Public Opinion Warfare: Understanding Communities on Twitter and Their Influence on the Web

Yueshuwei Wu[1], **Kaela Hasan**[1], **Bing Lin**[1], **and Yuxin Qian**[1]

[a]Department of Mathematics, University of California, Los Angeles, CA, 90024

This manuscript was compiled on June 13, 2020

With the increasing use and impact of social media, the network structure of such platforms poses a unique opportunity to study the makeup of the information ecosystem, and how information travels. In this paper, we explore this topic by analyzing a retweet network composed of 105K Twitter users where 5K users are marked as "hateful" or "normal" users by the data provider. We compared these marked users to the network's remaining set of users, finding compelling differences in terms of their general behavior and the content they post. Using network visualizations and the Louvain method for community detection, we detected possible communities' structures and quantified the influence that specific communities had on Twitter. Overall, our findings indicate that hateful users are more connected and tend to be more efficient at spreading information. Conservative users, regardless of label, have a relatively looser retweeting connection with each other but still have a high impact on their retweeters, which could indicate that in addition to hateful users, conservative users are very influential and efficient in spreading information.

Twitter | Social Network | Community Behavior | Network Structure

Social media is becoming especially abundant today, leading to the formation of increasingly complex social networks. These networks can be used to study and model a variety of topics. Although we have seen various communities on social media, there is little understanding of how these communities behave, what type of content they share, and, most importantly, their influence on the information ecosystem.

Overall, many aspects of the structure of online communities remain unclear, e.g., how do some specific communities operate? Is it possible to quantify their influence? In this paper, we aim to address these questions by relying on the set of 105K accounts released by Manoel Ribeiro as a dataset on Kaggle.(1) We first characterized the activity of different kinds of users by comparing their statistical results. Then we investigated the structure and behavior of different communities. Finally, we quantify the influence of specific dominant communities by examining their retweeters in the network.

Main findings.Our study leads to several key observations:

1. The "hateful" users in this retweet network are more connected and efficient at spreading information.
2. Hateful users have higher “loyalty” as compared to normal users but have a lower tendency to add hashtags to their posts.
3. Conservative users play important roles in the retweet network, regardless of whether they are labeled as normal or hateful.
4. In terms of conservative users, a high impact on retweeters resulted from extreme cases of connection.

Background

In this section, we provide a brief overview of the dataset used and the social network that is dissected in this paper.

Twitter. Twitter is a social media platform based on the sharing of short posts. It serves as a way to facilitate discussion and share information and opinions with millions of users. Users can post and interact with short messages known as "tweets" that other users can react to by *liking* the post or by further sharing the information by *retweeting*. Users can also *follow* other users to keep up-to-date with any information they share. Tweets may contain *hashtags*, which categorize tweets by keyword, and *mentions*, which refer to other users on Twitter.

Dataset. Since Twitter has banned crawling users' data, our retweet network was obtained from Kaggle. We started from the dataset of 105K Twitter users posted by Manoel Ribeiro on Kaggle. This data does not include user identification or screen names. Instead, the users were assigned unique number identification.

The dataset contains a directed retweet network constructed of 544 users who are labeled as “hateful”, 4.4k “normal” users, and 100K “other” users whose tendency is not clear. For each user, several content-related, network-related, and activity-related features were also provided.

Note that the criteria used by Ribeiro to identify these troll accounts are not public. It is also worth mentioning that this information only accounts for a small portion of the Twitter network. Consequently, we have no control over the data and do not have access to the actual network so that biases could occur.

Significance Statement

This study suggests that conservative Twitter users are very influential and efficient in spreading information and opinions in general. Highly connected hateful conservative users and relatively loosely connected normal conservative users are two extreme cases in community subgraph density, but both are very efficient in spreading information and opinions. The results suggest a quadratic relationship between the community's subgraph density and its influences.

Statistical Analysis

Methods. In this section, we present an in-depth analysis of the activities and the behavior of hateful, normal, and other users. To determine if there are any statistically significant differences between the behavior of different types of users, we decided to examine the relationship between user count (referring to the total 105K users) and several activity-related parameters. We always studied normalized characteristics to account for the different sizes of subgroups.

General Characteristics.

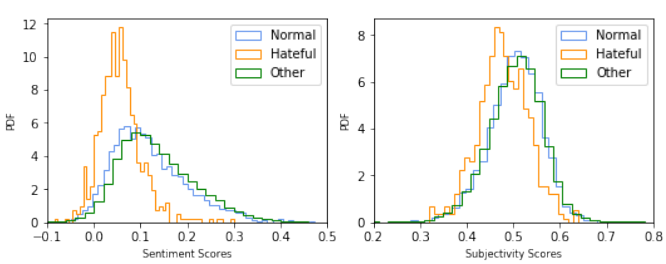
Likes Received From Each Follower. As most real-world social networks, the normalized distribution of followers count, followees count, and received like count all followed the power law. To estimate the influence of a user, we calculated the average of likes received from one follower for all three groups. We were aware that non-followers could like a tweet as well, but in this bold estimation, we assume followers contributed all the likes.

