Social Browsing on Flickr

Kristina Lerman [*]
USC Information Sciences Institute
4676 Admiralty Way
Marina del Rey, California 90292
lerman@isi.edu

Laurie A. Jones [†]
Mills College
5000 MacArthur Blvd.
Oakland, California 94613
lajones@mills.edu

ABSTRACT

The new social media sites — blogs, wikis, del.icio.us and Flickr, among others — underscore the transformation of the Web to a participatory medium in which users are actively creating, evaluating and distributing information. The photo-sharing site Flickr, for example, allows users to upload photographs, view photos created by others, comment on those photos, etc. As is common to other social media sites, Flickr allows users to designate others as "contacts" and to track their activities in real time. The contacts (or friends) lists form the social network backbone of social media sites. These social networks facilitate new ways of interacting with information, e.g., through what we call *social browsing*. The contacts interface on Flickr enables users to see latest images submitted by their friends. Through an extensive analysis of Flickr data, we show that social browsing through the contacts' photo streams is one of the primary methods by which users find new images on Flickr. This finding has implications for creating personalized recommendation systems based on the user's declared contacts lists.

Keywords
Social networks; social filtering; social browsing

1. INTRODUCTION

Flickr[1] is one of the crop of new "social media" sites, along with blogs, wikis and their kin, that are transforming the Web to a participatory medium where the users are actively creating, evaluating and distributing information. Flickr's interface is exceedingly simple. A user can upload images to Flickr or view and comment on other users' images. A user can annotate an image with tags or submit it to a variety of special interest groups. Like many other social media sites, Flickr also allows users to designate others as "friends" or "contacts" and offers an interface to see in one place the latest images submitted by these friends. The friends lists form the social network backbone of Flickr.

This paper examines how people use Flickr: specifically, how they find new images to view. We claim that the social networks on Flickr facilitate new ways of interacting with information — what we call *social browsing*. Rather than searching for images by keywords (tags) or subscribing to special interests groups, users can browse through the images created by photographers they had selected as being most interesting or relevant to them. We should that social browsing is an important method used for navigating Flickr.

2. ANATOMY OF FLICKR

Flickr consists of a collection of interlinked user, photo, tag and group pages. A typical user page shows the latest photos she has uploaded, the images she have marked as her favorite," and her profile, which lists her contacts and groups she belongs to. A photo page provides a variety of information about the photo: who uploaded it and when, what groups it has been submitted to, its tags, who commented on the image, how many times the image was viewed or marked as a "favorite". A tag page shows either that user's images tagged with a specific keyword, or all public images similarly tagged. A group page shows the photo pool, group membership, popular tags, discussions and other information about the group. A user can browse Flickr via the links between these pages or go to the Explore page, which shows the 500 most "interesting" images uploaded on any given day. Explore page is meant to showcase the best images on Flickr. Although the selection algorithm is kept secret to prevent gaming the system, certain metrics are taken into account: "where the clickthroughs are coming from; who comments on it and when; who marks it as a favorite; its tags and many more things which are constantly changing."[2] In addition to these browsing methods, Flickr allows users to browse geotagged images through the recently introduced map interface.

The datasets. We used the Flickr API to download a variety of data for our study. For the data not provided through the API (e.g., the number of views), we wrote specialized data scrapers to extract this information from the Web pages. Since scraping required a separate HTTP request, this had an effect on the image statistics (e.g., number of views is incremented by every HTTP request). We corrected for this effect in post-processing. We gathered the following data:

---

[2]http://flickr.com/explore/interesting/

Explore set: consists of the 500 "most interesting" images (as chosen by Flickr's Interestingness algorithm) uploaded on July 10, 2006.

Apex set: consists of the 500 most recent images added to the *Apex* group[3]. This group is one of "the best of Flickr" groups that are intended to to showcase a selection of the best images and photographers.

Random set: contains 480 most recent images uploaded to Flickr on July 10, 2006 around 4 pm Pacific Time. Although we started with 500 images, some were made private or deleted entirely from Flickr.

For each image, we collected the name of the user who uploaded the image; number of views and comments the image received; number of times it was marked a "favorite"; the number of tags; the number of groups it was submitted to. We also extracted the names of users who commented on or favorited the image.

3. SOCIAL BROWSING

How do users find new images on Flickr: do they find them through groups, or by searching by tags? Do they find them by browsing through the photo streams of their contacts? We believe that the latter effect, what we call *social browsing*, explains much of the activity generated by new images on Flickr. Below we present a detailed study of the images from the *Random*, *Apex* and *Explore* sets that help us to answer these questions.

