The Bottom-Up Formation and Maintenance of a Twitter Community: Analysis of the #FreeJahar Twitter Community

Eugene Ch'ng

School of Computer Science | International Doctoral Innovation Centre
University of Nottingham Ningbo China
Science and Engineering Building
199 Taikang East Road, 315100 Zhejiang Ningbo, China
eugene.chng@nottingham.edu.cn

Abstract

Purpose – The article explores the formation, maintenance and disintegration of a fringe Twitter community in order to understand if offline community structure applies to online communities

Design/methodology/approach – The research adopted Big Data methodological approaches in tracking user-generated contents over a series of months and mapped online Twitter interactions as a multimodal, longitudinal 'social information landscape'. Centrality measures were employed to gauge the importance of particular user nodes within the complete network and time-series analysis were used to track ego centralities in order to see if this particular online communities were maintained by specific egos.

Findings – The case study shows that communities with distinct boundaries and memberships can form and exist within Twitter's limited user content and sequential policies, which unlike other social media services, do not support formal groups, demonstrating the resilience of desperate online users when their ideology overcome social media limitations. Analysis in this article using social networks approaches also reveals that communities are formed and maintained from the bottom-up.

Research limitations/implications – The research data is based on a particular dataset which occurred within a specific time and space. However, due to the rapid, polarising group behaviour, growth, disintegration and decline of the online community, the dataset presents a 'laboratory' case from which many other online community can be compared with. It is highly possible that the case can be generalised to a broader range of communities and from which online community theories can be proved/disproved.

Practical implications – The article showed that particular group of egos with high activities, if removed, could entirely break the cohesiveness of the community. Conversely, strengthening such egos will reinforce the community strength. The questions mooted within the paper and the methodology outlined can potentially be applied in a variety of social science research areas. The contribution to the understanding of a complex social and political arena, as outlined in the paper, is a key example of such an application within an increasingly strategic research area - and this will surely be applied and developed further by the computer science and security community.

Originality/value – The majority of researches that cover these domains have not focused on communities that are multimodal and longitudinal. This is mainly due to the challenges associated with the collection and analysis of continuous datasets that have high volume and velocity. Such datasets are therefore unexploited with regards to cyber-community research.

Keywords – Social network analysis, big data, twitter, online communities, social media, multimodal network, longitudinal network, freejahar, centrality measure

Paper type Research Paper

1. Introduction

Can communities form within Twitter? The social networking and multiplatform micro-blogging service that allows a limited tweet of 140 characters is generally associated with the spread of information. Users however are able to follow, or subscribe to the posts of other users, which create a network of followers and followees. It must not be argued that the followers-followees are a community as observations of users are mainly inactive with outbursts of retweets around certain viral news. It may be argued that the followers-followees phenomenon is at most a network with equal ties. We must however probe deeper in order to identify and isolate potential community behaviour within which the Twitter environment can provide. This is important partly due to the fact that many recent world events that tipped the balance of powers of governments originated from coordinated activities within Twitter, and partly due to the need to understand collective behaviours in cyber-communities. Twitter as a social media however, is certainly a research environment (Golder & Macy, 2013) where we can explore our questions.

Twitter service interprets keywords prefixed by a hashtag '#' as topical, and users are preceded with a @ in the tweets. Unlike other social networks, the service does not allow formal group creations. The sequential presentation of tweets as viewed within Web browsers or on mobile devices is chronological. Can communities form under such a limited environment? If communities are able to form, in what way do they maintain and support their existence, especially when Twitter does not allow formal groups?

The interest of this research lies in the controversial online teen #FreeJahar movement calling for the freedom of the Boston bombing suspect Dzhokhar Tsarnaev because teenage girls believe he is "too beautiful to be a terrorist" (Nelson, 2013). Concerns were raised that in on-line forums the younger Boston Bombing suspect appeared to attract a cult teen following expressing affection and concern for him (DailyMail, 2013). Facebook, Tumblr tribute accounts were set up in support of the teen with the

#FreeJahar Twitter tag, which were trending when these activities first appeared. Teen activities in Twitter do need to be monitored (Wiederhold, 2012) as observed below:

> *"I can't be the only one who finds the suspected bomber to be sexy, can I?" 19 April 2013.*

> *"i don't even care if jahar is a terrorist he's cute i don't want him to die." @******, 20 April 2013.*

> *"I'm not gonna lie, the second bombing suspect, Dzhokhar Tsarnaev, is hot. #sorrynotsorry"*
> *@******, 20 April 2013.*
> *"Yes I like Justin Bieber and I like Jahar but that has nothing to do with why i support him. I know*

> *hes innocent, he is far too beautiful" @******, 25 April 2013.*

> *"How many RT's for our boy jahar look at that beautiful face #freejahar*
> *pic.twitter.com/K9xKFvv5HT" @******, 4 May 2013.*
> *"Getting one of Jahar's tweets tattooed on me tomorrow. Guess you could say I'm a #FreeJahar*

> *supporter," @******, 7 May 2013.*

The term community has two major uses. The first being 'territorial' and 'geographical', which refers to the notion of community-neighbourhood, town and city, whereas the latter refers to the relational aspect of a community (Gusfield, 1975). Whilst geography plays an important role in the formation of social ties within online communities (Takhteyev, Gruzd, & Wellman, 2012), it is believed that online social networks such as Twitter can be highly relational. Such communities are concerned with the quality of character of human relationship (spiritual, professional, ideological, etc.) without reference to territories. It is noted early that modern society develops community around interests and skills more than around locality (Durkheim, 1964), the consistency of such a notion of community has been maintained in the Information Age. Community is better defined by the nature of relationships between individuals rather than geographical proximity (Preece & Maloney‐Krichmar, 2005).

