Co-evolution of Social and Affiliation Networks

Elena Zheleva
Dept. of Computer Science
University of Maryland
College Park, MD 20742, USA
elena@cs.umd.edu

Hossam Sharara
Dept. of Computer Science
University of Maryland
College Park, MD 20742, USA
hossam@cs.umd.edu

Lise Getoor
Dept. of Computer Science
University of Maryland
College Park, MD 20742, USA
getoor@cs.umd.edu

ABSTRACT

In our work, we address the problem of modeling social network generation which explains both link and group formation. Recent studies on social network evolution propose generative models which capture the statistical properties of real-world networks related only to node-to-node link formation. We propose a novel model which captures the co-evolution of social and affiliation networks. We provide surprising insights into group formation based on observations in several real-world networks, showing that users often join groups for reasons other than their friends. Our experiments show that the model is able to capture both the newly observed and previously studied network properties. This work is the first to propose a generative model which captures the statistical properties of these complex networks. The proposed model facilitates controlled experiments which study the effect of actors' behavior on the network evolution, and it allows the generation of realistic synthetic datasets.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*Data Mining*

General Terms

Algorithms, Measurement, Experimentation

Keywords

evolution, social network, affiliation network, graph generator, groups

1. INTRODUCTION

In recent years, there has been a proliferation of online social networks. Many of the networks have millions of users, and allow complex interactions through linking to friends, public messaging, photo commenting, participating in groups of interest, and many others. Studies have been performed to characterize and explain the behavior of users, and most of them concentrate on modeling how users join the network and form links to each other. Little is known about how different types of interaction influence each other. In our work, we address the problem of modeling social network generation explaining both link and group formation.

In social networks, users are linked to each other by a binary relationship such as friendship, co-working relation, business contact, etc. Each social network often co-exists with a two-mode affiliation network, in which users are linked to groups of interest, and groups are linked to their members. In our study we use three large datasets from online social and affiliation networks, and discover a number of interesting properties. The datasets were from Flickr, LiveJournal and YouTube, collected by Mislove et al. [9].

Using the newly observed and previously studied statistical properties of these networks, we propose a generative model for social and affiliation networks. The model explains the complex process of forming the networks, and captures a number of affiliation network properties which have not been captured by a model before: power-law group size distribution, large number of singletons (group members without friends in the group), power-law relation between the node degree and the average number of group affiliations, and exponential distribution of the number of group affiliations for nodes of a particular degree. Our findings are important for understanding the evolution of real-world networks and suggest that the process is more complex than a naïve model in which groups are added to a fully evolved social network. They also show that users join groups for different reasons and having friends in the group is often not necessary. This suggests that information spreads in the network through channels other than the friendship links, and this observation has implications on information diffusion and group recommendation models.

In addition, this model can be used for synthetic network generation. This is an important application because real-world network datasets are often proprietary and hard to obtain. Controlling network parameters allows the generation of datasets with different properties which can be used for thorough exploration and evaluation of network analysis algorithms.

Our contributions include the following:

- We discover a number of new properties in social and affiliation networks.
- We propose the first generative model for network evolution which captures the properties of both real-world social and affiliation networks.

KDD'09, June 28–July 1, 2009, Paris, France.

- We provide a thorough evaluation of our model which shows its flexibility for synthetic data generation.

Notation. We study the interactions of two graphs, the social network graph, G_s, and the affiliation graph, G_a. For our purposes, a social network is a graph $G_s = \{V, E_s\}$ which has one type of node corresponding to the users that participate in it. Nodes can form links which can be directed or undirected; $e_s(v_i, v_j, t)$ denotes the link that v_i and v_j have formed at time t. A directed link is formed whenever one user links to another. An undirected link requires the approval of both parties in order to be formed.

In an affiliation network $G_a = \{V, H, E_a\}$, there are two types of nodes, the social network users V and groups H that they have formed. We represent the network as a bipartite graph in which undirected links $e_a(v_i, h_j, t)$ are formed between user v_i and group h_j at time t when this user becomes a member of the group. There are a number of reasons why groups are formed. For example, groups can exist because of a common interest, such as philately or book-reading clubs; they can be based on common business relation, such as an employing company, or they can be based on common personal traits, such as geographic location. What is common between the groups that we study is that users have voluntarily chosen to be parts of them, as opposed to clustered together by a group detection algorithm.

2. RELATED WORK

The evolution of social and affiliation networks exhibits a number of properties previously studied in the literature. We describe some of them in more detail in Section 4.2.

2.1 Evolution of social networks

The majority of literature on analyzing network properties has focused on friendship networks, or actor-actor networks in general. Studying the static snapshots of graphs has led to discovering properties such as the 'small-world' phenomenon [10] and the power-law degree distributions [2, 4]. Time-evolving graphs have also attracted attention recently, where interesting properties have been discovered, such as shrinking diameters, and densification power law [7].

There have been a number of models proposed to capture these properties. For a survey, one can consult the work by Chakrabarti and Faloutsos [3]. For example, unlike the random graph model, the preferential attachment model proposed by Barabasi et al. [2] captures power-law degree distributions. The forest fire model [7] also captures the power-law degree distribution together with densification and shrinking diameters over time. A more recently proposed, microscopic evolution model [6] is based on properties observed in large, temporal network data, providing insight into the node and edge arrival processes. Another recent model, the butterfly model [8], concentrates on capturing the evolution of connected components in a graph. In our work, we extend the microscopic evolution model by including processes of forming and joining groups of interest.

