Evolution of an Online Social Aggregation Network: An Empirical Study

Sanchit Garg
IIT Delhi
New Delhi, India

Trinabh Gupta
IIT Delhi
New Delhi, India

Niklas Carlsson
University of Calgary
Calgary, AB, Canada

Anirban Mahanti
National ICT Australia
Alexandria, NSW, Australia

ABSTRACT

Many factors such as the tendency of individuals to develop relationships based on mutual acquaintances, proximity, common interests, or combinations thereof, are known to contribute toward evolution of social networks. In this paper, we analyze an evolving *online social aggregator*, FriendFeed, which collates content generated by participating individuals on a variety of Web 2.0 services and allows easy dissemination of the aggregated content to other participants of the aggregator. Analyzing data collected between September 2008 and May 2009, we find that although preferential attachment captures the evolution of the network, its influence varies significantly based on how long ago a user joined the service. In particular, preferential attachment does not appear to apply to new entrants of the FriendFeed service. Analysis suggests that proximity bias plays an important role in link formation. We study the influence of common foci and find that individuals have a greater affinity toward those with similar interests.

Categories and Subject Descriptors

C.2 [**Computer-Communication Networks**]: Miscellaneous; C.4 [**Computer Systems Organization**]: Performance of Systems; H.3.5 [**Online Information Service**]: Web-based services; J.4 [**Computer Applications**]: Social and Behavioral Sciences

General Terms

Measurement

Keywords

Social Networks, Social Aggregation, Evolution, Web 2.0

1. INTRODUCTION

With the increasing popularity of online social networking (OSN) and content sharing services that enable Web users to share instant messages, blogs, videos, and photos, it is not surprising that users have accounts on many such services. This scenario results in the scattering of information, and has motivated the development of social aggregation services that seamlessly collate content posted by a user on various services and facilitate easy dissemination of the collated content. In this paper, we empirically study the evolution of one such aggregation service called FriendFeed [7].

FriendFeed is an online social networking service specializing in content aggregation. FriendFeed allows its users to aggregate content from services such as Twitter, Flickr, Reddit, StumbleUpon, and YouTube. At the time of writing, a total of 57 services were supported. By subscribing to a service, a user entitles FriendFeed to make API calls on the user's behalf and automatically obtain information on all activity of the user on their subscribed services. FriendFeed broadcasts this information to all "followers" of this user. For example, user A choosing to follow user B results in a unidirectional link from A to B (cf. Figure 1). In FriendFeed jargon, A is called the "follower" of B and B the "friend" of A. Similarly, A follows C. Now, any activity by B (or C) on her/his subscribed services can be viewed on a consolidated platform by A. Note that B may or may not choose to reciprocate and hence follow A. Essentially, the FriendFeed OSN may be viewed as a directed graph. In this paper, we refer to the initiator of a link as the *source node*, and the user being followed as the *destination node*.

FriendFeed enables link formation in several intriguing ways. For example, users can post comments on content posted by their "friends". Once a user makes a comment on a "friend's" post within the FriendFeed service, the post becomes visible to all those following the user. For instance, in Figure 1, if A posts a comment on a *tweet* by B, the tweet becomes visible to D and increases chances of D following B. Note that A need not have an account on Twitter to view B's tweets; users can view information posted by their "friends" on services on which they themselves do not have an account. A new user may also wish to follow a set of famous bloggers and active users, recommended by FriendFeed at the time of joining. In addition, a user can choose to follow people based on their email contacts. Links may also form because of social phenomena as discussed next.

The social science literature has documented many factors that influence the evolutionary dynamics of *offline* social networks [11, 18, 25]; here we consider these factors in the context of FriendFeed. For instance, *triadic closure* [25] wherein individuals form new relationships with friends of existing friends may apply to FriendFeed. More generally, relationships may be formed because of proximity between individuals [11, 25, 26]. Presence of *common interaction focus* or shared *group affiliation* between individuals is also known to increase the potential of forming new relation-

IMC'09, November 4–6, 2009, Chicago, Illinois, USA.

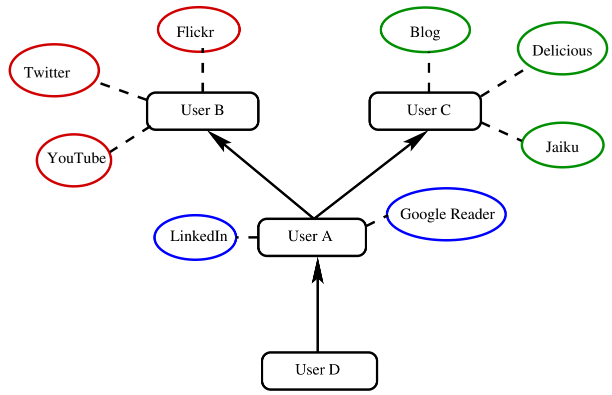


Figure 1: An illustration of FriendFeed's structure.

ships [18], and may result in relationships forming that are not purely guided by proximity.

