

Social Browsing on Flickr

Kristina Lerman and Laurie Jones
University of Southern California
Information Sciences Institute
4676 Admiralty Way
Marina del Rey, California 90292
lerman@isi.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. We claim that 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.

1. Introduction

Flickr¹ 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, view other users’ images, comment on them and mark them as favorite. A user can annotate an image (usually their own) with tags. A user can also submit images to existing

special interest groups, or create a new one. Flickr is transparent: every username, every group name, every descriptive tag is a hyperlink that can be used to navigate the site, and unless it has been designated private, all content is publicly viewable and in some cases, modifiable. Clicking on a user’s name in a comment will take one to that user’s photo stream, from where one can also browse the user’s favorite images (by other photographers). One can examine the user’s tags, and check the groups he or she belongs to. Like many other social media sites, Flickr allows users to designate others as “friends” or “contacts” and offers an interface to see in one place the latest images submitted by friends. The friends lists form the social network backbone of social media sites.

The basic elements of Flickr — transparency, social networking, tagging — are also present to varying degrees on other social media sites, whether they are used for sharing bookmarks (e.g., del.icio.us), news stories (e.g., digg.com), musical tastes (e.g., MySpace.com) or academic papers (e.g., CiteULike.org). The emergent social tagging structures on these sites have already attracted the interest of researchers [2, 5]. This paper examines how people use Flickr: specifically, how they find new images to view. We claim that social networks on Flickr facilitate new ways of interacting with information — through what we call *social browsing*. Rather than searching for images by keywords (tags) or subscribing to special interests groups, users browse through the images created by their friends — the photographers they had selected as being most interesting or relevant to them.

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 [3] used by many popular commercial recommendation systems attempts to find users with similar interests by comparing their opin-

¹<http://www.flickr.com>

ions about products. They will then recommend new products that were liked by other users with similar opinions. Researchers have recognized [6] 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. [4] shows that social filtering to be an effective recommender system on the social news aggregator site Digg.com. 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 others within the user’s social network found interesting.

This paper shows that although Flickr offers users many ways of finding images — through tags, groups, calendar, maps, etc — social browsing explains the bulk of user activity. One of the consequences of social browsing is that images by photographers with large social networks are more likely to be selected for Flickr’s front page. Section 2 describes Flickr in more detail. We describe our data collection methods in Section 3 and analyze the impact of social networks on users’ browsing behavior in Section 4. We conclude, Section 5, by describing how social networks can be used for personalized image recommendation.

2. Anatomy of Flickr

A typical Flickr photo page is shown in Figure 1. It 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 and when, how many times the image has been viewed or bookmarked as a “favorite”. Clicking on the favorites link gives a list of users who have saved the photo in their “favorites” folder. Clicking on a user’s name brings one to their photo stream, which shows the latest photos they have uploaded, the images they have marked as their “favorite,” and their profile, which gives information about the user, which includes a list of their contacts and the groups they belong to. Clicking on the

tag shows user’s images that have been tagged with this keyword, or all public images that have been similarly tagged. Finally, the group link brings the user to the group’s page, which shows the photo pool, group membership, popular tags, discussions and other information about the group.

Flickr offers the user a number of ways to browse it. A user can browse by popular tags, through the groups directory, or by searching for a specific tag, group or user. In addition, one can browse Flickr through the Explore page and the calendar interface, which provides access to the 500 most “interesting” images on any given day. A user can also browse geotagged images through the recently introduced map interface. Finally, Flickr also allows for social browsing through the contacts interface that shows in one place the recent images uploaded by the user’s designated contacts. A user can continue browsing Flickr from any photo page by following (user, tag, group, etc) links.

Every day Flickr chooses 500 most “interesting” of the newly uploaded images to feature on the Explore page. Although the algorithm that is used to select the photos is kept secret to prevent gaming the system, certain metrics are taken into account: “where the click-throughs 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

Getting one’s image selected, especially as one of the top ten most “interesting” images, is a badge of honor to Flickr users that carries widely exercised bragging rights. Tracking the Explore rank of one’s photos has become a sport for some members, as getting in the top ten, or top one, allows one to submit the image to certain prestigious groups.