We observed that the hateful users received more likes from one follower on average. For hateful users, the average of likes received from one follower is 16.8. In contrast, normal and other users have much lower average ratios, 10.5 and 6.8, respectively, which indicates that the followers of hateful users exhibit higher "loyalty."

Mentions. We found that each hateful user mentioned ∼230 users on average in this dataset. While the other two groups mentioned ∼150 users on average, which is 42.1% lower. Depending on whom they are mentioning, this behavior may reflect a strategy of manipulating users' opinions regarding a particular topic.

Content Analysis.

Sentiment Analysis. Next, we assessed the sentiment and subjectivity of each user for all three groups. Fig.1 reports the probability distribution functions (PDF) of the sentiment and subjectivity scores of tweets posted by different user groups.



(a) (b)

Fig. 1. PDF of the score of (a) Sentiment and (b) Subjectivity.

The sentiment scores for all three groups follow a normal distribution with a positive skew. The mean of the hateful group, 0.056, is very close to the neutral point 0. In comparison, the averages of the other two groups, being 0.118 and 0.135 respectively, are more positive. Also, we observed that 8% of the hateful users received a negative sentiment score while the other two groups exhibit 0.1% of negative score users. Overall, we observed that hateful users tend to be more negative/neutral, while normal and other users are more positive/friendly. Interestingly, all three groups have a similar average subjectivity score of around 0.5.

Hashtags. Our next step is to study the use of hashtags by these users. Although we did not have access to the content sent by users, we were able to conduct content analysis by looking at the hashtags used. On average, the hateful users use significantly fewer hashtags (∼48) than normal (∼85) and other users (∼92) in the collected data. This implies that hateful users have a lower tendency of adding hashtags in their posts.

Hateful		Normal		Others	
Hashtag	(%)	Hashtag	(%)	Hashtag	(%)
MAGA	1.93%	MAGA	0.35%	ad	0.37%
Syria	1.03%	Syria	0.34%	MAGA	0.28%
WhiteLivesMatter	0.87%	ISIS	0.27%	Halloween	0.19%
JFKFiles	0.79%	Brexit	0.24%	WorldSeries	0.17%
agchatoz	0.75%	ad	0.23%	1	0.16%
AltRight	0.70%	Pakistan	0.22%	Brexit	0.16%
Iraq	0.70%	Somalia	0.21%	Pakistan	0.16%
RedPill	0.64%	WorldSeries	0.21%	Trump	0.13%
ma4t	0.64%	BREAKING	0.19%	PuertoRico	0.13%
Brexit	0.59%	Iran	0.18%	auspol	0.13%
Trump	0.54%	Iraq	0.18%	BREAKING	0.12%
Goldwater	0.52%	Rohingya	0.18%	MeToo	0.12%
UraniumOne	0.49%	Halloween	0.18%	AI	0.11%
SpencerAtUF	0.48%	Trump	0.17%	cdnpoli	0.11%
WhiteGenocide	0.46%	1	0.16%	PPP	0.11%

Table 1. Top 15 hashtags in tweets from three user groups.

In Table.1, we report the top 15 hashtags for all three groups. As seen in the table, while all three groups shared some conservative political topics, e.g., #MAGA, #Brexit, the hateful group embodied a lack of social life-related tags, e.g., #Halloween, #WorldSeries. We also observed that the hashtags were more diluted among normal and other users who exhibit less politically active behavior.

At this point, since the other statistical results of different groups are not particularly distinguishable from each other, we were unable to discern more differences between subgroups based solely on other statistical values. We decided to transition into network analysis to learn more about the structure of the network.

Network Analysis Methods

Centrality. Centrality measures are critical in network analysis because they help to define the importance of a given node. Based on the organization of Twitter, we chose to employ PageRank and the HITS Algorithm to examine the importance of particular nodes and communities.

Authority and Hub Centrality and the HITS Algorithm. Kleinberg(2) introduced the concept of hubs and authorities in networks and applied their notion to his centrality algorithm called hyperlink-induced topic search or HITS. The HITS algorithm assigns two different centrality scores to each node i: the authority centrality x_i and the hub centrality y_i. A node with high authority centrality is pointed to by many nodes with high hub centrality. Conversely, a node with high hub centrality points to many nodes with high authority centrality.

By definition, the authority centrality of a node is proportional to the sum of the hub centralities of the nodes that point to it:

$$x_i = \alpha \sum_j A_{ij} y_j$$

where α is a constant.

Similarly, the hub centrality of a node is proportional to the sum of the authority centralities of the nodes it points to:

$$y_i = \beta \sum_j A_{ji} x_j$$

where β is a constant separate from α.

