Cost of Collaboration vs Individual Effort in Social Networks

S. Anand, R. Chandramouli and K. P. Subbalakshmi
Department of ECE, Stevens Institute of Technology, Hoboken, NJ 07030
Email: {asanthan,mouli,ksubbala}@stevens.edu

***Abstract*—We study the dynamics of social networks in terms of population growth and control of user behavior. Most of the current research in social networks focus on static analysis through graph theoretic models to represent the networks or focus on modeling the traffic. Here, we study the cost of collaborative vs individualistic behavior of users in order to grow their network size in a social network. Each user incurs a cost (monetary or emotional) for collaboration. We formulate the behavior of the users as a non-linear optimization problem with a cost. The objective function of the optimization problem is obtained using a stochastic analysis of population growth in social networks, based on the first-passage time of a birth-death process. The stochastic model is validated by comparison with real data obtained from Twitter Results indicate that a homogeneous social network (in which users have similar characteristics) will be individualistic. However, heterogeneous social networks (users with different characteristics) exhibit a threshold effect, i.e., there is a minimum cost, below which the network is as collaborative as desired and a maximum cost above which the network is individualistic as required.** *To the best of our knowledge, this is one of the first analysis of dynamics of user behavior and temporal population growth in social networks.*

Index Terms – Social networks, growth, behavior, individualistic, collaborative.

I. Introduction

The advent of online social networks (OSNs) that links users based on professional interests (e.g., LinkedIn [1]) or personal interests (e.g., Facebook [2], Orkut [3], MySpace [4], Twitter [5], etc) have given rise to new avenues of research due to the availability of data to perform the analysis [6]-[21]. While most approaches in the literature discuss static analysis of social networks (e.g., user relationships [7],[8][10]) and traffic models [17]-[21], the dynamics of social networks in terms of user cooperation or the temporal growth of the network have not been done in detail. Here, we present

- a stochastic model to study the growth of social networks and
- its extension to a non-linear optimization model to control collaborative vs individualistic user behavior so that users grow their networks collaboratively or in an individualistic manner.

Social networks were traditionally modeled as graphs with each user representing a node and two nodes joined by an edge if the corresponding users share a friendship relation. This model later led to the analysis of user groups using the "friend of a friend" relation [7]. Recent studies on graph theoretic models for social networks can be found in [8], and most recently, in [10]. Dynamic behavior of social networks has been studied in terms of varying user relationships by modeling social networks as random graphs and using the spectral properties of the adjacency matrix to model the randomness. Chen [11] presented a set-covering approach to model the influence of users in a social network. Popularity models for users using EM algorithms can be found in [12].

The graph theoretic models for social networks were further investigated to study the properties of the graphs (e.g., [13]-[15]). Leskovec *et al* studied the evolution of node degrees in an online social network graph and showed that the diameters of the graphs decreases as $O(\log(\log N))$, where N is the number of nodes in the graph. This result was further reinforced by Kumar *et al* [14], where they studied data from Yahoo and Flickr and exhibited a power law behavior in the degree of nodes. In [15], Leskovec *et al* provided heuristics for the vertex cover problem to identify popular users and messages that cover maximum content on a topic in OSNs. Co-operation of users in Wikipedia was studied in [16]. It was shown that the number of edits in Wikipedia follows a log-normal distribution.

Another topic of research in social networks in has been the characterization of the traffic arriving into these networks, like e-mail [17] and video traffic [18]. In [17], Malmgren *et al* collect statistics on the e-mails sent from two different countries and universities. They fit a hidden Markov model to characterize the e-mail traffic during the busy and the inactive hours. Philipa *et al* [18] study the characteristics of video traffic in YouTube and its effect on social networking. In [19], Nazir *et al* study the impact of third party applications on Facebook. Heyman and Lucantoni [20] argued that although internet traffic exhibits self-similarity, they can be modeled as a Markov modulated Poisson process (MMPP). In [21], we presented an architecture to collect Twitter messages ("tweets") sent to particular users and showed the the number of tweets sent to a user can me modeled as a Poisson process. Most of the studies mentioned above focus on the static analysis of social networks, i.e., user relationships or the characterization of the traffic.

An important aspect of human society is the study of individual and collaborative group behaviors [23]. Users collaborate with others for mutual benefit such as in political scenarios [24], or for growth in professional careers [25] or for emotional support [26]. Some of these decisions could also involve ethical issues, e.g., [27]. While the large volume of data in online social networks has been applied to social network analysis [28], the dynamics of user behavior in the form of collaboration between users in groups and the growth of groups have not been studied in detail. In OSNs, users can benefit based on several factors, one of them being the size of their "friend" or "follower" network.

In this paper, we study the behavior of users in group in terms of growing the size of their online social network. This can be achieved by individual efforts or by collaboration with other users. We provide a non-linear optimization model according to which users collaborate with other users in the network to grow their individual network size. Users incur a cost (which could be monetary in terms of paying money or emotional in terms of an obligation) when collaborating with other users. The cost depends on a parameter, C. Results indicate that in a homogeneous network, i.e., a network in which all users have similar characteristics, users are individualistic and do not collaborate with other users. In heterogeneous social networks, i.e., networks of users with different characteristics, a threshold effect is observed where in bounds on the parameter, C, can be obtained that control the behavior of the group. Specifically we show the existence of a lower bound, such that for values of C less than the lower bound, the network can be as collaborative as desired. Similarly, we obtain an upper bound, on C, for values above which, network is as individualistic as required. Thus, user behavior can be controlled by controlling C.

The objective function of the optimization problem is obtained by a stochastic analysis that models the growth of social networks. We characterize the growth and decay of a social network using the first passage time of a birth-death Markov process. We validate our model by comparison with data collected from Twitter, using the APIs and architecture we developed in [21]. We choose Twitter because it is the OSN that has had the highest growth in popularity [29]. The results of the stochastic analysis indicate that less popular users benefit more through collaboration while more popular users benefit by individual efforts. *To the best of our knowledge, this is one of the first analysis of dynamics of user behavior and temporal population growth in social networks.*

The rest of the paper is organized as follows. Section II presents the stochastic model and the analysis to study the dynamics of the population growth in social networks. Numerical examples on the stochastic model are presented in Section III. In Section IV, we present the non-linear optimization problem that controls the behavior of users in a social network. Additional applications of our analysis are presented in Section V and conclusions are drawn in Section VII.