In this paper, we ask the following questions with reference to the evolution of the FriendFeed social network:

1. Can preferential attachment [2, 21, 26, 27] explain evolution of the network? If yes, what role does age of the involved nodes play.

2. What role social phenomena such as triadic closure and proximity bias (defined in terms of the shortest distance between two nodes in the network graph) have in network evolution? How do these social pulls coexist with preferential attachment?

3. How does presence/absence of common group affiliation influence link formation? In this work, we assume, somewhat simplistically, that subscription to common services is an indication of common interaction foci.

Social aggregators are arguably a niche service that potentially target the "information hungry" Web users. In this work, we provide insights on how the social network of this new type of service is evolving. Our work uses this service and its users to highlight the importance of social factors in link formation.

Our observations are as follows. First, we find that the tendency of edge formation based on source and destination degrees, as a phenomenon, depends significantly on the age of the nodes; linear preferential attachment may explain observed data only for well-established nodes. Second, we find that the tendency to form edges with those close by appears to act as tie breaker between nodes of similar degree. Third, we find that exposure of users to common foci appears to increase the probability of edge formation as compared to an absence of such interaction points. In particular, bias owing to group affiliation is a more dominant factor for edge formation for relatively new source nodes. In general, our analyses suggest that preferential attachment models may apply only to well-entrenched, relatively old users, whereas the new users are often forming links that are best explained by theories such as group affiliation, homophily, and triadic closure.

Much of the related work on online social networks, with the exception of [4, 16, 19] (discussed in detail in Section 4), has focussed on other aspects such as structural properties, user behavior, and information dissemination (e.g., see [1, 5, 6, 10, 12–15, 20]), and have not considered the social phenomena mentioned above. In particular, one contribution of our work is to demonstrate how age, proximity bias, and group affliation may augment the well-known preferential attachment model. We also note that social aggregators are a relatively new type of service, and, to the best of our knowledge, prior work has not considered how these aggregators evolve.

This remainder of the paper is structured as follows. Data collection methodology is discussed in Section 2, followed by our results in Section 3. Section 4 reviews related work. Concluding remarks are presented in Section 5.

2. MEASUREMENT METHODOLOGY

Study of social network evolution requires data to be collected over a relatively long time period. In this section, we describe our measurement methodology. Section 2.1 discusses our data collection approach, Section 2.2 presents high-level summary of the data sets, while Section 2.3 outlines limitations of the data sets.

2.1 Data Collection

We obtained regular snapshots of the FriendFeed network over a period of 70 days between 26 February and 6 May, 2009. This data collection was complemented with two additional snapshots of the FriendFeed social network, one from September 2008, and another from January 2009. Table 1 summarizes the dataset.

We crawl FriendFeed's network once within every five days between 26 February and 6 May, 2009. Our crawler uses FriendFeed's API. The API returns for any user under consideration, information such as the list of users being followed, the services subscribed, and recent comments posted. The crawl on 26 February was seeded using information from the service's *public timeline* which returns details of 20 most recent activities on FriendFeed, followed by a Breadth First Search (BFS) to discover other users. In addition, we probed the public timeline for new users once every five minutes, or whenever the list of users maintained by the crawler was exhausted (in order to restart the crawl).

All subsequent crawls of the FriendFeed network were seeded by the list of users generated by the most recent crawl. The basic strategy was identical to that used for the initial crawl in that the public timeline was probed at least once every five minutes or whenever the crawler's list of users became empty. Each of these crawls lasted no more than five days. We note that invariably the crawler finished crawling data for all the users seen previously within the first two days of the crawl and then frequently probed the public timeline (for new users to restart the BFS) for the remaining three days. In fact, for the latter three days, on average, the crawler ended up probing the public timeline once each second. In the latter three days, the total number of new users discovered was on an average only 6% of the users crawled in first two days. We find that 89% of these newly discovered users have degree < 2; we hypothesize that there are users who have just joined the network. We do not include in our analysis the newly discovered users with degree > 10 (0.1% of new users discovered) as these might be the users that existed in the network earlier but were not crawled.

In addition, we also have two other snapshots of the FriendFeed network. The first snapshot was taken between 14 and 23 September 2008, and second taken between 6 and 12 January 2009. We note that the users found in September 2008 are a proper subset of those discovered in January 2009, which in turn are a proper subset of users found on 26 February 2009. Also, for every other crawl, the set of users discovered is a superset of those in the previous dataset. The two datasets of September 2008 and January 2009 give us an opportunity to create two classes of nodes based on age while studying the influence of age bias on social network evolution. For example, the September 2008 dataset allows us to consider nodes that are at least 150 days old when we started the longitudinal data collection.

Crawl #	Collection Time	N	E	d	E_b (%)	P
1	14 Sep-23 Sep 2008	113,247	1,403,444	24.7	56.7	4.02
2	6 Jan-12 Jan 2009	130,603	2,024,344	31.0	55.3	4.05
3	26 Feb-3 Mar 2009	162,293	3,005,545	36.9	46.9	3.99
16	1 May-6 May 2009	218,441	3,750,632	34.3	45.7	4.04

Table 1: High-level statistics. N is the total number of nodes, E is the total number of edges, d is the average degree, E_b is link reciprocity, and P is average path length.

