Threshold Phenomena in the Evolution of Communities in Social Networks

Sreenivas Gollapudi*
Microsoft Research

Krishnaram Kenthapadi†
Microsoft Research

Rina Panigrahy‡
Microsoft Research

ABSTRACT

We provide a theoretical model to explain the evolution of communities in social networks where new friends are formed by looking at friends of friends. Specifically, a new friendship link is formed based on the number of common friends between the two nodes. Additionally, we also allow a fraction of the new links to be random. Furthermore, we allow deletion of links to model degradation of old friendships in such a way as to keep the average number of friends a person has nearly unchanged. The probability of addition of a new link based on the number of common friends is parameterized by an exponent α as in the preferential attachment model. We observe that as α increases, we obtain more refined clustering of the graph. Specifically, our simulations show that for values of $\alpha \in [0, 1]$, this process results in a uniform random graph regardless of the original structure of the graph. As we increase α from one, we observe sudden transitions in the cluster formation at specific values of α. Moreover, this threshold value lies in the range $[1, 2]$. We study several tree-like backbone networks which are representative of nascent social networks and demonstrate that this process mimics the evolution of communities in such networks.

1. INTRODUCTION

Societies are complex phenomena: Given a collection of individuals, as time progresses, new relationships form between the individuals and old relationships disappear, and communities form as alliances appear. However, the formation of alliances are complex and often unpredictable. Some pertinent questions that arise include: *How do communities evolve given an initial structure of a nascent social network? Given the structure of a nascent social network, can we predict what communities are likely to form over time? How will the shape of the social network change over time? What is a good model for such predictions?* These questions have received a renewed interest in the context of online social networks such as the web, or friendship networks such as `myspace`, `orkut`, and `facebook`. The work in this area is motivated, in part, by the huge opportunity such social networks present in the context of online advertising markets and other web services such as online recommendation systems such as `amazon.com` and online auction markets like `ebay.com`. The structure of the social network often encodes a lot of information about interactions between nodes in the network. Understanding the structure and the evolution may also help us protect such networks from attacks by spammers who may be exploiting the network links to propagate malicious spam. In other examples such as viral marketing, understanding the evolution of the topology of the network often forms the basis for studying diffusion of information through such networks.

Several theoretical models have been proposed for how social networks evolve over time [12, 6, 14, 13, 7, 8]. The essential underlying principle behind most approaches is that new friendships are formed through existing common friends. For instance one popular method adds new friendship links with probability proportional to the number of common friends. Formation of new links may also be compensated by randomly deleting existing friendship links. Surprisingly we find that if new friendships are formed in this fashion, the social network soon loses any correspondence to its initial structure and eventually degenerates into a uniform random graph, one without any distinct communities. On the other hand, if we bias the friendship formation further by making the probability proportional to the number of common friends raised to a certain power α, then depending on the value of α community structures begin to form. The larger the value of α, the more refined the communities become.

Contributions of this study: Given a structure of a social network, wherein the edges represent friendships between the nodes in the network, we present a model for evolution of communities in such networks that is based on the following principles. For simplicity, we assume that the number of nodes in the network is fixed and study the evolution of links with this set of nodes. The new connections are formed based on the number of common friends between the nodes. In addition, we also allow a small fraction of connections to be added randomly between any two random nodes. At the same time, there is an upper bound on the maximum number of friends each node can have [1]. If the new links are formed with probability proportional to the number of common friends, surprisingly, we find that eventually regardless of the initial structure of the graph, it converges to a uniform random graph. However, if the probability of the formation of a new link is proportional to the number of common friends raised to the power of a parameter α, then we find that, above a certain threshold, depending on the structure of the initial network, communities begin to form in the network. We apply our model in backbone networks that are usually a tree connecting several star like components. Many nascent social networks seem to have this structure. We describe the model

*e-mail: sreenig@microsoft.com
†e-mail: krishnaram.kenthapadi@microsoft.com
‡This work was done when the author was at Stanford University. e-mail: rina@microsoft.com

WWW2008, April 21–25, 2008, Beijing, China.
.

[1]If the maximum degree of a node is not constrained and instead, edges are deleted randomly so that the total number of edges is fixed, we observe the rich-get-richer phenomena in which most nodes end up with no friends and a few rich nodes form a clique. This leads to an undesirable scenario where no communities are formed.

in Section 3 and summarize our results in Section 4.

