Social Interaction in Flickr: Mechanisms and Role in Link Formation

Karthik Gopalakrishnan
Indian Institute of Technology
Patna
Patliputra Colony
Patna, India
karthik.cs11@iitp.ac.in

Arun Pandey
Indian Institute of Technology
Patna
Patliputra Colony
Patna, India
arun.cs11@iitp.ac.in

Joydeep Chandra
Indian Institute of Technology
Patna
Patliputra Colony
Patna, India
joydeep@iitp.ac.in

ABSTRACT

Online social networking sites such as Facebook, Twitter and Flickr are among the most popular sites on the Web, providing platforms for sharing information and interacting with a large number of people. The different ways for users to interact, such as liking, retweeting and favoriting user-generated content, are among the defining and extremely popular features of these sites. While empirical studies have been done to infer network growth processes in these sites, few studies have characterized social interaction behavior and the role of social interaction in network growth.

In this paper, we analyze large-scale data collected from the Flickr social network to understand favoriting behavior and study the effect of a favorite on network growth. We do this using a systematic formulation of Flickr as a two-layer temporal multiplex network: the first layer describes the follow relationship between users and the second layer describes the social interaction between users in the form of favorite markings to photos uploaded by them. We shall henceforth refer to the social interaction layer on link formation in the follow layer through multiplex triadic closure.

Categories and Subject Descriptors

C.4 [Computer Systems Organization]: Performance of Systems—Measurement techniques; H.3.5 [Information Storage and Retrieval]: Online Information Services—Web-based services; J.4 [Computer Applications]: Social and Behavioral Sciences—Sociology

General Terms

Human factors, Measurement

Keywords

Flickr, social networks, measurement, interaction, growth

1. INTRODUCTION

Online social networking sites such as Facebook, Twitter and Flickr are among the most popular sites on the Web today. Users of these sites connect with each other by becoming friends, followers and so on. They also interact with each other by various means, such as liking, retweeting and favoriting user-generated content. At a macro level, the amount of such interaction occurring in these sites is enormous: over 1.8 million ‘likes’ were done on Facebook every minute [10] in 2013 and the number is likely much higher now. The availability of such data provides an excellent opportunity to analyze interaction dynamics in large-scale social systems. However, few studies [6, 13] have characterized the underlying mechanisms of social interaction in these sites. The growth of social networks has also been the subject of a number of research efforts [4, 5, 7], but the role of social interaction in network growth is not well understood as most efforts have exclusively focused on structural properties of nodes in social networks such as degree and clustering coefficient to infer growth processes.

In this paper, we analyze large-scale data collected [3] from the online social networking site Flickr, one of the most popular photo-sharing sites on the Web. For our analyses, we systematically formulate Flickr as a two-layer temporal multiplex network: the first layer describes the follow relationship between users and the second layer describes the social interaction between users in the form of favorite markings to photos uploaded by them. We shall henceforth refer to the first layer as the follow layer and the second layer as the social interaction layer. We analyzed the changes between snapshots of the network at different points in time to answer the following key questions:

1. What are the underlying processes or mechanisms governing the growth of the social interaction layer?

2. What is the effect of a favorite marking occurring in the social interaction layer on link formation in the follow layer?

Our investigation of social interaction in Flickr reveals that users create and receive favorites in direct proportion to the current number of favorites created and received respectively. Additionally, we find that users create favorites to new users in direct proportion to the current number of
users favorited. Further, we find that users receive favorites from new users in direct proportion to the current number of users who have favorited them. We also find that favorites are reciprocated quickly if at all they are reciprocated and observe a difference in reciprocation times for two different kinds of favorites that we define in a later section. We finally observe that favoriting in the social interaction layer leads to link formation in the follow layer through multiplex triadic closure.

We believe our work is an important step towards understanding interaction behavior in social networks. Our work can also lead to the development of new growth models for social networks which consider both social interaction and structural properties of nodes to predict the formation of new follow links. Overall, our work contributes new insights about social interaction in content-based communities like Flickr.

The rest of the paper is organized as follows. We present our formulation of Flickr in Section 2. We present analyses of favoriting mechanisms in Section 3 and link formation after a favorite in Section 4. We summarize related work in Section 5 and conclude in Section 6.

