How to Identify Successful Actors of the Flickr Community and How to Determine Their Attributes

Marc Egger, Kai Fischbach, Peter Gloor, Andre Lang, Mark Sprenger

Seminar für Wirtschaftsinformatik und Informationsmanagement
Universitat Köln
Pohligstraße 1
50969 Köln

egger@wim.uni-koeln.de
lang@wim.uni-koeln.de
sprenger@wim.uni-koeln.de
fischbach@wim.uni-koeln.de
pgloor@mit.edu

Abstract: The emergence of Web 2.0 technologies has greatly enhanced capabilities of online social networking. Flickr is a social networking platform that enables users to share digital pictures. In contrast to other social networking sites, photos are the basic unit of information exchange in Flickr. We analyse relationships between users with a data set that covers 35% of the whole Flickr community. We define metrics of success for content, content providers, the relationship between both and identify implications of having information about successful content. We establish performance criteria and examine the correlation between the performance and the connectivity of users. Based on these key metrics we identify high quality content producers for a given category.

1 Introduction

Under the term "Web 2.0" the Internet is currently going through a new growth phase where end users create content and communities are built for user interaction. One of these Web 2.0 services is Flickr, a photo-sharing platform that allows users to upload photos, tag, comment and add them to favourite lists and build a personal social network. Using the relationships between users, pictures, tags and groups, analysis of Flickr's social structure gives a hint about the way people interact on a digital platform. Our project is part of a larger research project that aims to predict the success of social structures inside communities. Therefore we introduce a new approach to measure the performance of social entities inside the community and work out managerial implications.

Although Flickr has become one of the most active Web 2.0 platforms, little research has been done so far. Mislove et al. [Mi07] state that online social networks including Flickr are so different from real world ones that the same measurements cannot be used. Kumar et al. [KNT06] from Yahoo Research have investigated the evolution of Flickr and Yahoo!360 by using an empirical model to prove their classification of three user groups.

The process of pictures becoming popular through social network relationships is explored by Lerman & Jones [LJ06]. Advertising on social networking websites has also been a point of interest in Clemons, Barnett & Appadurai [CBA07] who suggest a change of marketing and advertisement efforts caused by new requirements.

This paper is divided into five sections. After a first look into the research questions and literature in section 1, section 2 deals with the analysis of Flickr's network structure. Section 3 describes the measurement of success of the community members. Section 4 contains the data analysis and the discussion including managerial implications. Section 5 gives a summary of the work and provides an outlook on further research.

2 Analysing the structure of the Flickr network

To get a view on the fundamental structure of the community it is necessary to map the network relationships. Our crawling of links between users follows an iterative approach, walking the users' friend lists. In the data gathering process, we were able to obtain the data of about 3.7 million user accounts. While the current number of registered users is unknown, a guess of 10.7 million users for December 2007 can be made using Polynomial Extrapolation based on the reports of Yahoo [Ya07], Elbies [El06] and Flickr [Fl06]. Possible reasons for missing almost two thirds of the users will be given in section 4. For the 3,690,312 discovered users, we found 46,928,035 directed friendship links. These can be further divided into friendship links that make up a unidirectional relation i.e. those who do not have a reciprocal link, and the ones forming a bidirectional relation together with its reverse link. This happens when the friendship request gets acknowledged by the recipient. Unlike other platforms, Flickr allows users to have unidirectional, unconfirmed friend relations. The directed links or unidirectional relations sum up to 18,999,909, whereas the reciprocal ones sum up to 27,928,126, building 13,964,063 bidirectional friendship relations. Seen from a friendship relation perspective, we therefore have 32,963,972 friendship relations with a fraction of 57.6% unidirectional and 42.4% bidirectional ones.

Kumar et al. [KNT06] state that the reciprocity of friendships in Flickr is around 70.2%, which appears misleading, because they define reciprocity as the fraction of edges having a reciprocal edge, neglecting that a bidirectional friendship consists of two edges. Therefore, bidirectional links are double counted in their analysis. Applying their method, we find 59.5% reciprocity in our data. 513,372 users have no incoming and 970,305 users no outgoing friendship links, while the average number of reciprocal friendships per user is 13.6. The average number of outgoing friendships is 17.3 and the number of incoming amounts to 14.8, counting only users with at least one incoming or outgoing friendship respectively. The user with the most incoming friend relations is "_rebekka" having 26,168 directed links. Implications from this will be examined later. We found 160,114 groups with a total number of group memberships of 18.305.400. The 34% percent of the Flickr users who are at least in one group are on average members of 17 groups. The group size averages 114.3 members, while the biggest group "Flickr Central" has 54,724 members.

