Camera Brand Congruence and Camera Model Propagation in the Flickr Social Graph

ADISH SINGLA, Bing Search, Microsoft Corporation
INGMAR WEBER, Yahoo! Research Barcelona

Given that my friends on Flickr use cameras of brand X, am I more likely to also use a camera of brand X? Given that one of these friends changes her brand, am I likely to do the same? Do new camera models pop up uniformly in the friendship graph? Or do early adopters then "convert" their friends? Which factors influence the conversion probability of a user? These are the kind of questions addressed in this work. Direct applications involve personalized advertising in social networks.

For our study, we crawled a complete connected component of the Flickr friendship graph with a total of 67M edges and 3.9M users. 1.2M of these users had at least one public photograph with valid model metadata, which allowed us to assign camera brands and models to users and time slots. Similarly, we used, where provided in a user's profile, information about a user's geographic location and the groups joined on Flickr.

Concerning brand congruence, our main findings are the following. First, a pair of friends on Flickr has a higher probability of being congruent, that is, using the same brand, compared to two random users (27% vs. 19%). Second, the degree of congruence goes up for pairs of friends (i) in the same country (29%), (ii) who both only have very few friends (30%), and (iii) with a very high cliqueness (38%). Third, given that a user changes her camera model between March-May 2007 and March-May 2008, high cliqueness friends are more likely than random users to do the same (54% vs. 48%). Fourth, users using high-end cameras are far more loyal to their brand than users using point-and-shoot cameras, with a probability of staying with the same brand of 60% vs 33%, given that a new camera is bought. Fifth, these "expert" users' brand congruence reaches 66% for high cliqueness friends. All these differences are statistically significant at 1%.

As for the propagation of new models in the friendship graph, we observe the following. First, the growth of connected components of users converted to a particular, new camera model differs distinctly from random growth. Second, the decline of dissemination of a particular model is close to random decline. This illustrates that users influence their friends to change to a particular new model, rather than from a particular old model. Third, having many converted friends increases the probability of the user to convert herself. Here differences between friends from the same or from different countries are more pronounced for point-and-shoot than for digital single-lens reflex users. Fourth, there was again a distinct difference between arbitrary friends and high cliqueness friends in terms of prediction quality for conversion.

Categories and Subject Descriptors: H.1.1 [**Models and Principles**]: Systems and Information Theory

General Terms: Human Factors, Experimentation

Additional Key Words and Phrases: Brand congruence, brand loyalty, flickr, social networks

20

This is an extended version of a previous conference publication [Singla and Weber 2009].
I. Weber was partially supported by the Spanish Centre for the Development of Industrial Technology under the CENIT program, project CEN-20101037, "Special Media" http://www.cenitsocialmedia.es.
Authors' addresses: A. Singla, Bing Search, Microsoft Corporation, One Microsoft Way, Redmond, WA 98052; email: adishs@microsoft.com; I. Weber, Yahoo! Research Barcelona, Avinguda Diagonal 177, 08018 Barcelona, Spain; email: ingmar@yahoo-inc.com.

DOI 10.1145/2019643.2019647 http://doi.acm.org/10.1145/2019643.2019647

ACM Reference Format:
Singla, A. and Weber, I. 2011. Camera brand congruence and camera model propagation in the Flickr social graph. ACM Trans. Web 5, 4, Article 20 (October 2011), 25 pages.
DOI = 10.1145/2019643.2019647 http://doi.acm.org/10.1145/2019643.2019647

1. INTRODUCTION

How much information about personal brand preferences can be derived from knowledge about the brand preferences of one's friends? Which aspects affect the probability to have similar camera preferences to my friends? Are there factors that are crucial when it comes to brand loyalty? How do new models propagate through the friendship graph? Are there distinct differences between the growth and decline of models concerning the influence of neighboring users? These are the kinds of questions we study in this work in the setting of the Flickr[1] online photo sharing site.

Understanding how friends, either in the settings of social networks or defined through personal contacts, affect our decision regarding the purchase of any kind of item, is interesting for a number of reasons. First, it is interesting to see if there is any correlation at all between an online relation (a friendship link on Flickr) and an offline property (such as the camera brand or model owned by a user). In other words, can we use the "virtual" world as a (distorted) mirror for the "real" world or are the two unrelated? Second, there are clearly potential applications for more effective, targeted and personalized marketing campaigns, especially in social networks. If there is indeed a correlation between online and offline behavior then knowing a user's public online friends can give companies direct access to users who can influence purchase decisions. Third, it makes it possible to quantify the affect of sociological phenomena such as "peer pressure" concerning purchase decisions.