2.2 Evolution of affiliation networks

To the best of our knowledge, there is no model that captures the evolution of affiliation networks in online communities. However, there are studies that describe the relationship between friendship links and group formation properties [1, 9]. They show that the probability of a user joining a group increases with the number of friends already in the group [1], and that higher degree nodes tend to belong to a higher number of groups [9].

Group detection is a related problem (for a survey, see [5]). Its goal is to find new communities based on node features and structural attributes. Unlike group detection work, our work concentrates on unraveling the process according to which existing communities were formed.

3. OBSERVATIONS

Though affiliation groups constitute a major part of many social networks, very little work in the literature focuses on analyzing group memberships and evolution. In this section, we analyze different affiliation networks and try to characterize some properties of affiliation groups that are consistent across various datasets. For our analysis, we used three large real-world datasets from LiveJournal, Flickr and YouTube.

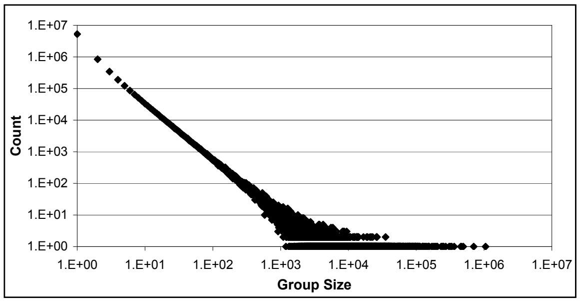
LiveJournal is a popular blogging website whose users form a social network through friendship links. Users also form affiliation links to various 'communitites,' which are groups of users with similar interests. We used a LiveJournal dataset with over 5.2 million users, 72 million links, and over 7.4 million affiliation groups. The second dataset is from Flickr, a photo-sharing website based on a social network with friendships and family links. Groups in Flickr are also formed on the basis of common interest. The Flickr dataset contains over 1.8 million users, 22 million links, and around $100,000$ groups. The third dataset is from YouTube, a popular video-sharing website with an underlying social network based on users' contacts. Users also form an affiliation network by joining social groups where they can post and discuss videos. The YouTube dataset contains over 1.1 million users, 4.9 million links and around $30,000$ groups. The full dataset descriptions can be found in the work of Mislove et al. [9]. Now, we describe the observations that we discovered by analyzing the datasets, and we relate them to previously observed properties.

3.1 Group size distribution

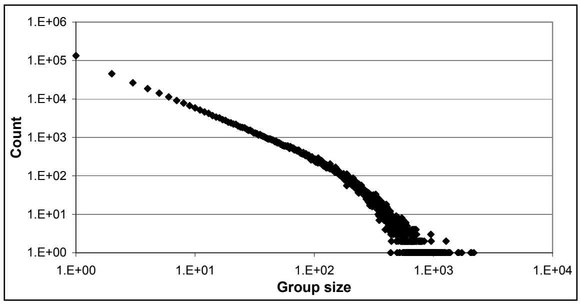
We begin by characterizing the relationship between the size of the affiliation group and its frequency of occurrence. The main observation is that, analogous to the degree distribution, the group size distribution follows a power law, with a large number of small groups and a smaller number of large ones. This has also been observed by Mislove et al. [9]. The results are illustrated in Figure 1.

3.2 Node degree vs. average number of group affiliations

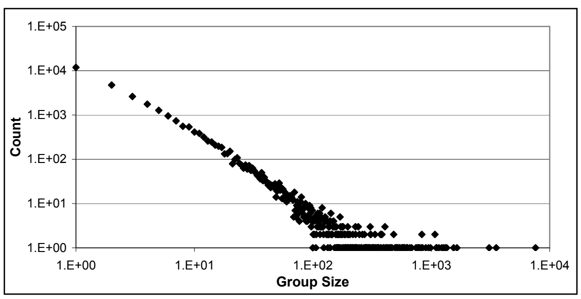
Looking at the relationship between the degree of a node and the number of its group affiliations, we observe that the nodes of lower degree tend to be members of fewer number of groups than the nodes with higher degree. However, the relation starts declining after a certain point, yielding lower number of group memberships for very high degree nodes. The relationship is illustrated in Figure 2, where the x-axis represents the node degree and the y-axis represents the average number of group affiliations for nodes with that degree. The nodes in the declining part represent a very small portion of the overall number of nodes (<1% of the size of the network in all cases), which is why we fitted only the increasing part of the data points to a function. We compared against over 55 different distributions including



(a) LiveJounral



(b) Flickr



(c) Youtube

Figure 1: Distribution of the number of groups of a particular size on log-log scale.

logistic, Dagum and Laplace, using EasyFit [1], a software for distribution fitting. A power-law relation was the best fit according to the Kolmogorov-Smirnov ranking coefficient.

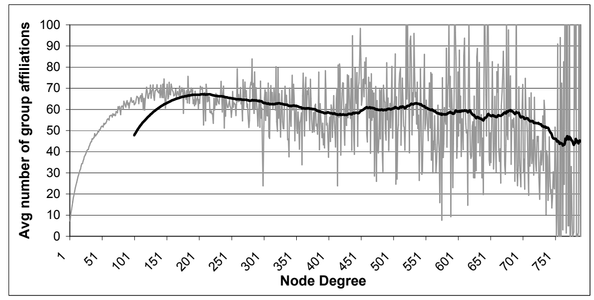
3.3 Distribution of the number of group affiliations

The previous observation was about the average number of group affiliations for nodes with different degrees. Here, we look at the actual distribution of the number of group affiliations with respect to the node degree. It turns out that the number of group affiliations for nodes of a certain degree k follows an exponential distribution. Figure 3 reports on $k = 50$ for LiveJournal and Flickr, and on $k = 25$ for YouTube but this was true for other degrees as well.

