On the Rise and Fall of Online Social Networks

Arnab Basu *† Ajit Brundavanam ‡
Simrat Singh Chhabra § Saswata Shannigrahi ¶

December 20, 2021

Abstract

The rise and fall of online social networks recently generated an enormous amount of interest among people, both inside and outside of academia. Gillette [Businessweek magazine, 2011] did a detailed analysis of MySpace, which started losing its popularity since 2008. Cannarella and Spechler [ArXiv, 2014] used a model of disease spread to explain the rise and fall of MySpace. In this paper, we present a graph theoretical model that may be able to provide an alternative explanation for the rise and fall of online social networks. Our model is motivated by the well-known Barabási-Albert model of generating random scale-free networks using preferential attachment or ‘rich-getters-richer’ phenomenon. As shown by our empirical analysis, we conjecture that such an online social network growth model is inherently flawed as it fails to maintain the stability of such networks while ensuring their growth. In the process, we also conjecture that our model of preferential attachment also exhibits scale-free phenomenon.

1 Introduction

The last few decades have seen the emergence of highly popular online social networks like MySpace, Orkut and Facebook. MySpace was founded in 2003 and it gained its peak popularity in 2008 [6]. However, most of the users started abandoning it since 2008 [3, 6]. Cannarella and Spechler [3] modified the SIR model of disease spread (see, e.g., [1]) to explain this phenomenon. The disease spread model has also been used to study the arrival and departure dynamics of the users in online social networks [11].

In this paper, we propose a model that may be able to provide an alternative explanation for the rise and fall of online social networks. To begin with, we note that people join online social networks due to the presence of their friends

*†Bilkent University, Ankara, Turkey. Email: arnab@bilkent.edu.tr
‡Corresponding Author
§Indian Institute of Technology Kharagpur, India. Email: ajitb@iitkgp.ac.in
¶University of Southern California, USA. Email: simratsc@usc.edu

Saint Petersburg State University, Russia. Email: saswata.shannigrahi@gmail.com
in these networks and leave due to the inactivity of their friends [11]. It is also known that a large percentage of online social network friendships are not active or strong friendships between the users [5, 8, 9], where a strong friendship between a pair of online social network friends is indicated by regular communications between them. In fact, the well-known Dunbar’s number [4] says that an individual can comfortably maintain stable relationships with at most 150 other people only. It means that the number of strong friendships of a person is likely to be limited to 150 in any social network, even though there is no bound on the number of friends one can have in most online social networks.

In order to identify strong friendships between people in a mobile communication network, Onella et. al [10] did an empirical study to establish that a greater neighborhood overlap between a pair of friends corresponds to a stronger friendship between them. We assume that such a result is also true for online social networks. The neighborhood overlap between a pair of users $A$ and $B$ in such a network (represented as an undirected graph) is defined as the number of users that are neighbors of both $A$ and $B$ divided by the number of users that are neighbors of either of them. We define a pair of users to be strong friends if their neighborhood overlap is at least a given constant.

In this paper, we study the change in the size of the largest connected component (LCC) in the strong friendship subgraph of an evolving random graph defined below in Section 2. This is motivated by the observation that the nodes in the core of an online social network are more likely to survive than the nodes at the periphery [11]. We consider the LCC in the strong friendship subgraph of an online social network to be its core that is important to retain most of its users.

## 2 Model

Our model is motivated by the well-known Barabási-Albert (BA) model [2] that can be used to generate random scale-free networks based on preferential attachment and growth. The preferential attachment is the property that a node with higher degree than another node is more likely than the other node to get connected to new nodes as the network grows, i.e., a manifestation of the ‘rich-getting richer’ phenomenon.

Starting with an initial collection of $m_0 > 0$ nodes where $m_0$ is a given natural number, the BA model adds one node at a time to the network. In each time step, a new node gets connected to $m \leq m_0$ existing nodes. The probability $p_i(t)$ that the new node at time $t + 1$ is connected to an existing node $v_i$ is

$$p_i(t) = \frac{d_i(t)}{\sum_j d_j(t)}, \quad (1)$$

where $d_i(t)$ is the degree of node $v_i$ at time $t$ and $\sum_j d_j(t)$ is the corresponding sum of the degrees of all nodes in the current network. We assume that the nodes $v_{m_0+1}, v_{m_0+2}, \ldots, v_{m_0+t}, \ldots$ are ordered in the sequence of their additions to the
network, following an arbitrary given ordering \( < v_1, v_2, \ldots, v_{m_0} > \) of the initial \( m_0 \) nodes.