The availability of the Flickr data offers the possibility to address such issues, not just for a few hundred people via personal surveys, but for millions of users and using information for several years. The main observation required to perform such an investigation lies in the fact that the a large fraction of pictures uploaded to Flickr come with machine related metadata, which contains information about the brand of the camera used, the exact model specification and also the date when the image was taken. This then allows us to assign a brand and a model to a user for a given period of time. The exact details will be explained in Section 3. Furthermore, Flickr profiles (optionally) contain the geographic location of a user and indications of the degree of online activity, such as the number of images uploaded by a user. All this information creates a plethora of dimensions to explore.

Our study focused on two things. First, we looked at factors influencing the probability that two users own cameras of the same brand. Second, we looked at the propagation of new models through the friendship graph. Note that Flickr displays information about the camera used to take a picture on the picture's main page under "Additional information", so that even users without any direct personal contact can take note of this.

Concerning the question of brand congruence, our main findings are the following. First, a pair of friends on Flickr has a higher probability of being congruent, that is, using the same brand, than two random users (27% vs. 19%), where these numbers refer to the time period of March to May 2008. Second, this effect can *not* be solely explained by geographical factors. Friends are more likely to be in the same country, but even random users in the same country are still less likely (23%) to be congruent

[1] `http://www.flickr.com`

than friends, and in particular than friends in the same country (29%). Third, the degree of congruence goes further up for pairs of friends (i) who both only have very few friends (30%), and (ii) with a very high degree of cliqueness[2] (38%). Fourth, given that a user changes her camera model between March-May 2007 and March-May 2008, high-cliqueness friends are more likely than random users to do the same (54% vs. 48%), and 38% of the high cliqueness friends who do change their model in the same period, change to a camera of the same brand. Fifth, users using high-end cameras cameras are far more loyal to their brand than users using point-and-shoot cameras, with a probability of staying with the same brand of 60% vs. 33%, given that a new camera is bought. In both cases, users are more likely to buy the same brand again than it is that a random user would buy the particular brand. Sixth, for users of high-end cameras the cliqueness is most relevant, increasing the baseline probability for two friends using such cameras to be congruent from 47% to 66%. All of these observations are statistically significant at 1% or lower using a two-sided Student's t-Test for equality of means with different sample variances [Kirk 2007].

Concerning the propagation of models through the friendship graph, we observed the following. First, the growth of connected components of users converted to a particular, new camera model differs distinctly from random growth in that existing components tend to grow, rather than users being converted in an isolated fashion. Second, the decline of dissemination of a particular model is however close to random decline. This illustrates that users influence their friends to change to a particular new model, rather than from a particular old model. Third, having many converted friends increases the probability of the user to convert herself, with an apparent effect of "diminishing returns". Here differences between friends from the same or from different countries are more pronounced for point-and-shoot than for digital single-lens reflex users, probably caused by more localized marketing campaigns for the former than for the latter. Fourth, there was again a distinct difference between arbitrary friends and high cliqueness friends in terms of prediction quality for conversion, where one high cliqueness friend was "worth" about as much as 3-4 arbitrary friends.

Finally, as high-cliqueness friendship turned out to be the single strongest factor in determining brand congruence, we tried to understand Flickr's social graph better by looking at factors that correlate with high cliqueness. High cliqueness among pairs of friends increases the probability for pairs of users to be (i) family members (as indicated by a common last name), (ii) share interests (as indicated by Flickr group membership), (iii) live in the same country, and (iv) they reciprocate their friendship links. We believe that these fragmentary but consistent results point at a link between high cliqueness and close friendship in the "real" world.

The rest of this article is organized as follows. In Section 2, we discuss work related to our analysis. Section 3 gives details about our data set and how it was obtained. Section 4 contains the actual results of our work. This section is split into four parts. First, in Section 4.1, we will present the techniques used by us. Then, in Section 4.2, we focus on (i) static brand analysis for a single time period and on (ii) more straightforward properties to investigate. Section 4.3 then goes beyond this by looking at the evolution over time and at more subtle features influencing brand congruence. Section 4.4 takes a different angle and looks at how new models disseminate through the friendship graph. Finally, in Section 5, we discuss possible extensions of our work and give a summary of our main findings.

[2]This term refers to the overlap between two sets of friends and is formally defined in Section 4.3.

2. RELATED WORK

In this section, we will discuss work, which is related to our study, as it (i) also looks at Flickr, (ii) deals with analysis of social networks more generally, and (iii) looks at brand congruence and related marketing analyses.

The present article is an extended version of previous work by Singla and Weber [2009]. The results and tables concerning the brand congruence are identical, but (i) all results concerning the propagation of new models, (ii) the description of the (surprising) distribution of point-and-shoot (P&S) vs. digital single lens reflex (DSLR) cameras in terms of developed and developing countries, (iii) the observations concerning the tendency of users to stick to the same camera type (P&S vs. DSLR) given a model change, and (iv) the detailed analysis of what constitutes a high-cliqueness friend are presented for the first time here.

