# Twitter Structure as a Composition of Two Distinct Networks

Meng Tong, Ameya Sanzgiri, Dimitrios Koutsonikolas and Shambhu Upadhyaya
Computer Science and Engineering, University at Buffalo, Buffalo, New York 14260
{mengtong, ams76, dimitrio, shambhu}@buffalo.edu

***Abstract*—In this paper, we study the structure and formation of Twitter and attempt to answer the role of Twitter – as news media or social network. Our analysis indicates that the Twitter network can be formally modeled as a composition of two main networks that have distinct roles in information propagation. Following this we also propose a concise configurable two-step model that generates a Twitter-like network to facilitate the development of a simulation platform for future research. We verify the validity of the proposed model by empirically analyzing two large datasets containing the topological information of Twitter and study its properties by means of mathematical analysis and simulation.**

## I. Introduction

Online social networks (OSNs), such as Facebook, Twitter, Instagram, etc., are becoming an important part of our daily lives. Twitter is especially interesting in this context since its content has been widely used in business and marketing [1], [2]. However, there have been also cases where Twitter has been misused, such as the hijacking of the Associated Press account and subsequent bogus tweet, that resulted in the loss of millions of dollars in the stock market. To design mitigation techniques that prevent misuse of Twitter, it is imperative to understand the role of Twitter as well as the mechanisms of information propagation.

There is a lot of interest from the research community to understand the Twitter structure and information dissemination mechanisms [3]–[6]. However, the research is often dependent on datasets of Twitter structure which are not always easily available. This is further exacerbated by the lack of formal modeling or understanding of the Twitter structure, which are the main reasons for the lack of suitable simulation platforms. Two of the key challenges in creating a simulation platform are the quantification of the role of Twitter and the characterization of its structure. The public notion of Twitter is that it is an OSN used to interact with friends and also is a micro-blogging site to disseminate information [7]. The authors of [8] investigated if Twitter is a social network or a news medium. Towards this end, they topologically analyzed a large dataset of the social graph of Twitter and one of their key findings was that the social graph did not fit power-law distribution, a key attribute of social networks. However, they were not able to draw any conclusions regarding the structure of Twitter.

The objective of this paper is to formally model Twitter's structure by analyzing the formation of its network, in order to obtain a better understanding of the information propagation. Based on the purpose and formation of the Twitter network and its structure we hypothesize that it is indeed a composition of two distinct networks and analyze each of them. We then propose a two-step configurable model to create a Twitter-like structure and identify key parameters necessary to maintain the properties of the network. This model can serve as a first step towards creating a scalable simulation platform that can be used to analyze information propagation and user behavior.

## II. Preliminaries

### A. *Twitter Network Formation Process*

When a user creates a new Twitter account, a new network creation is initiated for the user, divided into three steps:

1) A list of popular users (e.g., celebrities or news media) is provided and the user is asked to "follow" five of them (need not necessarily be the suggested ones).

2) Twitter's Who to Follow algorithm [9] analyzes the areas of interest based on the selections in the first step and provides a categorized list so that the user can follow another five entities.

3) Twitter asks for permission to access the user's contact list from the email used to sign up to find people who are already on Twitter and suggests five of them from the list to follow.

The user $\rightarrow$ follower entity abstracts the dissemination model of information from a user to its followers (other users who "follow" a user). The information is propagated via Twitter specific messages called *tweets*. In terms of relationship, unlike other social networks, the relationship between a user and its follower in Twitter can be asymmetric. Specifically, when a user gains a follower, they both do not automatically follow each other, thus a user does not necessarily gain access to *all* the tweets of its followers.

### B. *Power Law Distribution*

Power law distribution (PLD) is often used in the understanding and analysis of complex network structures, and especially in the purview of social networks.

**Definition and Properties:** Mathematically, the probability density function (pdf) of PLD is

$$p(x) = Cx^{-\gamma} \tag{1}$$

where $C$ is a constant and $\gamma$ is the scaling parameter. It has been shown via real world datasets that the scaling parameter $\gamma$ typically falls in the range between 2 and 3, although there are some exceptions [10], such as Twitter [11].