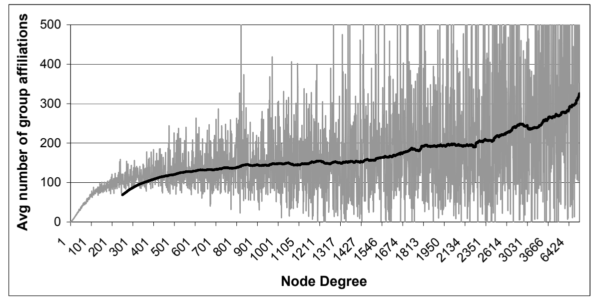
3.4 Properties of group members

According to Backstrom et al. [1], nodes are more likely to join groups in which they have more friends. However, it turns out that, in our datasets, there is a large portion of group members without friends in the group (*singletons*), meaning that they did not join the group because of a friend.

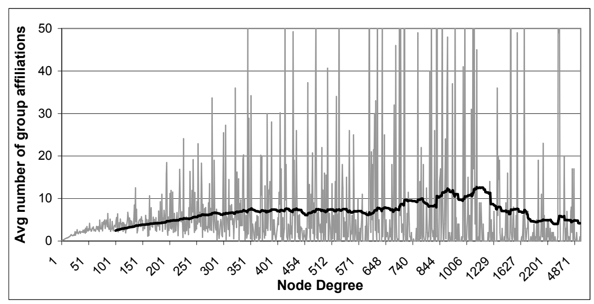
[1]At `http://www.mathwave.com`



(a) LiveJounral



(b) Flickr

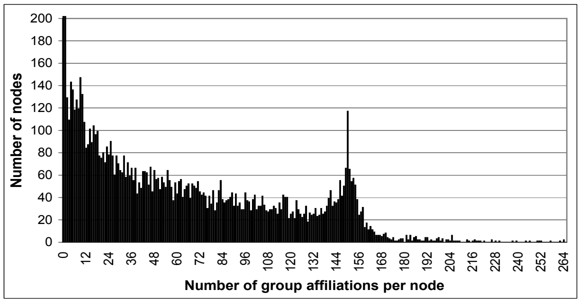


(c) Youtube

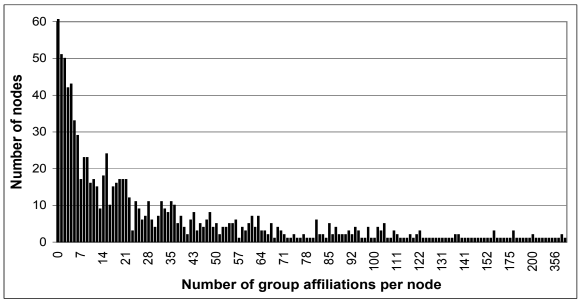
Figure 2: Node degree vs. average number of group affiliations

This is surprising because it shows that users join groups for various reasons, friendship being only one of them.

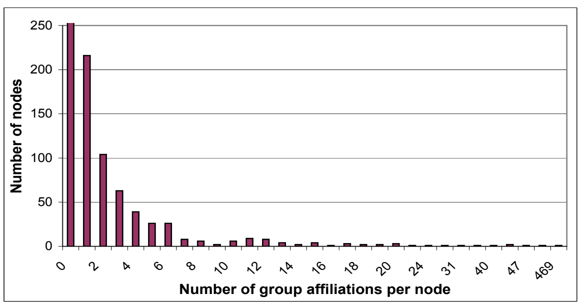
We measure the maximum node degree within groups of various sizes in our datasets. For all groups of a given size, we measure the average maximum degree per group and the average number of singletons (nodes with no friends within this group) as a percent of the group size. The results show a large number of singletons overall, especially in small groups, indicating that a large percentage of the members of a specific group do not have any friends within this group. This conclusion was confirmed by analyzing the average maximum degree per group. It turned out that the friends of the maximum-degree node within a group do not constitute a large percentage of the group size, even in small groups. The numbers are illustrated in Figure 4, where the *upper* series shows the average ratio of the number of singletons to the group size, and the *lower* series represents the average ratio of the maximum degree to the group size. This result shows that the larger the group a user belongs to, the more likely it is for him/her to have a friend in the group. For example, in Flickr, 76% of the members of groups of size 50 are singletons, while for groups of size 500, this number drops to 29%.



(a) LiveJounral - Degree = 50



(b) Flickr - Degree = 50



(c) Youtube - Degree = 25

Figure 3: Distribution of the number of group affiliations for nodes with specific node degrees.

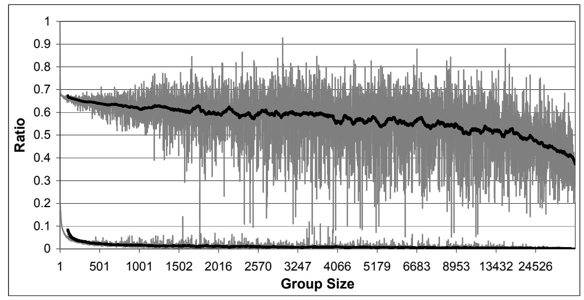
4. CO-EVOLUTION PROPERTIES AND MODEL

A model which describes the evolution of a social network together with the evolution of an affiliation network needs to capture a number of simple events, as well as statistical properties of both networks. Here, we present the events of our co-evolution model and desired properties, some of which have been presented in other work. Then, we present our co-evolution model, which extends the node arrival and link formation processes of the microscopic evolution model [6] to dynamic social and affiliation networks.