2.1 Previous Work on Flickr

As Flickr has a very rich and interesting set of data and as this data is accessible via a public API, it has been used for several studies. Among other things investigated, people have looked at general graph properties [Mislove et al. 2007], and they have extensively used the *tags*. The usage pattern of tags was investigated in Marlow et al. [2006], place and event event names were automatically derived in Rattenbury et al. [2007], hierarchies of interesting locations were discovered in Crandall et al. [2009], and the Flickr tags were used to design and evaluate tag recommendation systems [Garg and Weber 2008; Sigurbjornsson and van Zwol 2008]. Social aspects for the Flickr data have also been studied previously. The growth of the Flickr social graph was studied in Mislove et al. [2008] where the authors observed that a model such as preferential attachment is not a good fit and something more "local" is going on, where users tend to link to users already close to them in the graph. The propagation of information in the Flickr social graph was recently studied in Cha et al. [2009]. Here the term "information" essentially refers to the image popularity. It was found that "(a) even popular photos do *not* spread widely through the network, (b) even popular photos spread *slowly* through the network, and (c) information exchanged between friends is likely to account for over 50% of all favorite-markings, but with a significant delay at each hop". Gender and home locations were deduced from tags in Popescu and Grefenstette [2010] and that line of work could be used to supplement our set of per-user features.

2.2 Previous Work on Social Network Analysis

The study of user behavior in large, online social networks in general is also by no means new. Most relevant to our study of brand congruence is the work in Singla and Richardson [2008], as they also consider the degree of similarity between two users who are connected. In their setting, the network studied is the MSN network and the links they used correspond to chat sessions between users. Similarity between two users is measured with respect to web queries made and with respect to personal information such as age or location. They do not consider any product-related information, nor do they consider any graph related properties, such as cliqueness. The same data set was also used in Leskovec and Horvitz [2008], where the focus is on the actual instant messaging behavior. That is, they investigate which factors influence the number and length of conversations. They also consider general graph properties and verify the "six degrees of separation". In Mislove et al. [2007], the focus is on properties related to link distributions in several large-scale online social networks. In particular, they investigate inlink and outlink distributions for Flickr. The work in Ahn et al. [2007] is similar to this, but uses different data sets also includes the factor of evolution over time. More related to our study is the work presented in Backstrom et al. [2006]. Here, the authors look at factors that govern the growth of communities. Their analysis uses LiveJournal (with its communities) and DBLP (with scientific conferences). In their setting, the single most influential feature turned out to be the "proportion of friends in a community who are friends with each other", which is similar to our notion of "cliqueness". Although we are unsure if their findings will apply to setting where (i) a user can only join a single community and (ii) there is a financial cost associated with joining a community, we plan to use a framework related to theirs in the future to investigate the evolution over time more closely.

2.3 Previous Work on Brand Congruence and Marketing

There is also work related to brand congruence in social networks and to "viral marketing". First, Flickr itself offers rudimentary, aggregate statistics about the cameras used by its users.[3] Then, there has been previous work on mining social networks for targeted advertising. Yang et al. [2006] look at a social network derived from email messages between 427 faculty members of a university. The products they study are books, where the data refers to book loans from the library. The "brands" of interest are topical categories. The authors then find that highly cliqued[4] groups of users, supposedly loosely corresponding to faculties, are more likely to borrow books on the same category. Though similar in spirit, our study differs in size, an actual use of product-related information, the techniques used and the thorough analysis of various factors influencing the strength of links.

Related to but different from our work are studies looking at recommendation networks such Epinions.[5] Here users explicitly rate certain products, they often have an explicit network of trusted users (concerning product reviews) and they often have the explicit possibility to recommend a product to other users. Richardson and Domingos [2002] explicitly raises the question how to identify the "best" users to target for viral marketing. To answer this question, given certain game theoretical modeling assumptions, both a hardness result and an approximation algorithms are presented in Kleinberg [2007]. Leskovec et al. [2006] discussed the algorithms to enumerate the recommendation patterns, and our analysis of the size distribution of connected components is similar to their approach of finding typical patterns of influence. Leskovec et al. [2007] presents simple stochastic models to explain the patterns and size of recommendation cascades. These works are also more concerned with modeling the evolution over time, assuming that the strength of influence between pairs of nodes is known. So, in a sense, our work is somewhat orthogonal to this group of work, as we are mostly interested in understanding which factors govern the strength of links.