3 Measuring the success of the community members

In order to measure success we have to define appropriate metrics. The first thing needed is an indicator that represents some kind of performance or success in the community. Since Flickr is a photo sharing community it seems adequate to measure the performance based on the photos a user has produced. Using the quantity of photos is less appropriate because it doesn't say anything about the quality and the appreciation of the community. To get a qualitative measurement indicator, we use the "Flickr favourites" function. Every produced photo on the platform can be added to a user's list of favourites, which represents some kind of appreciation of the photographer's work. Our strategy is to sum up the number of times the photos of one photographer are referenced in this way. This makes up the photographer's indicator called "FavCount".

$$FavCount(u_j) = \sum_{i=1}^{I} \sum_{k=1}^{K_i} FavPic(u_i, p_{ik}, u_j)$$

$$FavPic(u_i, p_{ik}, u_j) = \begin{cases} 1 \text{ if favorite picture } p_{ik} \text{ of user } u_i \text{ belongs to user } u_j \\ 0 \text{ else} \end{cases}$$

u_j = user for which the favcount is to be computed

u_i = user with favourite list

p_{ik} = picture k on favourite list of user u_i

The next indicator refers to the connectivity of users. We count the number of incoming and outgoing relations of users and use the "contribution index" indicator from SNA.

$$\text{contribution index} = \frac{\text{outgoing friend relations - incoming friend relations}}{\text{outgoing friend relations + incoming friend relations}}$$

The contribution index is defined in an interval of [-1; 1]. The more the index moves towards 1, the more outgoing friend relations a user has. If the index is negative the observed user has more incoming connections from other users.

4 Data analysis and discussion

To identify the most successful Flickr users, we determine the FavCount based on the favourite lists of a subset of one million random users. The top 5 users in reference to FavCount are _rebekka with a favCount of 87,543, _LaraJade (79,271), imapix (61,371), soleá (60,548) and T Glow (54,941) as seen in Figure 4-1. "_rebekka", or Rebekka Guðleifsdóttir from Iceland, who was even mentioned in the Wall Street Journal according to Shopify [Sh07], is by far the most appreciated and popular user on the network. She has 1,144 outgoing and 26,168 incoming friendship links, of which 1,029 are bidirectional. The first noticeable fact is that the five best performing users are members of an above-average number of groups and have a high number of friends. The much higher count of incoming than outgoing friendship links originates from its secondary function as bookmarks. New pictures posted by the users on one's friend list are shown immediately on the profile page. Therefore, a high incoming friendship count indicates that

many other users have "subscribed" to a user's photostream. To get a closer look at this, Figure 4-2 shows the users with the most incoming friendship relations. It is apparent that all of these users are performing well, having a FavCount of at least 20,000. The fraction of bidirectional friendship is low, which underlines the previous explanations. One exception from the rule is "merkley???", who confirms every friendship invitation he receives. This may be his unique way to success, because although being member of even fewer groups than the average Flickr user, he still gets about the same attention as the other top users.
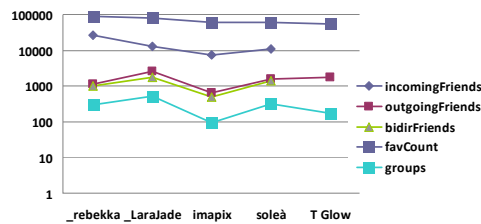

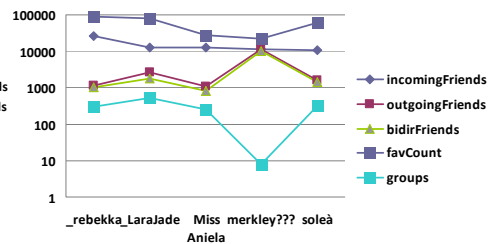Figure 4-1: Most performing users by FavCount

Figure 4-2: Most incoming friend relations

The usage of friendship links as bookmarks also makes the interpretation of the contribution index difficult. Because of the excess incoming friendship links, the contribution index for successful content producers is negative, as seen in Figure 4-3. It includes the 90,000 users with a FavCount greater than 1,000. The contribution index for these users is basically between 0.25 and -1 with only a few samples outside. This underlines the way people using friendship links to stay in contact. On the other hand there is no consistent behaviour of the best performers regarding answering friendship invitations. We also found users with a contribution index of 1, who have no publicly visible incoming friendship links, which may be due to their privacy settings.
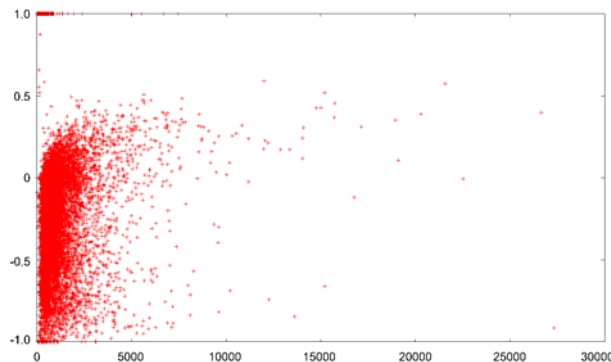