**PLD and Social Networks:** One important property is the heterogeneity among the possible values of a variable following a PLD which means that the average value cannot well describe the variables that follow a PLD. In the context of social networks, it can be used to analyze and gain understanding of the structure of the network. As has been shown in [11], many OSNs are likely to have a degree distribution following a PLD, with the scaling parameter falling in the range between 2 and 3, with the exception of Twitter [11]. As mentioned above many OSNs are likely to have a degree distribution following a PLD and one way to determine if a network is a social network, is to examine if the degree distribution follows PLD. This process is known as distribution fitting and involves examining if the plot of a dataset's complementary cumulative distribution function (CCDF) on a log-log scale is a straight line. Fitting PLD is a non trivial task [12] and while the method mentioned above is simple, it is not very accurate. This paper follows the method described in [12] which is an efficient implementation of the method in [10].

**Preferential Attachment Model:** This is one of the formation models that leads to a degree distribution that follows PLD as described in [13] and has two key features. First, the model assumes a growing network;second, when a new node joins, it has a greater affinity to connect to popular nodes (higher degrees) than unpopular nodes, thus the name "preferential attachment." If $m$ is the number of links a new node forms upon joining and $d_i(t)$ is the degree of an already existing node $i$ at time $t$, then when the new node joins, $i$ gains $m\frac{d_i(t)}{\sum_{j=1}^{t} d_j(t)}$ new links.

To ease the modeling process, we use the mean field approximation approach by assuming that every new node forms the same number of links; this average behavior has been proved to be a good approximation [14]. This gives the increasing rate of $d_i(t)$ to be

$$\frac{dd_i(t)}{dt} = m\frac{d_i(t)}{\sum_{j=1}^{t} d_j(t)} \tag{2}$$

Solving this differential equation with a start condition of $d_i(i) = m$ will give a PLD with $\gamma = 3$. Though the original preferential attachment works on undirected networks and results in a PLD with $\gamma = 3$, a modification of the model can be used to incorporate directed networks and to yield a wide range of values for the scaling parameter as shown in Sec. III-B1.

## III. Analysis of the Twitter Structure

For the precise understanding of the structure as well as the role of Twitter, the authors of [8] analyzed the topology of Twitter users and concluded that Twitter exhibited a non-power-law follower distribution, a short effective diameter, low reciprocity, which all mark a deviation from known characteristics of human social networks. However, the formation process as explained in Sec. II provides some insight into analyzing the network.

The process almost[1] clearly separates the formation of two subnetworks, say, information network (IN) and social network (SN). The first two steps in Sec. II-A can be regarded as helping the new user form the "information network", by suggesting popular users. This is very reasonable since users would want someone with public trust/credibility as information sources, thus popular users provide good choices as they are trusted by a large number of users. The formation of the IN is also a basic characteristic of the preferential attachment model which is further displayed by the "find and follow well-known people" as part of the second step.

The third step builds the "social network", by importing from other existing social relationships, typically people from the email contact list who are already using Twitter. The formation process reveals two other important parameters helpful in building a model of the Twitter structure. First is the total number of users that a user will follow upon joining Twitter, which if the user strictly follows the Twitter suggestions, is 15. Second is the ratio of the number of users a new user follows by searching his contact list to the total number of users the new user follows, which we define as the "social ratio" and denote as $\alpha$. The significance of $\alpha$ is explained in Sec. V.

[1]By almost we mean that there is a possibility for some of the users' contacts to appear in the suggested lists of steps 1 and 2 of Sec. II-A.

### A. Network Separation

Based on the preceding discussion, we hypothesize that the Twitter network is separable into two different networks based on their usage purposes. Thus, the two subnetworks can be formally defined as:

**Social Network (SN)**: a network containing all mutual relationships. This is an undirected graph where every pair of connections implies that the connected users mutually follow each other on Twitter. Nodes only in the social network correspond to the white circular node in the social network in Fig. 1. These nodes have only mutual relationships.

**Information Network (IN)**: a network containing all the one-way relationships. This is a directed graph where every pair of connections implies that one user follows the other but not vice versa. Nodes in the IN correspond to the triangular nodes in the IN in Fig. 1 and have only a one-way relationship. There are also those nodes that exist in both networks and have both mutual as well as one-way relationships (shaded nodes in Fig. 1).