3. Data analysis

We used Flickr’s API to download a variety of data for our study. For the data not provided through the API (for example, 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:

Explore set: consisted of the 500 “most interesting” images (as chosen by Flickr’s Interestingness algorithm) uploaded on July 10, 2006. We saved the image’s rank on the first day (the lower the rank, the more interesting the image).

²<http://flickr.com/explore/interesting/>

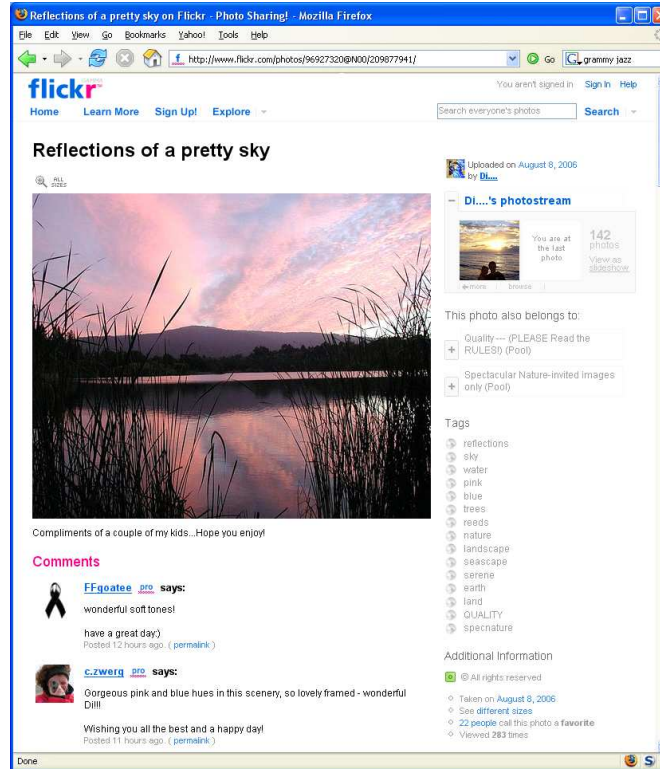


Figure 1. A typical photo page on Flickr

Apex set: consisted of the 500 most recent images added to the *Apex* group³. This group is one of “the best of Flickr” groups that are intended to showcase a selection of the best images and photographers. Photographs can be added to the group only by invitation from another group member.

Random set: contains 480 most recent of the 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, leaving us with a smaller set.

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.

For each image in the three sets, we tracked hourly the number of views, comments and favorites the image received over the period of eight days starting on July 10, 2006. While the views of the *Apex* images, some of them months old, did not change much, the

number of views received by the new images in the *Explore* and *Random* sets did change significantly. Figure 2 shows the number of views vs time received by select images from the *Explore* and *Random* sets. The curves are jagged because Flickr updates the counts of views every two hours. Images generally receive most of their views within the first two days, after which they were viewed less frequently. Some of the *Explore* set images show the “Explore effect” — the dramatic rise in the number of views received by images featured on Flickr’s Explore page.

The “Explore effect” is even more dramatic in Figure 3, which shows the total number of times the images in each set were viewed over course of eight days. While the images in the *Random* set received on average just 20 views, the *Explore* images received 450 views on average. *Apex* images show cumulatively more views because they are much older, sometimes months old, although there was no significant increase in the number of views over the course of the tracking period. It is worth noting that the top 20 *Explore* images show the biggest overall gain in views. This is probably caused by one of the following factors: (a) images ranked in the top 10 can be posted to the special *Interestingness*—

