Characterization of FriendFeed – A Web-based Social Aggregation Service

Trinabh Gupta
IIT Delhi
India

Sanchit Garg
IIT Delhi
India

Anirban Mahanti
NICTA
Australia

Niklas Carlsson
University of Calgary
Canada

Martin Arlitt
HP Labs
USA

Abstract

Many Web users have accounts with multiple different social networking services. This scenario has prompted development of "social aggregation" services such as FriendFeed that aggregate the information available through various services. Using five weeks of activity of more than 100,000 FriendFeed users, we consider questions such as what types of services users aggregate content from, the relative popularity of services, who follows the aggregated content feeds, and why.

Introduction

With the increasing popularity of online social networking (OSN) and content sharing services, many Web users have accounts on many of these 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. FriendFeed (`www.friendfeed.com`) is one such service.

FriendFeed allows aggregation of information from a number of services that include popular social networking, video sharing, photo sharing, and blogging services. A FriendFeed user can choose to aggregate content from among the supported services into the user's FriendFeed profile page. A FriendFeed user can "follow" the activity of other users of this service by subscribing them as "friends". A friend on FriendFeed is a unidirectional relationship. Note that a user following another need not be collating information from the same set of services, and that any user with a public profile can be subscribed to as a friend.

On FriendFeed users can comment and start discussions on the aggregated content, similar to functionalities provided by typical OSNs. Commenting on aggregated content facilitates information dissemination in the FriendFeed network. For example, consider three users A, B, and C. Suppose that A follows B, and B follows C, but A does not follow C. If B comments on some of C's aggregated content, A will become aware of both B's comment and the aggregated content under consideration on C's profile page. This process may also result in A deciding to follow C as well.

This paper studies the FriendFeed service, with emphasis on social aggregation properties and user activity patterns. We are interested in understanding what types of services users aggregate content from, who follows the aggregated content, and why. We are also interested in understanding the characteristics of the FriendFeed social network and how they relate to the characteristics of the social network services that it aggregates. Note that FriendFeed being an aggregation service enables us to study different services from one common observation point, and allows us to get a unique "sneak peek" on how these social networking and content sharing services are being used by a common set of users. We believe that social aggregation services are of interest to information hungry Web users, especially those that like to always stay connected with the ongoings in the cyber world.

Social aggregation services are a recent phenomena, and, to the best of our knowledge, there has been no prior work on characterizing such services. Our work fills this void and provides the following insights (Gupta et al. 2009):

- The FriendFeed social network has significantly smaller network distances than some of the OSNs it aggregates. This is likely due to its "celebrity" driven nature, wherein a large fraction of the users are friends to a small set of users (e.g., famous bloggers). The node degree distribution properties of the FriendFeed network, however, are very similar to those reported for other OSNs.
- The active usage of services (as measured by the number of posts, for example) is significantly more skewed than the passive usage of services (as measured by subscriptions to services, for example). Micro-blogging services are the most popular in terms of both active and passive usage. Services such as video sharing (e.g., YouTube) and photo sharing (e.g., Flickr) observe high subscription rates but relatively low activity.
- The activity level of users is independent of the number of followers (or friends) they have and also the number of services they aggregate content from. Instead, it is typically dependent on the type of services subscribed.
- We observe higher user retention rates for certain services than has been previously reported for them. This is likely due to low activity users being more reluctant to join a service like FriendFeed. This hypothesis is further supported by the short tail of low activity users that we observe.

Table 1: Summary of Data Set.

Number of users	111,877
Private users (%)	12.7
Total number of directed edges	1,404,446
Average degree of a user	25.21
Content data collection period	30 Sept - 4 Nov 2008
Total number of observed posts	10,678,120

Table 2: Properties of the FriendFeed Social Network.

Network	Reciprocity	Diameter	Path Length
FriendFeed	0.53	12	4.02
Twitter	0.58	-	6
Flickr	0.70	27	5.67
LiveJournal	-	20	5.58
YouTube	-	21	5.10

Related Work

Properties of the network graph formed by users of various OSNs have been studied. Kumar et al. (2006) studied how the density, average diameter, and effective diameter of the Yahoo! 360 and Flickr OSNs have evolved over time. Their work categorized the users as passive members (singletons), users who migrate their offline social network to online (isolated communities), and linkers who fully participate in the evolution of the network. Java et al. (2007) studied Twitter's social network and found that users generally reciprocate links, and that the node degree distribution follows power law. Characteristics of three social networking services were compared by Ahn et al. (2007); they report a multi-scaling behavior of the degree-distribution and attribute it to the presence of heterogeneous types of users on the network. Shi et al. (2007) compare several network properties of two network samples of Blogosphere, differing in time duration and data collection methodology, and find them to be remarkably consistent. Leskovec et al. (2008) analyzed link formation in four OSNs and then presented a model to synthesize networks of arbitrary scale.