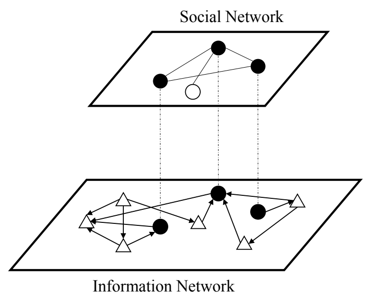



Fig. 1: Network Separation. An arrow from node A to node B in the IN indicates user A follows user B.

The three types of nodes correspond to the three different types of users discussed in the previous section. Although the nodes in the two networks may overlap, the links in the two networks will remain mutually exclusive. It should be noted that we only consider the follower network i.e, only the out degrees of all the nodes since a user's tweets will appear in all its followers' timelines. From the perspective of information propagation, the out degree of a node typically indicates its ability to spread information from the point of view of size. Hence, it is meaningful to study the follower network, rather than the friend (nodes it follows) network.

For our theoretical analysis, the overall Twitter follower network is denoted as a graph $G_a = (V_a, E_a)$. Here $V_a$ denotes the set of all the nodes appearing in the graph, $V_a = \{v|d(v) \geq 0\}$, where $d(v)$ is degree (in and out) of node $v$. As users cannot have negative followers and those with no followers have no ability to spread information, all $v \in V_a$ have a non-negative number of followers. $E_a$ is the set of all the follower relationships in the network. If $e_{ij} \in E_a$, then user $i$ follows user $j$.

We can then define the SN as the graph $G_s = (V_s, E_s)$ and the IN as the graph $G_i = (V_i, E_i)$, with the following relationships:

$V_s \cup V_i = V_a$; $E_s \cup E_i = E_a$; $E_s \cap E_i = \emptyset$ $e_{ij} \in E_s \Rightarrow e_{ji} \in E_s$

Using the idea of network separation, we also investigate if the two subnetworks have a more clear degree distribution. Fundamentally, we want to verify if either one or both of them would better fit a PLD, thus conforming to the characteristics of a human social network. The testing of this hypothesis is described in Sec. IV.

### *B. Generation of Proposed Models*

Based on the observations from the real world process in Sec. II-A and the above analysis, we propose two configurable models to generate a network capturing the degree distribution of the real Twitter network. It is important to note that our goal is not simply to generate a network with its degree following PLD; but rather to find a scalable process which could support a Twitter user behavior as much as possible and thus lead to a similar network distribution at the same time.

*1) Description of Proposed Models:* At each time step, a new node joins the network, making the model a growing network formation. Upon joining the network, the new node selects a subset of nodes, $m$, from the existing nodes to form a relationship. Among these $m$ nodes, some are selected as information sources, while others are selected as friends in the real world. The former ones (step 1) will appear in the IN and are named "information network nodes." The new node forms a directed link with each of these nodes; such links are called "information links." The latter ones (step 2) will appear in the SN and are named "social network nodes" or "mutual followers." These nodes will form two directed links, from the new node to each of them as well as in the opposite direction; such links are called "social links." The formation of links between existing nodes would be equivalent to changing some initial parameters in our proposed model. Thus, if $\alpha$ is the social ratio, there are $(1-\alpha)m$ IN nodes and $\alpha m$ social network nodes. The following two steps are used by a node to connect to the $m$ nodes.

**Select information network nodes**. Based on the preferential attachment scheme, the probability of selecting an existing node as an IN node by the new user is directly proportional to its current in-degree. This process is similar to Twitter recommending popular users to new users. However, it should be noted that the current in-degree of an existing node includes both its IN in-degree as well as its social network in-degree.

**Select social network nodes**. The principles of people selecting SN nodes are largely dependent on their real world social networks making it relatively difficult to model. We consider two options stated below.

(i) *Preferential attachment*: While selecting the SN nodes, the new node will also connect to popular nodes. However, here only the SN in-degree is considered. This is referred to as Model I and the degree distribution of this model should follow a PLD.