2. RELATED WORK

There is a plethora of work on the structural analysis and evolution of the web-graph [1, 11, 9]. These and other related works focused on different topological properties of graphs such as diameter, connectivity, small-world phenomena which is based on the hypothesis that everyone is connected to everyone else through a path of small length (six degrees of separation). Some studies addressed [14, 8] studied link formation in which new edges are favored between nodes connected by a small distance/paths. Newman [13], in another study, looks at the evolution when edges are added between nodes based on the number of common friends. In our study, we find that this model leads to a uniform random graph and is unable to explain any community formation. Other models include using the Jaccard coefficient which is a commonly used similarity measure between two nodes.

One of the most well-studied model of graph evolution is *preferential attachment* [2] wherein new connections are formed with higher probability among nodes with higher degree. Thus, in this model, the probability p of new node connecting a node i depends on the degree d_i of node i, so that $p = d_i / \sum_j d_j$. However, this ''rich-get-richer" model is not always appropriate to describe the dynamics of online social networks in which pre-existing social relationships in a community define how the community evolves as part of the bigger social network. Jackson and Rogers [6], in fact, present a model which is similar to the one we propose. However, they use their model to study the degree distribution and connectivity properties of the resulting networks. Liben-Nowell and Kleinberg [12] studied the link prediction problem in social networks in which they address the formation of new interactions between members of a social network. They propose that the probability of forming edges between two nodes is a function of the "proximity" of the two nodes. One of the proximity functions they study is the number of common neighbors two nodes share. Their results indicate that this is indeed a good function to predict the probability of future interactions between two given nodes. However, they do not explicitly characterize these interactions and the resulting effect on the topology of the network.

There have been other studies in the physics community that looked at understanding the evolution of networks from latent information in the network [15, 13, 7, 16]. These works have, in general, tried to study the structure of the social networks resulting from a defined set of interactions. In fact, Jin and others [7] proposed models that mimic the growth of social networks very similar to this work – namely, the rate of edge formation between two nodes is proportional to the number of their common friends, relationships decay over time, and the presence of an upper limit on the number of friends an individual can have. However, they do not analyze their model in terms of the topology of the resulting network like we do in this work. Moreover they do not consider the effect of randomness on the evolution of the network topology.

Threshold phenomena have been commonly observed in classical random graph models [3, 4]. For example, when the number of edges in a random graph exceeds a certain threshold, suddenly giant components begin to appear and the diameter of the graph shrinks. In fact, it is known that such sharp threshold phenomena hold for any *monotone* graph property [5].However, classical random graph models may not be appropriate for studying large networks such as the internet and social networks such as `flickr` and `myspace` [10, 12].

3. MODEL FOR FRIENDSHIP

We next formally describe our model for friendship. In each time interval, the probability that a new edge is added between two nodes is dependent on the number of common friends they share. We also allow a fraction of the new links to be random. Further, each node can handle only a bounded number of friends. If the number of links becomes large, the node starts to drop friendship links randomly so that its number of friends becomes bounded again. Although this is a discrete process, we will model it as a continuous process that intuitively captures essential features of the discrete process.

Thus we model the social network as an undirected edge-weighted graph on n nodes, $G = (V, E)$, in which the nodes correspond to people. Let A denote the weighted adjacency matrix of this graph. An edge (i, j) between two (distinct) nodes corresponds to friendship, with the edge weight, $A(i, j)$ indicating the probability (or the strength) of the friendship. As discussed in Section 1, we assume that the number of nodes is fixed while edges can be added or deleted over time. To understand the evolution of the social network, we would like to study the structure of the graph after a sufficiently long time.

In the continuous process, the probability that node k is a common friend of nodes i and j is equal to $A(i, k) \cdot A(k, j)$ (with the simplifying assumption that the probabilities are independent), so that the expected number of common friends that i and j share is equal to $\left(\sum_{1 \le k \le n} A(i, k) \cdot A(k, j)\right) = (A^2\ (i, j))$. Parameterizing by an exponent α, we model the change in $A(i, j)$, $\Delta A(i, j)$ during one time interval to be proportional to $\left(\sum_{1 \le k \le n} A(i, k) \cdot A(k, j)\right)^\alpha = (A^2\ (i, j))^\alpha$. As in the preferential attachment model, as we increase the exponent α, we favor "rich becoming richer", that is, edges are even more likely to form between nodes that have many common neighbors. Thus we increment each entry of the adjacency matrix by the corresponding probability as described above, after scaling suitably. In each time interval, we look at the total weight of all the edges in the graph and add a h fraction of this value as the number of such edges. In addition, we add a small fraction, h' (much less than h) of random edges: for the continuous process, this corresponds to incrementing each (non-diagonal) entry of the adjacency matrix by the same value.