Let \( G_0 \) be the initial (time \( t = 0 \)) graph with \( m_0 \) nodes, and \( G_t \overset{\text{def}}{=} (V_t, E_t) \) be the random graph with the set \( V_t \) of \( m_t = m_0 + t \) nodes after \( t \) nodes have been added along with the addition of \( mt \) edges \( \{e_1, \ldots, e_{mt} \} \) by the BA model, i.e., \( E_t = \{e_1, \ldots, e_{mt} \} \cup E_0 \) where \( E_0 \) is the initial set of edges between the starting \( m_0 \) nodes. For a given \( \epsilon > 0 \), we define two users \( v_i, v_j \in V_t \) to be strong friends at time \( t \) if their neighborhood overlap \( N_t(v_i, v_j) \) at time \( t \) is greater than or equal to \( \epsilon \), i.e.,

\[
N_t(v_i, v_j) \overset{\text{def}}{=} \frac{|\{v' \in V_t : e_{v_iv_j} = e_{v'_iv'_j} = 1\}|}{d_i(t) + d_j(t) - |\{v' \in V_t : e_{v_iv_j} = e_{v'_iv'_j} = 1\}|} \geq \epsilon, \tag{2}
\]

where, for a pair of nodes \( v, v' \), \( e_{vv'} = 1 \) implies the existence of an edge between them and 0 implying otherwise.

Given \( \epsilon > 0 \), the time-\( t \) strong friendship subgraph \( G^*_t \subseteq G_t \) is defined as follows: \( G^*_t \) contains the same set of nodes as that of \( G_t \), and a pair of nodes \( i \) and \( j \) are connected by an edge in \( G^*_t \) if and only if they are connected by an edge in \( G_t \) and their neighborhood overlap in \( G_t \) is greater than or equal to \( \epsilon \), i.e., \( G^*_t \overset{\text{def}}{=} (V_t, E^*_t) \) where \( E^*_t \overset{\text{def}}{=} \{\{v_i, v_j\} \in E_t : N_t(v_i, v_j) \geq \epsilon\} \). At time \( t \), the corresponding LCC in \( G^*_t \) is denoted by \( L_t \). For any set \( S \), \( \|S\| \) denotes its cardinality.

Before proceeding further, we would like to make some clarifications. As is well-noted in the literature (see, e.g., [7, Chap. 8] and references therein), the description of the BA model as originally proposed in [2] is quite imprecise and does not explain several basic facts, namely, how the first edge is connected, what the dependencies are between the \( m \) edges added at any given time \( t \) and, more importantly, how to address the issue that the expected number of edges connected a new incoming vertex at any time \( t \) to earlier existing vertices is not \( m \) (as the BA model claims it to be) but actually 1.

To address these issues, we propose an algorithm which we henceforth refer to as the *Parallel Preferential Attachment* or the PAPA Algorithm or simply PAPA. It works as follows: To start with, it mandates that \( G_0 \) with \( m_0 \) nodes has some given nonzero degrees \( d_{i}(0), \ i = 1, \ldots, m_0 \). If \( m_0 = 1 \), then we initialize \( d_1(0) = 1 \) as a boundary case so that, naturally as required, the new (second) node coming in will surely \( (p_1(0) = 1) \) get connected to this initial existing node, i.e., exactly \( m = m_0 = 1 \) edge will be created between these two nodes. In general, we use the following iterative setup to simultaneously (in parallel) add the \( m \leq m_0 \) edges from a new node \( v_{m_0+t+1} \) at time \( t+1 \) to the \( m_0 + t \) nodes existing till time \( t \). At any time step \( t + 1 \), PAPA runs as follows:

1. We start with a set \( S \equiv \emptyset \).

2. The interval [0, 1] is first divided into \( m_0 + t \) non-overlapping sub-intervals \( [p_0(t) = 0, p_1(t)], [p_1(t), p_2(t)], \ldots, [\sum_{i=0}^{m_0+t-1} p_i(t), \sum_{i=0}^{m_0+t} p_i(t) = 1] \) where \( p_i(t) \) is given by (1) above.
3. Now, a random number $X \sim U[0,1]$ is generated and suppose it falls in the interval $[\sum_{i=0}^{l-1} p_i(t), \sum_{i=0}^{l} p_i(t))$ for some $1 \leq l \leq m_0 + t$.