(ii) *Multiplicative process*: In [15] it has been shown that a multiplicative process will generate a lognormal distribution. This process is simulated by randomly selecting social nodes from the set of existing nodes and is referred to as Model II.

Note that the process of forming the SN is independent from the formation of the IN, but not vice versa. The effect of these two options for SN node selection is tested in Sec. V. We now sketch a mathematical analysis for the above two models.

*2) Mathematical Analysis:* We assume that $d_k^i(t)$ is the in-degree of node $k$ in the IN at time $t$; similarly, $d_k^s(t)$ is the in-degree of node $k$ in the SN at time $t$, and $d_k(t)$ is the total in-degree of node $k$ at time $t$. When a new node joins at time $t$, the number of new information links an existing node $k$ will gain is

$$\frac{dd_k^i(t)}{dt} = (1-\alpha)m\frac{d_k^i(t)+d_k^s(t)}{\sum_{j=i}^{t} d_j(t)} \tag{3}$$

Similarly, the number of new social links an existing node $k$ will gain, assuming Model I is:

$$\frac{dd_k^s(t)}{dt} = \alpha m \frac{d_k^s(t)}{\sum_{j=i} d_j^s(t)} \tag{4}$$

Solving this equation for the SN,

$$d_k^s(t) = \alpha m(\frac{t}{t_k})^{0.5} \tag{5}$$

where $t_k$ is the time at which node $k$ was added to the system.

Substituting (5) back to (3) gives the rate of increase as

$$\frac{dd_k^i(t)}{dt} = \frac{d_k^i(t)}{At} + \frac{\alpha m}{At_k^{0.5}} \times \frac{1}{t^{0.5}} \tag{6}$$

where $A = \frac{1+\alpha}{1-\alpha}$.Solving the differential equation for the IN gives

$$d_k^i(t) = \frac{2\alpha m}{2-A} \times ((\frac{t}{t_k})^{\frac{1}{A}} - (\frac{t}{t_k})^{0.5}) \tag{7}$$

Comparing with (5), we can observe that there are two power law components (consistent with our hypothesis) and that the resulting scaling parameter is affected by $\alpha$ as well as the network structure of the SN part. This means that a larger $\alpha$ will lead to a larger scaling parameter of the IN. Similarly, for Model II the change in number of social links will be given by:

$$\frac{dd_k^s(t)}{dt} = \frac{\alpha m}{t} \tag{8}$$

The above analysis is similar to the one in [13] and is customized to our context. We omit the details of the derivations for the sake of brevity.

## IV. Empirical Study

### *A. Dataset Information*

The Twitter datasets from [8] and [16] are used in this paper, and are denoted as $D1$ and $D2$, respectively. Table I lists the basic information from the two datasets. From Table I, we can see that the number of Twitter users increased by more than $25\%$ in just three months. Further, it should be noted that these two datasets contain the whole topology of Twitter network at the time they were crawled. The datasets provide a measure of ground-truth of Twitter network, since a large network like Twitter is shaped by all its users. Thus any conclusions reached from a partial or sampled dataset would not be convincing enough. Further, by studying the difference between the two datasets crawled within a close time period of each other could provide insights into the topological analysis as well as the evolution of Twitter.

### *B. Fitting and Results*

To test our hypothesis using real data traces, the SN and the IN are first extracted from the originally unseparated datasets. To our surprise in both the datasets there are only about $50\%$ of Twitter users who have at least one mutual link with another user. This corroborates the conclusion reached in [8] that Twitter has a lower level of reciprocity. The value of reciprocity of $22\%$ in $D1$ ($21.6\%$ in $D2$) is low when compared to other OSNs like Flicker and Yahoo! 360. The difference between the maximum and the average degree in the SN also suggests that the heterogeneity among nodes could possibly lead to the SN fitting a PLD. Table I presents the statistics for the information and the social networks. From it we can see that only a small fraction of Twitter users in the IN have no followers and there is a significant difference in the average degree and maximum degree. Although $D2$ is larger in size than $D1$, both exhibit the same basic properties in their separated networks.