PageRank Centrality. PageRank centrality(3) is often employed in calculating the importance of a citation or website because it dilutes centrality based on the number of neighbors a given node has. For example, Google uses PageRank as an essential part of listing the more important website at the beginning of the results, with the importance of a given site related to its links to other important pages. In PageRank, the centrality given by the association to its neighbors is inversely proportional to the number of neighbors a node has, which can help to normalize centrality based on the out-degree of a node. The formula for PageRank is defined as:

$$x_i = \alpha \sum_j A_{ij} \frac{x_j}{k_j^{out}} + \beta$$

The Louvain method for community detection. In order to detect and analyze communities, the network should be partitioned into communities of densely connected nodes, with only a sparse connection between nodes of different communities.

To further examine the structure and connectivity of the network, we employed an algorithm created by Blondel(4). The algorithm splits the nodes in a network into separate communities such that the modularity score of the network is maximized within the range between [-0.5, 1]. The higher the final score, the more reasonable the community detection. Recall, modularity measures the extent to which like nodes are connected in a network and is calculated by:

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) + \delta_{g_i g_j}$$

Network Data Trimming Method. The original retweet network includes 105K users and 2.29M edges. Due to the large size of the dataset and limited access to computational power, we decided to trim the network in order to enable the analysis and visualization of the network.

Since we do not have access to the original tweets, we cannot track the spread of opinions directly. To estimate the influence of users that we are interested in, we assumed the effectiveness of a user's opinion spread is proportional to the number of its direct retweeters. Therefore, we trim the network by removing the majority of the "other" users from the original network, leaving only the selected users and their direct retweeters.

Network Analysis

In this section, we present an in-depth analysis of the structure of this retweet network.

The HN-Network. We began our network analysis by constructing a network of the normal and hateful users and their retweeters, which we called the HN-network. By using the network trimming method mentioned in the previous section. This resulted in a trimmed network with ∼35K nodes, ∼124K edges.

To help us visualize the network, we further omitted the retweeters (dangling nodes) by restricting the range of in-degrees to be higher than one. This left us with 4,646 nodes and 9,575 edges with a network density close to 0. This small density value indicates that the HN-network is sparsely connected.

We then applied the Louvain method for community detection with a standard resolution value of 1, resulting in a modularity value of 0.644, allowing us to construct reliable communities of users. It is worth mentioning that this algorithm often overestimated the size of communities. To aid in visualization, we applied the ForceAtlas2 algorithm for the network layout and located five large communities as seen in Fig.2.

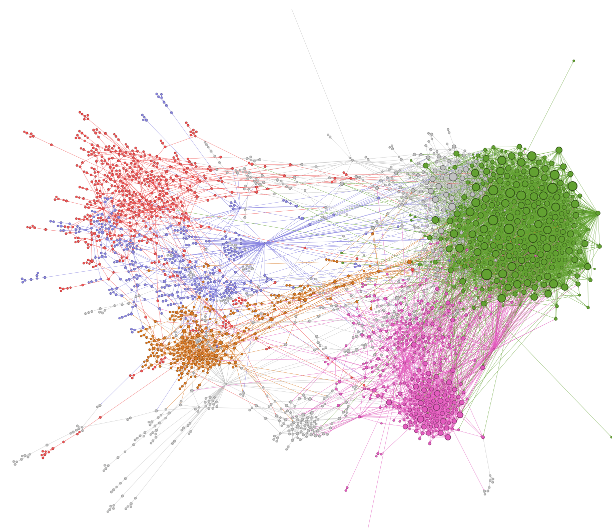


Fig. 2. ForceAtlas2 layout of the HN-Network. Five largest communities were colored.

To better visualize the interaction between communities, we applied a circular layout that positioned nodes clockwise according to their modularity classes. We then measured their authority, hub, and PageRank centralities. The result plots are shown in Fig.3.

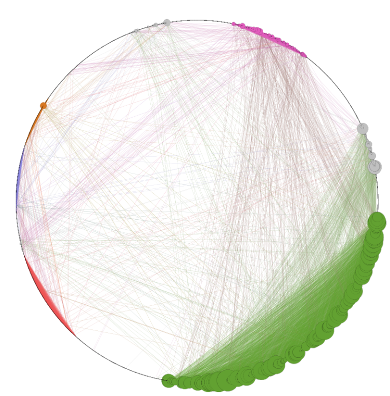


Fig. 3. HN-Network Community Interaction-Authority Rank Plot.

The results for all three measures of the centralities indicated that the majority of the most important users resided in the green community. [Note: If not specified otherwise, the larger the size of a node, the larger the centrality score of the node. Other centrality measures can be find in the Supporting Information Appendix (SI).] Also, we observed the interactions between the green community and two other smaller communities.