An interesting, very detailed study on brand congruence in real-life social networks was investigated in Reingen et al. [1984]. Here, the "nodes" of the social network were 49 members of a sorority at a university. Link types ranged from "roommate", via "sharing a bathroom" to "joint sports activities". Product types ranged from shampoo, via TV shows to pizza. Although the dataset was very small and sparse, the authors could find significant effects for, for example, room neighbors sharing pizza preferences or people sharing a bathroom using the same brand of shampoo. Though this kind of detailed study via surveys certainly does help to identify certain effects, it has clear scalability problems when it comes to the analysis of global networks with hundreds of thousands of users.

[3] `http://www.flickr.com/cameras/brands/`
[4] Our definition of "cliqueness" is identical to their definition of "cohesion".
[5] `http://www.epinions.com`

Table I. Some Basic Numbers Describing Our Dataset

Number of users	
- before pruning	3.9M
- after pruning	2.1M
Number of edges	
- before pruning	67M
- after pruning	44M

Number of ...	
brands	96
models	1785
countries	168
groups	203K
users w. country	510K
users w. group	850K

More brands and countries were present, but they were too insignificant to be picked up by our hand-crafted mappings to IDs.

There has also been work on the study of explicit brand communities, that is, clubs centered around a certain product brand and often directly sponsored by the corresponding corporation. One such study, focusing on car clubs, is presented in Algersheimer et al. [2005]. The authors show that, for example, the most active group members are those who feel most positively toward the brand. However, it is also argued that simply trying to market the club membership, in an effort to then boost sales for the new members, will generally not work as a positive attitude has to precede an active role in the club. Finally, there is also the aspect of brand loyalty, both in the sense of purchase and attitudinal loyalty. The relevance of psychological factors such as brand trust and brand affect was investigated in Chaudhuri and Holbrook [2001]. We also look at brand loyalty, but our focus is more on user types. In particular, we looked at whether “expert” users are more or less loyal to their brands than other users.

3. DATASET

All of our data was obtained using the public Flickr API.[6] Basic statistics about the data are given in Table I. We crawled a complete connected component of the Flickr friendship graph, starting with a very active user with over 100 contacts.[7] As part of this crawl, we downloaded the complete list of public photos for each user. In this initial phase, we obtained a graph containing 3.9M users and 67M friendship edges, as well as a list of 500M images with their “date taken” information. Note that, while it is possible to get the list of (public) pictures of a user, along with the corresponding “date taken” information, via a single API call, other information has to be obtained on a per-image basis.

We now pruned this initial set of users, as we were only interested in users for which we could extract brand information, which in turn was derived from information for uploaded images. Therefore, we removed all users from the data set who had not uploaded any public pictures. This left us with 2.1M users and 44M friendship links. Note that friendship edges in Flickr are not necessarily symmetric, as adding a friend does not require authorization. This means that user A can have B as a friend, without user B having A as a friend. Of the 44M links, 30M belong to pairs of bidirectional links and the remaining 14M links are not reciprocated by the other person.

After we had obtained the complete list of users, we then went on to obtain additional information for each user of interest. In particular, where provided, we obtained the country of the user, as provided in her user profile. Here, quite a bit of care had to be taken as the location could be entered in a free text format so that all of “California”, “San Fransisco”, “USA”, “America” and even “Canada’s neighbor” needed to be mapped to the same, unique country ID. Out of the 2.1M users investigated, 581K specified some country and 519K of these cases were mapped to a valid country

[6] http://www.flickr.com/services/api/
[7] http://www.flickr.com/people/acastellano/

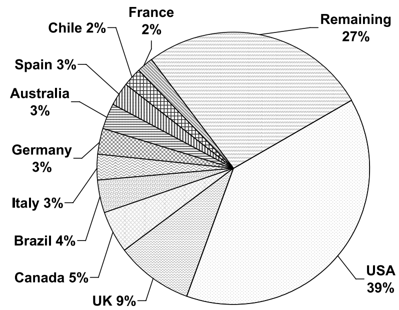

Fig. 1. User distribution across countries for the 24% (out of 2.1M) of users who specified valid country information.

ID. Unmapped cases include, for example, “land of Putin” or “I am Italian”. Figure 1 shows the distributions of valid countries found.

Furthermore, for each user, we obtained the groups joined on Flickr. Groups on Flickr represent communities of users sharing common interest, usually related to photography. All of this data is (i) static in the sense that it does not relate to a particular period of time and (ii) not directly connected to the brand or model information, which is our focus of interest.