2.2 Summary of Data Sets

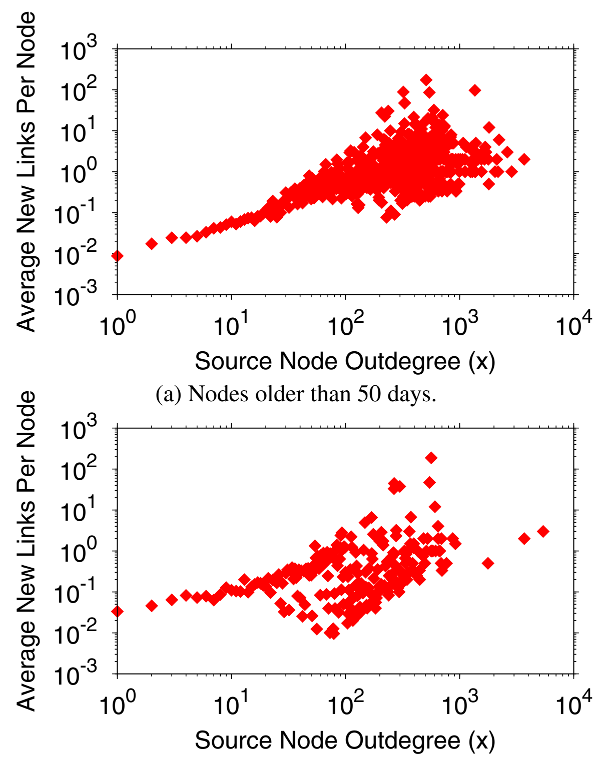
Table 1 presents a high-level summary of the collected data. At the end of our data collection, more than 200 thousand users were found with close to four million directed edges among them. In general, a majority of the users, approximately 60%, subscribed to between two and ten services. Those users that are subscribed to only one service are typically subscribed to Facebook, Twitter, or a blog. For further details on service aggregation and their usage, we refer the reader to our earlier work [9].

2.3 Limitations

It is important to note the limitations of the collected data to allow proper interpretation of our results. One limitation is a byproduct of the information returned by the APIs. The APIs do not provide us timestamps on when a user joined or when new links are formed between users, and therefore, we have to estimate these from the data collected. For the new users that join the network we use the time at which we found them as an estimate of their time of entry into the system. Any new link found between users that were present in an earlier crawl, however, is proof of relationship formed since the previous crawl. Here again, we associate the time at which we find a new link to the time stamp of the link formation.

It is also likely that we do not capture the entire FriendFeed network. However, because of our frequent polling of the public timeline, we believe that a very large fraction of the active users must have been captured.

Finally, we also note that approximately 12% of the nodes discovered during the crawl had *private* profiles. For these users, we are not able to obtain the list of users they follow, and any other information pertaining to their activities. Our analyses omits these private users.



(a) Nodes older than 50 days.

(b) Nodes younger than 50 days.

Figure 2: Preferential attachment for source node selection when forming new links.

3. RESULTS

There exists a vast, growing, and somewhat controversial literature on evolution of networks (e.g., see [2, 21, 26, 27]). Preferential attachment models, which received much attention following the work by Barabási and Albert [2] on "scale-free" networks [17, 21–23, 27], are considered in Section 3.1. Proximity bias, particularly with regard to its coexistence with preferential attachment is discussed in Section 3.2. Section 3.3 discusses our results on role of shared group affiliations in link formation. The results are based on the link formations seen in our continuous data collection between February and May, 2009.