II. Stochastic Analysis

We collect the tweets on the topic, "Haiti" from Twitter, using the data collection architecture described in [21]. We start with an initial size of members, Q, that tweet on "Haiti". We then count the number of new users that tweet on Haiti every minute. In [21], it was shown that the arrival of tweets to a user (or on a topic) is a Poisson process. Here, we first measure the arrival rates for varying values of the initial population, Q. Fig. 1 shows that the arrival rate of tweets on a topic depends on the number of users that currently tweet on the topic, i.e., the arrival process of tweets is a state-dependent Poisson process. We use this to model the user population on

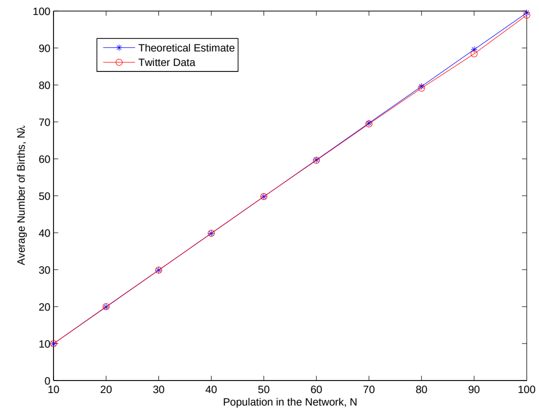


Fig. 1. Arrival rates of the Poisson process corresponding to tweets on Haiti, for various values of the initial population, Q.

a topic as a continuous time Markov chain (CTMC) which represents a birth-death process with state-dependent transition rates. The arrival rate of users at state, N, is $N\lambda$. The interval times between successive deaths (corresponding to users being inactive, i.e., users ceasing from tweeting on the topic) is modeled as an exponentially distributed random variable, with mean $\frac{1}{\mu}$. The death-rate at state, N is therefore, $N\mu$. Let $N(t)$ be the population of the network at time, t, which is the number of active users tweeting on a topic at time, t. Based on the arguments presented above, $N(t)$ as a CTMC corresponding to a birth-death process, with transition rates as shown in Fig. 2.

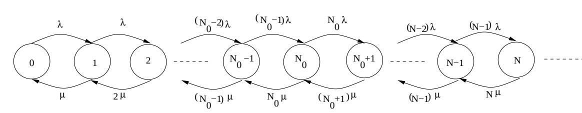


Fig. 2. CTMC model for the number of active users in an OSN group.

Let $\pi_n = \lim_{t\to\infty} \Pr\{N(t) = n\}$ and let $\boldsymbol{\pi} = [\pi_k]_{0\le k<\infty}$ be the steady state stationary probability vector of the occupancy of the CTMC. The local balance equations of the CTMC shown in Fig. 2 are $\lambda\pi_0 = \mu\pi_1$ and for $M \geq 2$,

$$(M-2)\lambda\pi_{M-2} + M\mu\pi_M = (M-1)\left[\lambda\pi_{M-1} + \mu\pi_{M-1}\right]. \qquad (1)$$

From the system of equations listed in (1) and the normalization condition, $\sum_{k=0}^{\infty} \pi_k = 1$, we obtain

$$\begin{aligned} \pi_0 &= \left[1 + \ln\left(\frac{1}{1-\rho}\right)\right]^{-1} \quad \text{and} \\ \pi_k &= \frac{\rho^k}{k}\left[1 + \ln\left(\frac{1}{1-\rho}\right)\right]^{-1} \quad k \geq 1, \end{aligned} \qquad (2)$$

where $\rho \stackrel{\triangle}{=} \frac{\lambda}{\mu}$. The necessary and sufficient condition for the CTMC shown in Fig. 2 to be positive recurrent is $\rho < 1$.

The mean time taken for the population to decay to a specified value, can be determined from the first passage time distribution of the birth-death processes. Consider a general birth-death process where λ_k is the arrival rate or the birth rate when the population of users is k. Let μ_k be the death rate of users in state, k. Let $\Delta_0 = 1$ and for $k \geq 1$,

$$\Delta_k \stackrel{\triangle}{=} \frac{\lambda_0 \lambda_1 \cdots \lambda_{k-1}}{\mu_1 \mu_2 \cdots \mu_k}. \tag{3}$$

Then, the mean time taken for a birth-death process to reach a state, $\tilde{N}$, starting from a state, $N_0 > \tilde{N}$, denoted as $\overline{D}(\tilde{N}, N_0)$, is the *first passage time* from the state, N_0 to $\tilde{N}$, defined as

$$\overline{D}(\tilde{N}, N_0) = E\left[\inf_{\tau > 0}\{\tau; N(\tau) = \tilde{N} | N(0) = N_0\}\right]. \tag{4}$$

If $N_0 = k$ and $t_k \stackrel{\triangle}{=} \inf_t\{t; N(t) = k-1 | N(0) = k\}$, then t_k is obtained as follows. The CTMC stays in state, k for a time, δ_k, which is exponentially distributed with rate, $\lambda_k + \mu_k$. The event that leads the CTMC into leaving state k could be

1) a departure, with probability, $\frac{\mu_k}{\mu_k + \lambda_k}$ or
2) an arrival, with probability, $\frac{\lambda_k}{\lambda_k + \mu_k}$.

When event 1) listed above occurs, it takes the CTMC to state, $k-1$. If event 2) mentioned above occurs, it takes the CTMC to state, $k+1$. The first time instant the CTMC would reach state, $k-1$, would then be the sum of the time taken to first hit state, k starting from state, $k+1$ (i.e., t_{k+1}) and a time, $\hat{t}_k$, which has the same distribution as t_k. Combining these arguments, t_k can be written as

$$t_k = \begin{cases} \delta_k, & \text{with probability } \frac{\mu_k}{\lambda_k + \mu_k} \\ \left(\delta_k + t_{k+1} + \hat{t}_k\right), & \text{with probability } \frac{\lambda_k}{\lambda_k + \mu_k}, \end{cases} \tag{5}$$

where δ_k is an exponentially distributed random variables with rate, $\lambda_k + \mu_k$ and $\hat{t}_k$ is a random variable with a distribution identical to that of t_k. Let $\tilde{f}_k(s)$ be the Laplace transform of the probability density function (pdf) of t_k. From (5), $\tilde{f}_k(s)$ can be written as

$$\tilde{f}_k(s) = \frac{\mu_k}{s + \lambda_k + \mu_k} + \frac{\lambda_k}{s + \lambda_k + \mu_k} \tilde{f}_{k+1}(s) \tilde{f}_k(s). \tag{6}$$

Let $E[t_k^m]$ be the m^{th} moment of t_k. Then [30],

$$E[t_k^m] = (-1)^m \lim_{s \to 0} \frac{d^m \tilde{f}_k(s)}{ds^m}. \tag{7}$$

Combining (6) and (7) for $m = 1$,

$$E[t_k] = \frac{1}{\lambda_{k-1} \Delta_{k-1}} \sum_{n=k}^{\infty} \Delta_n, \tag{8}$$

where Δ_k is as defined in (3). Since $\overline{D}(\tilde{N}, N_0) = \sum_{l=1}^{N_0 - \tilde{N}} E[t_l]$,

$$\overline{D}(\tilde{N}, N_0) = \sum_{l=1}^{N_0 - \tilde{N}} \frac{1}{\lambda_{l-1} \Delta_{l-1}} \sum_{n=l}^{\infty} \Delta_n. \tag{9}$$

From (3) and (9), the mean time taken for a birth-death process with transition rates shown in Fig. 2, to decay to $\tilde{N}$ starting from state, N_0, can be written as

$$\begin{aligned} \overline{D}(\tilde{N}, N_0) &= \frac{1}{\lambda} \sum_{l=1}^{N_0 - \tilde{N}} \frac{1}{\rho^{l-1}} \sum_{n=l}^{\infty} \frac{\rho^n}{n} \\ &= \frac{1}{\lambda} \sum_{l=1}^{N_0 - \tilde{N}} \frac{1}{\rho^{l-1}} \left[\ln\left(\frac{1}{1-\rho}\right) - \sum_{n=1}^{l-1} \frac{\rho^n}{n}\right]. \end{aligned} \tag{10}$$

Note that the expression in (10) is valid only for $\rho < 1$.