The key observation that allowed us to obtain brand and model information, is that when images are uploaded to Flickr, the so-called Exif[8] metadata is usually preserved. This metadata contains information about the manufacturer, the model, as well as other information related to resolution or exposure time, which was not used by us. It is also available on a picture’s main page on Flickr under the header “Additional information”. Given this Exif data, we then assigned brands and models to a user for a particular period of time as follows. First, we chose three time slots to investigate, namely, the period of March 1 to May 31 for the years 2006, 2007, and 2008. This period was chosen as (i) it was close to the date the data was obtained (June 2008) and (ii) it lies closely after, but does not overlap, with the Christmas season for which we expect a change in the model distribution. Then, for each of these three time slots, we tried to obtain the Exif data for up to 10 public images for each of the 2.1M users of interest. The reason that we used only 10 images per user is that for each Exif data a separate call to the API had to be issued and one such call took, roughly, 1 second. Obtaining this data for all of the public 500M images discovered during the crawl, would not have been feasible in an acceptable time frame.

If a user had uploaded more than 10 images for a time period of interest, we first tried to obtain the relevant Exif information for the image closest to the center of the slot (April 15) and then worked our way toward the ends of the interval in a symmetric fashion. In the vast majority of the cases (94%), all the metadata obtained (if there was any) for a user in a time slot was consistent in the sense that only a single brand occurred. For the cases where this was not the case, we used a simple majority voting scheme and assigned the strongest brand within a time slot to each user, where ties were broken at random. In the same manner, users were assigned a model for each time slot, if at least one of the (up to) 10 images contained a valid model information. Mapping brands to unique IDs again required some manual labor. For example, we took care to map “Minolta,” “Konika,” and “Konica” to the same unique brand ID. As for mapping camera models to IDs, we used the list of 1,785 cameras available at `http://www.flickr.com/cameras/`. This list also contains information about the

[8] `http://en.wikipedia.org/wiki/Exchangeable_image_file_format`

Table II. An "Active" User is One Who Uploaded a Public Picture with Brand Information in the Respective Time Period

	Mar-May06	*Mar-May07*	*Mar-May08*
Active users	470K (350K)	670K (520K)	630K (500K)
... w. country	160K (118K)	210K (164K)	200K (159K)

A large fraction of users did not upload pictures in the time periods and were hence "inactive". Numbers in parentheses refer to the number of users with valid model information. The reason that the numbers for the slot in 2008 is *lower* than for 2007 is probably that we obtained the data in mid-June, when not all users had uploaded their pictures for the corresponding period yet. 152K users were active in all three time periods (68K with country information) and 1.2M were active during at least one time period (326K with country information).

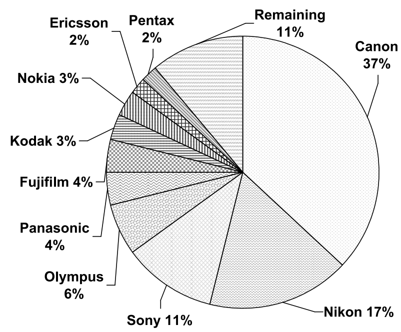

Fig. 2. Brand distribution in 2008 for the 630K users who uploaded a picture taken between March and May with a valid brand in the Exif data.

camera's type, in particular whether it is a point-and-shoot (P&S) or a digital single-lense reflex (DSLR) camera. We use this type classification as given by Flickr and the names are not meant to imply something about the actual camera use. We again tried our best to ensure that different camera names (such as "Maxxum 7D" and "Dynax 7D") referring to the same model were mapped to the same ID. We ignored cameras not included in the online list, which appeared to be comparatively old models. Table II shows the number of users for the three time slots considered and Figure 2 shows the distribution among the various brands.

It should be noted that, although we have time-related information about the users' camera brands, the underlying friendship network, the geographical information and the groups information is only for June 2008, the time of the crawl. Hence, the fact that two users are friends in our dataset does not necessarily mean they were friends in 2006. However, we do assume that the friendship network in 2008 is still a good approximation of the network in 2006, and so we also include results for this period. But our focus is always on the most current time slot. Also note that although 2.1M users in our crawl have at least one public image, only 1.2M of them uploaded a picture with valid brand information during our time periods of interest.

Although not directly related to brand congruence or the propagation of new models through the friendship graph, we also tried to investigate whether there is a "digital divide" with respect to the types of cameras used. Table III shows that the percentage of users from the poorest countries in the world that use a high-end DSLR camera is similar to the percentage in the richest countries. The most likely explanation for this is a pronounced bias in terms of access to the internet and to Flickr in the first

Table III. Differentiaton of Users with Known Country According to Whether They Used a P&S or a DSLR Camera in 2008

Country Type	*Total Users*	*P&S*	*DSLR*
Low Income	486	60%	40%
Lower Middle Income	5614	61%	39%
Upper Middle Income	19K	69%	31%
High Income	120K	56%	44%

Other camera types were not included and so the percentages are relative to each other. Countries were categorized into four classes according to the world bank's classification. Representatives of the classes are, in increasing order of income, Bangladesh, India, Russia and Switzerland.

place. Whereas in a rich country Flickr users are more likely to be at least somewhat representative of the population, this is unlikely to hold even approximately in poor countries.