However this may cause the total number of friends of a node to become too large. In fact, simulation results show that the graph evolution becomes too skewed (see footnote in section 1) for any $\alpha > 1$. For $\alpha = 1$, the graph converges to a uniform random graph. Essentially, we saw a total absence of community structure in this case. We assume that every node can handle a total friendship strength of at most a constant. Hence each node scales down its edges so that the total friendship strength of this node is at most the above constant. In effect an edge (i, j) gets scaled down to the smaller of these two values. Note that all weights we use in describing the graphs and our simulations of the model are relative weights and may need to be scaled down to satisfy the degree constraints.

In summary, the evolution of edges in each time interval can be expressed in the following two steps:

1. **Add new edges** - $\Delta A(i, j) = \beta \cdot (A^2\ (i, j))^\alpha + \mu$ (for $i \neq j$) where β and μ are chosen in such a way that the first term corresponds to h fraction and the second term corresponds to h' fraction of the total weight of existing friendships respectively. That is, $\beta \cdot \sum_{1 \le i < j \le n} (A^2\ (i, j))^\alpha = h \cdot \sum_{1 \le i < j \le n} A(i, j)$ and $\mu \cdot \binom{n}{2} = h' \cdot \sum_{1 \le i < j \le n} A(i, j)$.
2. **Scale down edge weights** - After applying the increment computed in step 1, we scale the edges as follows to keep the

maximum weighted degree of a node at most 1. The scaling is computed as $A_{new}(i,j) = \frac{A(i,j)}{\max(\sum_{1\le k\le n} A(i,k), \sum_{1\le k\le n} A(j,k))}$ for each edge (i,j) and then each edge weight $A(i,j)$ is updated to $A_{new}(i,j)$.

Steps 1 and 2 are repeated iteratively until we converge to a fixed point. The natural question to ask is what happens to certain backbone networks under this process. For example, we would like to understand whether new communities (clusters) form, existing communities disappear, or there is a drastic change in the structure of the network. We will first study simple graphs such as the star graph and the dumbbell graph (see Figure 1). We will then study more complex graphs that are obtained by connecting star graphs using an underlying tree backbone network (such as the graphs in Figure 2 and Figure 3).

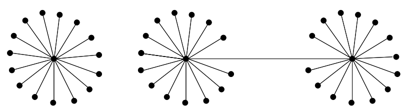

Figure 1: Example star and dumbbell graphs

4. RESULTS

We now present a summary of the main observations of our study.

1. Surprisingly, we find that for parameter α equal to one (new links are added in proportion to the number of common friends), regardless of the original structure of the graph, all graphs tend to converge to a uniform random graph.

2. As α is increased from 1, at certain threshold α_0 that depends on the initial graph, communities begin to appear. For $\alpha \in [0, \alpha_0]$, the graph converges to a uniform random graph and for $\alpha > \alpha_0$, the graph converges to clusters (communities) where the intra-cluster edge strength is much higher than the inter-cluster edge strength. Furthermore, as α is increased above α_0, again at certain thresholds, more and more refined clusterings are obtained. The number of clusters increases at higher thresholds.

3. In all our observations, the threshold values for α lie in the range 1 and 2, i.e., for $\alpha > 2$, we observed no refinement of the communities.

4. We observe that these thresholds are independent of the amount of randomness (parameter h') and depend only on the initial graph structure.

5. THEORETICAL ANALYSIS FOR SPECIFIC BACKBONE NETWORKS

In this section, using theoretical analysis, we show that for $\alpha = 1$ and for any graph, the only fixed point for the above process corresponds to a uniform random graph (where all edges have the same probability) under some simplifying assumptions. Next, for larger values of α, we theoretically analyze the evolution of simple graphs such as the star graph and a variant of the dumbbell graph.