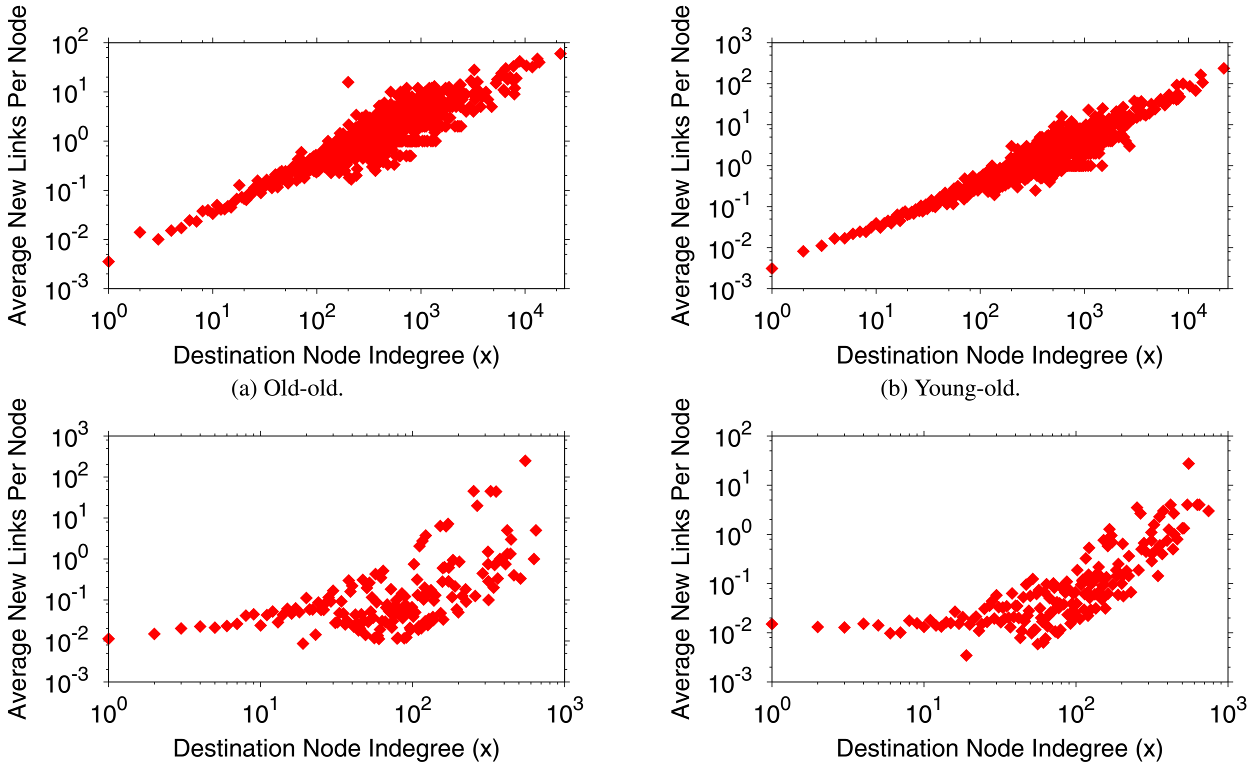
3.1 Preferential Attachment Models

Preferential attachment models for network evolution state that new nodes tend to connect to nodes with higher degrees rather than to nodes with lower degrees [2, 21, 26]. The general class of preferential attachment models require the probability p_i of selecting node i with k_i links to be proportional to ${k_i}^{\alpha}$, where α is a constant with $\alpha \approx 1$ implying linear preferential attachment and $\alpha > 1$ implying super-linear preferential attachment [26]. Linear preferential attachment is known to result in scale-free networks [2, 26]. Empirical studies have found linear preferential attachment to apply to online social networks such as Delicious, Flickr, and Yahoo! Answers [16, 19] but not to LinkedIn [16].

We use Maximum Likelihood Estimation (MLE) principle to estimate the parameter α [3, 28]. Specifically, we assume $p_i \propto {k_i}^{\alpha}$ and for every new edge that forms in the network we measure the log-likelihood that this new link was formed under this model. We vary α between zero to two and report the α value that produces the highest likelihood value.

For the case of source node selection, we take node out-degrees as the k_i and obtain a maximum likely value of 0.8 of α, for the complete network. This suggests that linear preferential attachment reasonably applies to the selection of source nodes in formation of new links. We are interested in understanding whether the observed behavior depends on age of the nodes. Therefore, we divide nodes into different classes based on their age (estimated as the difference between time of link formation and the time since arrival of node in the system) and label the nodes as "younger than (x) days" and "older than (x) days". We study how age bias x affects preferential attachment.

Figure 2(a) and (b) plot the average number of new links formed by source nodes with out-degree k for the case when the source



(a) Old-old. (b) Young-old.

(c) Old-young. (d) Young-young.

Figure 3: Preferential attachment for destination node when forming new links. Average number of new links received as a function of node in-degree for four example cases.

Age Threshold (x)	Nodes older than x days	Nodes younger than x days
0	0.8	-
10	0.85	0.1
20	0.85	0.3
30	0.85	0.5
50	0.9	0.55
80	0.9	0.6
180	0.9	0.65

Table 2: α for different source node classes.

Age	α			
Threshold (x)	(old,old)	(young,old)	(old,young)	(young,young)
0	0.9	1.05	-	-
10	0.95	1.05	0.05	0.15
20	1.0	1.1	0.15	0.3
30	1.0	1.1	0.25	0.35
50	1.0	1.1	0.4	0.5
80	1.05	1.15	0.4	0.5
180	1.05	1.15	0.55	0.65

Table 3: α value for preferential attachment; columns correspond to groups $\langle$ source node class, destination node class $\rangle$ for new links, where groups are based on age threshold x.

nodes are older than and younger than 50 days old, respectively. While both plots suggest strong positive correlation, linear preferential attachment is especially evident for source nodes that are at least 50 days old at the time they form new links. (The respective α values were 0.9 and 0.55.) Our analyses suggests that linear preferential attachment applies to the observed data for "older" nodes but not for "younger" nodes. Further evidence is provided in Table 2 which presents estimated α values for different source node classes based on age thresholds.