Recent work has considered how users use OSNs. Krishnamurthy et al. (2008) studied Twitter usage. They found that users can be classified as broadcasters, acquaintances, or spammers, based on indegree, outdegree, and number of *tweets* transmitted. Java et al. (2007) characterized Twitter users into information source, information seeker, and friends, based on the content of the tweets. Caverlee and Webb (2008) studied characteristics of MySpace users such as the number of friends they have, the number of comments they make, their gender, and their privacy preferences.

Data Collection Methodology

Our data collection consisted of two phases. The first phase captured the network of FriendFeed users, while the second phase captured the activity of the users identified in the first phase over a period of five weeks. Please consult (Gupta et al. 2009) for details of the data collection.

At the end of the first phase, for all users with publicly visible profiles, we had information on the users they followed and the services they were subscribed to. Because the FriendFeed API does not directly provide the users following a particular user (i.e., only directed edges emanating out of an user is available via API calls), we estimate the "followers" of a user from the social network graph by the users' respective out neighbors. Therefore, it is possible that we underestimate the indegree to users.

The second stage involved collection of data on the posts made by each user between 30 September and 4 November 2008. Each post contains the time when it was published, the service on which it was published ("internal" being FriendFeed itself) and comments (if any) made on this post.

Table 1 summarizes our data set. We found 111,877 users, with over 1 million directed edges among them. Among the users discovered, 12.7% had "private" profiles. For these users, we could not gather any information on posts and the users they follow; we ignore these private users in our analyses. In total, there were more than 10 million posts aggregated from various services; non-stationarities with respect to the amount of activity per day were seen, with there being more activity on weekdays than on weekends.

Characterization Results

Network Properties

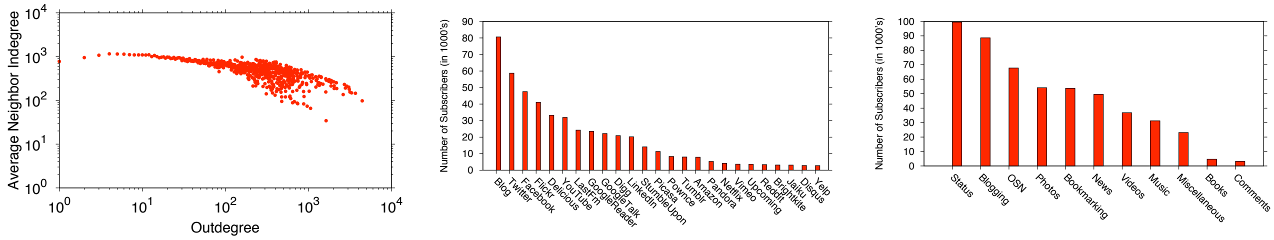
Table 2 summarizes some fundamental metrics capturing important network properties of FriendFeed's social network, along with known results for some OSNs (Ahn et al. 2007; Mislove et al. 2007; Kumar, Novak, and Tomkins 2006; Java et al. 2007).

FriendFeed's social network follows power-law (Gupta et al. 2009). Specifically, the number of users subscribed to an user (indegree) and the number of users followed by an user (outdegree) both follow power laws. Using techniques described by Clauset et al. (2007), we find that both indegree and outdegree distributions follow power law with shape parameters 2.28 and 2.12, respectively. Although FriendFeed allows for directional connections, these results are starkly similar to those typically observed for OSNs in which bidirectional relationships are required.

Another dimension of interest is the degree of reciprocity in relationships between nodes (Kumar, Novak, and Tomkins 2006). The FriendFeed network has a reciprocity of 0.53, which suggests that it is 53% probable for a directed edge to be present between two nodes in both directions, if there is an edge in a single direction. Similar reciprocity values have been observed in other non-aggregating networks.

We studied whether or not the "small world" phenomena applies here and if it does, to what extent and why. We evaluated the diameter and average path length of the FriendFeed network (Gupta et al. 2009). The FriendFeed network has a diameter of 12 and an average path length of 4.02. The average path length is shorter than those reported for Twitter, Flickr, and YouTube. We explain the difference in average path length by looking at the structure of the network.