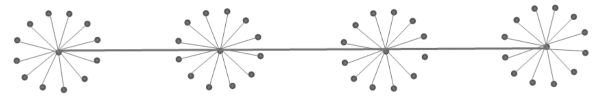

Figure 2: Star graphs connected by a line (line star graph)

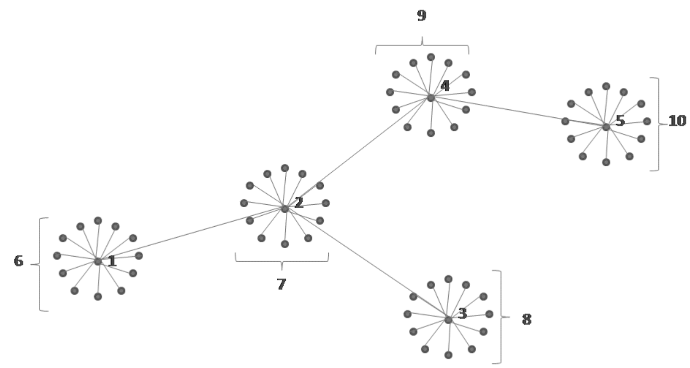

Figure 3: Star graphs connected by a tree (tree star graph)

5.1 Analysis for $\alpha = 1$ for arbitrary graphs

We now show that for any network, if the parameter α is set to 1, the only fixed point for the process is a uniform random graph. We make a few simplifying assumptions. We capture friends of friends edges using the matrix A^2 (that is, we do not require the diagonal entries $A(i,i)$ to be zero) and include a non-zero fraction of random edges. Further, instead of scaling down each edge locally as in step (2) described in Section 3, we scale down each edge by a fixed amount at the end of each iteration. For example, the scaling could be done such that the Frobenius norm of the matrix remains fixed.

Lemma 1. For the above process, the only fixed point corresponds to identical weights for all edges (i.e., that of a uniformly random graph).

PROOF. We will denote the matrix with all ones as R, the scaling constant as η, and the weights for friends of friends edges and random edges as h_1 and h_2 respectively. Then we can express the condition for the fixed point, A_{fixed} as:
$A_{fixed} = \eta(A_{fixed} + h_1 A^2_{fixed} + h_2 R)$
or equivalently,

$$\frac{1-\eta}{\eta h_2} A_{fixed} - \frac{h_1}{h_2} A^2_{fixed} = R$$

Expressing the singular value decomposition of A_{fixed} as $A_{fixed} = UDU^T$ and multiplying the above equation by U^T on the left side and U on the right side, we get:

$$\frac{1-\eta}{\eta h_2} D - \frac{h_1}{h_2} D^2 = U^T R U$$

Let x_i denote the sum of entries of the i^{th} column of U, i.e., $x_i = \sum_{1\le k\le n} U(k,i)$. Then, we can observe that $U^T RU(i,j) = x_i x_j$. Since $U^T RU$ is a diagonal matrix, it follows that all but one (say, x_1) of x_i's are equal to zero. Hence the eigen vectors of A_{fixed} except the first one are normal to the all ones vector. As the columns of U are orthonormal, the first column vector of U must be proportional to the all ones vector. Further, the entries of D except $D(1,1)$ are equal to zero. Consequently A_{fixed} is determined entirely by the first column vector of U and $D(1,1)$. Since the first column of U has identical entries, it follows that all entries of A_{fixed} must be identical. □

5.2 Analysis for star graph

Consider a star graph (see Figure 1) on n nodes (that is, a center node joined to $n-1$ leaf nodes). We show that, independent of the parameter α, this graph converges to the uniform random graph on all the nodes, with the possibility of the center being disconnected.

For this analysis, we assume that the fraction of random edges added is negligible and only consider edges added through common friends. By symmetry, there will be only two different types of edges. Let a denote the strength of each edge between the center and a leaf node and b denote the strength of each edge between two leaf nodes.

At the fixed point, we would like to obtain the ratio of the strength of a leaf-leaf edge to that of a leaf-center edge since this ratio would indicate how close the fixed point graph is to the initial tree as against the (uniform) complete graph.

Lemma 2. For any constant $\alpha \geq 1$ and for asymptotically large n, the fixed point for the above process corresponds to one of the following cases:

- $a = b = \frac{1}{n-1}$. The resulting graph is the (uniform) complete graph.
- $a = 0, b = \frac{1}{n-2}$. The resulting graph is the complement of the star.