For a general birth-death process with birth-rate, λ_k in state, k and death-rate, μ_k, in state, k, the mean time taken for the population to grow to a critical mass, N^*, starting from an initial mass, $N_0 < N^*$, defined as

$$\overline{U}(N_0, N^*) = E\left[\inf_{\tau > 0}\{\tau; N(\tau) = N^* | N(0) = N_0\}\right], \tag{11}$$

can be obtained as follows. Let $\omega_k \stackrel{\triangle}{=} \inf_{t>0}\{t; N(t) = k+1 | N(0) = k\}$ and let $\tilde{g}_k(s)$ be the Laplace transform of the pdf of ω_k. Then, following the arguments used in obtaining (5) and (6),

$$\tilde{g}_k(s) = \frac{\lambda_k}{s + \lambda_k + \mu_k} + \frac{\mu_k}{s + \lambda_k + \mu_k} \tilde{g}_{k-1}(s) \tilde{g}_k(s). \tag{12}$$

Then $E[\omega_k] = -\lim_{s \to 0} \frac{d\tilde{g}_k(s)}{ds}$, i.e.,

$$E[\omega_k] = \frac{1}{\lambda_{k-1} \Delta_{k-1}} \sum_{n=0}^{k-1} \Delta_n, \tag{13}$$

where $\Delta_0 = 1$ and Δ_k is as in (3) for $k \geq 1$. Since $\overline{U}(N_0, N^*) = \sum_{l=1}^{N^* - N_0} E[\omega_l]$, it can be written as

$$\overline{U}(N_0, N^*) = \sum_{l=1}^{N^* - N_0} \frac{1}{\lambda_{l-1} \Delta_{l-1}} \sum_{n=0}^{l-1} \Delta_n, \tag{14}$$

Applying (14) to the CTMC in Fig. 2,

$$\overline{U}(N_0, N^*) = \frac{1}{\lambda} \sum_{l=1}^{N^* - N_0} \frac{1}{\rho^{l-1}} \left(1 + \sum_{n=1}^{l-1} \frac{\rho^n}{n}\right). \tag{15}$$

Note that the expression in (15) is valid both for $\rho < 1$ as well as for $\rho \geq 1$.

It is also observed from (10) and (15) that the mean time to decay and build, depend on the initial population size, N_0 and the final population ($\tilde{N}$ for decay and N^* for growth). However, it is also noted that the mean times depend only on the differences, $N_0 - \tilde{N}$ and $N^* - N_0$. Therefore, for the same amount of decay (i.e., same value of $N_0 - \tilde{N}$) or growth (i.e., value of $N^* - N_0$), the mean time to decay or grow does not depend on the initial population, N_0.

III. Numerical Examples

We run $Q = 100$ queries each minute to Twitter using the experimental set up described in [21], on the topic, "Haiti". At the k^{th} minute, let there be A_k number of arrivals and D_k number of deaths. The number of actively tweeting users in the k^{th} minute, N_k, is obtained as $N_k = N_{k-1} + A_k - D_k$, with $N_0 = Q = 100$. From this, we compute the number of times the population builds to a critical mass, N^* or to decays to a specified value, $\tilde{N}$ and the time taken each for each such occurrence. We use this to obtain the average time for the population to build to a critical mass, N^* or decay to a specified value, $\tilde{N}$. This gives the parameters corresponding to the Twitter data. We compute the same parameters using the expressions in (15) and (10), respectively and compare them with those obtained from our experiments.

Fig. 3(a) presents the mean time taken for a social network to grow to a critical mass, N^*, starting from an initial mass, $N_0 = 100$. Fig. 3(b) presents the mean time for the population for decay to a specified value, $\tilde{N}$, starting from $N_0 = 100$. The arrival rate or the birth rate and the death rate are measured from the data and are found to be 0.89 per user per minute and 0.94 per user per minute, leading to $\rho = 0.95$. It is observed that the values obtained from the Twitter data matches closely with those obtained by the stochastic analysis presented in Section II.

The analysis discussed in II can be applied to individual users instead of networks. A user typically has an initial network size (e.g., an initial list of friends or an initial list of followers). The user may want to expand the list of friends. This could be done by inviting other users (outside of its existing list of friends) to be a friend or by collaboration with the existing list of friends by seeking referrals. "Growth" for individual users then represents the growth in the size of their personal network, which could be the number of friends they have in the network or the number of friends following them.

The activity of a user that can add to the population in the network of the user is the invitation to other users to join the network. This activity level is analogous to the arrival rate of tweets on a topic. Similarly, some people who join the user's network may eventually leave, e.g., some people may stop following the user. This is analogous to the deaths or lack of tweets on a topic. The ratio between the rate at which people join the user's network and that at which people leave, is analogous to ρ of a topic in Twitter. Based on the discussion above, we present the following definitions

Definition 3.1: The *activity level*, λ, of a user is the rate at which people arrive into the network of a user.

Definition 3.2: The *popularity* of a user is the ratio, ρ, of the rates at which people enter the network of a user and that at which people leave the network of a user.

The popularity is the ratio between the success and failure rates of the user. This is analogous to the ratio, $\rho = \lambda/\mu$ of tweets on a topic in Twitter. Thus, $\rho > 1$ indicates a *more popular* user because people join the network of the user faster than they leave. Similarly, $\rho < 1$ indicates that people leave the network of the user faster than they join and hence, a *less popular* user. The case, $\rho = 1$ represents a *popularity-neutral user*. The expression in (15) can therefore be used to determine the mean time taken for a user to grow its network to a desired size, starting from an initial mass. It is also observed from (15) that the mean time for the network of a user to build or decay depends both on the activity level of the user as well as the popularity of the user.

In order to study the growth process of more popular and less popular users (as defined in Definition 3.2) in detail, we consider growth of two types- (a) growth by "individual effort", which is the case when a user aims to grow from a network size, N_0, by a size, N^* (i.e., to a size $N_0 + N^*$) and (b) "collaborative" growth, which is the case when network of a user grows from an initial mass, N_0, by an intermediate size, $\hat{N}$, and then goes on to grow from $N_0 + \hat{N}$ to the desired critical mass, $N_0 + N^*$, i.e., grows by a size, N^* in two steps, first by a size $\hat{N}$ and then by a size, $N^* - \hat{N}$.