Preferential attachment in the context of selecting destination node for new links refers to the preference of higher in-degree nodes to be chosen as the destination nodes. For the evolution of complete network, α is 0.9, suggesting that linear preferential attachment applies to selection of destination node for new links. As in the preceding analysis, we investigate the role of age. The nodes are divided into classes based on the age threshold (x). "Old" class consists of nodes that are x days or older, and "young" class consists of nodes that are not older than x days. Because destination node selection depends on the source node initiating the link, we not only categorize the destination nodes but also the source nodes from which the edges emanate.We consider four classes of "source-destination" nodes as discussed next.

Figure 3 plots the average number of new links falling on nodes of a particular in-degree, separately for each class of new links when the age threshold is 50 days. Figure 3 (a) and (b) suggest that the number of new links attaching to "older" destination nodes exhibits strong positive correlation with in-degree; however, when the destination nodes are "younger" the link formation is not linear in node in-degree, more so for nodes with smaller in-degrees, as shown in Figure 3 (c) and (d). Table 3 presents the α values for various thresholds.

We summarize the results. First, age of nodes plays a significant role in link evolution. In particular, unlike established nodes, the observed data of edge formation for relatively younger nodes cannot be explained based on the linear preferential attachment mechanism. As the average age of the *young* nodes increases, linear preferential attachment applies. Second, for the case of *old* desti-

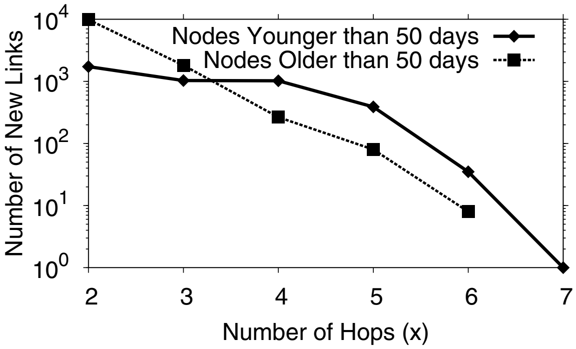


Figure 4: Evidence of proximity bias in link formation.

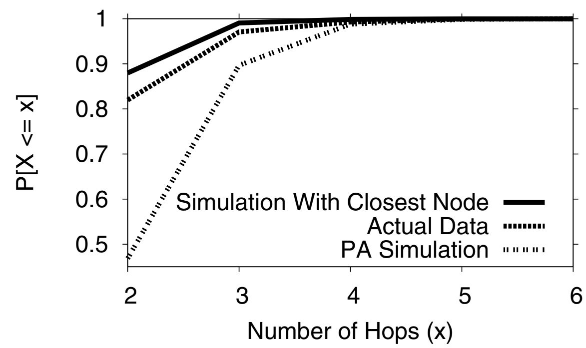


Figure 5: Proximity bias in link formation (empirical data and simulated networks)

nation nodes, the most likely values of α, when source nodes are *young* are consistently more than when they are *old*. Hence, for the selection of destination nodes belonging to *old* class, *young* source nodes attach more preferentially than *old* source nodes.

3.2 Proximity Bias

Social science literature has shown that proximity bias influences node selection when new relationships are formed [11]. In simple terms, proximity bias is the tendency of nodes to link with those nearby in the network graph. Triadic closure, which states that if two nodes are two hops apart (as, for example, B is from D in Figure 1), they are more likely to form a link; this is a special case of proximity bias.

In the preceding section, we provided empirical results that suggest applicability of linear preferential attachment to the selection of destination nodes when new links are formed, especially for the older nodes in the system. The preferential attachment models in Section 3.1 are proximity oblivious. Here, we empirically study the role proximity plays in destination node selection within the FriendFeed system.

Figure 4 shows the average number of new edges formed in a period of five days as a function of the distance x (measured in hops) between the nodes prior to link formation. A significant bias towards closer nodes in the network is seen in the link formations. We next investigate how preferential attachment and proximity bias potentially coexist.