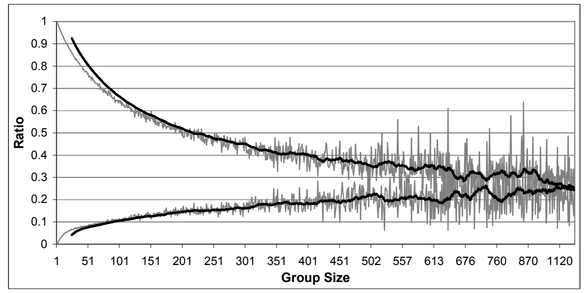
4.1 Events

The possible events that our model allows are:

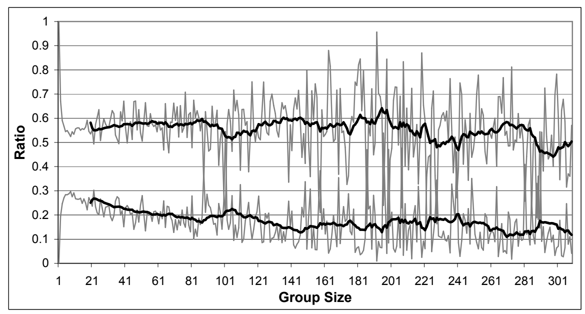
- a node joins the network and links to someone
- a new group is formed with one member
- a node joins an existing group
- a new link is formed between existing users



(a) LiveJounral



(b) Flickr



(c) Youtube

Figure 4: Ratio of the number of singletons to the group size (upper series) and ratio of the maximum degree to the group size (lower series).

4.2 Desired properties

A co-evolution model needs to capture properties of both social and affiliation networks. Here, we show three types of properties: properties of the social network alone, properties of the affiliation network alone, and properties of both.

Properties of the social network. The properties are:

- *power law degree distribution* - the node degrees are distributed according to a power law with a heavy tail. This property has been observed in many other studies.
- *network densification* - the density of the network increases with time [7].
- *shrinking diameter* - the effective diameter of the network decreases as more nodes join the network [7].

Properties of the affiliation network. We would also like to capture the following affiliation network property:

- *power law group size distribution* - the group sizes are distributed according to a power law with a heavy tail.

Properties involving both the social and affiliation networks. These properties describe the relationship between a social network and an affiliation network:

- *large number of singletons* - many nodes do not have any friends inside the groups they are affiliated with.
- *power-law relation between the node degree and the average number of group affiliations* - see Section 3.2.
- *exponential distribution of the number of group affiliations for a particular node degree* - see Section 3.3.

4.3 Co-evolution model

We now propose a co-evolution model which captures the discussed desired properties. Our model is undirected, and it has two different sets of parameters: one is concerned with the evolution of the social network, and the other determines the factors of development of the affiliation network. We also present a naïve model which assumes that the evolution of the affiliation network is independent of the evolution of the social network. Both models utilize the microscopic evolution model [6] for generating the social network because that model is based on observing the temporal properties of large social networks. We present its main components first.

Microscopic evolution model. The main ideas behind the microscopic evolution model are that nodes join the social network following a node arrival function, and each node has a lifetime a, during which it wakes up multiple times and forms links to other nodes. These are the set of parameters needed for the microscopic evolution model: $N(.)$ is the node arrival function, λ is the parameter of the exponential distribution of the lifetime, and α, β are the parameters of the power law with exponential cut-off distribution for the node sleep time gap. Further details of the model can be found in the paper by Leskovec et al. [6]. We utilize these parts:

Node arrival. New nodes $V_{t,new}$ arrive at time t according to a pre-defined arrival process $N(.)$.

Lifetime sampling. At arrival time t, v samples lifetime a from $\lambda.e^{-\lambda.a}$: v becomes inactive after time $t_{end}(v) = t + a$.

Algorithm 1 Naïve model

Set of nodes $V = \emptyset$
for each time period $t \in T$ **do**
  Set of active nodes at time t, $V_t = \emptyset$
end for
for each time period $t \in T$ **do**
  Node arrival. $V = V \cup V_{t,new}$
  for each new node $v \in V_{t,new}$ **do**
    Lifetime sampling
    First social linking
  end for
  for each node $v \in V_t$ **do**
    Social linking
  end for
  for each node $v \in V_t \cup V_{t,new}$ **do**
    Sleep time sampling
  end for
end for
Set of groups $H = \emptyset$.
for i=1:number of groups **do**
  Group creation. New group h_i is created and its size s is sampled from s^{-k}. $H = H \cup \{h_i\}$.
  for j=1:s **do**
    Group joining. Pick a random node $v \in V$ and form an affiliation link to it $e_a(v, h_i, null)$.
  end for
end for

First social linking. v picks a friend w with probability proportional to degree(w) and forms edge $e_s(v, w, t)$.

Sleep time sampling. v decides on a discrete sleep time δ by sampling from $\frac{1}{Z}.(\delta^{-\alpha}).e^{-\beta.\text{degree(v)}.\delta}$. If the node is scheduled to wake up before the end of its lifetime ($t + \delta \leq t_{end}(v)$), then it is added to the set of nodes $V_{t+\delta}$ that will wake up at time $t + \delta$.