4. RESULTS

Here, we present the results of our analysis. First, in Section 4.1, we introduce the basic technique of analysis which was used for most of our study. Sections 4.2 and 4.3 then present our main findings, ranging from rather basic brand congruence analysis to a more advanced analysis of how camera changes of one user affect her friends.

4.1 Techniques

Our basic technique of data analysis for brand congruence, also used in Singla and Richardson [2008], is the following: we consider pairs of users of a certain type and measure the percentage of them sharing the same camera brand, an event we refer to as brand congruence, or simply as being congruent. For example, we look at pairs of random users and compare their probability of brand congruence (as a baseline) to the probability for pairs of friends. We then consider more and more relevant conditions for possible pairs of users, such as whether they are in the same country, whether they share many common friends or whether they both use high-end cameras.

We evaluated all of these numbers for the full list of 44M friendship links. However, due to the obvious problems of scalability, we did not compute these numbers for all 4.4 trillion possible pairs of random users. Here, we once sampled uniformly at random a set of 44M random pairs, irrespective of any existing friendship links, and then computed all the relevant properties for this collection of pairs. No "self-links" were allowed for this. The 44M pairs were then further conditioned to, for example, only make statements about pairs of random users in the same country. Both for pairs of friends and for the pairs of random users, we always include the absolute numbers of pairs which have a certain property, along with the percentage of these which are congruent. Note that for certain statistics this sampling was not required. For example, if there are only two brands, brand A and B, and 60% of users use brand A and 40% brand B, then one can directly compute that the probability of a random pair of users being congruent is 36% + 16% = 52%. Here, 36% is the probability that two users are congruent on A and 16% is the probability that they are congruent on B. Wherever such a closed-form solution was possible, the results agreed for at least three significant digits.

To study the propagation of models through the friendship graph, we mostly used the following two techniques. First, a method similar to one used in Leskovec et al. [2006]. Here, we describe the changes of the network topology as nodes convert to/from

Table IV. Probability that a Pair of Two Users Will Share the Same Camera Brand, Depending on Whether They Are *Random* Users or *Friends*

	Mar-May06	*Mar-May07*	*Mar-May08*
random users	0.16 (2.0M)	0.17 (4.2M)	0.19 (3.7M)
friends	0.22 (5.9M)	0.25 (11M)	0.27 (11M)

Here and throughout the article, the number in parentheses give the absolute numbers of pairs used.

a new model in terms of the distribution of connected components of certain shapes. These distributions are then compared to distributions obtained by random propagation of new models. Second, we use more local techniques, similar to those used for brand congruence, and try to understand which factors influence the probability that a particular node will convert to a new model between two time steps.

4.2 Basic Brand Analysis

Differences between Pairs of Friends and Random Users. As a baseline experiment, to compare effects of various factors against, we measured the probability of brand congruence between any two random users in a given time slot. Table IV gives the results for the three time slots. To see a first relevance of the friendship factor when it comes to brand congruence, we measured the probability of congruence on our Flickr friendship network in a given time slot. Table IV gives the results for the three time slots as well.

Looking at numbers of Table IV, we see an increase of 80% in the probability of sharing the same brand in the friendship graph as compared to random pairs. This result is observed for all the timeslots. However, it could well be that two friends are simply more likely to be in the same geographical area and that the higher degree of congruence can be solely explained by geographical changes of the predominant brand. Our next experiment will show that this is not the underlying reason for the observed effect.

Random Pairs with Country Conditions. To see if the underlying explanatory factor is the common country of two friends, we carried out experiments, where we split pairs of random users into two classes, depending on whether they come from the same country or not. Only users who had provided valid information concerning their country were used for this experiment.

Table V gives the probabilities that we get by conditioning on random user pairs in the two ways previously mentioned. Note that this is not a proper breakdown of the results in Table IV, as the majority of users do not have any country information. These users are still included in the previous results, but are not used for the country-related analysis. The fact that conditioning on the same country leads to higher congruence probabilities (see Table IV for comparison) means that the higher congruence for friendship links can at least partly be explained by regional congruence. However, (i) the congruence probabilities are still *lower* than for a randomly selected pair of friends and (ii) the congruence among friends increases further, when we also condition on both users having the same country (and a valid country to begin with), as shown in Table V.