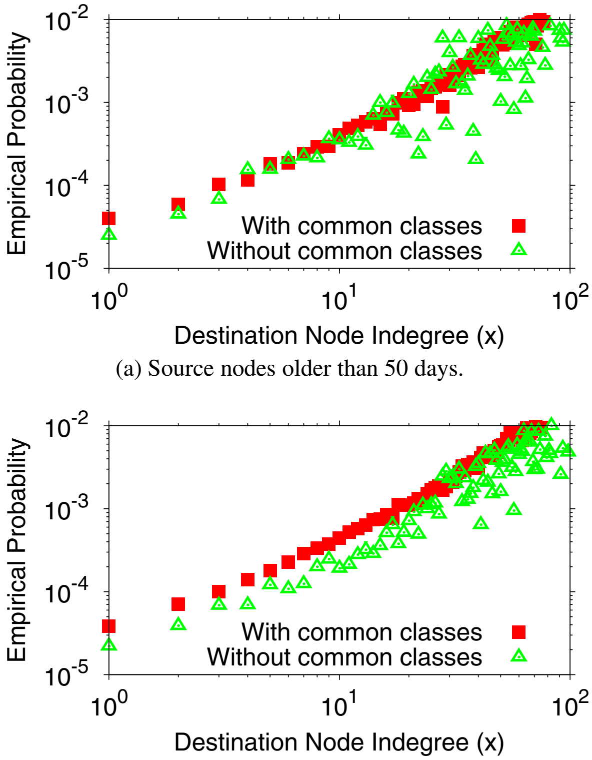
First, we consider whether or not preferential attachment results in some degree of proximity bias, for the FriendFeed network. The possibility exists because the FriendFeed network is characterized by a very small average path length of 4 (cf. Table 1) and preferred high degree nodes may be close to many nodes. For our purpose, we focus on link formations for destination nodes which are older than 50 days; these account for over 70% of new links found in our data set. We pick the source nodes of all the new edges formed and simulate the edge formation assuming linear preferential attachment [2]; i.e., for the source node forming new links, the destination node is selected based on the probability which is directly proportional to the in-degree of the destination node. Figure 5 shows the cumulative distribution function of the new links formed between nodes at a distance x hops for the simulated network graph. For the ease of comparison, the cumulative distribution corresponding to the collected dataset is also presented. Clearly, the two plots are significantly different. Empirically, the triadic closure property applies to 82% of the new links, compared to 47% of the links in the simulated data. As the number of hops increases to three and beyond, the plots for the two cases start to overlap. It appears that in actual network growth, proximity contributes much more than what would have been the case if only linear preferential attachment was applicable.

We now consider how proximity bias and preferential attachment may complement each other. Consider a simple model of link formation which utilizes both preferential attachment and proximity bias. Our model is as follows. For a new link to be formed, the source node first selects the destination node in-degree using linear preferential attachment and then among all the nodes of this in-degree, it selects the closest node to form the link (i.e., proximity bias is used as a tie breaker). To confirm the applicability of our model, we again simulate link formation and this time, unlike the previous simulation, after selecting the destination in-degree we pick the closest node rather than any random one; the results from this simulation are also shown in Figure 5 and these exhibit a closer match to the empirical data. We separately note that in our collected data, for over 80% of the cases the closest destination node of that in-degree is selected for link formation.

3.3 Group Affiliation

Group affiliation based evolution suggests that users exposed to common interaction foci form links with each other more often than those without [11, 24]. Kossinets and Watts [11], for instance, considered a social network of students, faculty, and staff, and found that students attending a common lecture were more likely to form links than those that did not. Similar to Kossinets and Watts, our objective is to discern the effect of group affiliations on evolution of the FriendFeed network. We use subscription to common services as evidence of common foci. In the ensuing discussion, we present evidence of group affiliation at work and analyze the coexistence of group affiliation with preferential attachment and proximity bias.

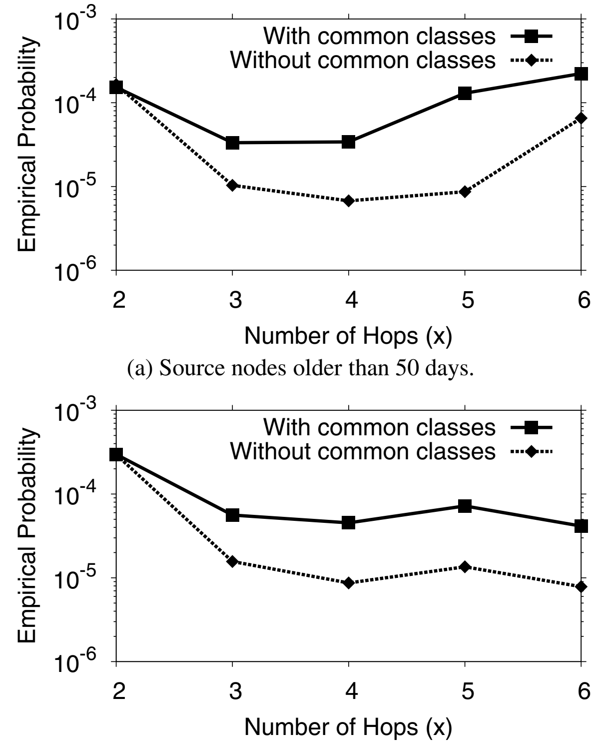
Figure 6 plots the empirical probability of edge formation between two nodes as a function of the in-degree of the destination node, corresponding to the two cases - the two nodes either share or do not share a common service. For the source nodes younger than 50 days (Figure 6(b)), the difference in the empirical probabilities is discernible till a threshold in-degree of 35. Note that the plot uses logarithmic scale and hence the differences are in orders of 10; for instance, for destination in-degree equal to 15, it is three times more probable that link forms if the nodes have a common service than if they do not. At first, it may seem that with respect to the maximum in-degree of 24,536, the threshold value of 35 is too low. However, it must be noted that edges falling on destination nodes with in-degree less than 35 comprise of more than 49% of the total new edges. The diminishing difference for high in-degrees indicates that here preferential attachment supersedes the effects of



(a) Source nodes older than 50 days.