Social linking. At wake up time t, v creates an edge $e_s(v, w, t)$ by closing a triad two random steps away (i.e., befriends a friend w of a friend).

Naïve model. Before we present our model, we present a naïve model which assumes that the evolutions of the social network and the affiliation network are two independent processes. As a first step, it creates the social network using the model of Leskovec et al. [6]. Then, it generates and populates groups in such a way that their sizes follow a power-law distribution with an exponent k. Algorithm 1 presents the naïve model in detail. We use this model as a baseline.

Co-evolution model. In this model, the affiliation network evolution co-occurs and depends on the social network evolution. When a node wakes up, besides linking to another node, it also decides on a number of groups to join. With probability τ, it creates a new group, else, it joins an existing group. There are two mechanisms by which it picks a group to join. In the first one, it joins the group of one of its friends. In the second one, it picks a group at random. Algorithm 2 presents the co-evolution model in detail.

Algorithm 2 Co-evolution model

Set of nodes $V = \emptyset$
Set of groups $H = \emptyset$
for each time period $t \in T$ **do**
  Set of active nodes at time t, $V_t = \emptyset$
end for
for each time period $t \in T$ **do**
  Node arrival. $V = V \cup V_{t,new}$
  for each new node $v \in V_{t,new}$ **do**
    Lifetime sampling
    First social linking
  end for
  for each node $v \in V_t$ **do**
    Social linking
    Affiliate linking. v determines n_h, the number of groups to join, sampled from an exponential distribution $\lambda' e^{-\lambda' n_h}$ with a mean $\mu' = \frac{1}{\lambda'} = \rho.\text{degree(v)}^{\gamma}$.
    for $i = 1 : n_h$ **do**
      if $rand() < \tau$ **then**
        Group creation. v creates group h, and forms edge $e_a(v, h, t)$. $H = H \cup \{h_i\}$.
      else
        Group joining. v forms edge $e_a(v, h, t)$. Group h is picked through a friend with probability p_v; otherwise, or if no friends' groups are available, it joins a random group with prob. proportional to the size of h.
      end if
    end for
  end for
  for each node $v \in V_t \cup V_{t,new}$ **do**
    Sleep time sampling
  end for
end for

Here, we present the parameters of the affiliation network evolution part in more detail. The first parameter, ρ, represents a tuning parameter that controls the density of the affiliation links in the network. The second parameter, γ, is the exponent of the power law that relates node degree with number of group affiliations. The last parameter to our model, τ, represents the probability by which an actor creates a new group at each time point. All our parameter values range over the interval $[0, 1]$ except ρ which ranges between 0 and the average number of group affiliations per node. We provide some guidelines for picking the right parameter values in the experiments section.

As noted in Section 4.2, the relationship between node degree and average number of affiliations is a power-law relation. Even though one can vary the exponent γ of this function, for simplicity, we fixed its value to 0.5, utilizing a square root function to compute this average.

It is also worth noting that other, more sophisticated techniques can be utilized in both social and affiliation aspects of the model that might be able to capture stronger correlation between the evoultion of both kinds of networks. One possible modification for the social link creation is considering random steps but with group bias, such that the probability of choosing a node u to close the triad is proportional to the number of groups the two nodes share. Another possible modification is to specify the number of groups a node will join in advance using the estimated power-law function. A disadvantage of such approach is that the approximated degree is hard to compute because it depends on the expected value of a function which changes with the degree. A thorough investigation of the different alternatives is left as future work.

In the group joining step of the algorithm, a node decides to join a group and it has two choices for picking that group. One is through a friend, and the second one is by picking a random group with probability proportional to the size of that group. It follows the first choice with some probability p_v, else it resorts to the second one. The intuition behind this is that some nodes in each group are singletons while others have friends in it. The second choice is also based on the observation that the size of the groups follows a power-law distribution; on the principle of "rich get richer," groups with larger size should have a larger probability of getting picked.

There are many options for computing the probability p_v such as making it a constant or dependent on the node degree. One can test which one is most appropriate in the presence of temporal data for affiliation networks. Since such data is hard to obtain, we try different possibilities in our model. It turns out that using a constant for p_v yields a relationship between the group size and the singleton ratio that decreases at first but then stabilizes around $1 - p_v$ at higher group sizes. In contrast, what we had observed initially was a relationship which decreases with increasing group sizes (see Figure 4). When we use a p_v which is correlated with the degree, then we observe a relationship closer to the desired one. In particular, we compute:

$$p_v = \begin{cases} \eta * \text{degree(v)} & \text{if } \eta * \text{degree(v)} < 1 \\ 1 & \end{cases} \tag{1}$$

though other functions of the degree may be more appropriate. The parameter η represents the friends' influence on the actor's decision to join a group; i.e. the likelihood of an actor joining one of the groups of his/her friends increases by increasing the value of η. The main intuition behind using a degree-correlated probability is the fact that as a node has more friends, the probability that one of its friends belongs to one of the larger size groups increases. Thus, utilizing the friendship bias parameter η actually increases its chances of joining this larger size group of its friend, thus leading to the decreasing relationship noted in the observations.

5. EXPERIMENTS

We present three sets of experiments. The first set observes the properties of data, generated by our co-evolution model, and the second set shows that the model is able to produce a dataset, very similar to one of the real-world datasets. We also present results for the naïve model which adds groups on top of a social network, showing that this model is not able to produce the real-world affiliation network properties.