The joint degree distribution (JDD) gives insight into the structure of the neighborhood of a node with degree k (Li et al. 2006; Mahadevan et al. 2006). The JDD is the average



(a) Joint Degree Distribution (b) Subscriptions per service (c) Subscriptions per category

Figure 1: Joint Degree Distribution and Subscriptions.

neighbor connectivity k_{nn}, which maps outdegree to average indegree of all nodes connected to nodes of that outdegree. An increasing k_{nn} indicates a tendency of high degree nodes to connect to other high degree nodes and a decreasing k_{nn} indicates a tendency of high degree nodes to connect to low degree nodes. The JDD provides information about one-hop neighborhoods around a node. The JDD can be further summarized by the assortativity coefficient r $(-1 \leq r \leq 1)$ where networks with $r < 0$ tend to have excess of links connecting nodes of dissimilar degrees and networks with $r > 0$ have excess of links connecting nodes of similar degrees.

Figure 1(a) presents the JDD of the FriendFeed network. A decreasing k_{nn} along with a negative assortativity coefficient of -0.0949 indicates the tendency of high degree nodes to connect to low degree nodes in the network. The disassortative graph has greater proportion of radial links that shortens the distance from the fringe to the core and as a result reduces the average path length. This is due to the “celebrity” driven nature of FriendFeed, in particular, the role played by FriendFeed’s 24 *recommended* users. We note that almost all the 24 *recommended* users have bidirectional relationship with each other, and over 25% of FriendFeed population is directly connected to at least one of these 24 nodes.

Flickr, LiveJournal, and Orkut’s social networks report an increasing value of k_{nn} and positive assortativity coefficients, and hence higher average path lengths. In these networks high degree nodes tend to connect to other high degree nodes and form a “core” of the network, different to “celebrity” driven nature of FriendFeed’s network.

Subscription to Services and Aggregation

This section dives into the social aggregation properties of FriendFeed. For simplicity, we use the number of subscriptions to indicate passive use of a service and content generation (i.e., posts) to indicate active use of a service.

Figure 1(b) presents the number of subscriptions to each of the 25 most subscribed services. We note that “blog” is the most popular service. This is due in part to all blogs being counted as the same service. Other highly subscribed services are Twitter, Facebook, and Flickr. Twitter is subscribed by around 60% of the users of FriendFeed. Approximately 50% of the user population subscribe to Facebook. Flickr is the most popular photo sharing service with around 40% of FriendFeed population subscribed to it. Other notables include Delicious, the most popular bookmarking service; YouTube, the most popular video sharing service; GoogleReader and Digg are the most popular news services.

Figure 1(c) shows the number of users subscribed to each category of service. Our service category classification uses the nomenclature specified by FriendFeed. In general, microblogging (status) and blogging appear to be the most popular services, closely followed by social networking and content sharing services. While subscriptions to individual services appear somewhat more skewed in their popularity than subscriptions to categories of services, the popularity distribution still is not that skewed. (Note that the y-axis is on linear scale.) It is possible that the ability to select services from multiple categories flattens the distribution.

The number of services each FriendFeed user subscribes to (i.e., the level of aggregation selected by each user) is also of interest (Gupta et al. 2009). We find that approximately 11% of the users have subscribed to no service and another 16% of the users have subscribed to only one service. Analysis of our data shows that almost all these users have very few incoming or outgoing links. Furthermore, if only one service is subscribed, the user is most likely to subscribe to Facebook. Facebook provides a customized FriendFeed toolbar that appear on the user’s Facebook profile, which allow them to take advantage of some of the aggregation features of FriendFeed. Approximately 73% users subscribe to two or more services, thus suggesting that the FriendFeed aggregation service serves its purpose.

User Activity across Services

While the findings above provide insights into the services that are popular to subscribe (and passively monitor), subscribing to a service does not necessarily imply that the user is actively using that service. To answer which services are popular to use, we turn our attention to the activity within each service and/or class of services. Because FriendFeed is an aggregation service, we believe that our results on user activity provide insights into the use of various Web-based social services and their relative popularity (at least for the information hungry Web users).

Figure 2 shows the number of posts through different services and Figure 3 shows the number of posts in different categories of services in our measurement period. As with subscriptions, micro-blogging services contribute to the most posts. In fact, it is the dominating service category with roughly five times as many posts as any other service. With the differences being significantly smaller in terms of number of subscriptions (cf. Figure 1(b) and (c)), this indi-

Table 3: User Retention Values.