Figs. 4(a) and 4(b) compare the collaborative growth and that by individual effort, for the cases, $\rho = 0.95$ (i.e., a less popular user) and $\rho = 1.1$ (i.e., a more popular user), respectively. For both the cases, we show the results for $\lambda = 0.89$ per user per minute and $\hat{N} = \lfloor N^*/2 \rfloor$. It is observed from Fig. 4(a), that for less popular users, the growth by collaboration is much faster than that by individual effort. For instance, it takes around 2000 minutes (36 hours) for a growth by individual effort by a size, 70 where as, the growth by 35 (i.e., 100 to 135) and then by another 35 (135 to 170) takes about 200 minutes (about 3 and half hours). The intuitive reasoning behind this is as follows. Unpopular users require longer time to create a trust in people thus grow to a desired mass. Here, collaboration help people build the trust. On the contrary, popular users (those with $\rho > 1$, already find the trust of people and reach their desired mass faster by individual efforts than by collaboration, as observed from Fig. 4(b). As an example, it is observed that individual effort from an initial mass of 100 by 70 takes about 60 minutes while that by collaboration takes about 85 minutes. It is also observed from Fig. 4(b), that for small values of N^* collaborative growth is faster than individual effort for the case, $\rho > 1$ as well. This is because, the popularity of a more popular user plays a significant role only when the desired size in growth is large. For small values of N^*, the popularity of the individual does not play a significant role.

In general, in a network with multiple users, depending upon the relative activity and popularity levels of the individual users, each user can grow faster either by individual effort or by collaboration with other users in the network. It then raises the following questions

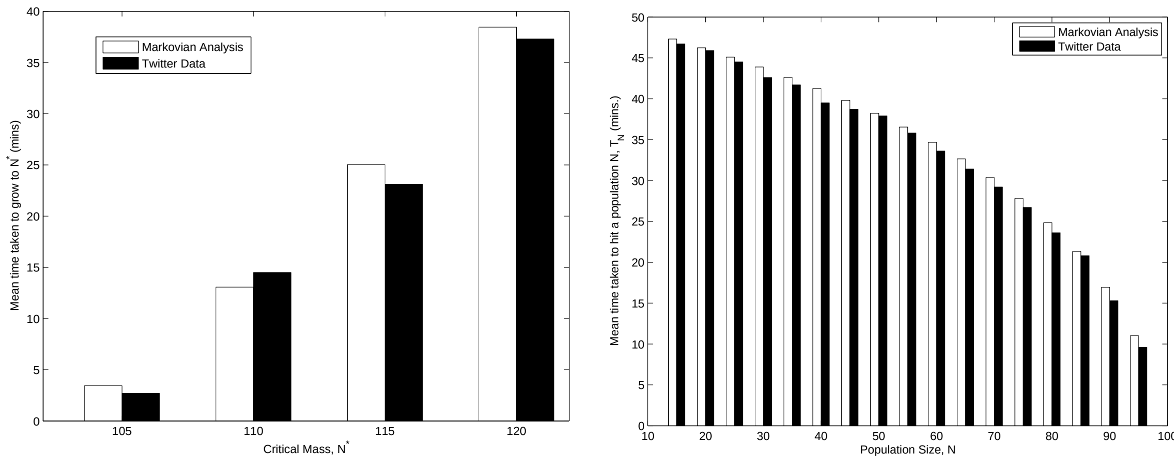
1) *When do users collaborate?*
2) *Who do they collaborate with?*
3) *To what extent do they collaborate?*

We answer these questions in the following section, where we formulate the behavior of users (i.e., collaboration vs individualistic behavior) as a non-linear optimization problem with a cost.

IV. Controlling Group Behavior

Consider a network in which a user can collaborate with K other users. This is represented as a node of degree, K, in a graph.

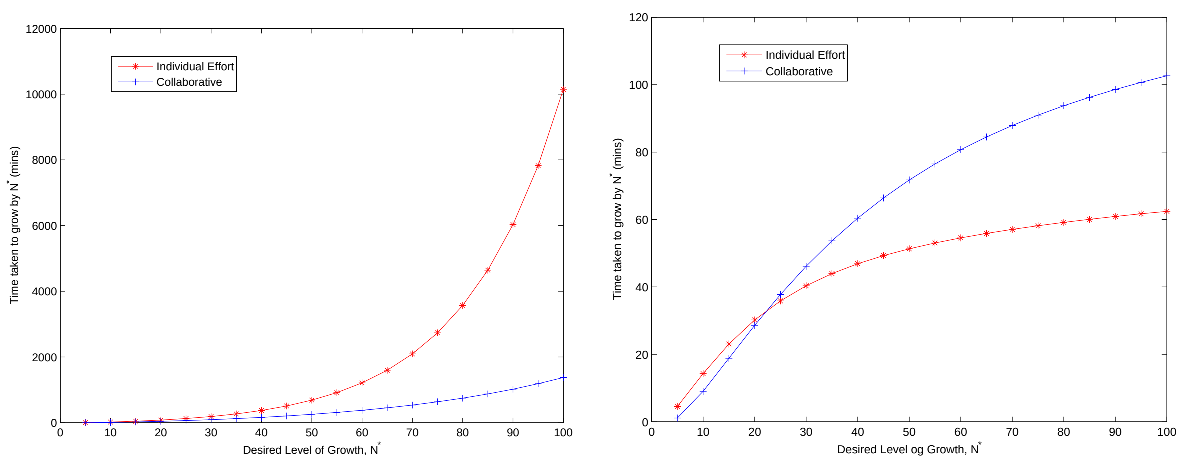
Definition 4.1: The j^{th} user with whom the user of interest can collaborate $(1 \leq j \leq K)$ is called its j^{th} *neighbor*.



(a) Growth process

(b) Decay process

Fig. 3. Mean time taken for the population in a social network to grow/decay.



(a) $\rho = 0.95$

(b) $\rho = 1.1$

Fig. 4. Mean time taken for a user to grow by a strength, N^*, starting from an initial mass, $N_0 = 100$.