TABLE I: Twitter Network Statistics

| Attribute | $D1$ | $D2$ |
|---|---|---|
| Total users | 41,652,230 | 52,579,682 |
| Total Links | 1,468,365,182 | 1,963,263,821 |
| Social Network Characteristics | | |
| Users in social network | 22,580,393 | 26,866,589 |
| Average degree in social network | 23 | 25 |
| Maximum degree in social network | 698,112 | 713,207 |
| Information Network Characteristics | | |
| Users in information network | 38,355,089 | 47,175,611 |
| Average degree in information network | 24 | 26 |
| Maximum degree in information network | 2,997,304 | 3,503,476 |

Our next step is to check whether the degree distribution of these two subnetworks follows a PLD, by plotting the CCDF in the normal scale. As mentioned in Sec. II-B, the fat tail feature is expected to be observed in the normal scale. Figures 2(a) and 2(b) show the plot of the CCDF of the individual subnetworks of $D1$; the CCDF plots of $D2$ are similar and hence are not shown.

In order to further confirm our hypothesis, the exact scaling parameters need to be calculated. Following the process described in [12], the datasets are fitted into PLD, and the scaling parameter as well as the goodness of fit compared to other candidate heavy-tail distributions (exponential/lognormal), are calculated. The fitting results are shown in Table II, where the comparison with two alternative distributions is indicated by the unnormalized likelihood ratio of PLD to exponential and lognormal shown by columns three and four respectively. As described in [12], a positive value indicates that the fit follows PLD, whereas a negative value indicates that the fit follows the other heavy-tail distribution, namely, exponential or lognormal.

TABLE II: Power Law Fitting of Social & Information Network

| Network | Scaling Parameter | Exponential | Lognormal |
|---|---|---|---|
| Social Network in $D1$ | 1.87 | 293 | -18 |
| Social Network in $D2$ | 1.88 | 309 | -24 |
| Information Network $D1$ | 2.24 | 34 | 28 |
| Information Network $D2$ | 2.15 | 155 | 10.7 |

Basically, the IN is a good fit of PLD with the scaling parameter equal to 2.24 in $D1$ and 2.15 in $D2$ compared to the exponential and lognormal distributions. However, for the SN the power law fitting does not showcase a better fit over lognormal distribution. In fact, as the fitting algorithm is not deterministic, the power law is a better fit to the lognormal only some of the times. This is another reason why two different models are proposed for these two possibilities in Sec. III-B1.

## V. Simulation and Results

**Simulation Setup:** All simulations start with an initial network containing $m_0$ nodes, fully connected with each other, in order to mimic the launch process of Twitter. As described in Sec. III-B1, the network formation continues by adding one node at each time step. On joining the network, a new node selects information and social type nodes to connect to, and the network gets updated. Once a node is selected as one of the two types it cannot be selected again. All simulations stop when the network size reaches 0.6 million, as after this the network structure is observed to be stable.

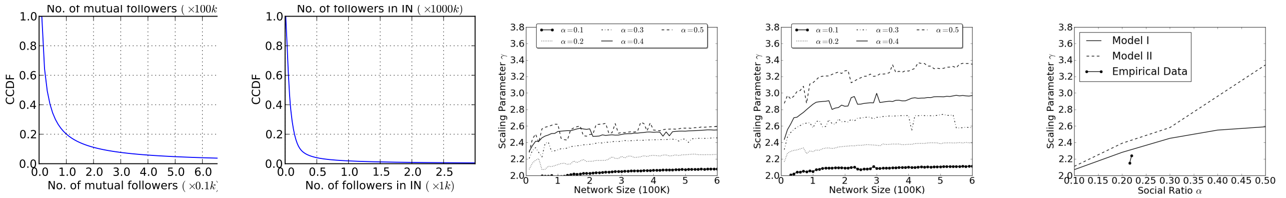
**Effect of Fixed $\alpha$:** Fig. 2(c) and Fig. 2(d) show the effect of $\alpha$ as the network evolves, for the models proposed in Sec. III-B1. In these simulations, both $m$ and $m_0$ are set to 20. Different values of $\alpha$ are selected such that both $\alpha m$ and $(1-\alpha)m$ are integers and tested on both the models. Although there are some oscillations in the curves, in general, all of them show the same trend for the scaling parameter $\gamma$. $\gamma$ increases as the network begins to evolve and eventually saturates at a stable value $\gamma_s$. The saturation scaling parameter $\gamma_s$ is our focus since it occurs when the network size approaches infinity. Fig. 2(e) shows $\gamma_s$ for the two models as well as for the empirical datasets $D1$ and $D2$ [2]. We can observe from the figures that a larger $\alpha$ produces a larger scaling parameter $\gamma$, which is consistent with our mathematical deduction. To compare the two models, closer attention should be paid in Fig. 2(c) and Fig. 2(d) to the lines when the "social ratio" $\alpha$ is equal to 0.2 since they are closest to the overall $\alpha$ in the empirical dataset, which is reported to be 0.22 in [8]. In Fig. 2(e), $\gamma_s$ is 2.28 for Model I and 2.39 for Model II. Since the empirical IN has a scaling parameter of 2.24 when $\alpha = 0.22$ (in $D1$) and 2.15 when $\alpha = 0.216$ (in $D2$), it can be concluded that Model I is a better fit for generating the desired IN.