4. Then, we update $S \leftarrow S \cup \{l\}$ and delete this interval of length $p_l(t)$ from $[0,1]$ subsequently.

5. Now, we stretch each of the remaining intervals by multiplying with $\frac{1}{1-p_l(t)}$.

6. If $m > 1$, the next update to $S$ is done in the same way using these new rescaled intervals for each subsequent edge. These steps are repeated from Step 3 above until all $m$ updates are made.

7. Finally, for each $l \in S$, we add an edge $\{v_{m_0+t+1}, v_l\}$ to the current configuration of $G_t$, i.e., make $e_{\{v_{m_0+t+1}, v_l\}} = 1$.

We refer to the graph generated by this Algorithm as the PAPA Random Graph and note three important points below about this Algorithm:

- Unlike most of the existing literature in this domain (see, again, [7, Chap. 8] and references therein), PAPA eliminates any possibility of multi-edges and self-loops as is obvious from Step 4 above. This property is novel and very crucial for real-life online networks where there is either a real online connection between two persons or there is none, and, physical justification of multiple connections between two persons is unrealistic. Since our primary aim is to study online social networks, we adopt this approach in this paper.

- The second important point is that we do not re-normalize our $p_i(t)$-s during the addition of the $m$ edges at any time point $t+1$, i.e., we hold those values constant at time $t+1$; hence, it does not reflect the inter-dependencies of the $m$ edges being added at that time $t+1$ through probability re-scaling. This actually provides an exact parallel or simultaneous preferential attachment scheme as is realistically done by a new user coming in to join an existing online social network. Moreover, PAPA has lower run-time complexity compared to other (sequential-attachment) algorithms in this domain because the ‘inter-edge-scaling’ of the attachment probability need not be done at every time-step $t$. However, it might be an interesting exercise to adapt our Algorithm to this inter-edge-scaling of the attachment probability and check if there is any significant difference in the results. We believe that as $t$ grows larger the effect of this re-scaling dies of asymptotically and hence the asymptotic limits might behave similarly.

- Finally, as is clear from Step 6 above, we add exactly $m$ edges at every time step $t$ as mandated by the BA Model but was imprecisely defined therein.

We are now in a position to state our first Conjecture of this paper.
Conjecture 1. Given a graph $G_0$ with $m_0 \geq 1$ nodes and an integer $1 \leq m \leq m_0$, the PAPA Random Graph $G_t$ generated from it is without any multi-edges or self-loops and is scale-free, i.e., there exists an integer $k_0 > 1$ such that for all $k \geq k_0$ it holds that

$$\limsup_{t \uparrow \infty} \frac{\| \{ v_i \in V_t : d_i(t) = k \} \|}{\| V_t \|} \leq \frac{1}{k^\gamma} \ a.s. \quad (3)$$

for some $\gamma > 1$.

3 Simulations and Results

We run our simulations with three different values of $\epsilon = 0.01, 0.05, 0.1$, the threshold for the strong friendship between a pair of users. In each of the three cases, we start with a complete graph with $m_0 = m$ nodes, where $m$ is the number of nodes that every new node is connected to among the existing set of nodes. Through subsequent addition of nodes done by PAPA as above, we obtain and plot the expected size (number of nodes) of $L_t$ in $G^*_t$ against the size (number of nodes) of the graph $G^*_t$ for a given $m_0$ and $\epsilon$. For any graph $G$, $\| G \|$ denotes the cardinality $\| V \|$ of the set $V$ of its nodes.

We observe from the Figures 1, 2 and 3 (see after References) that the expected size of the LCC in $G^*_t$ increases till a point (peak) and starts decreasing thereafter in each of the plots. Based on these observations, we now state our second Conjecture of this paper.