5.1 Synthetic data

In our first set of experiments, we vary the parameters of the model in order to generate a few synthetic datasets. Then, we check whether each dataset has the properties described in Section 4.2.

We have fixed the parameters of the social evolution part throughout this set of experiments, and varied the parameters of the affiliation part of the network. We assume an exponential node arrival function, to achieve higher growth rate in our generated network, which is in accordance with what Leskovec et al. [6] showed in some social networks, such as Flickr. However, other arrival functions can also be utilized within our model. The other parameters of the social evolution aspect were fixed as reported by Leskovec et al. for Flickr data: $\lambda = 0.0092$, $\alpha = 0.84$, and $\beta = 0.002$. We also fix the value of the second parameter to the affiliation model, γ, to 0.5.

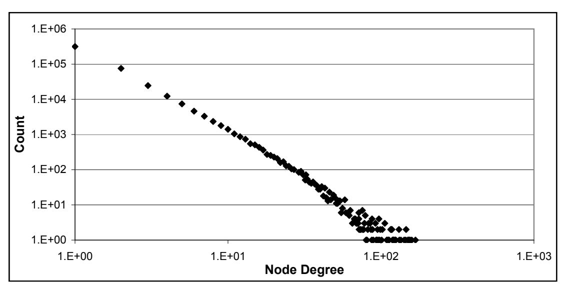


Figure 5: Degree distribution in a synthetic network

We first illustrate the results for the social network generated using the specified parameters. The model was run for 400 time steps, resulting in a network with 140,158 actors and 245,043 social links. The degree distribution in the resulting network follows a power-law, as Figure 5 shows. The network densification property also holds, as illustrated in Figure 6 which represents the number of nodes and number of edges at each time point on a log-log scale.

In order to test the affiliation aspect of our evolution model, we investigated the effect of each parameter in the model on the properties of the resulting affiliation network. We start with our first parameter ρ, which represents a tun-

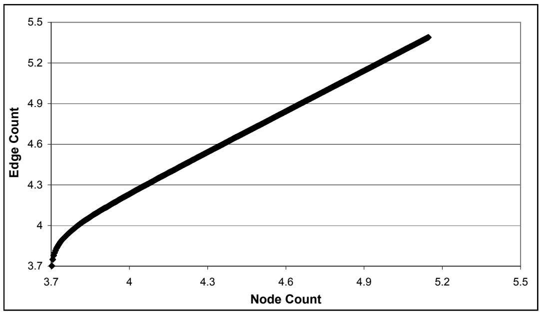


Figure 6: Densification in a synthetic network

ing factor of the affiliation links' density. The main properties that are affected by varying the value of ρ are the total number of affiliations and the distribution between the node degree and average number of group affiliations. As illustrated in Figure 7, we can note that the general power distribution persists among different values of ρ, but the main effect is the scale of the distribution; as increasing the value of ρ, more affiliation links are created, and correspondingly increasing the average number of group affiliations per node. Theoretically, the values for this parameter can vary from 0, where no affiliation links are created in the network, to the maximum number of groups, where fully connected affiliation network emerges. Practical values for ρ varies between 0 and 25. The total number of affiliation links for each value of ρ is reported in Table 1.

ρ	**Affiliation Count**
3	285,536
10	2,411,710
20	4,771,072

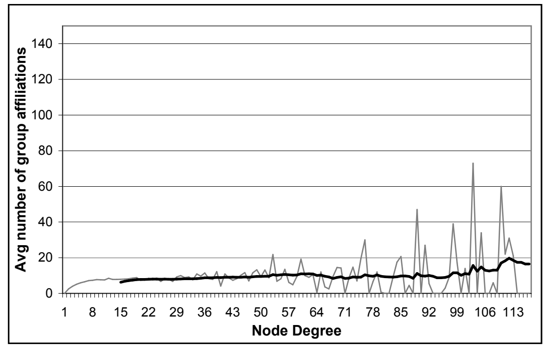
Table 1: Number of affiliation links with varying ρ

Our next parameter, τ, represents the probability with which a node creates a new group. This parameter directly affects the number of groups in the resulting network, as well as the group size distribution. As illustrated in Figure 8, we note that although the power law distribution of the group sizes holds for various values of τ (which is one of the desired properties), the maximum group size decreases significantly with increasing the value of τ. This decline in the maximum group size is caused by the fact that for higher values of τ, nodes tend to create new groups more often than joining existing ones, which leads to the existence of a large number of groups with relatively small sizes. This conclusion is also clear in the results illustrated in Table 2, where the resulting number of groups in the network and the maximum group size vary significantly with changing the parameter value.

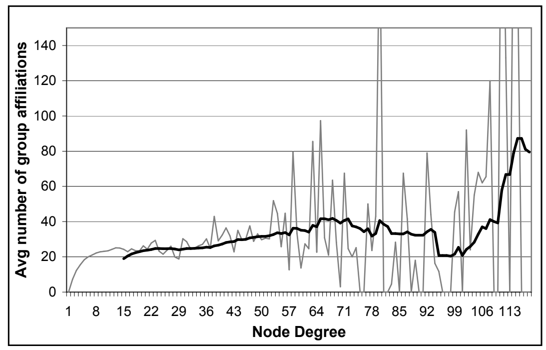
τ	**Groups Count**	**Max Group Size**
0.1	66,887	39,753
0.5	245,143	560
0.9	332,437	32