However, the fitting of the SN part remains an open question. As analyzed in Sec III-B2, although the SN is formed independent of the IN, it does influence the fitting of the scaling parameter on the IN. Since changing the selection of SN nodes from random to preferential attachment decreases $\gamma_s$ from 2.39 to 2.28 (from Fig. 2(e)), it is reasonable to hypothesize that a SN with a lower $\gamma$ will further decrease $\gamma_s$ progressively, yielding a value even closer to 2.24 or 2.15. The preferential attachment scheme cannot produce an undirected network with scaling parameter around 2, suggesting that human factors outside the scope of degree should be considered.

**Different $\alpha$ for Different Users:** In Sec. II we assumed that all users have the same social ratio $\alpha$.We test our assumption's validity and the goodness of this approximation, by randomly picking an $\alpha$ value from the set [ 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.2, 0.2, 0.5, 0.5, 0.4, 0.4, 0.6, 0.8, 1, 1, 1, 1 ] for a new node when it joins the network (Fig. 3(a)). This distribution is similar to that calculated from the empirical dataset. We observe that a large fraction of the users have no or very small percentage of mutual friends, and some values of $\alpha$ are more frequent than others. Fig. 3(b) shows that the plot of $\gamma_s$ is almost the same when all nodes either have the $\alpha$ value fixed at 0.3 or when they draw this value from a distribution with an expected value of $\alpha = 0.3$.

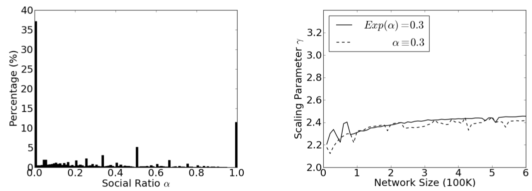
**Effect of Initial Network Size:** Fig. 4 shows the effect of initial network size $m_0$ on the resulting scaling parameter during network formation. We observe that the initial network size does not affect $\gamma_s$ but does influence the speed of reaching the stable stage. A large initial network size takes a longer time to reach a saturation scaling parameter. This influence is obvious in Figure 4; when $m_0$ is equal to 80, the network did not reach a saturation stage before there are 0.6 million nodes.

[2]The two points correspond to the $\gamma_s$ values calculated by obtaining the $\alpha$ values from datasets $D1$ and $D2$.




(a) Social Network (b) Information Network (c) Effect of $\alpha$ on $\gamma$ (Model I) (d) Effect of $\alpha$ on $\gamma$ (Model II) (e) $\gamma_s$ for the models, $D1$ & $D2$

Fig. 2: 2(a) and 2(b) represent the CCDF of networks in $D1$ zoomed to 1000x; 2(c), 2(d) and 2(e) show the effect of social ratio ($\alpha$) on the scaling parameter ($\gamma$).




(a) Distribution of $\alpha$ in $D1$ (b) Fixed $\alpha$ compared with different $\alpha$

Fig. 3: Analysis of social ratio.