Figure 4-3: Contribution Index vs. total friendships of a user

4.1 Managerial implications of "FavCount" and incoming friend relations.

In contrast to other social platforms such as Facebook or MySpace, Flickr provides a simple way users can express appreciation for the work of other users. If users favour pictures of other users, conclusions about the performance of the photographer can be

drawn. Since high-performing users tend to acquire a lot of incoming friend relations, it is interesting to know which users and topics are attract high attention. This could be a great chance for advertising; the photos of prominent photographers could be used for product placement of banner ads and for sponsoring. One example is the above-mentioned "high performing" user _rebekka who owns approximately 26.000 incoming friend relations. The scenes in her photos sometimes include cars whose licence plate was exchanged by a label of the model-type. _rebekka states that she works for a large automotive company since her popularity on Flickr has become public [Sh07]. The company seems to use the photostream to conduct product placements, which implicates that the most performing users offer the most interesting advertising space. Although this might be good for mass marketing issues, advertising markets could be segmented by enhancing the FavCount with the tag function in order to find the best performing photographers of a certain topic. Based on the assumption that tags describe topics, the strategy is only to consider photos matching certain tags when determining the FavCount. Photographers with the highest FavCount for a certain topic have the best performance and get the most attention of their theme group. This information helps finding target groups, opinion makers and, of course good photographers of a certain topic. In general, finding lead members is valuable. Because social networks grow by participation, lead users contribute considerably to the network's activity and attraction. Having a community that consists of more active members who could even encourage members of other platforms to switch over is a valuable competitive advantage. Knowing good-performing members in the platform gives a clue which users should be encouraged to maintain their degree of participation in order to protect this advantage.

4.2   Discussion

In our data gathering process, we were able to obtain the data of about 3.7 million user accounts or 35% of the network, corresponding to everyone who has at least one incoming friendship relation coming from the core, or who is a member of at least one public group. There are various reasons why we miss almost two thirds of the users. First, missed users might be users who do not actively participate in the network. Second, these could be users not employing the social networking functions or trying to protect their privacy by hiding their profile, which can be done possible through the privacy settings.

Another potential limitation of our work stems from the definition of the performance indicators: only users with similar attributes are comparable. For example, two users who both have a FavCount of 1000 are barely comparable if one user owns 1000 and the other 10 pictures. The use of tags for defining topics is limited as well because of the ambiguity of synonyms and homonyms, so the application of FavCount to determine best performing photographers may not always be feasible depending on the topic.

5 Summary and Outlook

We analysed a large social network and defined performance indicators to explore success of users. The results show that Flickr users take advantage of the social network functions and form a densely connected network in respect to contact relations and group memberships. Using FavCount and friend relations proved that members with high performance tend to have more incoming friend relations and group memberships. Since friend relations are often used as some kind of bookmark to photostreams, the bookmarked user attracts attention that might be commercially used. The focus of future research lies in the closer analysis of the network structures using methods from the field of social network analysis. One goal would be to analyse the social structure of successful and unsuccessful groups to determine structural characteristics that might have a positive effect on the performance of a group. This would provide a possibility to make predictions of a group's performance. For example if there is a sub-community inside Flickr whose social structure evolves in a certain direction, it might be possible to make a prediction if the group tends to produce successful content.

While the present analysis only uses a small fraction of the available information, we anticipate using other cues in Flickr to determine the closeness of the relationships among users. Not only the contact list and group membership, but also the comments or tags posted to another user's photo can be used to define a correlation between users and their performance.

6 Bibliography

[CBA07] Clemons, E.K.; Barnett, S.; Appadurai A: The Future of Advertising and the Value of Social Network Websites: Some Preliminary Examinations. In: Proc. of the 9th Int. Conf. on Electronic commerce. Minneapolis, 2007. ACM, New York, 2007; pp. 267-276.

[El06] Elbies, J.: 2006 Innovators: Technology - PopPhotoJanuary/February 2007. [Online]. Hachette Filipacchi Media, 2006. Available from: http://www.popphoto.com/inamericanphoto magazine/3514/2006-innovators-technology.html [Accessed: 30th January 2008].

[Fl06] Flickr: Discussing Scout Statistics Update July 31, 2006 in fd's Flickr Toys. Scout statistics update July 31, 2006. [Online]. Available from: http://flickr.com/groups/flickrtoys/ discuss/72157594218765130/ [Accessed: 31st January 2008].

[KNT06] Kumar, R.; Novak, J.; Tomkins, A.: Structure and Evolution of Online Social Networks. In: Proc. of the 12th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, Philadelphia, 2006. ACM Press, Philadelphia 2006; pp. 611-617.

[LJ06] Lerman, K.; Jones, L.A.: Social Browsing on Flickr. In: Proc. of Int. Conf. on Weblogs and Social Media, Boulder, 2006. Extended version, 2006, available from: arXiv Cornell University Library at http://arxiv.org/abs/cs.HC/0612047 [Accessed: 31st January 2008].

[Mi07] Mislove, A. et al.: Measurement and Analysis of Online Social Networks. In: Proc. of the 7th ACM SIGCOMM conf. on Internet measurement, San Diego, 2007. ACM Press, San Diego, 2007; pp. 29-42.

[Sh07] Shopify - Rebekka Guðleifsdóttir, Homepage, 2007. [Online]. Available from: http://www.shopify.info/clients/rebekka/#interview [Accessed: 31st January 2008].

[Ya07] Yahoo! präsentiert deutsche Version von Flickr. Press release on German Yahoo site, 2007. [Online]. Available from: http://yahoo.enpress.de/pressemeldungen.aspx?p=1202 [Accessed: 30th January 2008].