Growth Models of Social Networks

SURF 2007 Final Report
Olga Mandelshtam
Advisers: Mason Porter and Michael Cross

September 26, 2007

In the past several years, online social networks have gained great popularity and are becoming increasingly useful in many communities. In this paper, we study the structures of online networks, focusing on the growth mechanisms by which they evolve. We consider growth models based on preferential attachment - the tendency of a node to connect to a node with higher degree. Using Facebook as a dataset, we develop a mechanism to model the Facebook network into a simple framework. We employ numerical and analytical methods to investigate and compare the structures of the model networks.

1 Introduction

Many real-world phenomena, including cellular processes, the internet, food webs, friendships, and scientific collaborations, can be modelled by complex networks.[3] A *network*, or graph, consists of a set of nodes connected by edges. A social network takes individual people or organizations as its nodes and various social connections such as friendship, social contact, or other types of interactions as its edges. Traditional studies of social networks were based primarily on surveys and questionnaires, which limited the size of the data that could be studied. However, with the advent of powerful computers, the focus of network studies has shifted from smaller-scale properties of relatively small networks (such as features of specific vertices) to statistical (macroscopic) properties of larger graphs containing thousands or even millions of nodes and links.[6] Such networks have been found to have a number of complex and interesting structural properties.[3]

A growth mechanism is a set of rules by which new nodes and edges are added to a network. In this work, we study the different network properties that are generated by various growth mechanisms, with the intent to design a mechanism that would produce a desired structure. In particular, we use data from Facebook, an online community popular among college students, to design and test these mechanisms.

In the next sections, we introduce network properties and growth mechanisms. We present a mechanism to model the Caltech Facebook network, and outline some experiments done with the model. In the section "Future Work" we discuss analytical calculations and explore directions in which the work will be continued.

2 Network Properties

Degree Distribution
The *degree*,or valency, of a node is the number of edges it has. A network where nodes are connected by constructing edges uniformly at random would have a normal (Poisson) degree distribution. On the other hand, a network based on real data may have other distributions, such as a *power-law* degree distribution, which contains a few individual nodes with very high degrees and large number of nodes with lower degrees. Such distributions are heavy-tailed, in that there is a relatively large proportion of nodes with very large degree. A power-law becomes apparent in a graph of the cumulative degree distribution (for example in Fig **??**), when the tail of the graph tends to a straight line. The *power*, or exponent of the distribution, refers to the slope of that line.

Assortativity
Assortativity is the tendency of similar nodes to congregate. For example, in a social network containing both men and women, the coefficient of assortativity is close to 1 if the chance of like genders to be linked is high, and close to -1 if it is low.

Transitivity
Transitivity is the likelihood of a node to be linked to the neighbor of its neighbor.

The presence of these properties in a network is highly dependant on the mechanism of its growth.

3 Growth Models

In this work, we focus primarily on growth mechanisms based on *preferential attachment*, by which a node is more likely to connect to a node with a high degree. For example, in the network of the World Wide Web, a site would be more likely to link to Google, which has billions of links, than to a private website with very few links. In a social network, preferential attachment makes sense because it is more likely for a person to connect to someone with a large number of connections, as those people tend to be more social and popular, which is why numerous models of social networks include elements of preferential attachment.[6] To select a node by preferential attachment, each node is assigned an unnormed probability that is directly proportional on its degree; in the models we study, for simplicity we use the degree itself as the unnormed probability.

The simplest model we consider is the Barabasi-Albert preferential attachment model, in which at each timestep, a connection forms between a (uniformly) randomly selected source node and a destination node with unnormed probability of selection of

$$d(1 + \frac{1}{2\log d}) \tag{1}$$

where d is the degree of the node (here the probability of selection is very close to d; however, a more complicated equation was used, perhaps to simplify analytical calculations).[3]

Kumar et. al. present another simple growth model in their investigation of the online social networks of Flickr and Yahoo! 360.[5] Kumar et. al. categorize the network users as "passive" (P) - the users who display minimal activity, "linkers" (L) - the users who actively connect themselves to other members, and "inviters" (I) - the users who are mostly interested in recruiting their friends to recreate an already existing offline community in the online network. In the model, at every timestep, when a node enters the network, it becomes a P, L, or I node with a predetermined probability. Simultaneously, a number of new edges are created, with their origins chosen at random from the L or I nodes using preferential attachment. If the origin of an edge is L, the edge connects to an L or I node. We use a similar categorization of users in the mechanisms we develop to model Facebook.

To form a connection on Facebook, a user must request friendship with ("friend") another user, who must subsequently confirm the friendship. For an arbitrary Facebook user, there are several ways to friend others:

1. To link to someone from an already existing outside network, one can:
   (a) Invite a new member
   (b) Directly search for a member already on Facebook.
2. To link to someone one is not well acquainted with, one can:
   (a) Browse through the connections of adjacent users,
   (b) Add users from one of their groups (in which members typically have common interests)
   (c) Add users by performing a random search.