(b) Source nodes younger than 50 days.

Figure 6: Evidence of group affiliation mechanisms.



(a) Source nodes older than 50 days.

(b) Source nodes younger than 50 days.

Figure 7: Group affiliation and proximity bias.

group affiliation. Qualitatively similar results were obtained for age thresholds less than 50 days.

For nodes older than 50 days (Figure 6(a)), on the contrary, no gap was seen in the empirical probabilities for the two cases. In general, it appears that as nodes age, their behavior is less dependent on factors such as group affiliation.

Figure 7 shows the effect of group affiliation with proximity bias by graphing the empirical probability of edge formation corresponding to the number of hops between the source and destination, for the two classes considered earlier; i.e., with and without common services. For number of hops equal to two, the empirical probabilities are nearly identical. This is perhaps due to the dominance of triadic closure. For hops greater than two, the role of group affiliation is evident, as can be seen from the empirical probabilities in the figure. For instance, for nodes younger than 50 days, two nodes which are four hops away are 8 times more likely to form an edge if a service is common than if not.

For nodes that just joined the FriendFeed network it appears that the group affiliation mechanism is significantly more influential than proximity bias. For instance, when both source and destination nodes are younger than 10 days, in over 95% of the cases, the destination node is not reachable from source node (hence proximity cannot be considered a factor for these); however, for 84% of these cases a common service exists. If we simulate the destination node selection (for these younger nodes in the system) and chose the destination node randomly then, on average, in only 49% of the cases a common service was found between the nodes. This suggests that group affiliation is an important factor for link formation, especially for relatively new nodes in the system.

4. RELATED WORK

Structural properties of online social networks such as degree distributions, diameter, and clustering coefficient have been studied [1, 20]. Kumar *et al.* [13] studied structural changes of Flickr and Yahoo! 360 networks due to their evolution. They observe that connected components that are not part of the giant component (i.e., isolated communities) evolve one user at a time, often merge with the giant component, and seldom merge with each other.

Evolution of online networks and models that may explain the observed growth has been considered [4,16,19]. Mislove *et al.* [19] indicate the applicability of preferential attachment and triadic closure on the Flickr network. Capocci *et al.* [4] evaluate the applicability of preferential attachment on evolution of Wikipedia. Among these, the work by Leskovec *et al.* [16] is closely related to ours. Leskovec *et al.* [16] show that linear preferential attachment [2] can model the source and destination node selection for new links reasonably well for Flickr, Delicious, and Yahoo! Answers complete networks. They compare the applicability of different models based on parameters like node degree, age, and their combination for source and destination node selection for the complete networks. We, on the other hand, focus on the variance in applicability of preferential attachment model (based on node degree) on nodes of different ages. For the networks considered by Leskovec *et al.* [16] they find that most of the new links lead to triadic closure [25]. They compare different triangle closing models using measures like tie strength [8] but do not explain "how preferential attachment and triadic closure co-exist?", especially for classes of nodes where preferential attachment exactly models selection of destination node for new links. We present a simple model, based on co-existence of preferential attachment and proximity bias, which concurs well with the observed data.

5. CONCLUSIONS AND FUTURE WORK

This paper examined the evolution of an online social aggregation network. Our analysis shows that age of nodes, proximity between nodes, and subscription to common services are factors that influence formation of new links in our data set. In particular, our analysis shows that by categorizing nodes based on age one can gain better insights into behavior of nodes, especially for applicability of preferential attachment models. We also find that proximity bias acts as a tie breaker, when preferential attachment explains the observed data reasonably well. Similarly, our analysis shows that group affiliation influences link formation, especially for nodes which have recently joined the network.

While the findings of our work may be specific to a social aggregation network, our analyses are fairly general and can find use in a variety of other settings, and there remains many interesting avenues to pursue in the future. One important direction is to consider the evolution of other online social networks and study how the factors discussed in this work influence network evolution. Study of other online social networks may help identity which factors are intrinsic to the evolution of online networks, thus leading to development of calibrated models for network evolution.

6. ACKNOWLEDGEMENTS

The authors are grateful to the anonymous reviewers for their constructive suggestions, which helped improve the presentation of the paper. This work has greatly benefited from discussions with Siddharth Mitra; we also thank him for help with developing the data collection tools used in this work. Martin Arlitt, Derek Eager, Max Ott, and Henrik Petander provided useful feedback on earlier versions of this paper. Brian Gallaway at the University of Saskatchewan kindly provided access to machines for data collection. Financial support for this research was provided by NICTA, iCORE, and NSERC. NICTA is funded by the Australian Government as represented by the Department of Broadband, Communications and the Digital Economy and the Australian Research Council through the ICT Centre of Excellence program.