It was noted a decade ago that "the Internet has altered our sense of boundaries, participation, and identity" (Renninger & Shumar, 2002), and much ink has been filled on the topic of whether online

communities are really communities (Bruckman, 2006). Twitter, unlike other services such as Usenet newsgroups, Internet Relay Chat (IRC), Facebook, SecondLife, etc., that allows a formal formation of communities (Wellman & Gulia, 1999) is different as it was originally created as a messaging service. It therefore may not be suitable to study Twitter using a similar approach in the literatures, see(Preece & Maloney‐Krichmar, 2005; Wellman & Gulia, 1999).

How then does a community look like in Twitter, if the definition of a community that we know can be formed at all? To answer this question, we must first look at communities familiar to us, i.e., those that have been formulated in the literatures. A community should exhibit a 'Sense of Community' (D. W. McMillan & Chavis, 1986; D. McMillan, 1976) among its members, which McMillan and Chavis originally explored and defined:

1. Membership – a feeling that one has invested part of oneself to become a member and therefore has a right to belong, this includes a boundary from which members are separated from out-groups

2. Influence – a sense of mattering, of making a difference to a group and of the group mattering to its members. (See Seo, Houston, Knight, Kennedy, & Inglish, 2013 for a particular example of the need of a sense of belonging for teens).

3. Reinforcement – the feeling that members' needs will be met by the resources received through their membership in the group

4. Shared emotional connection – the commitment and belief that members have shared and will share history of common places, time together, and similar experiences

It may be worth the effort to formulate concepts on the form in which a Twitter community would assume and thereby look for such an expression in the data. In this context, however, whilst it is helpful to

take a nominalist approach in defining the concepts, particularly of community boundaries, in reality, the automated computational approach that is employed in the research gathers datasets within the realist strategy (Laumann, Marsden, & Prensky, 1989). Considering the volume of Twitter data that this research collects, it will be extremely tedious to attempt to find a boundary using qualitative approaches, thus, a computational, quantitative approach in social networks analysis is used. Comparisons thus can be made and hypotheses tested when data has been analysed. We may begin to see patterns of community as we explore the datasets in the subsequent sections.

2. Methods

As opposed to the conventional method of mapping the follower-followee network, the approach defined here practically maps the actual evolution of instantaneous activities occurring within a timescale, over a series of days encompassing both large and small events. This is more useful as activities define the true interactions of active members.

A Big Data Twitter streaming software was used (Ch'ng, 2014) for mapping Twitter users (egos) and tweets as *nodes*, and *edges* representing links between egos and their tweets. Tweets are represented as nodes so that the flow of information is made obvious. Only tweets containing the keywords #Dzhokhar, #FreeJihad, #FreeJahar, #Tsarnaev are recorded. The reason for using only four keywords was that these were the keywords that were consistently used in the Tweets. In fact, #FreeJahar would have been sufficient as it appeared in all of the tweets. This resulted in 60 longitudinal datasets, each containing 5 hours of continuous data from 17 May 3.00pm and to 31 May 12.05pm (15 days). The data were recorded from 17 May onwards as news of the 'movement' were not reported until then. An additional 30 days of data records the decline of the #FreeJahar activities.

The file sizes of the series are shown in Figure 1. Peaks and valleys are consistent with the time of activities during the hours spanning both days and nights, except when the keywords were trending.

Figure 1.

The relative importance of nodes uses Betweenness, Closeness (Freeman, 1979; Newman, 2005; Sabidussi, 1966) and Eigenvector (Bonacich, 1987) centralities measures. Betweenness is a measure of information brokerage between parties, Closeness measures the spread of information from a node to all other nodes, where lower closeness has shorter distance to other nodes. High Eigenvector demonstrates increased numbers of egos who were connected to important egos in the network, an indication of heightened activities.

3. Results

An analysis of the datasets shows that news events were all related to the Boston Bombers with the topics: 'triple murder', 'FBI kills man', 'Ibrahim Todashev', 'Al Qaeda Mag praises Tsarnaev brothers', 'Dzhokar Recovers', 'Mother and Father', 'Russians provides info about brothers'. Does the #FreeJahar community exists within the graph? Visualising the mapped networks will reveal this information.

Figure 2 visualises 16 graphs, 5 hours each (ranked by file size from the smallest 5708 to the largest 803, left to right, top to bottom). The graphs were reconfigured using Gephi's ForceAtlas algorithms so that connected nodes due to higher interactions, appear closer together (clusters in Figure 2). Each graph shows a different signature as they carry varying explosions of information when news went viral. Egos that have a larger degree distribution (larger nodes) are news channels and their correspondents (CNN, Guardian, BBC, NYT, etc.). Some networks have large populations of individual nodes (516 and 425), and the satellite clusters of dyads and triads surrounding 356 and 442.

When unconnected nodes are removed, a large single cluster of connected egos and tweets remained (Figures 3-5). A closer inspection of conversations showed the clusters as a community of #FreeJahar

supporters. Communities in the graph have a unique expression as compared to circular retweet networks (black nodes), they are characterised by disorderliness and isolation in the overall network. The same community is observed in all the datasets as a large cluster of connected nodes. Within the community are egos with associated membership symbols (see the next subsection for examples). These members persist across all data points in Figure 1.