PROOF. We can express the change in the strengths of each type of edge by considering the number/strengths of length two paths.

$$\begin{aligned}\delta a &\propto ((n-2)ab)^\alpha \\ \delta b &\propto \left(a^2 + (n-3)b^2\right)^\alpha\end{aligned}$$

At the fixed point,

$$\begin{aligned}\frac{a+\delta a}{\max((n-2)(b+\delta b)+a+\delta a,(n-1)(a+\delta a))} &= a \\ \frac{b+\delta b}{\max((n-2)(b+\delta b)+a+\delta a,(n-2)(b+\delta b)+a+\delta a)} &= b\end{aligned}$$

As these equations hold for infinitesimally small values of δa and δb, we notice that if a and b are positive,

$$\begin{aligned}a+(n-2)b &= 1 \\ b &\geq a\end{aligned}$$

Suppose $a = b$. Then we get the solution $a = b = \frac{1}{n-1}$.

Suppose $b > a$. Then, we have $1 = a+(n-2)b < (n-1)b$ and $1 = a + (n-2)b > (n-2)b$ so that $\frac{1}{n-2} < b < \frac{1}{n-1}$.

We now show that any value of b in this range will not be stable (due to the presence of the δ's) and hence not correspond to a fixed point. As $b > a$, the denominators in the fixed point equations will be the same and hence we get: $\frac{\delta a}{a} = \frac{\delta b}{b}$

Denoting $\gamma = b/a$ and $\beta = 1/n$ and dividing throughout by $n^\alpha a^{2\alpha}/b$, we get:

$(1-2\beta)^\alpha \gamma^{\alpha+1} = \left(\beta + (1-3\beta)\gamma^2\right)^\alpha$

For the choice of $\gamma > 1$, as $1/n\beta \to 0$, we get $\gamma \to 1$. Hence there is no solution for γ exceeding 1 by a constant.

Now consider $a = 0$ or $b = 0$. The fixed point equations imply that the only feasible solution is $a = 0, b = \frac{1}{n-2}$. □

5.3 Analysis of a two-clique graph

For the ease of analysis, we consider a simpler variant of the dumbbell graph. This graph consists of two cliques of n nodes each, which are connected by weak edges (compared to the intra-clique edges). We denote the strength of an edge in either clique by a and that of inter-clique edge by b. As in the star network analysis, we only consider edges added through common friends.

At the fixed point, we would like to obtain the ratio of the strength of an inter-group edge to that of an intra-group edge since this ratio would indicate how close the fixed point graph is to the initial dumbbell graph as against the (uniform) complete graph.

Lemma 3. For any constant $\alpha \geq 1$ and for asymptotically large n, the fixed point for the above process corresponds to one of the following cases:

- $b = a = \frac{1}{2n-1}$. The resulting graph is the (uniform) complete graph.
- $b = 0, a = \frac{1}{n-1}$. The resulting graph is the union of two disjoint (uniform) complete graphs on n nodes each.
- $\gamma = b/a$ is a positive root of the equation:
$$(1+\gamma^2) = 2\gamma^{1-1/\alpha}$$

PROOF. We can express the change in the strengths of each type of edge by considering the number/strengths of length two paths.

$$\begin{aligned}\delta a &\propto \left((n-2)a^2 + nb^2\right)^\alpha \\ \delta b &\propto (2ab(n-1))^\alpha\end{aligned}$$

At the fixed point,

$$\begin{aligned}\frac{a+\delta a}{(n-1)(a+\delta a)+n(b+\delta b)} &= a \\ \frac{b+\delta b}{(n-1)(a+\delta a)+n(b+\delta b)} &= b\end{aligned}$$

As before, these equations hold for infinitesimally small values of δa and δb, and hence we notice that if a and b are positive,

$$(n-1)a + nb = 1$$

As the denominators in the fixed point equations are equal, we get:

$\frac{\delta a}{a} = \frac{\delta b}{b}$

Denoting $\gamma = b/a$ and $\beta = 1/n$,

$$\gamma^{1/\alpha}\left(1 - 2/n + \gamma^2\right) = 2\gamma\left(1 - 1/n\right)$$

Clearly $\gamma = 1$ and $\gamma = 0$ are possible solutions. For asymptotically large n, the other solutions are positive roots of the equation:

$$(1+\gamma^2) = 2\gamma^{1-1/\alpha}$$

□

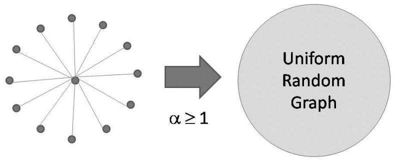

Figure 4: Transformation of a star graph into to a clique

6. EXPERIMENTS

We ran simulations of our model using Matlab on various topologies described in the previous sections to study the threshold phenomena in the evolution of social networks. We denote the threshold values of the parameter α by $\alpha_0, \alpha_1, \ldots, \alpha_k$ where $\alpha_0 \leq \alpha_1 \leq \ldots \leq \alpha_k$, and α_i, $1 \leq i \leq k$ denotes a threshold at which the graph converges into more finely-refined components than it did at threshold α_{i-1}.

