	No	No	No	No	No
flyheight.com	No	Yes	No	No	No	No	No	No	No
fprnradio.com	Yes	Yes	No	No	No	No	No	No	No
freewoodpost.com	No	No	No	No	No	No	Yes	No	No
geoengineeringwatch.org	Yes	Yes	No	No	No	No	No	No	No
globalassociatednews.com	No	No	No	No	Yes	No	Yes	No	No
globalresearch.ca	Yes	Yes	No	No	No	No	No	No	No
gomerblog.com	Yes	No	No	No	No	No	No	No	No
govtslaves.info	Yes	Yes	No	No	No	No	No	No	No
gulagbound.com	Yes	Yes	No	No	No	No	No	No	No
hangthebankers.com	Yes	Yes	No	No	No	No	No	No	No
humansarefree.com	Yes	Yes	No	No	No	No	No	No	No
huzlers.com	Yes	Yes	No	No	Yes	Yes	Yes	No	Yes
ifyouonlynews.com	No	Yes	No	No	No	Yes	No	No	No
infowars.com	Yes	Yes	Yes	Yes	No	Yes	No	No	No
intellihub.com	Yes	Yes	No	No	No	No	No	No	No
itaglive.com	Yes	No	No	No	No	No	No	No	No
jonesreport.com	Yes	Yes	No	No	No	No	No	No	No
lewrockwell.com	Yes	Yes	No	No	No	No	No	No	No
liberalamerica.org	No	Yes	No	No	No	No	No	No	No
libertymovementradio.com	Yes	Yes	No	No	No	No	No	No	No
libertytalk.fm	Yes	Yes	No	No	No	No	No	No	No
libertyvideos.org	Yes	Yes	No	No	No	No	No	No	No
lightlybraisedturnip.com	No	No	No	No	Yes	No	No	No	No
nationalreport.net	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
naturalnews.com	Yes	Yes	Yes	Yes	No	No	No	No	No
ncscooper.com	No	Yes	No	No	No	No	No	No	Yes
newsbiscuit.com	Yes	Yes	Yes	No	No	No	No	No	No
newsexaminer.com	No	Yes	No	No	No	No	No	No	Yes
newslo.com	Yes	Yes	Yes	Yes	No	No	No	No	No

newsmutiny.com	Yes	Yes	Yes	No	No	No	No	No	No
newswire-24.com	Yes	Yes	No	No	No	No	No	No	No
nodisinfo.com	Yes	Yes	No	No	No	No	No	No	No
now8news.com	No	Yes	No	No	No	Yes	No	No	Yes
nowtheendbegins.com	Yes	Yes	No	No	No	No	No	No	No
occupydemocrats.com	No	Yes	Yes	No	No	No	No	No	No
other98.com	No	Yes	Yes	No	No	No	No	No	No
pakalertpress.com	Yes	Yes	No	No	No	No	No	No	No
politicalblindspot.com	Yes	Yes	No	No	No	No	No	No	No
politicalears.com	Yes	Yes	No	No	No	No	No	No	No
politicops.com	Yes	Yes	No	No	No	Yes	No	No	No
politicususa.com	No	Yes	No	No	No	No	No	No	No
prisonplanet.com	Yes	Yes	No	No	No	No	No	No	No
private-eye.co.uk	Yes	Yes	Yes	No	No	No	No	No	No
react365.com	No	Yes	No	No	No	Yes	No	No	Yes
realfarmacy.com	Yes	Yes	No	No	No	No	No	No	No
realnewsrightnow.com	Yes	Yes	Yes	No	No	Yes	No	No	No
redflagnews.com	Yes	Yes	No	Yes	No	No	No	No	No
redstate.com	No	Yes	Yes	No	No	No	No	No	No
rilenews.com	Yes	Yes	Yes	No	No	Yes	No	No	No
rockcitytimes.com	Yes	No	No	No	No	No	No	No	No
satiratribune.com	No	Yes	No	No	No	No	No	No	Yes
stuppid.com	No	No	No	No	No	No	No	No	Yes
theblaze.com	No	Yes	No	No	No	No	No	No	No
thebostontribune.com	No	No	No	No	No	Yes	No	No	No
thedailysheeple.com	Yes	Yes	No	No	No	No	No	No	No
thedcgazette.com	Yes	Yes	Yes	Yes	No	Yes	No	No	No
thefreethoughtproject.com	No	Yes	Yes	No	No	No	No	No	No
thelapine.ca	Yes	No	No	No	No	No	Yes	No	No
thenewsnerd.com	Yes	Yes	No	No	Yes	No	No	No	No
theonion.com	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
theracketreport.com	No	No	No	No	No	No	Yes	No	No
therundownlive.com	Yes	Yes	No	No	No	No	No	No	No
thespoof.com	Yes	No	No	No	No	No	Yes	No	No
theuspatriot.com	Yes	Yes	No	No	No	No	No	No	No
truthfrequencyradio.com	Yes	Yes	No	No	No	No	No	No	No
twitchy.com	No	No	Yes	No	No	No	No	No	No
unconfirmedsources.com	Yes	Yes	No	No	No	No	No	No	No
USAToday.com.co	No	No	No	No	No	No	No	Yes	Yes
usuncut.com	No	Yes	Yes	No	No	No	No	No	No
veteranstoday.com	Yes	Yes	No	No	No	No	No	No	No

wakingupwisconsin.com	Yes	Yes	No	No	No	No	No	No	No
weeklyworldnews.com	Yes	No	No	Yes	No	No	Yes	No	No
wideawakeamerica.com	Yes	No	No	No	No	No	No	No	No
winningdemocrats.com	No	Yes	No	No	No	No	No	No	No
witscience.org	Yes	Yes	No	No	No	No	No	No	No
wnd.com	No	Yes	No	No	No	No	No	No	No
worldnewsdailyreport.com	Yes	Yes	Yes	No	No	No	Yes	No	Yes
worldtruth.tv	Yes	Yes	No	Yes	No	No	No	No	No
yournewswire.com	No	No	No	No	No	Yes	No	No	No