Figure 2.

3.1.   The Boundary and Activities of a Twitter Community

What defines the boundary and membership of this particular Twitter community? A sample of more prominent graphs was selected for analysis here (Figure 3-5). An inspection of the egos indicates that a common symbol exists, They carry these #FreeJahar associated symbols in their Twitter names and profile pictures: 'dzhokar', 'jahar', 'troy', 'crossley' or 'tsar', or mutations of the name ('Truy', 'Crussley', 'Croosley', etc) and related verbs and nouns: 'free', 'innocent', 'let_him_go', 'family', etc. A closer inspection of the community showed that it has separate communities within the larger network. This reinforces the observation that online communities are also recursive, where one community may be split into more internal communities (Sales-Pardo, Guimera, Moreira, & Amaral, 2007). There are two larger communities within the network flanked by a smaller group. The larger community are #FreeJahar opposers with associated keywords such as 'faithful', 'usa', 'bible', etc. The symbol of the opposing community is not as distinct as the #FreeJahar supporters and therefore require the reading of the tweets and the profiles. As expected, drawing a catchment area (coloured overlays) around the analysis of the communities shows distinct boundaries between the groups. The internal interaction alone is sufficient to form a 'protective' boundary around the #FreeJahar group. The smallest group was identified as having conspiracy theory oriented tweets, tweets which catalyses the #FreeJahar movement's activities for

ideological gains. Here we discovered the boundary in terms of interactions between nodes and a common symbol.

Figure 3.

Dataset 759 (Figure 3) is a network with large activities and a very prominent community expression. The graph showed a large community cluster consisting of #FreeJahar supporters (no overlay), opposers (transparent green patch), and conspiracy theorists (transparent orange patch) supporting the #FreeJahar movement's activities. They are clearly separated into two opposing but cohesive communities. Opposer egos closest to the #FreeJahar boundary have more intense interactions with the group. Sentiments in the tweets range from statements, quotes and retweets of news and intense emotional arguments between the two groups. Most of the tweets from the opposing community were criticisms. Discussions and retweets took place within the communities, which maintained the cohesiveness of the connection throughout the datasets.

Figure 4.

Figure 5.

3.2.   The Bottom-Up Organisation of the Community

Communities do not form and then disintegrate. Efforts are needed to maintain the boundary and reinforce membership bonds so that the community becomes stronger over time. The limited and sequential nature of the Twitter environment makes it difficult to maintain an active community boundary, there is a higher probability of disintegration unless members assume some form of leadership. Conversely, as the #FreeJahar group is formed from disparate actors with strong similar ideology, the community may be organised from the bottom-up where all members have equal importance.

Figure 6 compared the top 10 egos in dataset 759 who were highly active within both communities over 5 hours of ego-centrality measures. Coloured lines and graphical symbols represent individual egos. The average score of all the egos is represented as a single solid black line accompanied with circle symbols. Each ego is affected by other ego centralities. The first graph (top) shows the growth of each Betweenness centrality over 5 hours at 30 minutes intervals. E and F initiated the communication between two communities in the first hour, with the latter being more active in discussions in the first three hours of activity. All ego Betweenness centralities grew proportionately after the third hour, #FreeJahar members were more active. There is consistency of Closeness centrality (middle graph) of all egos, indicating an equal distribution of social proximity with other members in the network.

The bottom graph measures the Eigenvector centrality of the egos in the communities. Egos having higher Eigenvectors are those that are connected to other highly active members. The #FreeJahar members appear to be more active in this case. The user U ended up being the most important ego in the community, but as observed, the Eigenvector centrality fluctuates over time in a consistent trend. In summary, the community does not have a central ego showing that leaderships are distributed. It appears that all the members are equal and that the community is maintained from the bottom-up. Egos from the opposing community occupy this social space and their activities as an out-group act as external conflict, which indirectly maintains the boundary of the #FreeJahar group via arguments, criticisms, and ridicule. This reinforces the theory that positivity and success in the interactions create cohesion (Cook, 1969), external conflict increase internal cohesion(Stein, 1976).

Figure 6.

The centrality measures of the cluster corresponds to Leavitt's observation (Leavitt, 1951), that "where high centrality, and hence independence are evenly distributed, there will be no leader, many

errors, high activity, slow organisation, and high satisfaction". The edges that play a central role in connecting the small-world network can be traced within 2 steps of egos with high Betweenness centrality within the social movement, the same agent with high Betweenness centrality makes the community cohesive. Removing these egos will invariably disrupt the entire community. Moody and White (Moody & White, 2003) observed that "a group is structurally cohesive to the extent that multiple independent relational paths among all pairs of members hold it together". In this context, removing the egos that keep the cluster alive will disrupt the community. Twitter's removal of highly active members confirmed Moody and White's concept of 'structural cohesion', defined as "the minimum number of actors who, if removed from a group, would disconnect the group". The removal of #FreeJahar members was due to the infringement of Twitter policies. As a result of the infringements, these accounts have since been suspended or deactivated. There was a Twitter post on the 30 July 2013 by one of the active members – "Aint nobody wanna #freejahar no more?" and 31 July – "Why is everyone deactivating their accounts? This battle is just beginning! #freejahar". This member's account has also been suspended. Figure 5 presents samples from the final decline of the #FreeJahar community.

4. Discussion

In this article, a longitudinal Twitter dataset associated with the #FreeJahar group calling for the freedom of the Boston bombing suspect were explored. The datasets consist of 5 hourly tweets over 45 days mapped as a network of activities present opportunities for discovering global behaviours from instantaneous contents produced by collective social actors as they interacted desperately at the local level within the confines of a digital display.