7. REFERENCES

[1] Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong. Analysis of Topological Characteristics of Huge online Social Networking Services. In *Proc. ACM WWW*, Banff, Canada, May 2007.

[2] A. Barabási and R. Albert. Emergence of Scaling in Random Networks. *Science*, 286:509–512, October 1999.

[3] I. Bezáková, A. Kalai, and R. Santhanam. Graph Model Selection using Maximum Likelihood. In *Proc. ICML*, Pittsburgh, USA, June 2006.

[4] A. Capocci, V. D. P. Servedio, F. Colaiori, L. S. Buriol, D. Donato, S. Leonardi, and G. Caldarelli. Preferential Attachment in the Growth of Social Networks: The Internet Encyclopedia Wikipedia. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 74, 2006.

[5] J. Caverlee and S. Webb. A Large Scale Study of MySpace: Observations and Implications for Onl ine Social Networks. In *Proc. AAAI ICWSM*, Seattle, USA, April 2008.

[6] M. Cha, A. Mislove, B. Adams, and K. P. Gummadi. Characterizing Social Cascades in Flickr. In *Proc. ACM WOSN*, Seattle, USA, August 2008.

[7] FriendFeed. http://www.friendfeed.com.

[8] M. Granovetter. The Strength of Weak Ties. *American Journal of Sociology*, 78(6):1360–1380, May 1973.

[9] T. Gupta, S. Garg, N. Carlsson, A. Mahanti, and M. Arlitt. Characterization of Friendfeed: A Web based Social Aggregation Service. In *Proc. AAAI ICWSM*, San Jose, USA, May 2009.

[10] A. Java, T. Finan, X. Song, and B. Tsing. Why we Twitter: Understanding Microblogging Usage and Communities. In *Proc. Joint 9th WEBKDD and 1st SNA-KDD Workshop*, San Jose, USA, August 2007.

[11] G. Kossinets and D. J. Watts. Empirical Analysis of an Evolving Social Network. *Science*, 311:88–90, January 2006.

[12] B. Krishnamurthy, P. Gill, and M. Arlitt. A Few Chirps About Twitter. In *Proc. ACM WOSN*, Seattle, USA, August 2008.

[13] R. Kumar, J. Novak, and A. Tomkins. Structure and Evolution of Online Social Networks. In *Proc. ACM KDD*, Philadelphia, USA, August 2006.

[14] K. Lerman and A. Galstyan. Analysis of Social Voting Patterns on Digg. In *Proc. ACM WOSN*, Seattle, USA, August 2008.

[15] K. Lerman and L. A. Jones. Social Browsing on Flickr. In *Proc. AAAI ICWSM*, Boulder, USA, March 2007.

[16] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins. Microscopic Evolution of Social Networks. In *Proc. ACM KDD*, Las Vegas, USA, August 2008.

[17] L. Li, D. Alderson, J. C. Doyle, and W. Willinger. Towards a Theory of Scale-Free Graphs: Definition, Properties, and Implications. *Internet Mathematics*, 2(4):431–523, 2006.

[18] M. Mcpherson, L. S. Lovin, and J. M. Cook. Birds of a Feather: Homophily in Social Networks. *Annu. Rev. of Sociol.*, 27(1):415–444, 2001.

[19] A. Mislove, H. S. Koppula, K. P.Gummadi, P. Druschel, and B. Bhattarcharjee. Growth of the Flickr Social Network. In *Proc. ACM WOSN*, Seattle, USA, August 2008.

[20] A. Mislove, M. Marcon, K. P. Gummandi, P. Druschel, and B. Bhattacharjee. Measurement and Analysis of Online Social Networks. In *Proc. ACM IMC*, San Diego, USA, October 2007.

[21] M. Mitzenmacher. A Brief History of Generative Models for Power Law and Lognormal Distributions. *Internet Mathematics*, 1(2):226–251, 2004.

[22] M. Newman. The Structure and Function of Complex Networks. *SIAM Review*, 45:167–256, 2003.

[23] M. Newman. Power laws, Pareto distributions and Zipf's Law. *Contemporary Physics*, 46:323–351, 2005.

[24] M. E. Newman. The Structure of Scientific Collaboration Networks. *Proc. Natl. Acad. Sci.*, 98:404–409, January 2001.

[25] A. Rapoport. Spread of Information Through a Population with Socio-Structural Bias: I. Assumption of Transitivity. *Bulletin of Mathematical Biology*, 15:523–533, December 1953.

[26] D. Watts. The "New" Science of Networks. *Annu. Rev. Sociol.*, 30:243–270, 2004.

[27] W. Willinger, D. Alderson, and J. C. Doyle. Mathematics and the Internet: A Source of Enormous Confusion and Great Potential. *Notices of the AMS*, 56(5), May 2009.

[28] C. Wiuf, M. Brameier, O. Hagberg, and M. P. Stumpf. A Likelihood Approach to Analysis of Network Data. *Proc Natl Acad Sci*, 103(20):7566–7570, May 2006.