Appendix III - Top 100 users (tweeters) of Real News Sites, total tweets

CollectedN,4420
farhaadaarif,4297
XRIM,3855
MyDailyReporter,3836
mlnangalama,3453
Uskeus,3227
serge_poznanski,3144
Nano_coin,2960
nneusj,2930
worldnews_net,2874
InvestingLatest,2867
rashidaldosari,2768
bdnews24,2679
thus_spake,2668
Real_Infinity95,2615
business,2580
HotChkNewsTweet,2565
SilberWorldNews,2496
swissifg,2395
shailendra_nair,2392
break_xxi,2347
livechannelfeed,2257
BSAFunds,2129
libertad717,2106
NewsWhileFree,2101
Tofbalzy1,2086
usanews2017,2045
NarrendraM,1997
MarcoAn56507103,1997
mshusa,1967
HaitiNewsNet,1957
reek6,1953
iAmitKushwaha,1873
TheBFRoom,1815
sarakwon_,1789
zoe_nor,1789
WilliamMungall,1724
revolution_fr,1690
TimerWashington,1687
julyrose299,1668
Gamechanger22,1664
Angelus1701,1660
yuuji_K1,1638
GLOBALECOM1,1609
Apocrifos,1572
leoretweetbot,1567

news__eeuu,1538
newsly,1520
USA_Update_,1510
abdullah995211,1467
ALLinOne_info,1448
vrai777,1389
crewislife,1310
budrowb3,1304
KeyNoticias,1295
thetopnetworker,1294
TDailyNews4,1291
venyenteratehoy,1281
JasmineLliki,1277
WorldwideNewsTu,1264
latimesfirehose,1252
videoviralUS,1242
Politics4All,1233
alllibertynews,1217
hikakO_o,1210
csmonitor,1209
TidyStucco_com,1202
mshcnn,1201
QkTipcom,1199
zgorlami,1182
riskinfo,1179
sandyc1772,1171
fypjiyeon,1164
lookjardin,1158
mandazfr,1158
dronejava1,1155
dronejava,1151
dronejava2,1149
HappHopper,1147
NewsInMotion_,1144
YourNewsTweet,1144
NYT,1132
WorldNews7777,1124
iSupportInt,1115
bonnieD57063479,1108
SWS_EDU,1105
QxNews,1098
cssly,1082
miggsy90,1082
mix_your_music,1078
jamesgreid,1072
OpeningDayNFL,1068
SlowestPoison,1068
GlobalReporte,1066

spacefeedtweets,1051
cahulaan,1047
CKKaufmannServi,1043
WSJ,1041
Roodhouse,1036
WPolitics_,1035

Appendix IV - Top 100 users of fake news sites, total tweets

freedomforthwin
9032
JoeFreedomLove
3445
YOUNGFiREBRAND
3103
TCANewsFeed
2981
Ascension_Guide
2836
newstome72
2802
WorldTruthTV
2610
hamel1776
2602
AllAmericanGirl
2436
NM99791307
2427
shrimp_shrimpy
2425
anonymous5595
2422
inkme211
2412
Chris_1791
2403
RealDead67
2330
EagleStarNET
2330
qkode
2280
beforeitsnews
2256
syqau
2227

npnikk
2169
yceek
2150
nuiotwo
2117
TroyCoby
2115
vnuek
2083
NoahJamesBangs
2056
SCroixFreePress
1931
integra_66
1843
Figue_j
1820
news_liveworld
1791
DemocracyMotion
1787
gqforbes4
1774
RayWarnerShow
1706
breitbot
1562
altnewsheadline
1525
teflon1
1486
Col_Connaughton
1457
YodaCon
1429
Deplorable69er
1421
DumpMSM
1416
RightnewsNews
1402
TEEITHIGH
1393
Roses_4_Thorns
1393
OMGbreaknews
1385

NozNewz_com
1378
Alonsofg
1364
SpyServe
1361
DNoticie
1361
The_Reporter24
1360
TRobinsonNew
1360
conservamother
1358
ViolatedBrit
1358
RealJamesKist
1358
CarlBullock16
1358
manager_politic
1358
TRobinsonNews
1357
TNewViewer
1356
USAGOP
1354
Trump_Force1
1351
Jerichomarch
1350
PropOrNotApp
1348
WITIWYG
1347
1964Alvaradosky
1335
HWDRepublican
1335
VMastery
1316
Alt_Right_
1311
thefuckingnews1
1300
B4INMarkets
1299

thefuckingnews2
1296
rfairexperience
1294
RealFKNNews
1281
thefuckingnews3
1263
crewislife
1263
StFerdinandIII
1261
INCREDIBLEsnews
1258
1ClickBiz
1232
Dpoliticmanager
1217
3lectric5heep
1182
mzee26
1160
joej2020usa
1160
OccuWorld
1144
JSavoly
1136
josewhales
1104
ScribbledPages1
1098
B4INHealthcare
1068
GrovelandJohn
1051
jimmypedya
1037
SavageNation
1018
nanopatents
1017
dailyrapid
1014
PhilWaton
1009
Retired_Actor
1007

OurTroubledTime
996
RebootBill
996
RightOfCenterNC
987
B4INSurvival
978
SOTCJTF
974
FrankMa24398057
968
TrutherForever
964
Joe_America1776
948
RonsNewsFeed
946