The tracking of Twitter activities apart from the follower-followee network reveals distinct spatial expressions between tweets, retweets and conversations. Using this approach, tweet nodes and edges

constituting a conversational nature could be identified and isolated from characteristic retweets. The tracking of multimodal connections will give us a more accurate measure of information than a follower-followee network.

A number of questions were presented at the beginning, probing the possibility of communities forming and maintained within the limited Twitter environment. Data analysis shows that communities do form within Twitter, and as a consequence, raise specific issues on coordinated behaviour and information dissemination within the social media. Twitter community differs from offline community in many ways due to the limits of the Twitter environment, the most apparent is reinforcement, and the support needed amongst members. It is not clear if online Twitter community facilitates offline gatherings, or if Twitter social ties led to other online groups (FaceBook friendships, email and phone exchanges, and etc.) as data could not be obtained. However, to this end, we are at least able to describe the nature of Twitter communities and how they are formed and maintained.

We have learned that Twitter communities are relational, formed via a common ideology and justified by validation of the ideology and the commonality of symbols. These worked together to segregate the in-groups from the out-groups. Members fulfil their needs via discussions and defended their cause against conflicts from another community, which creates internal cohesion. MacMillan and Chavis stated that, "people possess an inherent need to know that the things they see, feel, and understand are experienced in the same way by others" Such a group norm validates their experience. Influence therefore is unidirectional – members influence the group. The community is organised from the bottom-up, with equal distribution of leading roles and activities over time. The eventual decline of the #FreeJahar community was due to the suspension of important egos from Twitter, resulting in the destruction of the community structure.

The #FreeJahar event is an exemplar case study that could be generalised to much broader scopes for this sort of work as it demonstrates rapid, polarising grouping, behaviour, growth, disintegration and

decline of an online community. The study shows that communities with distinct boundaries and memberships can form and exist within Twitter's limited user content and sequential policies, which unlike other social media services, does not support formal groups, demonstrating the resilience of desperate online users when their ideology overcome social media limitations.

Social networks can increase our range of human connectedness beyond the boundary of users' geographical location. Communications sent now may be retrieved and responded to, much later in time. This invariably opens up a broad range of opportunities as space and time, in the eye of a user are 'compressed' to within a digital display. The fact that communities can form where services that facilitate group formation are not supported is an interesting phenomenon to look at. It will be beneficial to collate extremely large datasets from ad-hoc communities within Twitter in the future, particularly where revolutions and socially mediated civil uprisings are concerned.

References

Bonacich, P. (1987). Power and Centrality: A Family of Measures. *American Journal of Sociology*, *92*, 1170–1182.

Bruckman, A. (2006). Learning in online communities. *The Cambridge Handbook of the Learning Sciences*, 461–472.

Ch'ng, E. (2014). The Value of Using Big Data Technology in Computational Social Science. In *The 3rd ASE Big Data Science 2014, Tsinghua University 4-7 August*. Beijing, China.: Association for Science and Engineering. doi:http://dx.doi.org/10.1145/2640087.2644162

Cook, S. W. (1969). Motives in a conceptual analysis of attitude-related behavior. In *Nebraska symposium on motivation*. University of Nebraska Press.

DailyMail. (2013, May 12). Boston suspect becomes teen hearthrob: Thousands of girls express their love for bomber in worrying online forums. *DailyMail Online*. Retrieved from http://www.dailymail.co.uk/news/article-2323342/Dzhokhar-Tsarnaev-The-teen-girls-crush-Boston-bomber.html

Durkheim, E. (1964). *The Division of Labor in Society*. New York: Free Press of Glencoe.

Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239.

Golder, S. A., & Macy, M. W. (2013). Social Media as a Research Environment. *CyberPsychology, Behavior and Social Networking*, *16*(9), 627–628.

Gusfield, J. R. (1975). *The Community: A Critical Response*. New York: Harper Colophon.

Laumann, E. O., Marsden, P. V, & Prensky, D. (1989). The boundary specification problem in network analysis. *Research Methods in Social Network Analysis*, *61*, 87.

Leavitt, H. J. (1951). Some effects of certain communication patterns on group performance. *The Journal of Abnormal and Social Psychology*, *46*(1), 38.

McMillan, D. (1976). *Sense of community: An attempt at definition*. George Peabody College for Teachers.

McMillan, D. W., & Chavis, D. M. (1986). Sense of community: A definition and theory. *Journal of Community Psychology*, *14*(1), 6–23.

Moody, J., & White, D. R. (2003). Structural cohesion and embeddedness: A hierarchical concept of social groups. *American Sociological Review*, 103–127.

Nelson, S. C. (2013, May 12). #FreeJahar: Boston Bombings Suspect Dzhokhar Tsarnaev's Teenage Fans Insist He's "Too Beautiful To Be A Terrorist." *Huffington Post*. Retrieved from http://www.huffingtonpost.co.uk/2013/05/12/freejahar-boston-bombings-suspect-dzhokhar-tsarnaevsteenage-fans-too-beautiful-terrorist_n_3262795.html

Newman, M. E. J. (2005). A measure of betweenness centrality based on random walks. *Social Networks*, 39–54.

Preece, J., & Maloney-Krichmar, D. (2005). Online communities: Design, theory, and practice. *Journal of Computer-Mediated Communication*, *10*(4), 0.