Table 2: Number of groups with varying τ

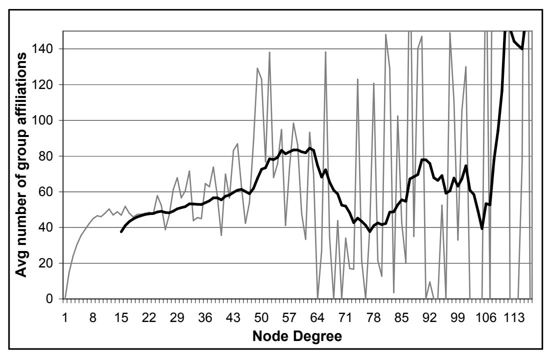
Finally, we investigate the parameter on which p_v depends, η. η represents the extent to which friends influence the decision of a node to join groups. The outcome of in-



(a) $\rho = 3$



(b) $\rho = 10$



(c) $\rho = 20$

Figure 7: Degree vs. average number of group affiliations with varying ρ.

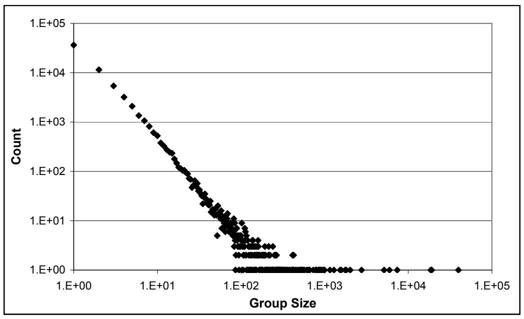
creasing the value of this parameter is a decreasing number of *singletons* and an increasing relative degree of the nodes within different groups. As illustrated in Figure 9, we can easily note that the general distribution captures the desired properties and the observations in real data. The value of η is highly dependent on the social network structure properties, such as the average node degree in the social network and the desired influence of friends on node's decision. For instance, if we have a value of $\eta = 0.1$ in a setting where the expected value for the average node degree is around 10, then we expect to see high percentage of nodes in the network being affected by their friends.

5.2 Real data

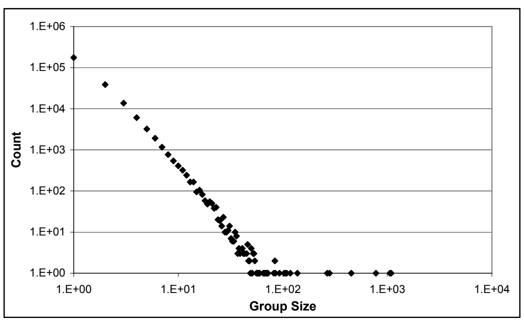
In this set of experiments, we look for the model parameters that will produce a network similar to one of the real-

	Real Network (Flickr)	Synthetic Network ($\rho = 2.5$, $\gamma = 0.5$, $\eta = 0.1$, $\tau = 0.03$)
Number of users	1,846,198	1,707,475
Number of groups	103,648	88,749
Number of affiliations	8,529,435	7,813,910
Average number of group affiliation per user	4.62	4.58
Number of groups/Number of users	0.0561	0.052

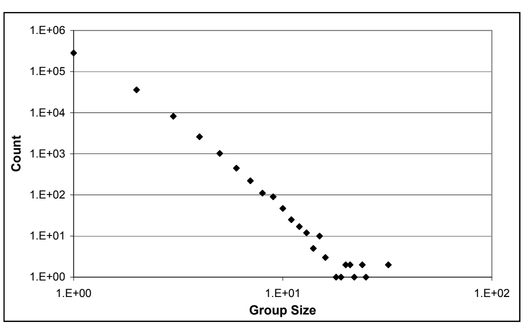
Table 3: Statistics of a real network (Flickr) vs. a synthetic one



(a) $\tau = 0.1$

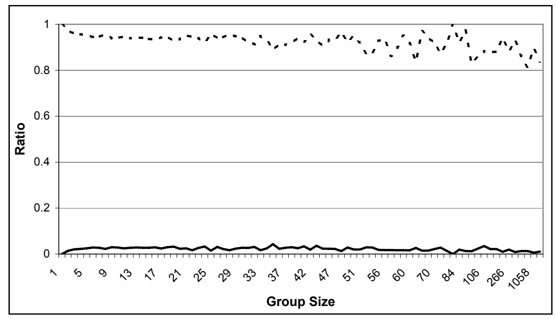


(b) $\tau = 0.5$

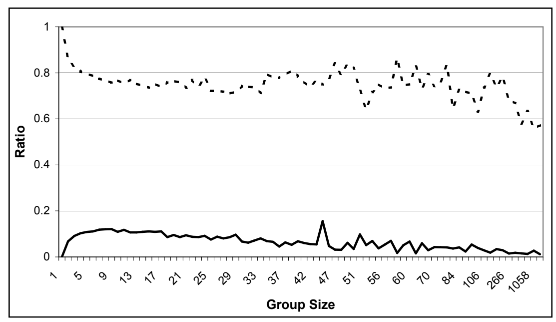


(c) $\tau = 0.9$

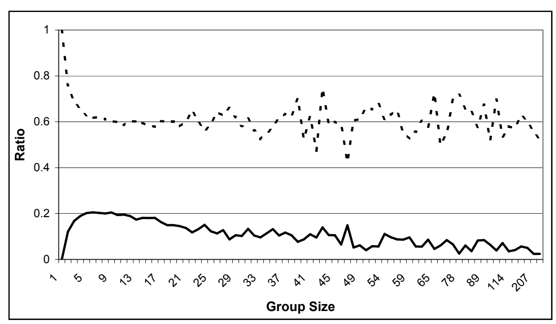
Figure 8: Group size distribution with varying τ



(a) $\eta = 0.01$



(b) $\eta = 0.05$



(c) $\eta = 0.09$

Figure 9: Group size vs. member attributes with varying η (dashed line: % ratio of singletons to group size, solid line: % ratio of maximum degree to group size).

world datasets we have used in the observations of Section 3. We searched for parameters that will produce an affiliation network resembling the actual one of Flickr since the social network evolution parameters for Flickr have already been reported by Leskovec et al. [6]. In order to get an initial seed of the search space for the evolution parameters of the affiliation network, we analyze the affiliation network properties of Flickr as observed in Section 3. A summary of the affiliation network statistics of Flickr is given in Table 3.