Conjecture 2. There exists a threshold real number $0 < \epsilon_0 < 1$ such that, for all $0 < \epsilon < \epsilon_0$, integers $m_0 \geq 1$, $m = m_0$ and a given complete graph $G_0$ with $m_0$ nodes, there exist finite integers $\tau(\epsilon, m_0) > m_0$ and $m_0 \leq \sigma(\epsilon, m_0) \leq \tau(\epsilon, m_0)$ satisfying the following property for $L_t$ generated by PAPA from $G_0$:

$$E[\| L_t \|]_{t \uparrow \tau(\epsilon, m_0) - m_0} \uparrow \sigma(\epsilon, m_0) \uparrow E[\| L_t \|]_{t \downarrow \tau(\epsilon, m_0) - m_0} \quad (4)$$

Note that the above Conjecture 2 automatically implies that

$$\sup_{t \geq 0} E[\| L_t \|] \leq \sigma(\epsilon, m_0), \quad (5)$$

i.e., $E[\| L_t \|] \leq \sigma(\epsilon, m_0)$ for all time $t \geq 0$.

4 Conclusions

Our observations suggest one possible explanation for the rise and fall of online social networks, i.e., the core (LCC) in such a network with increasing number of nodes starts reducing in size after reaching a peak. Since this might be the effect of preferential attachment where a popular user is likely to befriend a large number of other users, it would be interesting to see whether the same
observation holds for random networks that have a restriction on the number of
friends that each node can have, i.e., where we only allow ‘preferential attachment
with limit’ or ‘limited rich-gets-richer’ connections. Another interesting
open problem might be to check if the size of the LCC increases monotonically
with that of the underlying network if $\epsilon > \epsilon_0$ where $\epsilon_0$ is as in Conjecture 2
above. Estimating explicit bounds for $\tau(\epsilon, m_0), \sigma(\epsilon, m_0)$ in terms of these net-
work parameters might also be looked into. Also, studying the same problem
for $G_0$ to be a non-complete graph and/or $1 \leq m < m_0$ will be interesting. We
believe that Conjecture 2 should hold in such situations.

References

[1] N. Bailey. Mathematical Theory of Infectious Diseases. Mathematics in
Medicine Series. Oxford University Press, 1987.
[2] A. L. Barabasi and R. Albert. Emergence of scaling in random networks. Science
286 (5439): 509-512, 1999.
[3] J. Cannarella and J. A. Spechler. Epidemiological modeling of online social
network dynamics. arXiv:1401.4208 2014.
[4] R. I. M. Dunbar. Neocortex size as a constraint on group size in primates.
Journal of Human Evolution 22 (6): 469-493, 1992.
[5] D. Easley and J. Kleinberg. Networks, crowds, and markets: reasoning about
a highly connected world. Cambridge University Press, Cambridge UK, 2010.
[6] F. Gillette. The rise and inglorious fall of myspace. Businessweek magazine,
2011.
[7] R. van der Hofstad. Random Graphs and Complex Networks. Cambridge Univ.
Press, Cambridge UK, 2018.
[8] B. A. Huberman, D. M. Romero and F. Wu. Social networks that matter:
Twitter under the microscope, First Monday 14 (1), January 2009.
[9] C. Marlow, L. Bryon, T. Lento and I. Rosem. Maintained relationships on Facebook,
http://overstated.net/2009/03/09/maintained-relationships-on-facebook
2009.
[10] J. P. Onnela, J. Saramaki, J. Hyvonen, G. Szabo, D. Lazer, K. Kaski, J. Kertesz
and A. L. Barabasi. Structure and tie strengths in mobile communication net-
works. Proc. Natl. Acad. Sci. USA, 104: 7332-7336, 2007.
[11] S. Wu, A. Das Sarma, A. Fabrikant, S. Lattanzi and A. Tomkins. Arrival and
departure dynamics in social networks. Proceedings of the Sixth ACM Interna-
tional Conference on Web Search and Data Mining (WSDM), 233-242, 2013.
Figure 1: Expected size of the LCC in the strong friendship graph vs. the size of the graph, for $\epsilon = 0.01$.

Figure 2: Expected size of the LCC in the strong friendship graph vs. the size of the graph, for $\epsilon = 0.05$.

Figure 3: Expected size of the LCC in the strong friendship graph vs. the size of the graph, for $\epsilon = 0.1$. 