Renninger, K. A., & Shumar, W. (2002). *Building virtual communities: Learning and change in cyberspace*. Cambridge University Press.

Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, *31*, 581–603.

Sales-Pardo, M., Guimera, R., Moreira, A. A., & Amaral, L. A. N. (2007). Extracting the hierarchical organization of complex systems. *Proceedings of the National Academy of Sciences*, *104*(39), 15224–15229.

Seo, H., Houston, J. B., Knight, L. A. T., Kennedy, E. J., & Inglish, A. B. (2013). Teens' social media use and collective action. *New Media & Society*, 1461444813495162.

Stein, A. A. (1976). Conflict and Cohesion A Review of the Literature. *Journal of Conflict Resolution*, *20*(1), 143–172.

Takhteyev, Y., Gruzd, A., & Wellman, B. (2012). Geography of Twitter networks. *Social Networks*, *34*(1), 73–81.

Wellman, B., & Gulia, M. (1999). Net surfers don't ride alone: Virtual communities as communities. *Networks in the Global Village*, 331–366.

Wiederhold, B. K. (2012). As Parents Invade Facebook, Teens Tweet More. *CyberPsychology, Behavior and Social Networking*, *15*(8).
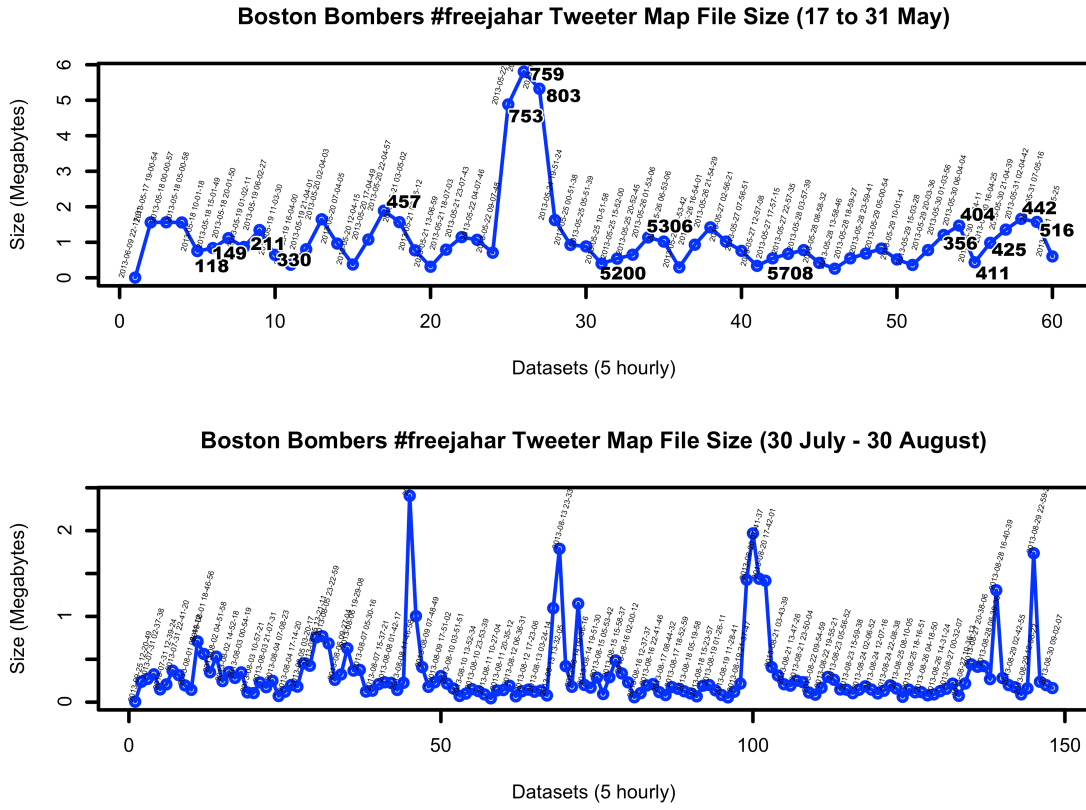
Boston Bombers #freejahar Tweeter Map File Size (17 to 31 May)

Boston Bombers #freejahar Tweeter Map File Size (30 July - 30 August)