2. METHODOLOGY

In this section, we briefly introduce the Flickr data set and present our formulation of Flickr as a two-layer multiplex network.

2.1 Flickr

Flickr is a social networking site which at its bare minimum enables users to make friends, share photos and interact by favoriting and commenting on photos. The data set used in this paper was generated by crawling Flickr with a time granularity of 1 day for a period of 104 consecutive days, recording 2.5 million users, 33 million links and 34 million favorite markings over 11 million photos. We refer the readers to [3] for a detailed description of the functionalities offered in Flickr and the data collection methodology used.

2.2 Formulation As A Multiplex Network

Multiplex networks are networks in which a fixed set of nodes are connected by different types of links [2]. We formulate Flickr as a two-layer multiplex network $G_t = (G_t^f, G_t^s)$ at any point in time $t$, where $G_t^f = (V, E_t^f)$ is a directed, unweighted graph that represents the follow layer at time $t$ and $G_t^s = (V, E_t^s, w)$ is a directed, weighted graph that represents the social interaction layer at time $t$. $V$ is time-independent and consists of the entire set of users recorded by the data set. $E_t^f$ and $E_t^s$ are the respective edge sets such that $e_{ab}^f \in E_t^f$ implies that as on time $t$, user $a$ follows user $b$ and $e_{ab}^s \in E_t^s$ implies that as on time $t$, user $a$ has favorited user $b$’s photos. The number of photos favorited is given by $w(e_{ab}^s)$, where $w : E_t^s \rightarrow \mathbb{N}^+$. Other forms of social interaction such as commenting on photos exist in Flickr, however, they are ignored in our formulation since information such as who commented on a particular photo is not captured by the data set. Note that since the data collection methodology used does not take into account the possible deletion of a favorite marking, $E_t^s \subseteq E_{t+1}^s \subseteq E_{t+2}^s \subseteq \ldots \subseteq E_n^s$, where $t_0 \leq t_1 \leq \ldots \leq t_n$ are different points in time. Figure 1 shows a representative sub-multiplex network of $G_t^s$.

Having formulated Flickr as a two-layer multiplex network, we study the growth patterns of the social interaction layer in the next section.

3. FAVORITING MECHANISMS

In this section, we present analyses to understand the underlying growth processes of the social interaction layer. For our analyses, we define two kinds of favorites: initiating and continuing. A favorite to a photo uploaded by a given user is termed as an initiating favorite if it is the first favorite done by the favoriter to the user. We call it initiating because the favoriter has initiated a directed social interaction relationship with the user through the favorite. A favorite to a photo uploaded by a given user is termed as a continuing favorite if it is not the first favorite done by the favoriter to the user. We call it continuing because the favoriter is continuing the directed social interaction relationship previously initiated with the user. In the next section, we empirically examine preferential attachment as a potential favoriting mechanism.

3.1 Preferential Attachment

Preferential attachment is a growth model in which nodes are more likely to create links to nodes with high degree [1]. In the case of directed graphs, this model is split into preferential creation and preferential reception, where nodes create new links in proportion to their outdegree and receive new links in proportion to their indegree respectively [9]. The reasoning given for this split is that users control who they link to but do not have control on who links to them. This is true for the action of favoriting as well. We empirically study whether preferential attachment can be used to describe the growth of the social interaction layer.

From the data set, we constructed the social interaction layer as on the first and the last day of the crawl period and analyzed how the node outdegree/indegree in the social interaction layer on the first day correlates with the number of favorites done/received per day until the last day. If preferential attachment is the growth mechanism of the social interaction layer, a necessary condition is that there should be a linear correlation between the two. However, this correlation would only serve as strong evidence of preferential attachment. This is because an observed linear correlation is not sufficient to claim that preferential attachment is indeed the model that governs the growth of the social interaction layer as other mechanisms could also result in the same correlation.

Figure 2 contains plots depicting this analysis for the four types of node degree in the social interaction layer: weighted
Reciprocity is a phenomena in which the creation of a link from one node to another causes the creation of a link in the opposite direction. The inherent directionality in reciprocation makes it valid only in the case of directed graphs. We wanted to see if favorite markings are reciprocated: if user A favorites a photo uploaded by user B, does that favoritism cause user B to later favorite a photo uploaded by user A?