Let the activity level of the user (as defined in Definition 3.1) be λ and that of its j^{th} neighbor be λ_j. Let the popularity of the user (as defined in Definition 3.2) be ρ and that of its j^{th} neighbor be ρ_j, $1 \le j \le K$. The user takes a time, $\overline{U}(N_0, N^*)$, given by (15) to grow from an initial mass, N_0 to a desired critical mass, N^*. Thus, then user desires to grow by a size $N_{\text{target}} = N^* - N_0$. The user can collaborate with the j^{th} neighbor to grow by a size, n_j. The amount of benefit obtained by the user, U_j, due to collaboration with the j^{th} neighbor for a strength, n_j, is the amount of time saved by collaborating with the j^{th} neighbor, i.e.,

$$U_j = \overline{U}(n_j) - \overline{U}_j(n_j), \tag{16}$$

where $\overline{U}(m) \triangleq \overline{U}(0, m)$ in (15) and $\overline{U}_j(m) \triangleq \overline{U}(0, m)$ in (15) with λ replaced by λ_j and ρ replaced by ρ_j. A user also incurs a cost, $C(\rho_j) = Cf(\rho_j)n_j$, where C is a non-negative parameter and $f(\rho_j)$ is some increasing function of ρ_j (a special case is the linear function, $f(\rho_j) = \rho_j$). The cost incurred is the price paid to the j^{th} neighbor for collaboration with that neighbor. This price could be monetary (by paying money to the neighbor for collaboration) or emotional (by having a moral obligation to collaborate with the neighbor later for the benefit of the neighbor). Therefore, the net utility obtained by the user due to collaboration with all its neighbors, $\hat{U}_j$, is

$$\hat{U}_j = \sum_{j=1}^{K} U_j - Cf(\rho_j)n_j, \tag{17}$$

where U_j is obtained from (16). Let $\mathbf{n} \triangleq \begin{bmatrix} n_1 & n_2 & n_3 & \cdots & n_K \end{bmatrix}$. Then the strength by which a user grows by collaboration with each of its K neighbors can be determined by solving the optimization problem

$$\max_{\mathbf{n}} \hat{U}, \tag{18}$$

subject to the constraints

$$\sum_{j=1}^{K} n_j \leq N_{\text{target}}. \tag{19}$$

In the above, the inequality occurs because the user may choose to achieve part of the desired strength, N_{target}, by individual effort. Let $\mathbf{n}^* = \begin{bmatrix} n_1^* & n_2^* & n_3^* & \cdots & n_K^* \end{bmatrix}$ represent the vector, $\mathbf{n}$ that solves the optimization problem in (18) subject to the constraints in (19).

Definition 4.2: The network is said to be *homogeneous* (or called a *homogeneous network*) if all users have similar characteristics (which could be the activity level, popularity, etc). In this case, it corresponds to $\lambda_j \approx \lambda_k \approx \lambda$, $\forall$ $1 \leq j, k \leq K$. A network that is not homogeneous is said to be *heterogeneous* or called a *heterogeneous network*.

Definition 4.3: A user is said to *collaborate with its j^{th} neighbor*, if $n_j^* > 0$.

Definition 4.4: A network is said to be *δ-individualistic* if a user collaborates with K_c of its neighbors, such that $0 \leq \frac{K_c}{K} \leq \delta \leq 1$. The network is said to be *individualistic* if it is δ-individualistic for $\delta = 0$.

Definition 4.5: A network is said to be *ϵ-collaborative* if a user collaborates with K_c of its neighbors, such that $1 \geq \frac{K_c}{K} \geq \epsilon \geq 0$. The network is said to be *collaborative* if it is ϵ-collaborative for $\epsilon = 1$.

The following theorem provides a characterization of homogeneous networks.

Theorem 4.1: A homogeneous network is individualistic.

Proof: If $\lambda_j \approx \lambda_k$, $\forall$ j, k AND if $\rho_j \approx \rho_k$, $\forall$ j, k, then, from (16), $U_j(x_j) \approx 0$, $\forall$ j and hence, from (17), $\hat{U}_j(x_j) < 0$, $\forall$ $C > 0$. Therefore, in such a scenario, $n_j^* = 0$, $\forall$ j, i.e., $K_c = 0$ and the network is δ-individualistic for $\delta = 0$, i.e., the network is individualistic. ■

Theorem 4.1 indicates that collaboration is not possible in homogeneous networks. It is then of interest to control the collaboration of users in heterogeneous networks, which is done as follows.

The optimization problem in (18) subject to (19) is an integer programming problem [31] and hence, is NP-hard [32]. Therefore, we simplify it as follows. The expression in (15) can be re-written as

$$\overline{U}(N_0, N^*) = \frac{1}{\lambda} \frac{1}{1 - \frac{1}{\rho}} \sum_{n=1}^{N^* - N_0} \frac{1 - \left(\frac{1}{\rho}\right)^{N^* - N_0 - n}}{n}. \tag{20}$$

The optimization problem in (18) subject to (19) can be re-written using the expression for $\overline{U}(n_j) = \overline{U}(0, n_j)$ in (20). We relax the integer programming problem as follows. Let $n_j = N_{\text{target}} x_j$. Here, $0 \leq x_j \leq 1$, represents the fraction of the desired strength that a user grows by collaboration with the j^{th} neighbor. The optimization problem in $\mathbf{n}$ now becomes an optimization problem in $\mathbf{x} \triangleq \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_K \end{bmatrix}$, written as

$$\max_{\mathbf{x}} \sum_{j=1}^{K} \left(\frac{1}{\lambda} \frac{1}{1 - \frac{1}{\rho}} \sum_{n=1}^{N_{\text{target}} x_j} \frac{1 - \left(\frac{1}{\rho}\right)^{N_{\text{target}} x_j - n}}{n} - \frac{1}{\lambda_j} \frac{1}{1 - \frac{1}{\rho_j}} \sum_{n=1}^{N_{\text{target}} x_j} \frac{1 - \left(\frac{1}{\rho_j}\right)^{N_{\text{target}} x_j - n}}{n} - Cf(\rho_j) N_{\text{target}} x_j \right), \tag{21}$$

subject to the constraints

$$\begin{array}{ll} \sum_{j=1}^{K} x_j \leq 1 & \text{and} \\ 0 \leq x_j \leq 1, & \forall j. \end{array} \tag{22}$$

The objective function in (21) can be further simplified by approximating the summation by an integral, to obtain

$$\max_{\mathbf{x}} \sum_{j=1}^{K} U_j(x_j) - Cf(\rho_j) N_{\text{target}} x_j = \max_{\mathbf{x}} \sum_{j=1}^{K} \left(\frac{1}{\lambda} \frac{1}{1 - \frac{1}{\rho}} \int_{z=1}^{N_{\text{target}} x_j} \frac{1 - \left(\frac{1}{\rho}\right)^{N_{\text{target}} x_j - z}}{z} dz - \frac{1}{\lambda_j} \frac{1}{1 - \frac{1}{\rho_j}} \int_{y=1}^{N_{\text{target}} x_j} \frac{1 - \left(\frac{1}{\rho_j}\right)^{N_{\text{target}} x_j - y}}{y} dy - Cf(\rho_j) N_{\text{target}} x_j \right), \tag{23}$$

subject to the constraints, (22). The Lagrangian corresponding to the optimization problem in (23) subject to the constraints, (22) can be written as

$$L(\mathbf{x}, \mu, \gamma) = \left(\sum_{j=1}^{K} U_j(x_j) - Cf(\rho_j) x_j \right) - \mu(\gamma^2 - 1 + \sum_{j=1}^{K} x_j). \tag{24}$$

The following lemma provides a framework to solve for the optimum $\mathbf{x}^* = \begin{bmatrix} x_1^* & x_2^* & x_3^* & \cdots & x_K^* \end{bmatrix}$.