We then calculated that the green community has a subgraph density of 0.037, while the other communities are not nearly as connected. Although the green community only accounts for around 8% of the total nodes, it generates more than 30% of the network's interactions.

To further study the green community, we repositioned the users in the circular layout according to their label of hate, beginning with the hateful users.

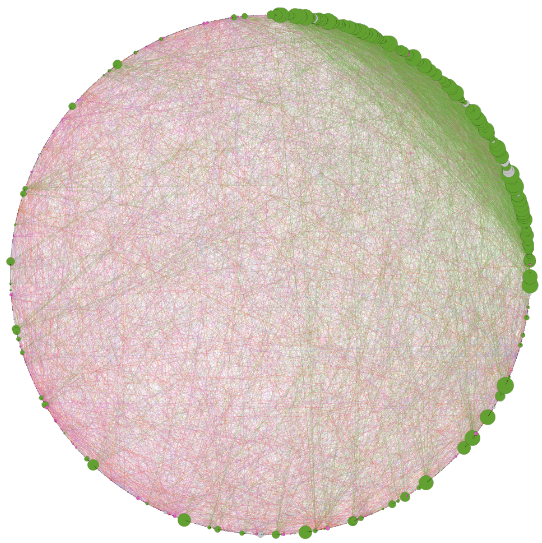


Fig. 4. HN-Network User Interaction-Authority Rank Plot. With nodes positioned by labels.

As seen in Fig.4, due to the nature of our design layout and the fact that the beginning of the circumference is heavily occupied by green nodes, we can conclude that the green community is composed of a majority of hateful users. This led us to hypothesize that hateful users possessed heavy influence within the retweet network, particularly in the hateful user community.

To further explore the influence of hateful users and their contributions to community interaction, we removed all hateful users. This led to a much sparser network, as seen in Fig.5. This extreme shift in network structure indicates that hateful users contribute significant interaction within the HNnetwork.

The H-Network. To further study the network of hateful users, we used the same trimming method to create a subgraph only containing hateful users and their retweeters. We call this the H-network.

Again, to help us visualize the network, we further omitted the retweeters (dangling nodes) by restricting the range of in-degrees to be higher than one. This left us with 347 hateful nodes and 2,900 edges with a network density close to 0.024.

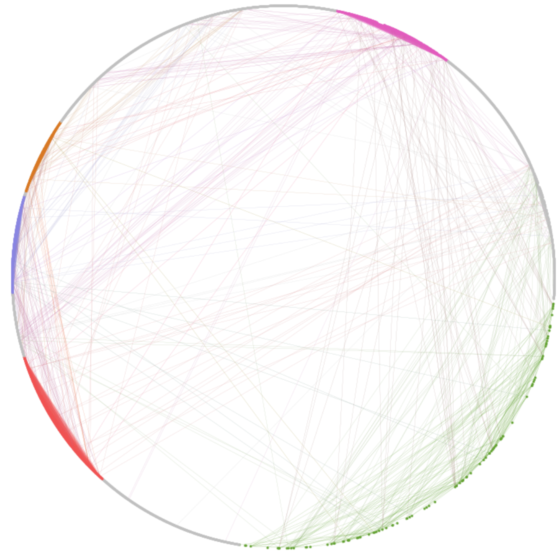


Fig. 5. HN-Network Community Interaction Network. With hateful nodes removed.

Compared to the HN network, the most drastic differences in these measures are the increase in the density from almost 0 to 0.024, indicating a notable increase in the connectivity of the network. The drop of density compared to the green community (0.037) indicates that some normal and other users connect the hateful users.

We then ran the Louvain method to identify communities within this subgraph. We arranged the nodes in a circular layout based on their modularity class to demonstrate the interactions between hateful user communities and calculated authority, hub, and PageRank centrality(Fig6).

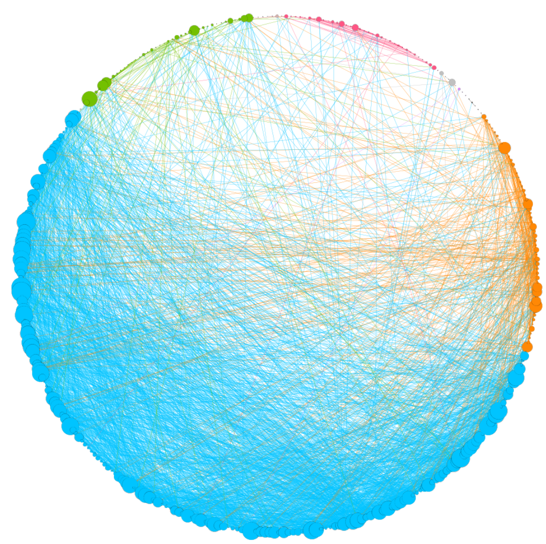


Fig. 6. H-Network User Community Interaction-Authority Rank Plot.