6.1 Star graph

In the case of a star network, we considered a network with 100 nodes. As we showed analytically in the previous section, we also show empirically that the graph converges toward a uniform random graph as edges are added and deleted as prescribed by our model. In fact, we find that the star graph becomes a uniform random graph for all values of α. Thus, we do not observe any threshold phenomena when this graph breaks apart any further. Since all edges have the same weight, we refer to this as a clique. We illustrate this transformation in Figure 4.

6.2 Dumbbell graph

We view this class of graphs as two star sub-graphs whose centers are connected by an edge[2]. The number of edges in the left and right star sub-graphs are n_a and n_b respectively. The corresponding edge weights are a and b respectively. The weight on the edge connecting the centers of the dumbbell graph is c. In the first set of experiments we set to characterize the evolution of the dumbbell graph as edges are added and deleted according to our proposed model. Figure 5 illustrates the evolution of the dumbbell graph as α is increased from 1 to 2. We observe that at $\alpha_0 \approx 1.35$, the dumbbell converges to two cliques without their centers and a single edge separately connecting the two centers. The values of the parameters are: $n_a = n_b = 100$, and $a = 0.8$, $b = 0.9$, and $c = 1.0$. The number of cliques does not change for values of $\alpha > 1.35$. The value of c in Figure 5 dominates the weights a and b at convergence because it is the most weighted edge that is common to all friends-of-friends edges in the graph.

We categorize these parameters as representing the *connectedness* of the graph (parameter c), *strengths* of the star sub-graphs (parameters a and b), and the *size* of star sub-graph (parameters n_a and n_b). We varied these key parameters in the experiments and cataloged their effect on the threshold value of α_0 at which the graph transforms from a uniform random graph into a set of cliques. We would like to remind the reader that the cliques are not defined in the true graph-theoretic sense. Basically, they are characterized by very strong intra-component edge weights compared to the inter-component edge weights. We provide very brief descriptions of the experiments due to lack of space:

- Effect of the number of nodes in the sub-graphs: We observed that as we increase the size of each sub-graph, the threshold value α_0 decreases and finally converges to a value around 1.35. This is because the intra-clique strength outweighs the inter-clique strength, resulting in a lower threshold at which the graph breaks up.

- Effect of the relative sizes of the sub-graphs on the threshold value: As the ratio n_a/n_b increases from 0 to 1, the threshold value increases first and then decreases, attaining a maximum when the ratio is around 1/4.

- We observed that the number of nodes required in each sub-graph to induce the formation of a center-center edge at convergence is proportional to the initial strength of the edge connecting the centers.

[2]Where it is clear, we refer to the star sub-graphs by sub-graphs

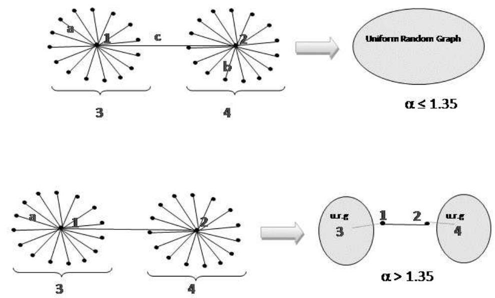

Figure 5: Transformation of a dumbbell graph into cliques. For $\alpha \leq 1.35$, the dumbbell graph converges into a uniform random graph and above this threshold value, the graph splits into two communities.

We also studied the effect of the size of each sub-graph on the strength of the uniform random graph. We further observed that the thresholds are independent of the amount of randomness (parameter h') and depend only on the initial graph structure.

6.2.1 Simulating a discrete model

Using a discrete model we verified for both the dumbbell and star networks that the graphs degenerate to a uniform random graph for $\alpha = 1$. The edges are deleted so as to keep the average degree constant. Specifically, an edge is deleted with probability $1 - d/\max(d_u, d_v)$, where d is the maximum degree, d_u and d_v are the degrees of the nodes u and v respectively connected by the edge. In this model, we counted the number of triangles and found that the number of triangles corresponds to that in a uniform random graph. Note that the number of triads, n_t, in a uniform random graph with n nodes and degree m is equal to $\binom{n}{3}p^3$ in expectation, where p is the probability of existence of an edge. Clearly, $p = \frac{mn/2}{\binom{n}{2}}$ giving $n_t \approx m^3/6$. Thus, for $n = 20$ and $m = 10$, we get $n_t \approx 165$. We verified that the number of triads in the graph at convergence is indeed close to this value.