Lemma 4.1: At the optimum $\mathbf{x}^*$, the user obtains equal net utility from all its K neighbors.

Proof: Applying the Karusch-Kuhn-Tucker (KKT) conditions [31] to the Lagrangian in (24),

$$\frac{\partial \left(U_j(x_j) - Cf(\rho_j) x_j \right)}{\partial x_j} = \mu,$$

$\forall$ j. The statement in Lemma 4.1 then follows. ■

Lemma 4.1 then implies that the optimal $\mathbf{x}^*$, can be determined by maximizing each $\hat{U}_j(x_j)i = U_j(x_j) - Cf(\rho_j) N_{\text{target}} x_j$ with respect to x_j. Applying the first order necessary conditions to the objective function in (23), the optimal x_j^* is obtain as a solution to

$$\left(\frac{1}{\lambda} - \frac{1}{\lambda_j} \right) \frac{1}{x_j^*} \int_{z=1}^{N_{\text{target}} x_j^*} \frac{\rho^z - \rho_j^z}{z} dz = Cf(\rho_j) N_{\text{target}}. \tag{25}$$

In order to meet constraints (22), we force $x_j^* = 0$ if $x_j^* < 0$. The optimal n_j^* is obtained as $n_j^* = \lfloor N_{\text{target}} x_j^* \rfloor$. Note that this could result in $\sum_{j=1}^{K} n_j^* = \tilde{N} < N_{\text{target}}$, in which case, the user grows by individual efforts directly (without collaborating with neighbors) for the amount, $N_{\text{target}} - \tilde{N}$.

Note that if $\lambda > \lambda_j$ and $\rho > \rho_j$, then $U_j(x_j) < 0, \forall\, x_j > 0$. Intuitively, this means that a user will have no incentive in collaborating with a neighbor who is less active as well as less popular. Therefore, we only analyze the case when $\lambda < \lambda_j$ OR $\rho < \rho_j$. Also consider the numerical example with $K = 5$ neighbors, $\rho = 0.01$, $\lambda = 0.89$, $N_{\text{target}} = 500$, $f(\rho_j) = \rho_j$ (i.e., linear pricing) and $C = 0.0001$. The values of λ_j, ρ_j and the corresponding optimal n_j^* are tabulated in Table I. It is observed that the most influential neighbor (the neighbor with the highest ρ_j) is over-loaded due to the collaboration with the user of interest. Additional numerical examples we tried also yielded similar results. This is because, from the definition of $\hat{U}_j$ in (17) (with U_j defined in (16)), it is observed that if $C \to 0$, $\sum_{j=1}^{K} \hat{U}_j$ is maximized when $x_{j'}^* = 1$, where $j' = \text{argmax}_j U_j$, which, typically is the user with largest ρ_j. Also, in order to avoid a user to depend only on collaboration, we impose a constraint that the maximum strength for which users collaborate, $\hat{N}_{\text{target}} = N_{\text{target}}/2$. The value of $\hat{N}_{\text{target}}$ can be fixed to any desired value in general. Here we show the results for $\hat{N}_{\text{target}} = N_{\text{target}}/2$. This is done to ensure that users expend at least as much personal effort as the amount of effort they expect from others for their personal growth. This does not affect the analysis as x_j^* is obtained by solving (25) with N_{target} replaced by $\hat{N}_{\text{target}}$.

TABLE I
THE OPTIMAL n_j^* WHEN A USER HAS $K = 5$ NEIGHBORS, $f(\rho_j) = \rho_j$ AND THE COST, $C = 0.0001$.

Neighbor (j)	Influence factor (ρ_j)	Activity level (λ_j)	Optimal n_j^*
1	0.2	0.67	0
2	11	0.43	500
3	0.4	1.1	0
4	1.6	0.79	0
5	2.8	0.41	0

Consider the case when the other parameters remain identical to those shown in Table I, but $C = 200$. In this case it was found that $x_j^* = 0, \forall\, j$. This is because the high value of the cost, C, causes $\hat{U}_j$ to become negative, $\forall\, j$. Intuitively, this means that if users incur a high cost for collaborating with their neighbors, then users tend not to collaborate and tend to grow individually, i.e., network tends to be individualistic. This then raises the following question, *Is there an upper threshold, $C_{\max}(\delta)$, on C such that for costs larger than $C_{\max}(\delta)$ the network is δ-individualistic for any δ, $0 \le \delta \le 1$?* The following theorem and corollary provides the answer to this question

Theorem 4.2: Let

$$C_0^{(j)} \triangleq \left[\frac{1}{f(\rho_j)} \left(\frac{1}{\lambda} - \frac{1}{\lambda_j} \right) \ln\left(\frac{\rho_j}{\rho} \right) \right]^+, \qquad (26)$$

where $[\alpha]^+ = \max(0, \alpha)$ and let $C_{\max} = \max_j C_0^{(j)}$. Then for $C > C_{\max}$, the network becomes individualistic, i.e., $x_j^* = 0$, $\forall\, j$.

Proof: Let $C_0^{(j)}$ be the value of the cost, C so that $x_j^* = 0$. From (25), $C_0^{(j)}$ is obtained as specified in (26) by applying the L'Hospital's rule. From (25) it is also observed that

$$\frac{\partial C}{\partial x_j^*} = \frac{1}{f(\rho_j)} \left(\frac{1}{\lambda} - \frac{1}{\lambda_j} \right) \frac{\rho^{\hat{N}_{rmtarget} x_j^*} - \rho_j^{\hat{N}_{\text{target}} x_j^*}}{x_j^*} < 0, \qquad (27)$$

if $\lambda < \lambda_j$ and $\rho > \rho_j$ or vice-versa (as mentioned earlier). In other words, x_j^* is a decreasing function of C. Therefore, $C > C_0^{(j)} \Rightarrow x_j^* = 0$, i.e., if $C > C_{\max}$, $C_{\max} \triangleq \max_j C_0^{(j)}$, $x_j^* = 0$, $\forall\, j$, i.e., the network becomes individualistic. ■

Corollary 4.1: $\forall\, \delta$, $0 \le \delta \le 1$, $\exists\, C_{\max}(\delta)$, such that for $C > C_{\max}(\delta)$, the network is δ-individualistic.