Friendship Graph with Country Conditions. As for two random users, we also split the results for two friends, according to whether they come from the same or different countries. Note that even two friends from different countries are still more likely to be congruent than two random users from the same country. This is a first indication

Table V. Probability of Congruence for Random Pairs and Friends, Depending on Whether They Are in the Same Country (SC) or in Different Countries (DC)

	Mar-May06	*Mar-May07*	*Mar-May08*
random users			
- same country	0.19 (64K)	0.22 (70K)	0.23 (44K)
- diff. country	0.16 (187K)	0.18 (336K)	0.19 (301K)
friends			
- same country	0.24 (908K)	0.27 (1.4M)	0.29 (1.5M)
- diff. country	0.21 (1.3M)	0.24 (2.1M)	0.28 (2.1M)

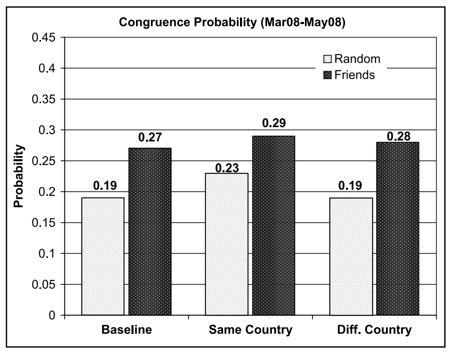

Fig. 3. Histogram summarizing the results of our basic analysis for the time period March-May 2008. Note that two friends who provided valid country information, even if they come from different countries, are still more likely to be congruent than two arbitrary friends.

of the relevance of friendship which will be investigated further in the next section. Figure 3 gives a summary of the results so far.

4.3 Advanced Brand Analysis

In the previous section, we saw that friendship information is clearly relevant for brand congruence. In this section, we dig deeper and try to find out "which kind of friendships" matter, "which other factors play a crucial role" and "how things evolve with time".

Varying Degree of Groups Similarity. One attempt to measure the closeness of two friends, is to find out if they share common interests. Though lists of interests are not available on Flickr, the groups joined provide some indication to the interests held by a particular user. Therefore, one could hypothesize that two friends who joined many of the same groups are "closer" in some sense and should be more congruent. To test this hypothesis, we classified the links between two friends according to the Jacquard coefficient of the groups joined. This coefficient is defined as $G_J(X, Y) = |G(X) \cap G(Y)|/|G(X) \cup G(Y)|$, where $G(X)$ denotes the group set joined by a user X. Only users who joined at least one group were used for this experiment.

Table VI does not show any correlation between the degree of group similarity and the probability of brand congruence. Interestingly, there appears to be a small drop for friends who "only share very few, but more than no groups". This is interesting because, a priori, one could expect a monotonicity: the more shared groups, the closer the friendship; the closer the friendship, the higher the brand congruence.

Table VI. The Degree of Overlap between the Groups Joined by Pairs of Friends Does Not Affect the Probability That They Use Cameras of the Same Brand, Although There Does Appear to be a Small Drop for Friends Having Some But Only Few Groups in Common

$G_J(X,Y)$	*Mar-May06*	*Mar-May07*	*Mar-May08*
baseline	0.22 (5.9M)	0.25 (11M)	0.27 (11M)
$0 \leq G_J < 0.2$	0.22 (5.3M)	0.25 (9.3M)	0.27 (10M)
$0.2 \leq G_J < 0.4$	0.20 (36K)	0.22 (86K)	0.24 (129K)
$0.4 \leq G_J < 0.6$	0.24 (4.6K)	0.26 (8.8K)	0.26 (9.7)
$0.6 \leq G_J < 0.8$	0.26 (559)	0.24 (1.0K)	0.27 (1.2K)
$0.8 \leq G_J \leq 1.0$	0.26 (3.4K)	0.24 (6.0K)	0.27 (6.8K)

Table VII. Probability of Brand Congruence for Two Friends, When the Friendship Type Is Broken Down into Mutual and One-Way Friendship

	Mar-May06	*Mar-May07*	*Mar-May08*
baseline	0.22 (5.9M)	0.25 (11M)	0.27 (11M)
Mutual	0.22 (3.9M)	0.24 (7.1M)	0.27 (8.0M)
One-way	0.22 (2.0M)	0.25 (3.4M)	0.28 (3.4M)

However, it seems that photography topical interests are not an indicator for closeness of friendship.

Mutual vs. One-Way Links. As opposed to most other social networks, such as Facebook or Myspace, Flickr allows one-way friendship links. Such links can serve the purpose of a "bookmark" for the other person's profile or they could be a sign of admiration. The "admired" person is informed about the created link and can then decide to reciprocate it or not. One could expect that reciprocated links are stronger indications for actual friendship and tend to lead to higher brand congruency than one-way links. Or one could hypothesize that one-way links, if they are indications of admiration, would lead to higher brand congruency. To test if there is any difference, we split the friendship links into two classes, mutual links and one-way links. Table VII shows that there are no big differences between the two types of friendship in terms of the degree of congruence.