The largest community colored in blue became visually pronounced and is the most important in terms of centralities. Furthermore, this blue community also exhibits a high degree of interaction among its members and between members of two other smaller communities, as discussed in the HN-network. The blue community density was calculated to be 0.053, indicating that the users are highly connected.

Since we found that hateful users behave similarly in both networks, we further investigated the makeup of the blue community in the H-network and the green community in the HN-network. We calculated a 70-80% of users overlap between these two communities, indicating that the green community in the HN-network is largely composed of the hateful users identified in the blue community in the H-network. Therefore, we conclude that the blue community in the H-network is a subset of the green community in the HN-network.

The N-Network. Following our examination of the hateful user network, we wanted better to understand the differences between normal and hateful users. We used the same trimming method to form a network of solely normal users and their retweeters. We call this the N-network.

Upon the further removal of dangling nodes, we were left with 1,946 hateful nodes and 2,327 edges with a network density of approximately 0.001, indicating that this normal user network is not nearly as dense or connected than the hateful users. The average degree of the network decreased to a value of less than one, meaning the original network was so sparsely connected that many nodes were disconnected following the removal process.

Similar to the previous network analysis, we ran the Louvain method for community detection and arranged the nodes in a circular layout based on their modularity class to demonstrate the interactions between normal user communities and calculated authority, hub, and PageRank centrality. (Fig.7)

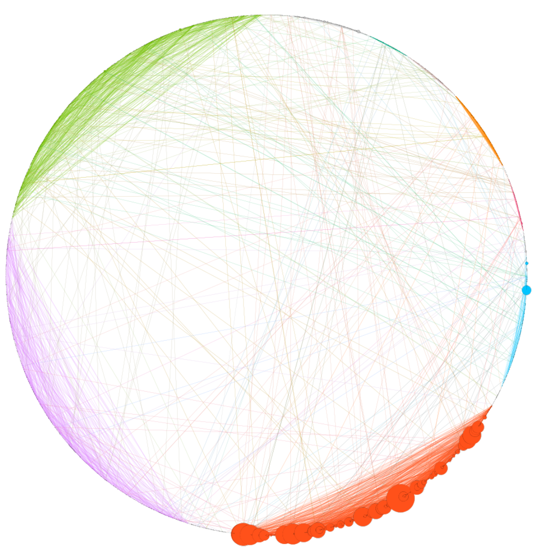


Fig. 7. N-Network User Community Interaction-Authority Rank Plot.

We observed that similar to the H-network, the authorities and hubs are gathered within a single community, which is colored in red this time. However, nodes with high PageRank measures were uniformly distributed across communities, which diverged from what we noticed in the hateful users' case.

On the other hand, we noticed that this red community behaves similarly to the hateful community in previous sections in terms of authority and hub centrality. We further investigated the hashtags used in the red community and reported in table.2.

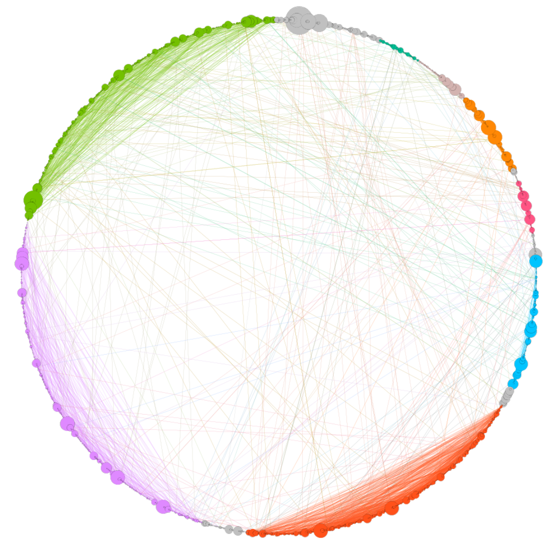


Fig. 8. N-Network User Community Interaction-PageRank Plot.

Hashtag	(%)	Hashtag	(%)
MAGA	3.68%	BoycottNFL	0.45%
UraniumOne	1.05%	FreeSpeech	0.45%
Brexit	1.00%	Iran	0.45%
Trump	0.78%	FakeNews	0.44%
JFKFiles	0.72%	LockHerUp	0.43%
cdnpoli	0.69%	bbcqt	0.41%
ThursdayThoughts	0.65%	DemocratLiesMatter	0.38%
DrainTheSwamp	0.65%	yyc	0.36%
Israel	0.55%	MoggMentum	0.36%
AnOpenSecret	0.51%	VoteGOP	0.36%

Table 2. Top 20 hashtags in tweets from red community users.

We discovered that the red normal-user community is posting similar content to the hateful users, and therefore they are also conservative users. Based on this observation, we were able to conclude that conservative users play essential roles in this retweet network, regardless of whether the users are hateful or normal.