³<http://www.flickr.com/groups/apexgroup/>

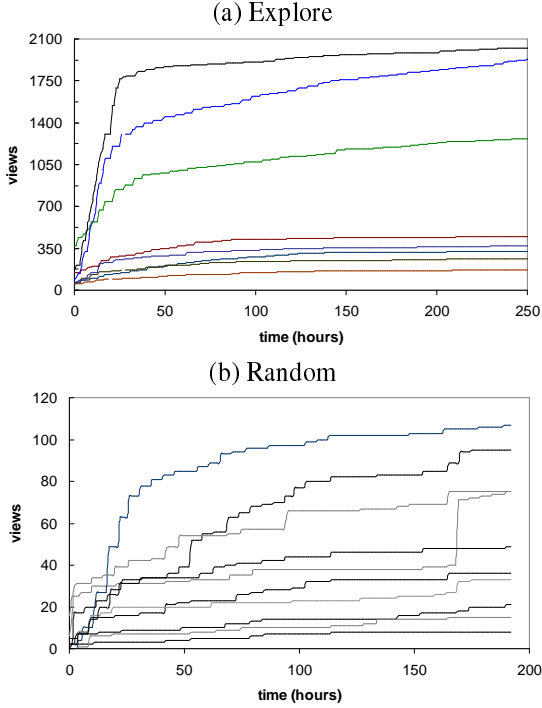


Figure 2. Cumulative number of times images in the (a) *Explore* and (b) *Random* sets were viewed over the time of the tracking period

Must be in Top 10^4 group, (b) people who browse *Explore* through the calendar interface probably scan the first two pages of images (10 images on each page) without paging further⁵, or most likely because (c) one of the top 20 images from the previous or current day is featured on the *Explore* page. Since Interestingness rankings are updated several times a day, it is possible images within the top rated 20 slip into the top 10, giving them access to the *Interestingness* group.

The number of times an image has been marked as a favorite (dotted lines in Figure 3) generally follows the number of views the image received. Although we do not show it, the number of comments closely tracks the number of times the image has been favorited.

In addition to image statistics, we extracted data about Flickr’s social networks. While the site shows a user’s list of contacts, one cannot easily get the list of user’s reverse contacts, i.e., other users who list the particular user as a contact. This is important information, since it shows how many people have access to the user’s photo stream. In order to reconstruct the so-

⁴<http://www-us.flickr.com/groups/interestingness/>

⁵Flickr Leech (<http://www.flickrleech.net/>) displays on a single page the thumbnails of all 500 “interesting” images for a specified day. It provides an additional portal for viewing *Explore* images.

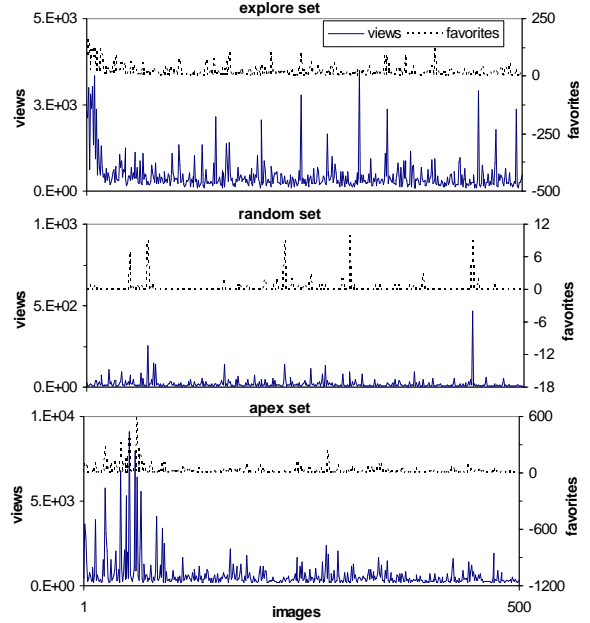


Figure 3. Number of times images in the *Explore*, *Apex* and *Random* sets were viewed and favorited by the end of the tracking period. Images in the *Explore* set are sorted by their rank, while *Apex* and *Random* images are shown in their chronological order of being added to the group or uploaded to Flickr respectively.

cial network, we crawled Flickr’s network of contacts. We limited the crawl to depth two due to the explosive growth of the network. Starting with about 1,100 unique users from our three datasets, we downloaded these users’ contacts, and their contacts’ contacts. This gave us a network with over 55,000 unique users and 5,000,000 connections. The resulting social network is not complete, but it allows us to estimate the number of reverse contacts a user has.

4. Social Browsing

Although getting selected for the *Explore* page boosts the number of views the image gets, *Explore* images had more views already after one day than most *Random* images attained after eight days. We believe that the more visibility an image has, the more likely it is to get more views, comments and be marked as a favorite by other users. How is the image’s visibility increased? This is related to how 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*, ex-

plains 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.