Facebook users can be differentiated into at least two different types. We identify the users as "active" or "passive". The passive users are mostly interested in maintaining their existing connections with people from an outside network, such as the real-world friendship network, and rarely participate in Facebook groups, or connect to strangers. On the other hand, active users are often involved in creating and participating in groups, and are much more interested in networking and connecting to others. Active users are also far less likely to reject a friendship request. In an article oriented on the social aspect of friendship networks, it is estimated that 50% of Facebook users are active.[8]

The mechanism we use to incorporate the most signifacant of the properties of Facebook attachment is precisely as follows. At each timestep:

1. A new user enters the network and is assigned "passive" or "active" based on the probability p of being passive.
2. The new user links to another user determined by preferential attachment, using the same formula as in the Barabasi-Albert preferential attachment model.
3. Add e edges. For each edge:
   (a) The source of the edge is chosen by preferential attachment; in addition, the unnormed probability that an active user is

selected is multiplied by b, since these users are more likely to participate in groups and forums.

(b) The destination of the edge is chosen by preferential attachment, and also as a function of the degree of separation from the source node: first a degree of separation s of 2, 3, or "random" is selected at random where s has an unnormed probability of $\log s^{-1}$. The probability that a passive user accepts the friendship is b^{-1}.

The parameters used in this model are the probability p of being a passive user, the relative likelihood b of an active user to form a friendship, and the number of edges e added at each timestep. We believe e can be calculated analytically from the average degree of a network, and p can possibly be estimated from psychological considerations.

TIME PLOT for FB4

This model is fairly sophisticated and scales as x^3,where x is the number of timesteps (see Fig **??**). Growing this model for 10,000 timesteps takes over a week. Since many of the Facebook networks we have data for have considerably more than 10,000 nodes, the resources and time required for this model make it non-ideal. Thus we explore a simplification of the model that still retains many properties of Facebook.

In the simplified model, an existing external network is used to model the real-world social network. In the same way that Facebook is to a great extent a subgraph of the real-world network, the model network is largely a subgraph of the external network that is used. The mechanism at each timestep remains similar to the sophisticated mechanism; however, when selecting a destination node to friend, a source node either draws from its neighbors in the external network, or selects a node at random, with some probability k. As in Fig **??**, jthe simplified model scales linearly with the number of timesteps and quadratically with the size of the external network used.

TIME PLOT for FB6

We consider the measurable network properties of average degree and exponent of the power-law degree distribution of the model networks. To avoid the effect of fluctuations in the system, it is necessary to grow the model until it reaches equilibrium, measured by the rate of change of the average degree and the power approaching zero. For the simplified Facebook model, the plots of the average degree and power vs time (Fig **??** and Fig **??**) demonstrate the onset of equilibrium at approximately ?? timesteps, when the rates of change fall below ??.

AVE DEG v TIME

POWER v TIME

4 Analytical Calculations and Future Work

We hope to analytically determine the asymptotic dynamics of the Facebook growth model we investigate as a function of the parameters. Analytical calculations would enable us to find the parameter values that correspond to the actual Facebook data. Moreover, we would be able to select the parameters that generate certain desired structural properties.

The calculation we are particularly interested in follows the principles of calculations performed by Krapivsky and Redner in their analysis of the connectivity of growing networks.[9]

We are also interested in considering the social and psychological implications of different user types in the Facebook (and other) networks. We are particularly interested in variations in behavior based on gender.

5 Conculsion and Discussion

The massive growth and increase in popularity of online social networks has greatly helped facilitate the spread of information, as it has become increasingly more common to use the online communities as tools for communication. The investigations into the structural properties and growth mechanisms of these communities has the potential to benefit, among others, the Caltech Alumni Association, which serves Caltech alumni by providing them with information and means of communication: an analysis of the alumni network may greatly facilitate and enhance these services.

We would like to thank SURF advisor Mason Porter, members of the research group for their input, the Caltech SURF program, and finally Andy Shaindlin and the Alumni Association for funding this work.

References

[1] K. Yuta, N. Ono, Y. Fujiwara, *A Gap in the Community-Size Distribution of a Large-Scale So-*

cial Networking Site. physics/0701168, 2007

[2] E. Kelsic, *Understanding complex networks with community-finding algorithms*. SURF 2005 Final Report. http://www.its.caltech.edu/ mason/research/kelsic.pdf

[3] R. Albert, A. Barabasi, *Statistical mechanics of complex networks*. Reviews of Modern Physics, V 74 (47-97), January 2002

[4] S. Strogatz, *Exploring complex networks*. Nature, V 410 (268-276), March 2001

[5] R. Kumar, J. Novak, A. Tomkins, *Structure and Evolution of Online Social Networks*. 2006. http://research.yahoo.com/publication/ structure_and_evolution_of _online_social_networks

[6] M. E. J. Newman. *The structure and function of complex networks*. SIAM Review, V 45(2) (167-256), 2003.

[7] http://netwiki.amath.unc.edu/

[8] Danah Boyd. *Viewing American class divisions through Facebook and MySpace*. Apophenia Blog Essay. Jume 2007. http://www.danah.org/papers/essays /ClassDivisions.html

[9] P. L. Krapivsky, S. Redner, F. Leyvraz *Connectivity of Growing Random Networks*. Phys. Rev. Lett. 85, 4629, 2000