Figure 1(a) shows the scatter plot of the number of contacts listed for the users in the *Random*, *Explore* and *Apex* datasets vs the number of their reverse contacts. A reverse contact is the user who lists a particular user as a friend and watches her activity. Since the number of user's reverse contacts is not directly available, we had to estimate it by crawling the contacts network to depth two. Generally, users in all three sets had contacts or were listed as contacts by other users, with *Explore* and *Apex* users being better connected than *Random* users. The points are scattered around the diagonal, indicating equal numbers of contacts and reverse contacts (possibly indicating mutual contact links).[4]

3.1 Social networks and views

Unfortunately, Flickr does not provide a record of users who viewed an image. Instead, we establish the link between the number of contacts and views indirectly by showing a correlation between the number of views generated by an image and the number of reverse contacts the user has. Figure 1(b–d) show the strength of the correlation between image statistics and features, such as the number of contacts and reverse contacts the user who uploaded the image has, the number of pools to which the image was submitted, and the number of tags it was annotated with.[5] The image statistics are: (1) the number of views the image received and (2) the num-

[3]http://www.flickr.com/groups/apexgroup/

[4]Interestingly, four of the images in the *Explore* set came from users with no reverse contacts, and two of these were not shared with any groups. Both of these images were tagged with "panda."

[5]All the correlations with correlation coefficient $C_r > 0.1$ are statistically significant at 95% significance level.
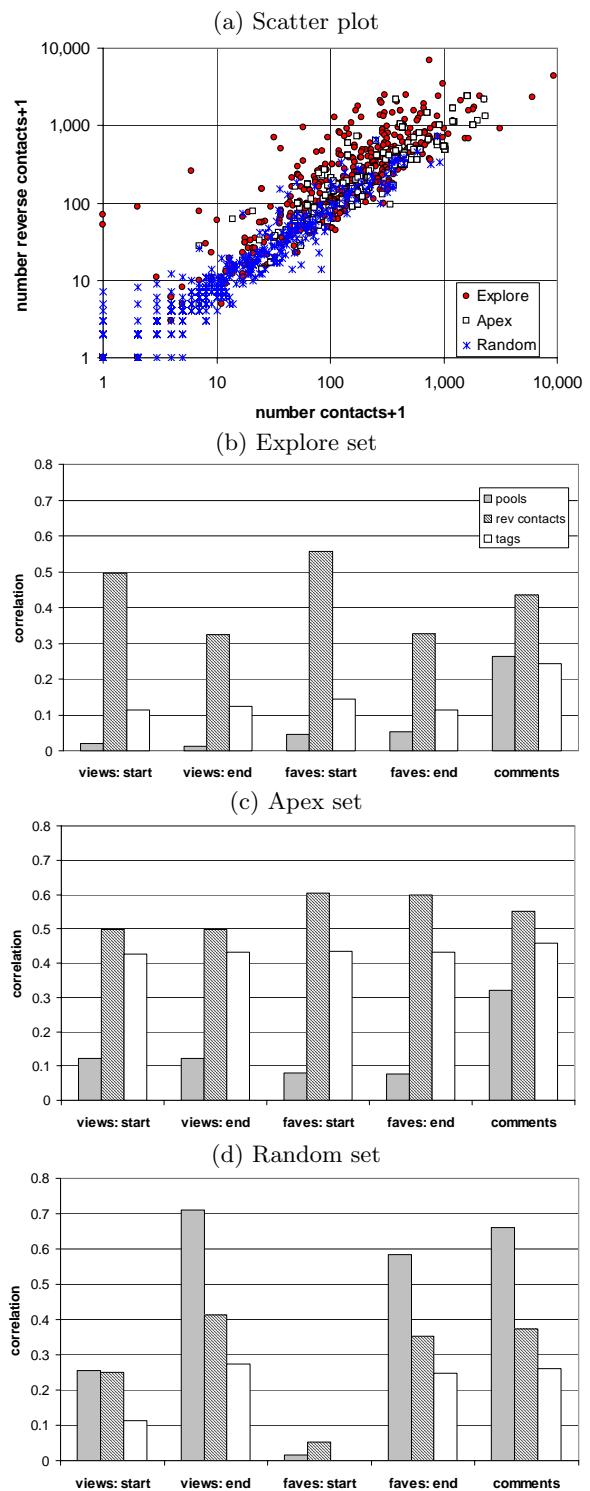
Figure 1: (a) Scatter plot of the number of contacts and reverse contacts of the users in the three datasets. (b–d) Strength of the correlation between image statistics (number of views and favorites at the beginning and end of the tracking period, number of comments) and image features (the number tags is has, pools it was submitted to and the size of the photographer's social network) for images in the three datasets.

ber of times it was favorited at the beginning and end of the tracking period and (3) the number of comments it received.