4.1. Pools and tags

When users upload images to Flickr, they have an option to share them with different groups, each with its own image pool. A large number of special interest groups already exist on Flickr, on a wide variety of topics — everything from Macro Flower Photography to one dedicated to the color orange — with new ones added daily. There is often a substantial overlap among group interests (there are more than a dozen groups dedicated to flowers alone), which results in images being posted to multiple groups. Figure 4(a) shows the distribution of the number of pools to which images in the *Explore*, *Apex* and *Random* sets have been posted. Although a typical user (*Random* set) does not share images with any groups, some users submit images to a surprisingly large number of groups — several users in the *Explore* and *Apex* sets have submitted their images to over 100, and on a few occasions over 200, groups.

Flickr also allows users to tag their images with descriptive keywords. Tagging is advocated by Flickr as a way to improve search of the user’s own, as well as other people’s, images. Figure 4(b) shows patterns in tagging usage across different data sets. Although very few *Random* users tag their images, *Explore* and *Apex* users do tend to use many tags, sometimes as many as 70. Interestingly, there seems to exist a preferred number of tags — around ten — for images in the *Explore* and *Apex* sets.

In both their tagging activity, as well as in submitting images to groups, *Explore* and *Apex* users are very similar to each other and different from *Random* users. There is considerable effort involved in sharing an image with a group, suggesting that social aspects of Flickr, such as sharing images with other users through groups, is very important to users, possibly more than being able to easily find them with tags.

4.2. Social networks

As explained above, Flickr allows users to designate others as “contacts,” and gives them instant access to the latest images their contacts upload to Flickr. The contact relationship is not symmetric. If user A designates user B as a contact, user A can see the photo stream of user B, but not vice versa. We call user A the “reverse contact” of user B. If user B also marks A as a contact, then they are each other’s “mutual contacts.”

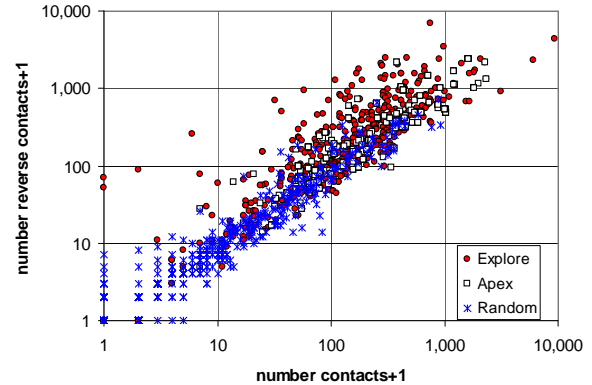
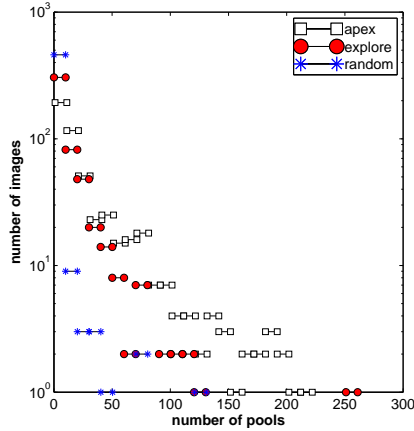


Figure 5. Scatter plot of the number of contacts and reverse contacts of the users in the *Random*, *Explore* and *Apex* datasets

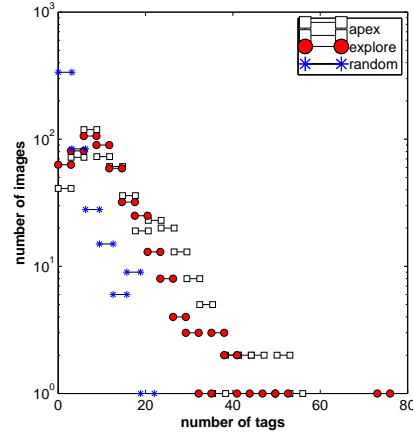
Do users take advantage of this feature of Flickr? Figure 5 shows the scatter plot of the number of contacts listed for the users in the *Random*, *Explore* and *Apex* datasets vs the number of reverse contacts they have. Since the latter number is not directly available, we had to estimate it by crawling the contacts network of users in our datasets to depth two, as explained above. Generally, users in all three datasets had contacts and were listed as contacts (reverse 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), although *Apex*, and especially *Explore*, users had greater numbers of reverse contacts.⁶