The Conservative Network. Following our previous conclusion of the importance of conservative users, we decided to investigate the structure of their network. Using the same trimming method, we constructed a new network of hateful users, normal users in the red community of the N-network, and their retweeters. We call this the conservative network.

Again, to help us visualize the network, we further omitted the retweeters (dangling nodes), leaving us with 556 conservative nodes and 3,907 edges with a network density of approximately 0.013. We found that these conservative users are surprisingly well-connected.

Applying the Louvain method once again, we found this conservative sub-network can be broken up into five smaller communities. We applied the ForceAtlas2 algorithm for the network layout(Fig.9) and observed a similar structure compared to the HN-network(Fig.2). This is to be expected due to the high centrality measures of these conservative users. We hypothesized that the majority of interactions within this retweet network revolve around them.

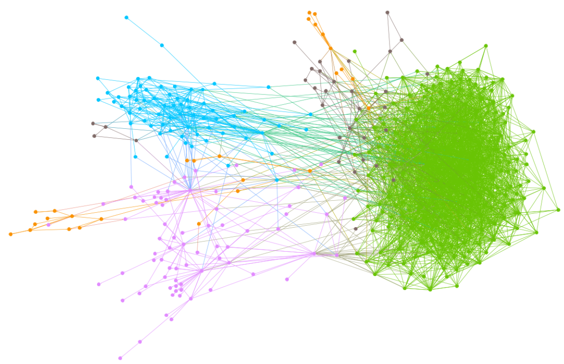


Fig. 9. ForceAtlas2 layout of the conservative community network. It has a similar structure compared to the HN-network in Fig.2

Influence Estimation

So far, we have discovered that most essential users in this retweet network are conservatives. Generally, a political group's goal on social media is to spread their opinion to all users ultimately. Therefore, we now further examined their influences in terms of the dissemination of their opinions on Twitter.

To assess their influence, we look at the correlation between the density and the average retweeted number for each community. The average retweeted number is calculated from the average out-degrees of the non-conservative retweeters of that community. Note, the higher the average retweeted number, the more likely to be retweeted.

In table 3, we report the density and average retweeted number for each community. In Fig.10, we present the scatter plots of data in table 3 .

	Community Subgraph Density	Ave Retweeted Number
Pink	0.011763	4.269912
Green	0.048369	4.97856
Blue	0.039185	2.704065
Orange	0.015514	1.409781
Black	0.022109	1.379921

Table 3. Communities' Subgraph Density and Ave Retweeted Number.

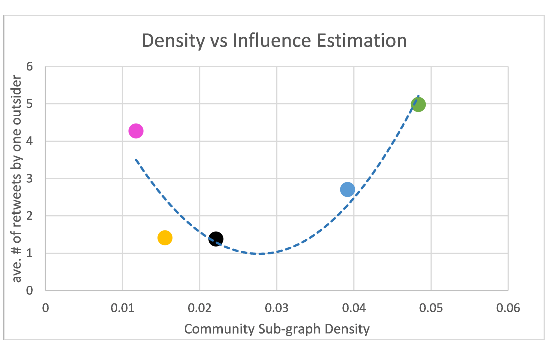


Fig. 10. Conservative community influence estimation with respect to community's subgraph density.

Examining the trend line we fit in the scatter plot, we observed an interesting, seemingly quadratic relationship between the community's density and the estimated average retweeted number. The conservatives users in the purple community display relatively loose connections with each other but have a high impact on their retweeters. In comparison, the green community possesses both high density within the subgroup and strong influence on their retweeters.

This result suggests two possible methods for maximizing the spread of conservative opinions in our Tweeter network: through a tightly connected group or a somewhat loosely connected one, but not something in between. However, note that due to the smaller size of the orange and black community (∼50 nodes), their numbers made this quadratic trend line less reliable.

Conclusions

Upon comparison of centrality and community structures between the main graph and its two subgraphs, we have concluded that it is possible to identify the influencers in a social network and that this group of users is highly connected. These users are particularly important when observing a variety of centrality measures. Although they do not make up a large percentage of the network, they account for the majority of the interactions. Additionally, upon a comparison between hateful and normal users, we found that hateful users are more interconnected and display higher levels of communication between users and follower "loyalty" than that of normal users. Upon further examination of the conservative user network, we determined that conservative users, regardless of label, have a high impact on their retweeters when they are either loosely connected or highly connected but not in the middle. This could indicate that in addition to hateful users, conservative users are very influential and efficient in spreading opinions.