Apex and the *Explore* sets show similar correlation values at the start of the tracking period, where the number of views, comments and number of times the image was favorited correlates strongly (or at least moderately) with the number of reverse contacts the user has. At the end of the tracking period, however, the number of views, favorites and comments for the images in the *Explore* set is less strongly correlated with the size of the user's social network. This is explained by the greater public exposure images receive through the Explore page. Groups seem not to play any role in the generating new views, favorites or comments for these images. Tags appear to be uncorrelated to the image activity for the *Explore* set, but somewhat correlated in the *Apex* set. This could be explained by users clicking on the "apex" tag (that all *Apex* photos are required to have) to discover new photos in that pool.

The data presented above shows that, at least until the image gets to the Explore page, the number of views (and favorites and comments) images produced by good photographers receive correlates most strongly with the number of reverse contacts the photographer has. This is best explained by social browsing, which predicts that the more reverse contacts a user has, the more likely his or her images are to generate views. Views gathered by *Random* images correlate most strongly to the number of pools the image was submitted to, and only moderately to the number of reverse contacts. Since users in the *Random* sets have smaller social networks, they get more exposure by posting images to groups.

3.2 Social networks and comments

Although Flickr does not keep a record of who *viewed* an image, there is a record of who *commented* on an image. We use this information to track what fraction of comments came from within the user's social network and what fraction came from outsiders.

We collected the names of users who commented on the images in the three sets and compared them to the names of users in the photographer's social network. Figure 2 shows the proportion of comments coming from the user's reverse contacts, mutual contacts and users outside of the given user's social network, averaged over images. For the images in the *Random* set (Figure 2(a)) that were not added to any pools, 55% came from users who list the photographer as a contact, 51% came from users who are mutual contacts of the photographer, while only 38% came from users outside of the photographer's social network. As the image is posted to more and more pools, its visibility to users outside of the photographer's social network grows. For *Random* images that have been posted to 20 or more pools, only 41% of the comments came from mutual contacts, while the proportion of comments coming from strangers grew to 49%.

These observations are even more pronounced for the *Apex* set, shown in Figure 2(b). For *Apex* images that appear in only one pool (*Apex* itself), the share of comments made by the photographer's mutual and reverse friends is 69% and 71% respectively. Only 29% of the comments came from
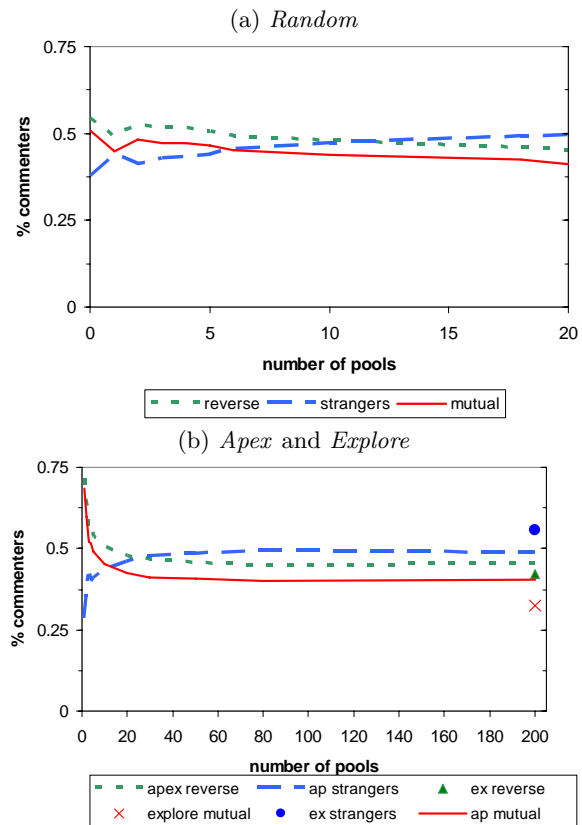

(a) *Random*

(b) *Apex* and *Explore*

Figure 2: Proportion of comments that came from the submitting user's reverse contacts, mutual contacts and strangers vs the number of pools to which the image was submitted for the three datasets

strangers. As the image gets shared with more groups, its visibility to outsiders increases, up to a point. After an image has been submitted to 30 groups, the share of the comments made by mutual contacts drops to 41%, reverse contacts drops to 47%, while the share of the comments coming from strangers grows to 48%. The image's visibility to strangers does not appear to increase by posting to additional groups. Sharing the image with 50 or more groups (up to 200) does not significantly change the distribution of comments coming from contacts and strangers. This seems to indicate that few of the groups are actively viewed (and commented on) by users.[6]

The symbols in Figure 2(b) are for the *Explore* images. We collected comments at the end of the tracking period, after they have been publicly shared through the Explore page. For this set, 56% of the comments came from strangers, far more than for the other two sets, reflecting the *Explore* images' greater public exposure. Still, about a third of the comments came from mutual and 42% from reverse contacts, showing that the user's social network is still active in commenting on and presumably viewing the images.