Since it is not possible to know whether a favorite caused the creation of the reverse favorite, we look at how long it takes to create a reverse favorite after the creation of a favorite. Figure 3 shows the cumulative distribution function (CDF) of the time taken for a reverse favorite to be created after a favorite. We analyzed the data separately for initiating and continuing favorites and additionally in the case of initiating favorites, we limited the analysis to users who have not favorited each other at all to see whether creating initiating favorites might cause discovery and promotion of one's own photos in the network. From Figure 3, we observe that over 50% of both kinds of favorites have reverse favorites created within 10 days. Thus, these favorites are likely to have been the cause for the creation of the reverse favorites. We also observe that for reverse favorites created within 10 days, reverse favorites are created slightly faster in response to initiating favorites than in response to continuing favorites. We postulate that this is the case because of a combination of the exploratory nature of users and homophily: the uploader is more inclined to view the profile of the favoriter when the favoriter has never favorited the uploader before (discovery), and the favoriter shares interests and uploads content similar to that uploaded by the uploader, thus causing a reverse favorite (promotion). While it is less likely that reciprocation occurs after 10 days, we see that for reverse favorites created after 10 days, reverse favorites are created faster in response to continuing favorites than in response to initiating favorites. We postulate that this is the case solely because of homophily: a repeat favoriter is more likely to share interests and be a part of the same community as the uploader than a one-off favoriter.

### 4. Link Formation After a Favorite

In this section, we briefly present a microscopic analysis of link formation in the follow layer after a favorite is created. For all users A, B and C such that user A follows user B before time $t$ and user B favorites user C at time $t$, we look at the follow layer after time $t$ to see whether user A follows user C. The creation of the A-C follow link can be termed as multiplex triadic closure, since the creation of the link closes a triangle over the two layers of the network.

Since the growth of the follow layer is captured during
the crawl period, we consider all favorites created during the crawl period and study the changes in the follow layer after each favorite is created. Figure 4 examines multiplex triadic closure for all favorites created during the crawl period. We observe that multiplex triadic closures occur in Flickr and the number of such closures occurring for a favorite after the favorite is created is directly proportional to the number of followers of the favoriting user.

5. RELATED WORK

In this section, we briefly review related work on social interaction. Wilson et al. [12] proposed the use of interaction graphs as a substitute for social graphs and found that such graphs are better validated by social applications. Valafar et al. [11] studied fan-owner interactions in Flickr and found that a small number of users in the friendship graph are responsible for most interactions. Yang et al. [13] studied the influence of factors such as user, tweet and time on retweeting behavior in Twitter. Lee et al. [6] compared favoriting reciprocity in Flickr and Twitter and found significant differences, which they postulated could be due to factors such as the kind of user and the type of content shared in these networks. Macskassy [8] recently studied social interactions in Twitter on several dimensions, including frequency of interactions and how the interactions are spread across different people.

6. CONCLUSION & FUTURE WORK

In this paper, we empirically studied social interaction in the form of favoriting in Flickr. We found that favoriting in Flickr is well-described by preferential creation and preferential reception. Further, most favorites are reciprocated within 10 days if at all they are reciprocated and we observed a difference in reciprocation times for initiating and continuing favorites. We also analyzed the effect of a favorite on follow link formation and observed the occurrence of multiplex triadic closure: when user A is a follower of user B and a favorite is created by user B to user C, a follow link is later established from user A to user C. We found that the number of multiplex triadic closures occurring for a favorite after it is created is directly proportional to the number of followers of the favoriting user.

We believe our work leads to new avenues for further research. The underlying sociological factors behind favoriting a photo can be studied to further understand the growth of the social interaction layer. New follow layer growth models can be developed and validated for content-based communities like Flickr that account for the role of social interactions in user discovery and follow link formation.

7. ACKNOWLEDGMENTS

The authors would like to thank Alan Mislove for graciously sharing a partly non-anonymized version of the Flickr data set described in [3].

8. REFERENCES