Figure 1. A series of Twitter datasets consisting of 5-hour activities over 45 day.
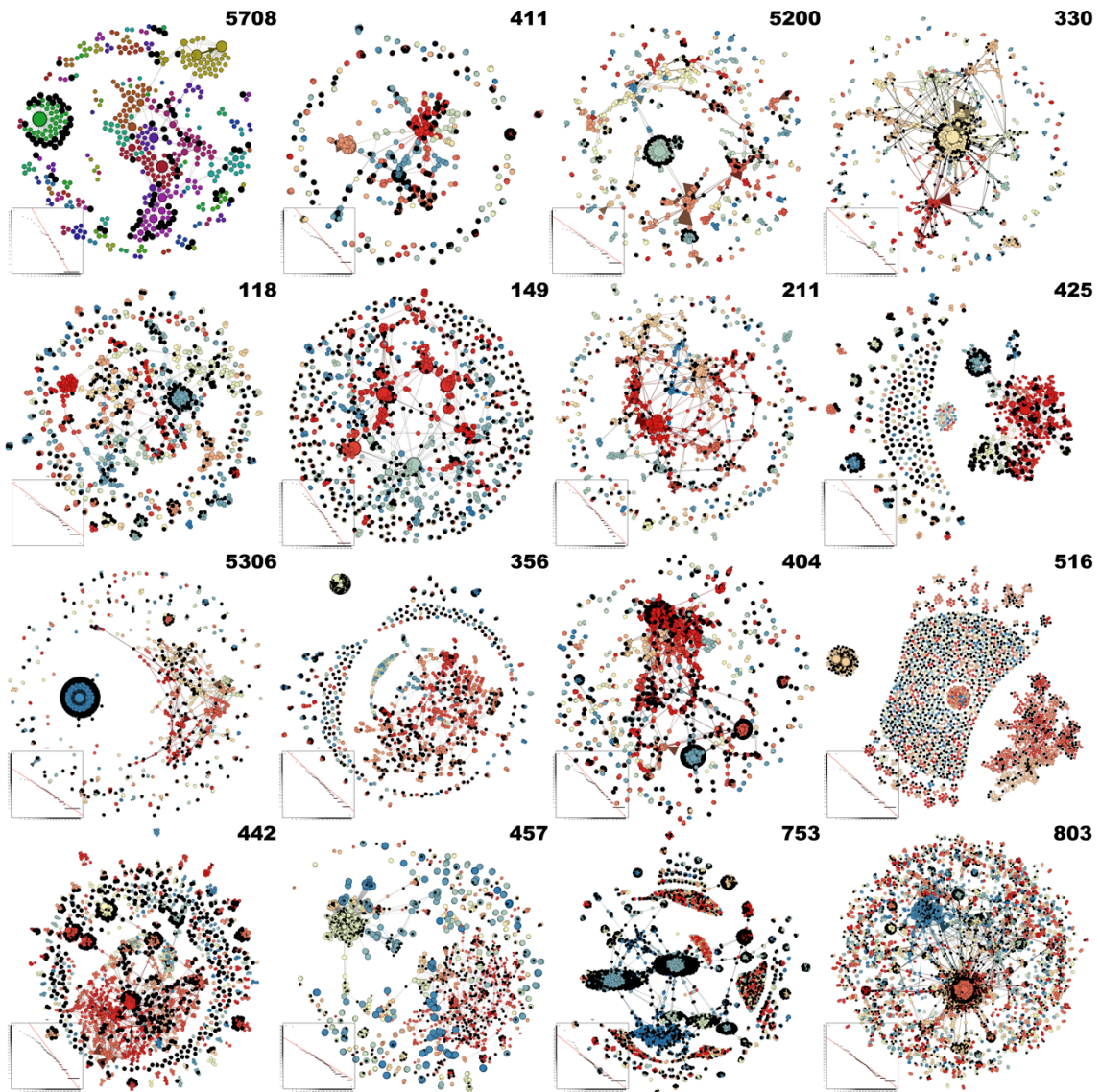
Figure 2. Comparative social information landscape of the sample datasets showing varying expressions but distinct signatures between retweets and conversations. The insets indicate the probability distribution of the degree centrality.
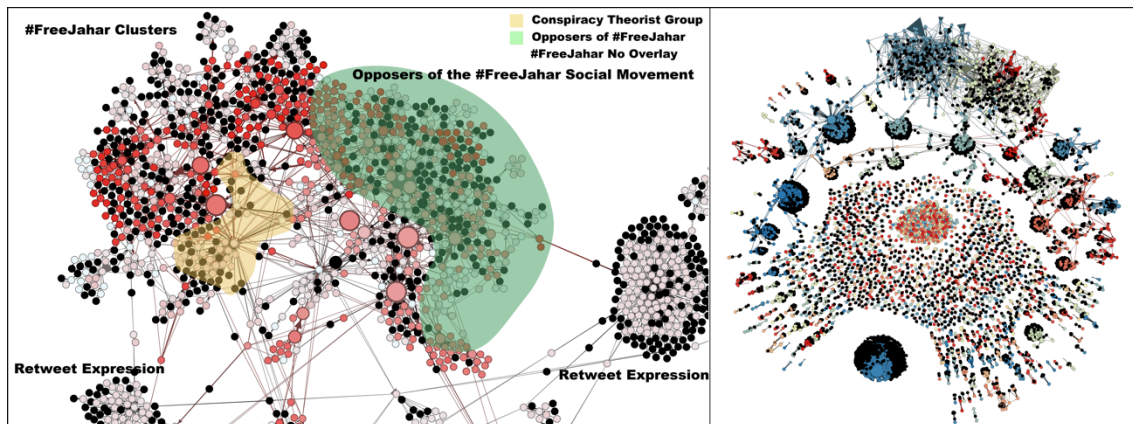
Figure 3. A visual reconfiguration of dataset 759 with distinct expression of retweets and community activities composed of conversations and retweets (black nodes). The majority of isolated dyads and triads (bottom right) The bottom image is a close-up of the #FreeJahar community consisting of supporters and opposers (transparent green patch). Opposer egos closest to the #FreeJahar group have more intense interactions with the supporters. The community maintain its boundary via interactions internally and with opposing group.
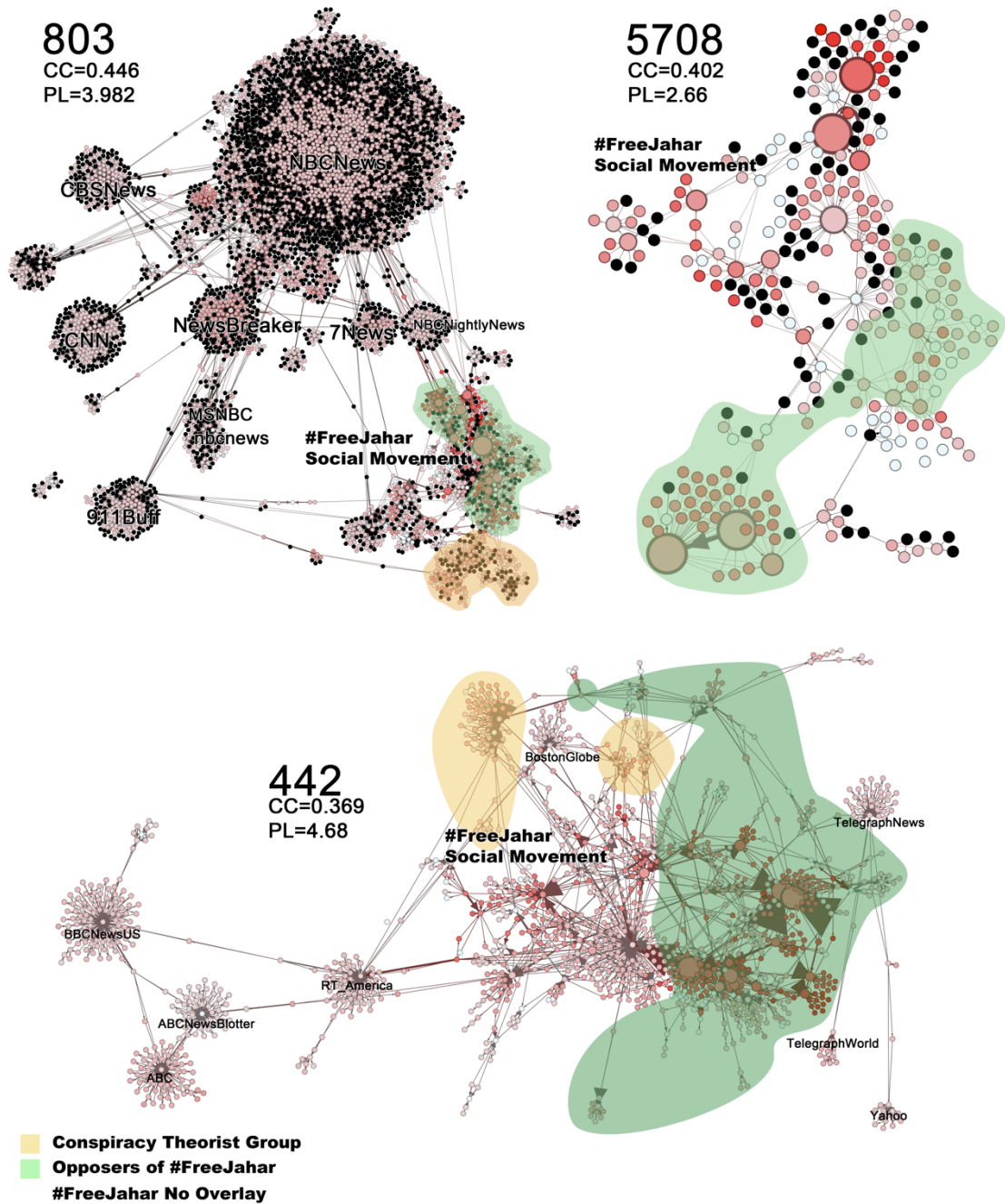
803
CC=0.446
PL=3.982