Proof: For any δ $0 \le \delta \le 1$, Let $K_c^* = \lfloor \delta K \rfloor$. It is possible to fix $\hat{C}$ such that $\hat{C} > C_0^{(j)}$ for atleast K_c^* neighbors. The value of $\hat{C}$ depends on K_c^*, i.e., on δ. Since C is a decreasing function of x_j, $\forall\, j$ from (27), for $C > C_{\max}(\delta) = \hat{C}$, $x_j^* = 0$ atleast for K_c^* neighbors, i.e., the network is δ-individualistic. ■

It is observed that an individualistic network does not represent a good network from a social point of view because for a stable social network, users should be willing to collaborate with one another [25]. Users collaborating with each other yields benefits in the form of the profit a user can obtain by levying a price for each collaboration. However, as observed from Theorem 4.2, a large price levied on neighbors could prevent a user from obtaining a profit because neighbors will not collaborate. Therefore, the cost, C levied on users must be smaller. From Table I it was observed that very small values of C result in users depending entirely on collaboration. *Does there then exist a lower threshold, $C_{\min}(\epsilon)$ on C below which, the network tends to become ϵ-collaborative (as defined in Definition 4.5)* $\forall\, \epsilon$, $0 \le \epsilon \le 1$.*?* The decreasing nature of x_j^* with respect to C also yields the following theorem and corollary which answers the question mentioned above.

Theorem 4.3: $\exists\, C_{\min}$, such that $\forall\, C < C_{\min}$, the network becomes collaborative.

Proof: From (27), x_j^* is a decreasing function of C. Therefore, as C decreases, x_j^* increases and for some $C = C_{\min}$, $x_j^* > 0$, $\forall\, j$ and $S = \sum_{j=1}^{K} x_j^* = 1$. For $C < C_{\min}$, users collaborate with all their neighbors. i.e., the network becomes collaborative. ■

Corollary 4.2: $\forall\, \epsilon$, $0 \le \epsilon \le 1$, $\exists\, C_{\min}(\epsilon)$, such that for $C < C_{\min}(\epsilon)$, the network is ϵ-collaborative.

Proof: The proof is identical to the proof provided for Corollary 4.1. ■

Theorems 4.2 and 4.3 yield the following theorem.

Theorem 4.4: If $C_{\min} < C < C_{\max}$, then the optimization problem in (21) subject to the constraints, (22) will always have a feasible solution.

Proof: The proof follows from Theorems 4.2 and 4.3. ■

Definition 4.6: For any ϵ and δ, $0 \le \epsilon, \delta \le 1$, the network is said to be *hybrid* if $C_{\min}(\epsilon) < C < C_{\max}(\delta)$, i.e., it is neither ϵ-collaborative nor δ-individualistic.

As a numerical example, we consider a user with $K = 100$ neighbors, $\lambda = 0.89$, $\rho = 0.01$ and generate values of ρ_j and λ_j uniformly distributed in $(0, 10)$ and $(0, 100)$, respectively.

We vary C and compute the percentage of collaborative neighbors (i.e., percentage of neighbors with $n_j^* > 0$ which is the fraction of the neighbors for which $n_j^* = 0$ scaled by 100). Fig. 5 presents the variation in the percentage of collaborative neighbors with respect to the cost, C. We fix $\epsilon = 0.85$ and $\delta = 0.1$. It is observed that for $C < C_{\min}(\epsilon) \approx 13$, the network is ϵ-collaborative Similarly, for $C > C_{\max}(\delta) \approx 40$ the network is δ-individualistic. For $C_{\min}(\epsilon) < C < C_{\max}(\delta)$, the network is hybrid. For this example, the network is individualistic for $C > C_{\max} \approx 50$ and collaborative for $C < C_{\min} \approx 10$.

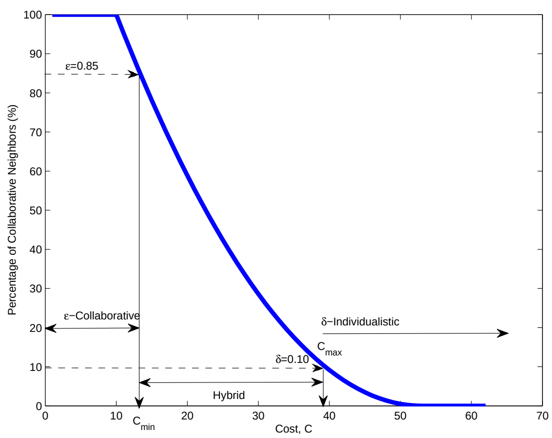


Fig. 5. Behavior of the users in the network with respect to the cost, C.

It is observed that the optimal x_j^* (and hence, the optimal n_j^*) depend on the relative values of ρ_j, λ_j of all the users. Therefore, one cannot conclude that maximum collaboration takes place with the neighbor with the largest influence factor (i.e., highest ρ_j) nor will minimum number collaboration take place with the neighbor with lowest influence factor (least ρ_j). This is illustrated by the numerical example shown in Table II for the scenario when a user has $K = 10$ neighbors, $\rho = 0.01$, $\lambda = \lambda_j = 0.89$, $\forall\ j$, $N_{\text{target}} = 500$, i.e., $\hat{N}_{\text{target}=250}$ and $C = 9.98$. It is observed from Table II, that maximum

TABLE II
THE OPTIMAL NUMBER n_j^* WHEN A USER HAS $K = 10$, $\rho = 0.01$, $\lambda = \lambda_j = 0.89, \forall\ j$ NEIGHBORS, $f(\rho_j) = \rho_j$ AND THE COST, $C = 9.98$.

Neighbor (j)	Influence factor (ρ_j)	Optimal n_j^*
1	11	28
2	0.05	8
3	133	34
4	2.1	18
5	29	37
6	7.8	23
7	61	48
8	0.021	11
9	0.48	23
10	6.1	20

collaborations take place with the $j = 7^{th}$ neighbor, where as the neighbor with the largest influence factor is $j = 3$. Similarly $j = 8$ is the neighbor with the least influence factor whereas minimum collaboration takes place with the $j = 2^{nd}$ neighbor.

V. Additional Applications

The analysis presented in Sections II and IV find (but not limited to) the following additional applications.

- *Growth of Organizations:* Consider an organization started by a group of N_0 number of people. Let new recruits join the organization and rate, λ and existing employees leave the organization at rate, μ. The joining and leaving of new employees represent a birth-death processes in an online social network (OSN) group. The analysis presented in Section II can therefore be applied to determine the mean time taken for an organization to grow to a desired strength, N^*. Depending on λ and μ, the organization could choose to grow by directly advertising itself or by referrals from existing employees or clients. The analysis in Section IV can be used to determine when the organization would prefer direct advertisements and when it would prefer to grow by referrals.
- *Defense:* Consider two armies, A and B attacking each other at rates, λ_A and λ_B, respectively from each unit of attacking equipment. Let $p_s^{(A)}$ and $p_s^{(B)}$ represent the probabilities of successful attacks from each unit of equipment from the armies A and B, respectively. If R_A and R_B represent the rates at which resources are replenished in armies A and B, respectively, then, the population of defense resources (i.e., defense personnel or equipments) in army A can be viewed as a birth-death process with birth rate, R_A and death rate, $\lambda_B p_s^{(B)}$. Similarly, the population of defense resources in army, B, is a birth-death process with birth rate, R_B and death rate, $\lambda_A p_s^{(A)}$. The analysis on the decay of population described in Section II can be applied to determine the mean time taken for either of the army to destroy the other to a specified population of defense resources. The analysis also provides a measure of the rates at which attacks or replenishment of resources or improvements in precisions of attacks must be made. If multiple armies decide to become each others' allies, then the analysis described in Section IV can be used to determine the amount of ammunition or resources an army can borrow from its allies. Theorems 4.3 and 4.2 and Corollaries 4.1 and 4.2 can be applied to determine the scenarios when armies would tend to form allies and when they would operate independently.
- *Fund Management:* Consider the funds available in the treasury of an organization. Larger amount of existing funds yield larger interests and indicates larger interests in investment in an organization. This then implies that the arrival rate of funds into the treasury of an organization is state-dependent. Similarly, larger amount of funds represents larger spending capability and the rate of expenditure is also state-dependent. We can then apply the analysis in Section II to determine the time

taken for the funds to grow to a required level or decay to a specified level and use this to provide a means for organization to plan their finances in advance. The analysis described in Section IV can then be applied to determine the mutual fund transfer policies between multiple firms.