The Flickr dataset is characterized by a relatively small number of groups in comparison to the number of users, where the actual ratio between the group count and the user count is 0.056. As a result, we expect to have a small value of τ close to this ratio. On the other hand, the average number of group affiliations per user in the real dataset is 4.62, and we assign this value to ρ. Finally, as observed in Figure 4, the average percentage of singletons in each group is lower than the average for the other datasets, indicating more friendship bias, thus increasing the value of η.

There are other factors to consider when specifying the affiliation network evolution parameters, such as the rate of node arrival and the probabilistic nature of the node's lifetime and sleep time gaps. For example, in Flickr's case, the exponential node arrival rate means that more nodes are created at later times. In this case, the distribution parameters should be a bit lower than the desired ones because many nodes will join towards the end of the evolution process but they will not have time to create many links and affiliations. By utilizing all these pieces of information to guide the parameter search, we were able to generate a network that has similar attributes to Flickr's, illustrated in table Figure 3. We argue that using a similar procedure for parameter selection can result in generating synthetic networks that have many of the properties of a real one.

5.3 Comparison with the naïve model

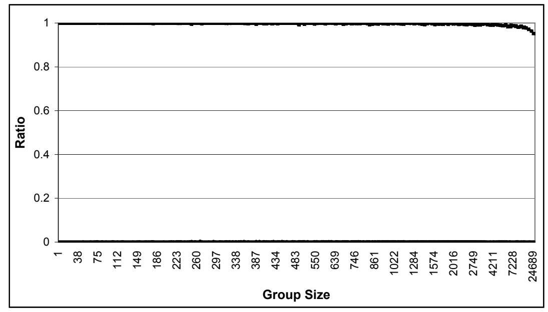
In this set of experiments, we were interested to learn whether we can produce the desired network properties by utilizing the naïve evolution model. The model can clearly capture the social network properties since the process of creating it is the same as in our co-evolution model. In terms of the affiliation network properties, we used the naïve model to produce a social network similar to Flickr, as described in the previous experiment. Then we created the desired number of groups and picked the size of each one from a power-law distribution with the parameters observed in Flickr. Each group was populated by picking random users from the social network. As a result, the naïve model is able to capture the group size distribution. However, Figure 10(a) shows that it is not able to capture the average number of singletons and the average maximum degree as a percent of the group size. By picking random members, almost all members in each group end up being singletons (except for groups with very large sizes), and the average maximum degree is close to 0. Figure 10(b) shows that the model is also not able to capture the relation between degree and average number of group affiliations for nodes with lower degrees. The naïve model generates a relation between them which is closer to linear than a power law.

6. CONCLUSION

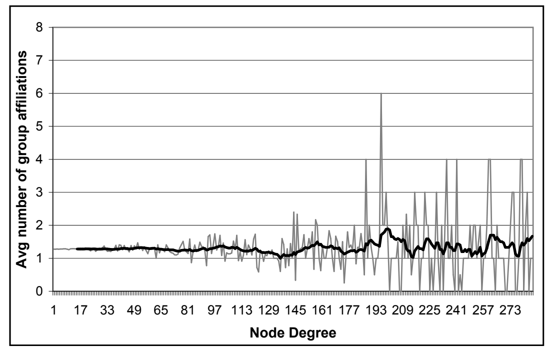
We presented a generative model for creating social and affiliation networks. The model captures important statistical properties of these networks, and provides new insights into the evolution of networks with both social and affiliation links. It shows that groups can be formed for various reasons and friendship links are not the only propagators of influence. We believe that this observation not only affects the design of network evolution models but it may have broader implications on other mechanism designs, such as group recommendation, information diffusion and viral marketing strategies.

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable feedback. This work was supported by NSF under Grant #0746930.



(a) Average number of singletons (dashed line) and average maximum degree (solid line)



(b) Degree vs. avg number of affiliation groups

Figure 10: The affiliation properties produced by the naïve model

7. REFERENCES

[1] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: Membership, growth, and evolution. *KDD*, 2006.

[2] A. L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[3] D. Chakrabarti and C. Faloutsos. Graph mining: Laws, generators, and algorithms. *ACM Computing Surveys*, 38(1), 2006.

[4] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. *SIGCOMM*, 1999.

[5] L. Getoor and C. P. Diehl. Link mining: a survey. *SIGKDD Explorations Newsletter*, 7(2):3–12, 2005.

[6] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins. Microscopic evolution of social networks. *KDD*, 2008.

[7] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: densification laws, shrinking diameters and possible explanations. *TKDD*, pages 177–187, 2005.

[8] M. McGlohon, L. Akoglu, and C. Faloutsos. Weighted graphs and disconnected components. *KDD*, 2008.

[9] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. *IMC*, 2007.

[10] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.