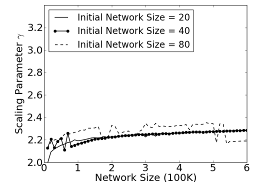



Fig. 4: Effect of Initial Network Size $m_0$ on $\gamma$ as the network evolves

## VI. Comparison to Related Work

Analysis based on topological datasets has led to inference on followers, friends, geographic distributions, etc. [3]–[5]. Similarly, content-based datasets' analysis has led to insights such as prediction of the content and users who retweet information, etc. [1], [6]. Authors of [17] use both types of datasets to understand the structure and interaction between students using Twitter without providing a formal model. Our work differs from the rest in two aspects. First, we use the formation process to propose a model and empirically validate it. Second, we present a concise two-step model that can generate a structure similar to Twitter and identify the parameters that can affect the structure of this network.

## VII. Conclusions and Future Work

In this paper, we challenge past researchers' conclusion that the Twitter network does not follow a PLD through mathematical and empirical analysis. We hypothesize that the Twitter network comprises of two subnetworks that follow the PLD and present a formal model as well as a two-step configurable model. We validate the hypothesis by extracting the two subcomponents from two large data sets and fit them into a PLD. Finally, we also identify and test the effects of parameters such as the social ratio $\alpha$ and scaling parameter $\gamma$ on the network. Our results suggest that despite being thought of as one, Twitter is different from other OSNs and does not possess their endemic characteristics. The formal model of the Twitter network we present can help in the creation of a comprehensive simulator to quantitatively analyze information propagation.

## Acknowledgment

This research is supported in part by National Science Foundation Grant No. DUE-1241709. The authors also thank Abhishek C Desikan for his help with this manuscript.

## References

[1] B. Suh, L. Hong, P. Pirolli, and E. H. Chi, "Want to be retweeted? large scale analytics on factors impacting retweet in twitter network," in *SocialCom, IEEE*, Aug 2010, pp. 177–184.

[2] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury, "Twitter power: Tweets as electronic word of mouth," *Journal of the American society for information science & technology*, vol. 60, no. 11, pp. 2169–2188, 2009.

[3] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter: Understanding microblogging usage and communities," in *Proc. of the 9th WebKDD and 1st SNA-KDD*, 2007, pp. 56–65.

[4] B. A. Huberman, D. M. Romero, and F. Wu, "Social networks that matter: Twitter under the microscope," *arXiv preprint arXiv:0812.1045*, 2008.

[5] B. Krishnamurthy, P. Gill, and M. Arlitt, "A few chirps about twitter," in *Proceedings of the first workshop on Online social networks*, 2008, pp. 19–24.

[6] D. M. Romero, W. Galuba, S. Asur, and B. A. Huberman, "Influence and passivity in social media," in *Proc. of WWW*, 2011, pp. 113–114.

[7] Wikipedia, "Twitter," http://en.wikipedia.org/wiki/Twitter, April 2014.

[8] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in *Proc. of 19th WWW*, 2010, pp. 591–600.

[9] P. Gupta, A. Goel, J. Lin, A. Sharma, D. Wang, and R. Zadeh, "Wtf: The who to follow service at twitter," in *Proc. of the 22nd WWW*, 2013, pp. 505–514.

[10] A. Clauset, C. R. Shalizi, and M. E. Newman, "Power-law distributions in empirical data," *SIAM review*, vol. 51, no. 4, pp. 661–703, 2009.

[11] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proc. of the 7th ACM SIGCOMM IMC*, 2007, pp. 29–42.

[12] J. Alstott, E. Bullmore, and D. Plenz, "powerlaw: a python package for analysis of heavy-tailed distributions," *arXiv preprint arXiv:1305.0215*, 2013.

[13] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[14] M. O. Jackson, *Social and economic networks*. Princeton University Press, 2010.

[15] M. Mitzenmacher, "A brief history of generative models for power law and lognormal distributions," *Internet mathematics*, vol. 1, no. 2, pp. 226–251, 2004.

[16] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, "Measuring user influence in twitter: The million follower fallacy," in *Proc. of ICWSM*, Washington DC, USA, May 2010.

[17] K. Stepanyan, K. Borau, and C. Ullrich, "A social network analysis perspective on student interaction within the twitter microblogging environment," in *ICALT IEEE*, 2010, pp. 70–72.