Service	Wk 1-2	Wk 2-3	Wk 3-4	Wk 4-5
Twitter	89.97	90.68	90.62	90.64
Blog	78.67	79.69	79.42	80.33
Delicious	77.96	77.83	78.67	78.50
Internal	67.09	69.66	69.22	70.11
Digg	67.71	69.19	68.70	69.61
Flickr	65.45	66.04	66.61	68.92

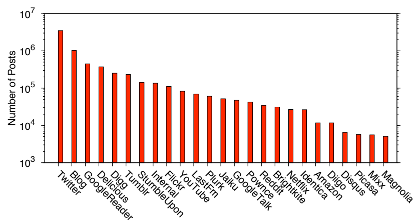


Figure 2: Activity across services.

cates that status messages update are used much more frequently compared to other feeds. Although much of the blog-service posts may be attributed to counting all blog services together, nonetheless note that Twitter has significantly more posts.

The results show relatively high active content generation rate through news services (e.g., GoogleReader and Digg). We also note that some of the most popular video sharing (e.g., YouTube), photo sharing (e.g., Flickr) and music services (e.g., LastFm) do not observe higher activity, although many users subscribe to these services. There is no content from Facebook as Facebook only allows a status bar to access FriendFeed but does not allow exporting of content on Facebook to FriendFeed. Finally, we note that a significant portion of posts are posted through FriendFeed's internal service and hence not attributed to any particular service.

Table 3 shows percentages of users retained for example services across each week-long interval of our measurement period. We consider a user active if the user has posted at least one entry during the week under consideration; this user is considered "retained" if the user posts at least one entry the following week (Java et al. 2007). The retention percentages are almost constant for the duration of measurement. Twitter retains an impressive 90% of its users, which is significantly higher than the other services. This can be attributed to the microblogging nature of the service.

Blogs and Delicious retain around 80% of their respective users, while Flickr and Digg retain around 70% of their users. Approximately 70% of users who post internally on FriendFeed follow on to post in the coming week.

Conclusion

This paper presents a measurement-based study of FriendFeed. We make several interesting observations that provide insights to its social aggregation properties and user activity patterns. For example, while we find the node degree distribution properties of the FriendFeed social network are

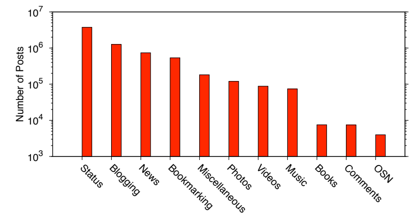


Figure 3: Activity across service categories.

very similar to those reported for other networks, its network has significantly smaller distances than some of the OSNs it aggregates. This is likely due to FriendFeed's "celebrity" driven nature. We also find that the active usage of services is significantly more skewed than the passive usage. Finally, we note that it appears that less active users are less likely to join an aggregation service such as FriendFeed.

References

Ahn, Y.-Y.; Han, S.; Kwak, H.; Moon, S.; and Jeong, H. 2007. Analysis of Topological Characteristics of Huge online Social Networking Services. In *Proc. ACM WWW*.

Caverlee, J., and Webb, S. 2008. A Large Scale Study of MySpace: Observations and Implications for Online Social Networks. In *Proc. AAAI ICWSM*.

Clauset, A.; Shalizi, C.; and Newman, M. 2007. Power-law Distributions in Empirical Data. arXiv e-prints.

Gupta, T.; Garg, S.; Mahanti, A.; Carlsson, N.; and Arlitt, M. 2009. Characterization of friendfeed - a web-based social aggregation service. Technical report, IIT Delhi.

Java, A.; Finan, T.; Song, X.; and Tsing, B. 2007. Why we Twitter: Understanding Microblogging Usage and Communities. In *Proc. ACM KDD*.

Krishnamurthy, B.; Gill, P.; and Arlitt, M. 2008. A Few Chirps About Twitter. In *Proc. ACM WOSN*.

Kumar, R.; Novak, J.; and Tomkins, A. 2006. Structure and Evolution of Online Social Networks. In *Proc. ACM KDD*.

Leskovec, J.; Backstrom, L.; Kumar, R.; and Tomkins, A. 2008. Microscopic Evolution of Social Networks. In *Proc. ACM KDD*.

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

Mahadevan, P.; Krioukov, D.; Fomenkov, M.; Huffaker, B.; Dimitropoulos, X.; kc claffy; and Vahdat, A. 2006. The Internet AS-Level Topology: Three Data Sources and One Definitive Metric. *CCR* 36(1):17–26.

Mislove, A.; Marcon, M.; Gummandi, K. P.; Druschel, P.; and Bhattacharjee, B. 2007. Measurement and Analysis of Online Social Networks. In *Proc. ACM IMC*.

Shi, X.; Tseng, B.; and Adamic, L. A. 2007. Looking at the Blogosphere Topology through Different Lenses. In *Proc. AAAI ICWSM*.