6.3 More complex graphs

To study the model in more complex graphs, we extend the graph to a backbone network connecting multiple star sub-graphs. For the sake of analysis, we consider two specific backbone graphs - a set of star sub-graphs connected by a line (see Figure 2) and a set of star sub-graphs connected by a tree as shown in Figure 3. We ran the same experiments in which we varied the relative strengths of connections in each star sub-graph and varied the number of nodes in each star sub-graph and studied the values of α_0.

6.3.1 Line star graph

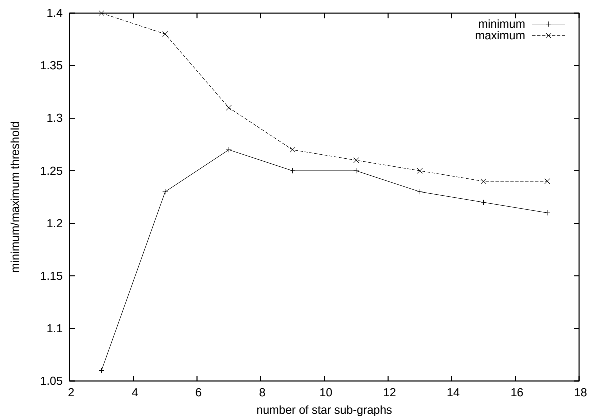

Figure 6: Minimum and maximum thresholds as a function of the number of star sub-graphs. The minimum threshold, α_0 is associated with the threshold at which there exist two components in the graph and the maximum threshold signifies a value above which the graph does not change much in terms of topology.

We considered different number of star sub-graphs connected by a series of edges in a line. For a number of line star graphs ranging from 4 star sub-graphs to around 20 star sub-graphs, we observed the following:

- Depending on the number of sub-graphs, we observed that the smallest threshold, α_0, at which the graph transforms from a uniform random graph increases initially and then decreases gradually with increase in the number of star sub-graphs. On the other hand, the largest threshold decreases with increase in the number of star sub-graphs. Significantly, we observe that both these thresholds converge to similar values when there are large number of star sub-graphs (see Figure 6). These experiments were run with each star sub-graph size equal to 100 and the strength on every edge set to 1.0. We also note that a common feature to all graphs with more than three star sub-graphs is that they all have non-leaf edges.

- At α_0, the uniform random graph decomposes into three components – the left-most star with its center, the right-most star with its center, and the reminder of the graph in one giant connected component. At α_1, this giant component disintegrates into individual star components, with all the centers by themselves in a component, i.e., these centers on non-leaf edges form a clique.

Effect of size of each sub-graph on the converged structure: In another set of experiments, we measured the effect of the size of each star sub-graph on the order in which the graph disintegrates into individual components. We chose a line graph with 7 star sub-graphs. The size of each star component is increased from 99 in sub-graph 1 to 699 in sub-graph 7. Figure 7 shows each threshold value α_i at which the largest remaining component in the graph disintegrates into two components. Clearly, the threshold value increases as the number of components in the graph increases. In fact, it is almost linear in the number of components. We observe that the relative strengths of each sub-graph plays an important role in the order in which the graph disintegrates. For example, sub-graphs with large number of nodes tend to stay connected more than the smaller sub-graphs which break away early on.

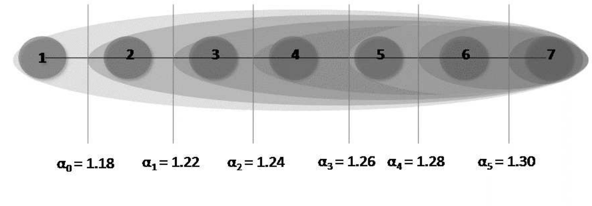

Figure 7: Order of disintegration of the graph. There are six different thresholds, each giving a more refined community structure than the previous one.