4.2.1. Social networks and views. Now that we have established that users do add contacts to their social networks, we will attempt to show they use them to browse Flickr. Unfortunately, Flickr does not make available the record of users who viewed an image. Instead, we establish this link indirectly by showing a correlation between the number of views generated by an image and the number of reverse contacts the user who uploaded the image has. Figure 6 shows the strength of the correlation between image statistics and social 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

⁶Interestingly, 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 about pandas, and were tagged with “panda.” This shows either that panda aficionados on Flickr are active and do use tags to search for new images of pandas, or people behind Interestingness algorithm chose pandas as the featured animal of the month.



(a)



(b)

Figure 4. Histogram of (a) the number of pools to which images from each set were submitted and (b) number of tags assigned to the images

tags it was annotated with.⁷ The image statistics are: (1) the number of views the image received and (2) the number 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 more 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.

4.2.2. 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 can use this record to track how many comments come from others within the user’s social network and how many come 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 their social networks. Figure 7 shows the proportion of comments coming from users’ reverse contacts, mutual contacts and strangers, or users outside of the given user’s social network. For the images in the *Random* set (Figure 7(a)) that were not added to any pools, 55% came from users who list the submitter as a contact, 51% came from users who are mutual contacts of the submitter, while only 38% came from users outside of the submitter’s social network. As the image is posted to more and more pools, its visibility to users outside of the submitter’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 7(b). For *Apex* images that appear in only one pool (*Apex* itself), the share of comments made by the submitter’s mutual and reverse friends is 69% and 71% respectively. Only 29%

⁷All the correlations with correlation coefficient $C_r > 0.1$ are statistically significant.

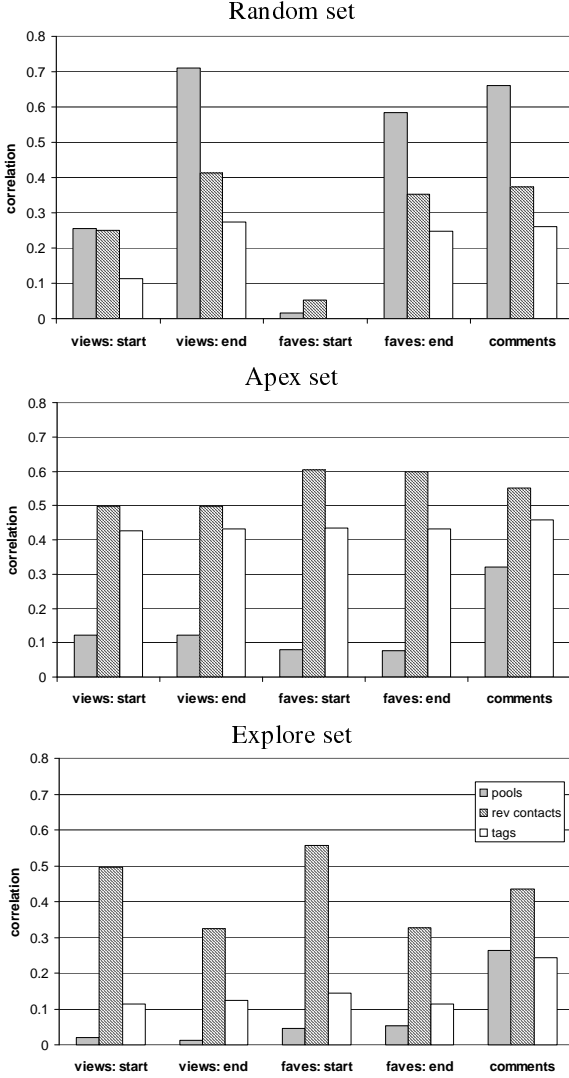


Figure 6. Strength of the correlation between image statistics (number of views and favorites at the beginning and end of the tracking period and the number of comments it received) and image features (the number tags it has, pools it was submitted to and the size of the photographer’s social network) for images in the three datasets.