Future Prospects & Discussions

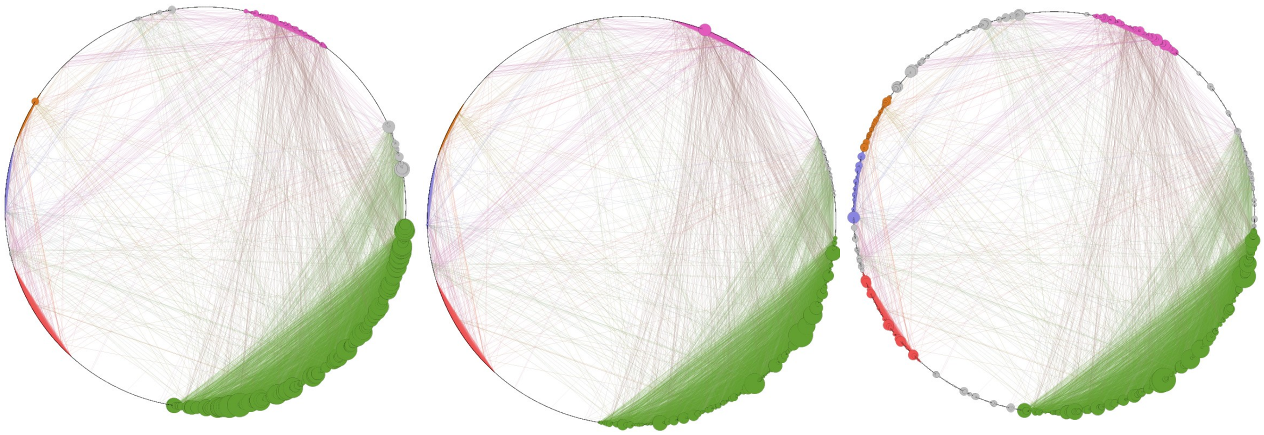
Due to the nature of the dataset, we had no access to the original tweets. In order to better understand and verify our findings, we should track tweets. To examine the spread of a single tweet, researchers could build a network starting with this tweet and working outwards. They could then compute density and betweenness centralities to examine the structure of this retweet network further.

We were also limited by computing power, forcing us to trim down our network. In order to authenticate our trimming method's viability, we should compare our data to that of the larger network and observe if it follows a similar pattern as we predicted. Other conservative retweet networks should be examined to verify our quadratic relation model. It would also be interesting to study a liberal retweet network and see how the two networks compare. Another possible consequence of our research is to gather enough data to quantify the relationship between density and influence estimation.

Acknowledgments. We would like to express our very great appreciation to Dr.Heather Brooks for her valuable and constructive suggestions during the planning and development of this research work. Her willingness to give her time so generously has been very much appreciated.

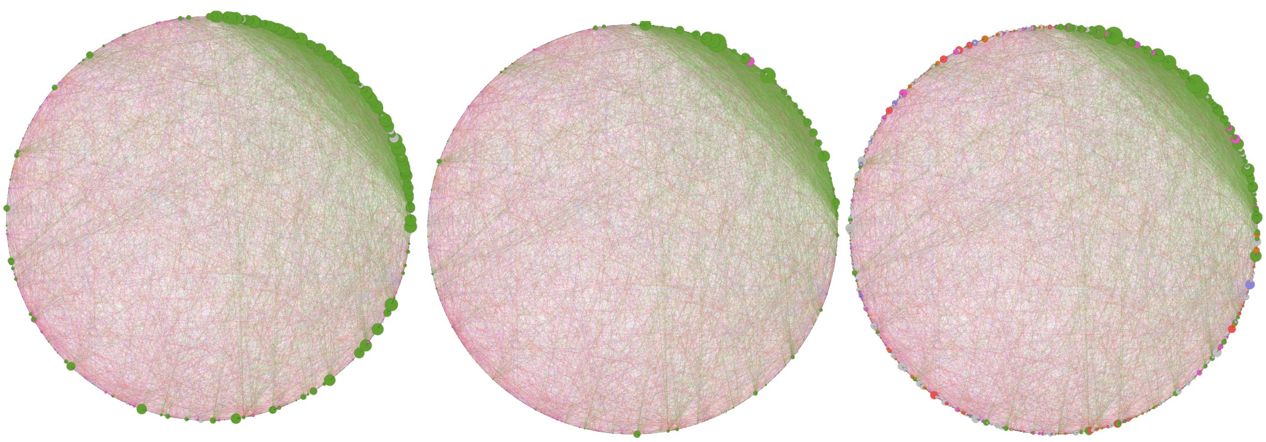
1. M Ribeiro, Hateful users on twitter detecting hate speech with context (2017).
2. JM Kleinberg, Authoritative sources in a hyperlinked environment. *J. ACM (JACM)* **46**, 604–632 (1999).
3. SBL Page, The anatomy of a large-scale hypertextual web search engine. *Comput. Networks* **30**, 107–117 (1998).
4. VD Blondel, JL Guillaume, R Lambiotte, E Lefebvre, Fast unfolding of communities in large networks. (2008).