VI. Acknowledgment

The authors thank Dr. Rohan Perera for developing the APIs to collect the required data from Twitter.

VII. Conclusion

We presented a non-linear optimization model with a cost, C, to control the dynamics of user behavior in social networks. The objective function for the optimization problem was derived using a stochastic analysis. The stochastic analysis was validated by verification with data obtained from Twitter. The key findings in this paper include the following.

- Users do not collaborate in a homogeneous network.
- In a heterogeneous network, the user behavior can be controlled by controlling the parameter, C. The network exhibits a threshold effect in that there are thresholds on C according to which, the network can be made as collaborative or as individualistic as required.
- Less popular users and networks benefit more by collaboration.
- More popular networks and users benefit more by collaboration for smaller desired strength and by individual effort for larger desired strength.

We also presented the applications of our analysis in organizational growth, defense and financial planning. Applications of our analysis to include the maximum referral capacity of users and the parochial nature of users are under investigation.

References

[1] [Online]. Available: http://www.linkedIn.com
[2] [Online]. Available: http://www.facebook.com
[3] [Online]. Available: http://www.orkut.com
[4] [Online]. Available: http://www.myspace.com
[5] [Online]. Available: http://www.twitter.com
[6] D. Jensen and J. Neville, "Data mining in social networks," *Proc., Symposium on Dynamic Social Network Modeling and Analysis*, Nov. 2002.
[7] L. Ding, T. Finin, and A. Joshi, "Analyzing social networks on the semantic web," *IEEE Intelligent Systems: Trends and Contreversies*, vol. 8, no. 6, Nov. 2004.
[8] Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong, "Analysis of topological characteristics of huge online social networking services," *Proc., ACM Conference on World Wide Web*, May 2007.
[9] X. Ying and X. Wu, "On randomness measures for social networks," *Proc., SIAM Conf. on Data Mining (SDM'09)*, Apr. 2009.
[10] Z. Liu, H. Hu, Y. Liu, K. W. Ross, Y. Wang, and M. Mobius, "P2p trading in social networks: The value of staying connected," *Proc. IEEE INFOCOM'2010*, Mar. 2010.
[11] N. Chen, "On the approximability of influence in social networks," *SIAM Jl. of Discrete Math.*, vol. 23, no. 3, pp. 1400–1415, Sep. 2009.
[12] T. Yang, Y. Chi, S.Zhu, Y. Gong, and R. Jin, "Directed network community detection: A popularity and productivity link model," *Proc., SIAM Conf. on Data Mining (SDM'2010)*, 2010.
[13] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over time: Dendification laws, shrinking diameters and possible explanations," *Proc., ACM Intl. Conf. on Knowledge, Discovery and Data Mining (SIGKDD'05)*, Aug. 2005.
[14] R. Kumar, J. Nowak, and A. Tomkins, "Structure and evolution of online social networks," *Proc., ACM Intl. Conf. on Knowledge, Discovery and Data Mining (SIGKDD'06)*, Aug. 2006.
[15] J. Leskovec, A. Krauss, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," *Proc., ACM Intl. Conf. on Knowledge, Discovery and Data Mining (SIGKDD'07)*, Aug. 2007.
[16] D. M. Wilkinson and A. Huberman, "Assessing the value of cooperation in wikipedia," *Informaiton Society Watch, First Monday*, vol. 12, no. 4, Apr. 2007.
[17] R. D. Malmgren, J. M. Hofman, L. A. N. Amaral, and D. J. Watts, "Characterizing individual communication patterns," *Proc., Intl. Conf. on Knowledge, Discovery and Data Mining (SIGKDD'09)*, Oct. 2009.
[18] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "YouTube traffic characterization: A view from the edge," *ACM SIGCOMM Conf. on Internet Measurement*, Oct. 2007.
[19] A. Nazir, S. Raja, D. Gupta, C. Chuah, and B. Krishnamurthy, "Network level footprints of facebook applications," *ACM SIGCOMM Internet Measurement Conf. (IMC'09)*, Nov. 2009.
[20] D. P. Heyman and D. Lucantoni, "Modeling multiple IP traffic streams with rate limits," *IEEE/ACM Trans. on Networking*, vol. 11, no. 6, pp. 948–958, Dec. 2003.
[21] R. W. D. Perera, S. Anand, K. P. Subbalakshmi, and R. Chandramouli, "Twitter analytics: Architecture, tools and analysis," *Proc. IEEE MILCOM'2010*, Oct. 2010.
[22] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," *Proc., IMC'07*, Oct. 2007.
[23] J. Boissevain, "The place of non-groups in social sciences," *Man: Journal of Royal Anthopology Institute*, vol. 3, no. 4, pp. 542–556, 1968.
[24] J. A. Barnes, "Networks and political processes," *Book Chapter: Social Networks in Urban Situations, Manchester University Press*, pp. 51–76, 1969.
[25] R. S. Burt, "The social structure of competition," *Explorations in Economic Sociology*, pp. 65–91, 1993.
[26] B. Cornwell, "The dynamic properties of social support: Growth, decay and stability and their effects on adolesent depression," *Social Forces*, vol. 81, no. 3, pp. 953–978, Mar. 2003.
[27] S. P. B. abd J. Molina, "Ethical and strategic issues in organizational social network analysis," *Journal of Applied Behavioral Science*, vol. 39, no. 3, pp. 337–349, Sep. 2003.
[28] S. Wasserman and K. Fraust, *Social Networks Analysis: Methods and Applications*. Cambridge University Press, 1994.
[29] "Twitter statistics: The full picture." [Online]. Available: http://thenextweb.com/socialmedia/2010/02/22/twitter-statisticsfull-picture
[30] S. M. Ross, *Introduction to Probability Models*. Elsevier Academic Press, 2007.
[31] D. G. Luenberger, *Linear and Non-linear Programming*. Kluwer, 1984.
[32] T. Cormen, C. Leisserson, R. Rivest, and C. Stein, *Introduction to Algorithms*. McGraw Hill, 2001.