---

[6]Groups such as the various 1-2-3 groups, Score Me or Delete Me groups require that the user view, favorite or comment on other images in the pool before submitting their own images. These groups are likely the ones driving most of the traffic associated with posting images to groups.

4. RELATED RESEARCH

Social browsing is a natural step in the evolution of technologies that exploit independent activities of many users to recommend or rate information for a specific user. Collaborative filtering [2] used by many popular commercial recommendation systems attempts to find users with similar interests by comparing their opinions about products. They will then recommend new products that were liked by other users with similar opinions. Researchers have recognized [4] that social networks present in the user base of the recommender systems can be induced from the explicit and implicit declarations of user interest, and that these social networks can in turn be used to make new recommendations. The advent of social media finally made social filtering — or recommending new products or documents based on whether the user's designated contacts found these products or documents interesting — feasible. Social filtering is being used by other social media sites, e.g. Digg.com [3], to recommend stories to users.

Social navigation, the concept closely linked to collaborative filtering, works "through information traces left by previous users for current users" [1]. Like footprints in the snow that help guide pedestrians through a featureless snowy terrain, social navigation systems help users evaluate the quality of information, or guide them to new information sources, by exposing activities of other users. Using a best seller lists, the popular or hot pages to find documents is an example of social navigation. Social browsing is more targeted, as it presents to the user only the documents that user's friends found interesting.

5. CONCLUSION

Social media sites such as Flickr are on the leading edge of the social Web revolution. Flickr, a social photo sharing site, allows users to post and tag their own images, view, comment on, and mark as favorite other people's images. More importantly, these sites allow users to designate other users as friends or contacts. The resulting social networks offer users new ways to interact with information, through what we call social browsing and social filtering.

In this paper we studied three groups of images: (a) images chosen randomly from those uploaded on a specific day (*Random* set), (b) images deemed by other photographers to be of exceptional quality (*Apex* set) and (c) images chosen by Flickr's Interestingness algorithm to be the best of those uploaded on a specific day (*Explore* set). We analyzed a number of metrics associated with these images — the number of views, comments and favorites they generated — and studied the relationship of these metrics to features such as the number of pools they were submitted to, the number of tags associated with the images, and the size of the users' social networks. *Explore* and *Apex* images appear very similar on a number of metrics, despite the fact that *Apex* images are months old (and presumably had more time to be submitted to more pools or accumulate more tags) and very different from the *Random* images. Judging by the size of social networks, photographers from these two sets are also very similar — and distinct from the *Random* photographers. This suggests that Interestingness algorithm does

as good a job of selecting good photographers as users do.[7]

We offered two sources of evidence to show that social browsing is an important mode by which users navigate Flickr. We showed that for the images produced by good photographers, the views and favorites they receive correlate most strongly with the number of reverse contacts the photographer has. We also showed this relationship directly by linking comments to the photographer's social network.

Tags are a less important way to share images, while pools don't appear to place a significant role, except for *Random* users, perhaps because they do not have social networks as large as those of the good photographers. We showed that users also check the Explore page to find new images. Those images generate large number of views, favorites and comments, with a significant fraction coming from users outside of the photographer's social network.

Just as Google revolutionized Web search by exploiting the link structure of the Web, produced by independent activities of many Web authors, to evaluate the contents of information, social media sites such as Flickr show the possibilities of harvesting independent activities of interconnected users to personalize information evaluation. As social networks grows, it will be impossible for users to keep track of their contacts through the kinds of simple interfaces now offered. Better interfaces, for instance, ones that create personal Explore pages by finding "interesting" images from among those produced by the user's contacts, are a feasible solution to the information overload.

6. REFERENCES

[1] Andreas Dieberger, Paul Dourish, Kristina Hk, Paul Resnick, and Alan Wexelblat. Social navigation: techniques for building more usable systems. *interactions*, 7(6):36 –45, Nov/Dec 2000.

[2] Joseph A. Konstan, Bradley N. Miller, David Maltz, Jonathan L. Herlocker, Lee R. Gordon, and John Riedl. GroupLens: Applying collaborative filtering to Usenet news. *Communications of the ACM*, 40(3):77–87, 1997.

[3] Kristina Lerman. Social networks and social information filtering on digg. In *Proc. of International Conference on Weblogs and Social Media (ICWSM-07)*, 2007.

[4] Saverio Perugini, Marcos Andr Gonalves, and Edward A. Fox. Recommender systems research: A connection-centric survey. *Journal of Intelligent Information Systems*, 23(2):107 – 143, September 2004.

---

[7]Interestingly, there is only a 10% agreement between Interestingness and photographer choices, since only 10% of the *Apex* images were featured on the Explore page in the past.