Supporting Information Appendix (SI)



(a) HN-Network Community Interaction-Authority Rank Plot. **(b)** HN-Network Community Interaction-Hub Rank Plot. **(c)** HN-Network Community Interaction-PageRank Plot.

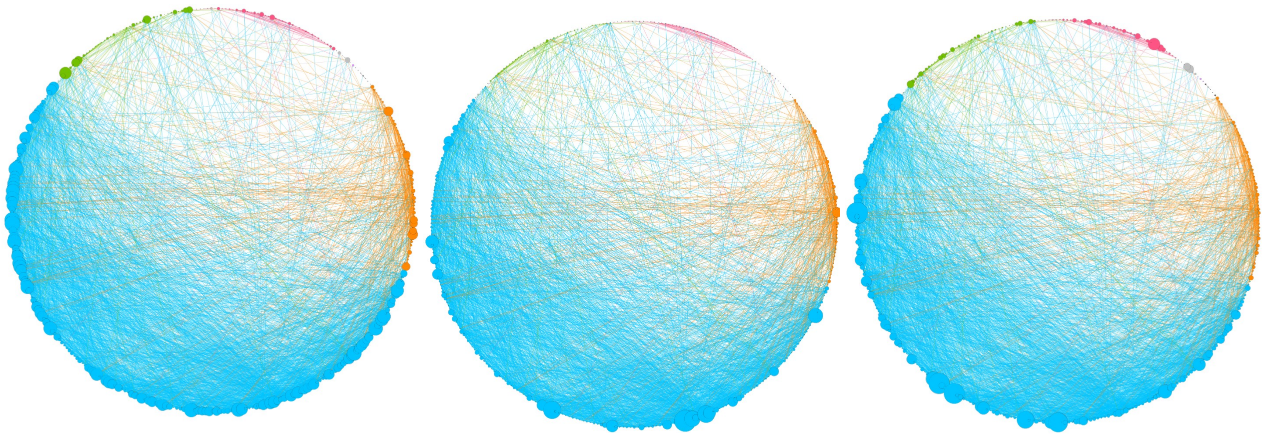
Fig. 11. HN-Network Community Interaction & Centrality Measurements



(a) H-Network Community Interaction-Authority Rank Plot. **(b)** H-Network Community Interaction-Hub Rank Plot. **(c)** H-Network Community Interaction-PageRank Plot.

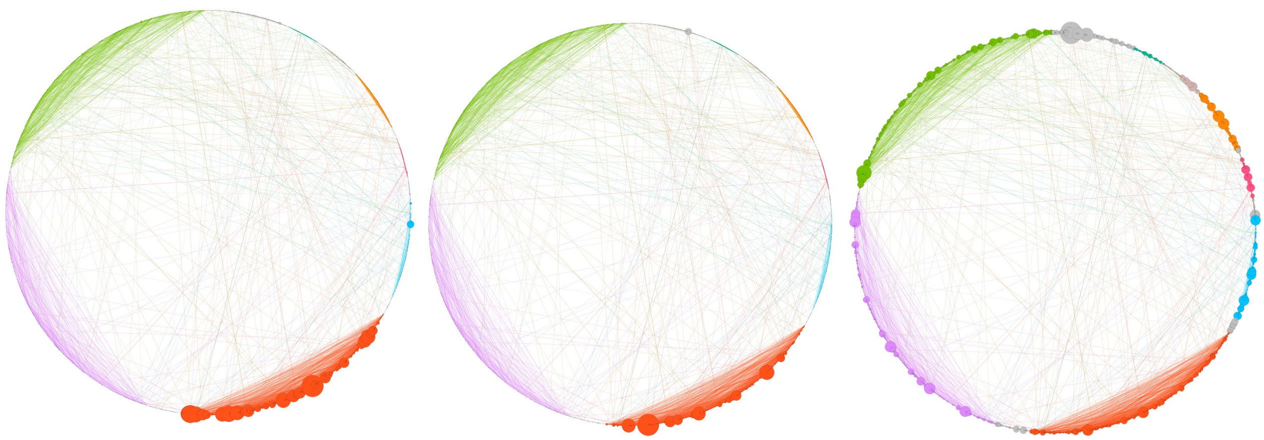
Fig. 12. H-Network Community Interaction & Centrality Measurements

[2]



(a) N-Network Community Interaction-Authority Rank Plot. **(b)** N-Network Community Interaction-Hub Rank Plot. **(c)** N-Network Community Interaction-PageRank Plot.

Fig. 13. N-Network Community Interaction & Centrality Measurements



(a) H-Network Community Interaction-Authority Rank Plot. **(b)** H-Network Community Interaction-Hub Rank Plot. **(c)** H-Network Community Interaction-PageRank Plot.

Fig. 14. H-Network Community Interaction & Centrality Measurements