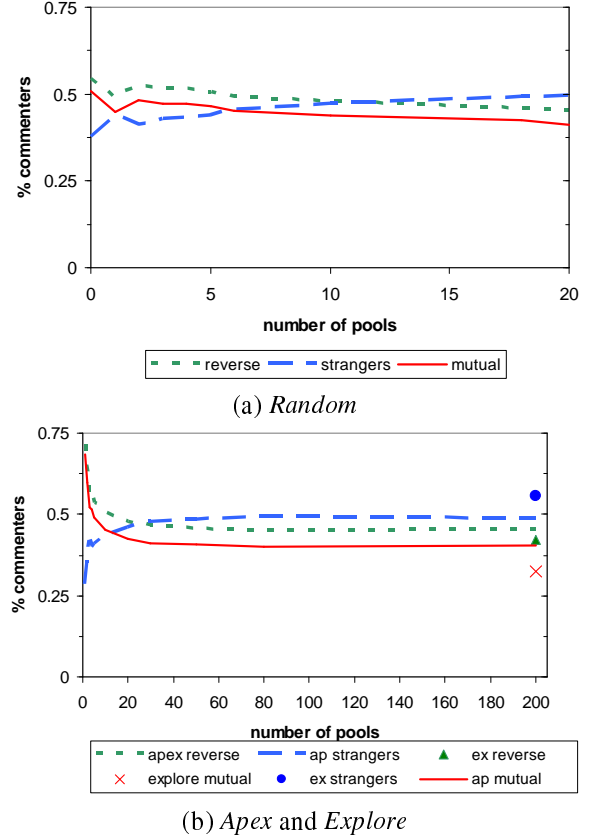


Figure 7. Proportion of comments that came from the submitting user’s reverse contacts, mutual contacts and strangers vs the number of pools to which the image was submitted for the three datasets

of the comments came from 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, even many more 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.⁸

The symbols in Figure 7(b) are for the *Explore* im-

⁸Groups 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.

ages. 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 come from strangers, far more than for the other two sets, reflecting the *Explore* images' greater public exposure. Still, about a third of the comments come 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.

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.⁹

We showed that social browsing is an important mode by which users use Flickr. We offered two sources of evidence for the claim that that users browse through the photo streams of their contacts. First, we showed that for images produced by good photographers, the views and favorites they receive correlate most strongly with the number of reverse contacts. We showed this re-

lationship directly by linking comments left on the images to the users in the photographer's social network. Again, almost 3/4 of the comments on the images of the good photographers, and 1/2 for of *Random*, come from other users within their social networks.

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. Still, the size of the photographer's social network appears to be the key to getting on the Explore page.

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, the social media sites such as Flickr show the possibilities of harvesting independent activities of interconnected users to personalize information evaluation.

Acknowledgements Laurie Jones wishes to thank CRA-W Distributed Mentor Program for support through a summer research fellowship. Kristina Lerman thanks Jim Blythe for insightful conversations. This research is based on work supported in part by the National Science Foundation under Award Nos. IIS-0535182 and BCS-0527725.

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] Scott A. Golder and Bernardo A. Huberman. The structure of collaborative tagging systems. Technical report, HP Labs, 2005. <http://www.hpl.hp.com/research/idl/papers/tags/>.
- [3] 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.
- [4] Kristina Lerman. Social networks and social information filtering on digg. In *submitted to International Conference on Weblogs and Social Media*, 2006.
- [5] Cameron Marlow, Mor Naaman, danah boyd, and Marc Davis. Ht06, tagging paper, taxonomy, flickr, academic article, toread. In *Proceedings of Hypertext 2006*, New York, 2006. ACM, New York: ACM Press.
- [6] Saverio Perugini, Marcos Andr Goncalves, and Edward A. Fox. Recommender systems research: A connection-

⁹Surprisingly, there is only a 10% agreement between Interestingness and photographers, because only 10% of *Apex* images were featured on the Explore page in the past.

centric survey. *Journal of Intelligent Information Systems*, 23(2):107 – 143, September 2004.