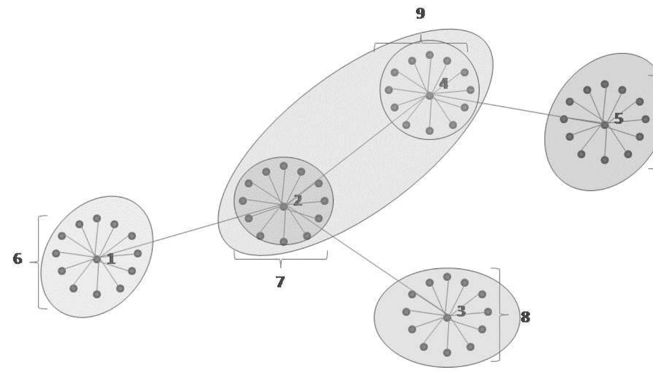

Figure 8: Transformation of a general backbone graph into cliques at $\alpha_0 = 1.3$ resulting in four communities. At the next threshold value of $\alpha_1 = 1.35$, one of these communities further splits into two, thereby resulting in five communities.

6.3.2 *Tree star graph*

Lastly, we ran the same set of experiments on a graph wherein the sub-graphs are connected by a tree backbone. The first threshold value α_0 was observed to be 1.3 when the graph starts to degenerate into smaller cliques from a uniform random graph. While the value of α was increased from 1, the connected backbone starts to break apart with some of the stars still connected to each other while the other star sub-graphs start to form their own cliques. For brevity, we show the final transformed graph at the threshold value in Figure 8. The backbone is connected to begin with and all the star sub-graphs have the same number of nodes. As in the case of the line graphs, we observed that the centers on the no-leaf edges stay together in a component while the star sub-graphs at the ends of the backbone become independent of other sub-graphs on the backbone. At higher value of threshold $\alpha_1 = 1.35$, the the larger component breaks up into its individual sub-graphs.

7. CONCLUSIONS

We propose a model for evolution of social networks where links are formed between nodes based on the number of common friends they share. We find that in our model, if an edge is added with probability proportional to the number of common friends, then the social network loses its identity and converges to a uniform random graph. However, if we introduce a parameter α, and make the probability of adding an edge proportional to the number of common friends raised to the power α, then we see the evolution of communities or clusters depending on the structure of the original graph. Further, such cluster formation follows a threshold phenomenon where clusters appear above a certain value of α depending on the structure of the original graph. Furthermore, there are multiple threshold values producing more-refined communities where the number of communities increase with higher threshold values. In our observations, these threshold values seem to lie between 1 and 2.

The main problem we faced is analyzing our model for complex graphs because of the non-linearity in the update of the friendship matrix in each iteration. It would be interesting to investigate a theoretical analysis for a larger class of graphs.

8. REFERENCES

[1] A Albert, H. Jeong, and A.-L. Barabási. Diameter of the world wide web. *Nature*, 401:130–131, 1999.

[2] A.-L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286(15), 1999.

[3] B. Bollobas. *Random Graphs*. Academic Press, 1986.

[4] P Erdös and A. Rényi. On random graphs i. *Publicationes Mathematicae Debrecen*, 5:290–297, 1959.

[5] E. Friedgut and G. Kalai. Every monotone graph property has a sharp threshold, 1996.

[6] Matthew Jackson and Brian W. Rogers. Meeting strangers and friends of friends: How random are social networks. *American Economic Review*, 2006. To Appear.

[7] E. M. Jin, M. Girwan, and M. E. J. Newman. The structure of growing social networks. *Physical Review Letters E*, 64(046132), 2001.

[8] Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, March 1953.

[9] Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. In *Proc. 9th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 668–677, 1998.

[10] Ravi Kumar, Jasmine Novak, and Andrew Tomkins. Structure and evolution of online social networks. In *Proc. 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 611–617, 2006.

[11] Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, D. Sivakumar, Andrew Tomkins, and Eli Upfal. The web as a graph. In *Proceedings of the 19th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, pages 1–10, 2000.

[12] David Liben-Nowell and Jon M. Kleinberg. The link prediction problem for social networks. In *Proc. ACM CIKM International Conference on Information and Knowledge Management*, pages 556–559, 2003.

[13] M. E. J. Newman. Clustering and preferential attachment in growing networks. *Physical Review Letters E*, 2001.

[14] M. E. J. Newman. The structure of scientific collaboration networks. *Proc. of National Academy of Sciences*, 98:404–409, 2001.

[15] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2002.

[16] Duncan J. Watts. *Small Worlds*. Princeton University Press, 1999.