NBCNews
CBSNews
CNN
NewsBreaker
7News
NBCNightlyNews
MSNBC
nbcnews
#FreeJahar
Social Movement
911Buff

5708
CC=0.402
PL=2.66

#FreeJahar
Social Movement

442
CC=0.369
PL=4.68

#FreeJahar
Social Movement

BostonGlobe
TelegraphNews
BBCNewsUS
RT_America
ABCNewsBlotter
ABC
TelegraphWorld
Yahoo

Conspiracy Theorist Group
Opposers of #FreeJahar
#FreeJahar No Overlay

Figure 4. Comparing similarities between datasets with the largest #FreeJahar activities. Dataset 442 is the second largest after 759, 803 is the third largest, and 5708 is the smallest dataset. Egos with highest

Betweenness centrality are larger, and reddest egos have the highest Closeness centrality. Isolated egos (dyads and triads) were removed prior to calculating the average clustering coefficient and path length. PL = path length, CC = clustering coefficient.
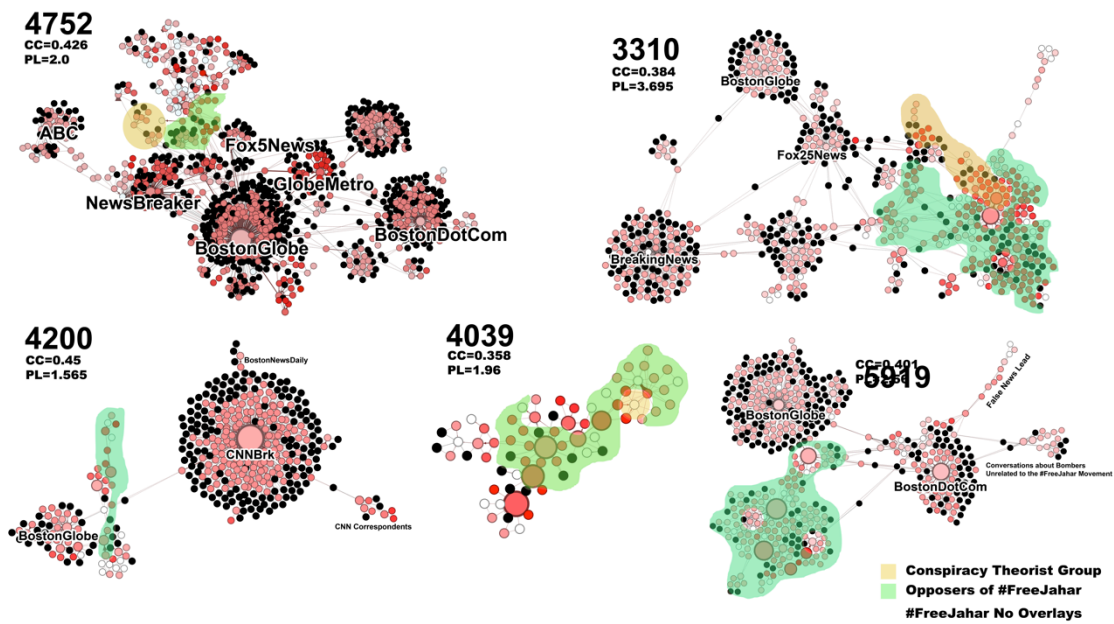

Figure 5. Larger datasets in the final collection of the #FreeJahar community showing the diminishing activities of the group in relation to news events. As the community declines, centrality measures of supporters decrease due to the disruption of Twitter accounts suspension, loss of interest, and community fractures as time progresses. PL = path length, CC = clustering coefficient.
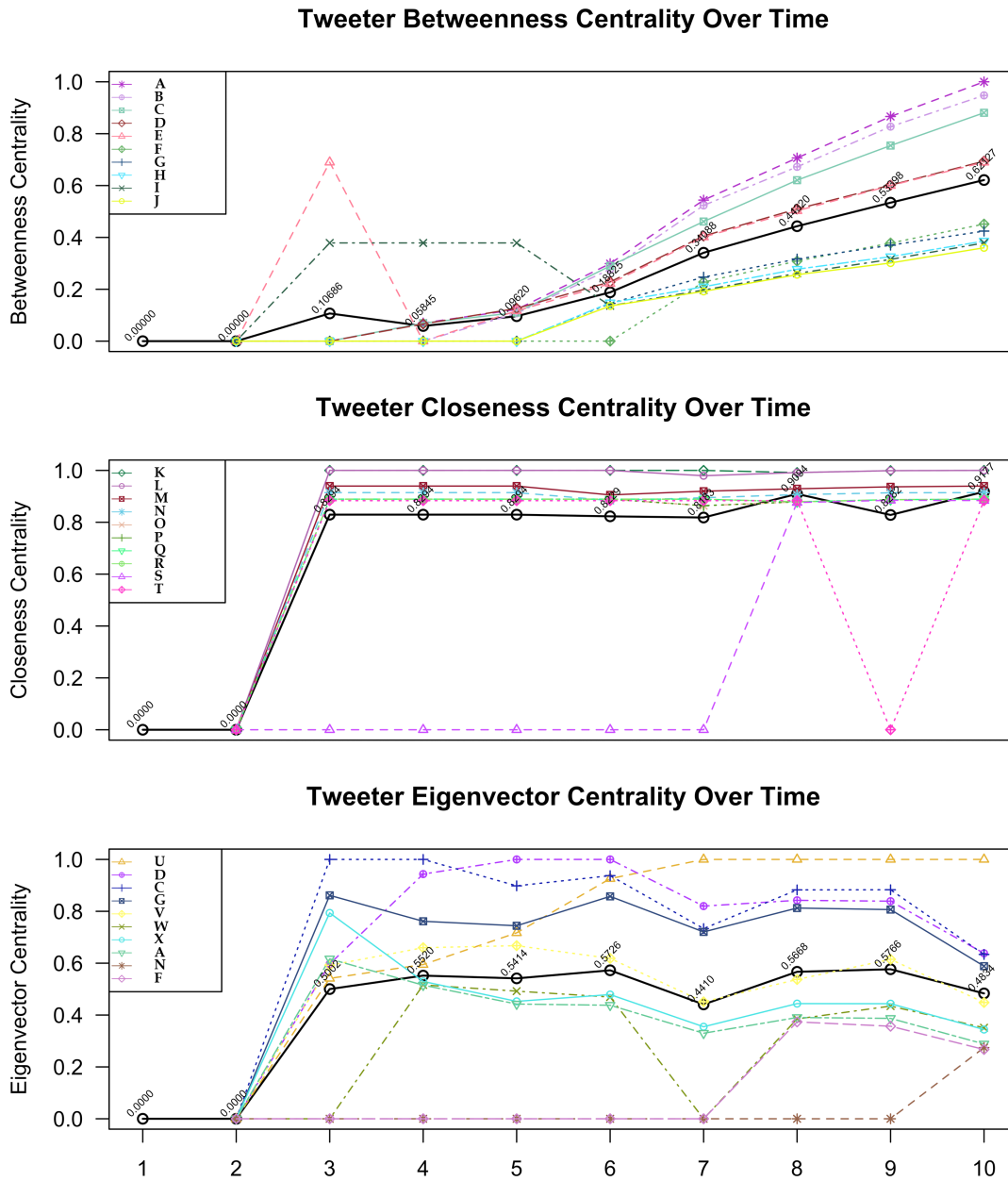
Figure 6. Dynamics of centralities demonstrating the bottom-up influence in the formation of the community in dataset 759. A-X are highly active individual Twitter users.