Many Friends vs. Few Friends. Another dimension to explore is the size of a person's list of friends. One could consider a friendship link from a person with few friends to be more "meaningful" than from a user with dozens of friends. To test if there are such difference between these two categories, we split users according to whether they have more than five friends (this is a "large" user in our terminology) or whether they have less than five friends (this is a "small" user). The threshold of 5 was chosen such that roughly one quarter of the users (28.8%) fell into the group of large "power users". Note that even though they account for less than half of the users, 88.5% of all friendship links are between "large" users. The distinction between "small" and "large" users was then combined with a breakdown into same or different countries.

Table VIII nicely shows that the congruence between pairs of friends, who are selective when it comes to adding friends on Flickr, is far stronger than between pairs of friends who have more than five friends. Furthermore, for pairs of "small" users the congruence is dramatically influenced by the fact whether the two friends are in the same country or not. Table IX shows the comparison results for random users

Table VIII. Brand Congruence Probabilities for Pairs of Friends when Conditioned (i) on the Size of the Friendship Lists of Both Friends and (ii) on a Common or Different Country of the Users

Type of friends pair	*Mar-May06*	*Mar-May07*	*Mar-May08*
baseline	0.22 (5.9M)	0.25 (11M)	0.27 (11M)
small-small			
- all	0.29 (43K)	0.28 (83K)	0.30 (83K)
- same country	0.28 (2.7K)	0.28 (4.1K)	0.31 (3.6K)
- diff. country	0.23 (714)	0.22 (992)	0.21 (745)
small-large			
- all	0.24 (143K)	0.25 (266K)	0.28 (258K)
- same country	0.25 (17K)	0.27 (27K)	0.28 (26K)
- diff. country	0.20 (8.2K)	0.22 (13K)	0.25 (11K)
large-small			
- all	0.22 (272K)	0.24 (480K)	0.26 (453K)
- same country	0.24 (28K)	0.25 (41K)	0.27 (38K)
- diff. country	0.20 (20K)	0.21 (30K)	0.24 (25K)
large-large			
- all	0.22 (5.4K)	0.24 (9.7M)	0.27 (11M)
- same country	0.23 (860K)	0.27 (1.4M)	0.29 (1.4M)
- diff. country	0.21 (1.3M)	0.25 (2.1M)	0.28 (2.1M)

A "small" user is one with up to five friends and a large user has more than five friends. Note that for a pair of small-small friends the congruence probability is greatly increased when they are in the same country (31% for 2008) compared to when they are not (21% for 2008).

of the same/different country, broken down the results into the sizes of each users. This shows that "small" users do not inherently have a strong tendency to share the same brand, even if they are in the same country, but that this effect is related to the friendship.

Varying Degree of "Cliqueness". Apart from "selectivity" one might also expect that the degree of "cliqueness" between two friends will have an effect. If two friends have many friends in common, they can be said to be in a social clique. Formally, we define the *cliqueness* F_J between between two users X and Y to be the Jacquard coefficient of their two sets of friends. That is, for a pair of friends (X, Y) we have $F_J(X, Y) = |\bar{F}(X) \cap \bar{F}(Y)|/|\bar{F}(X) \cup \bar{F}(Y)|$. Here, $\bar{F}(X) = F(X) \cup \{X\}$, which is the set of X's friends together with X itself. The cliqueness between two users is 1.0 if and only if they share all their friends (which includes the case where they do not have any friends apart from each other[9]), and it can never be zero.

Figure 4 shows how an increase in cliqueness (on the X-axis) leads to a higher brand congruence. Table X gives a breakdown of users according to low (≤ 0.5) and high (> 0.5) cliqueness in combination with conditioning on a common or different country. Here, one can see, by looking at the absolute numbers of links used in each cell, that high cliqueness friends have a far higher probability of being in the same country (91%) than low cliqueness friends (42%), given that both friends provided country information.

[9]But this case is not included in our dataset, as we would have never discovered this isolated component in our crawl.

Table IX. Brand Congruence Probabilities for Pairs of Random Users when Conditioned (i) on the Size of the Friendship Lists of Both Users (Who Will Most Likely Not be Friends) and (ii) on a Common or Different Country of the Users

Type of random pair	*Mar-May06*	*Mar-May07*	*Mar-May08*
baseline	0.16 (2.0M)	0.17 (4.2M)	0.19 (3.7M)
small-small			
- all	0.16 (657K)	0.16 (1.3M)	0.18 (1.1M)
- same country	0.17 (6.0K)	0.18 (9.2K)	0.21 (7.2K)
- diff. country	0.15 (25K)	0.15 (41K)	0.17 (31K)
small-large			
- all	0.16 (498K)	0.17 (1.0M)	0.19 (922K)
- same country	0.18 (10K)	0.21 (16K)	0.23 (14K)
- diff. country	0.16 (43K)	0.17 (76K)	0.19 (66K)
large-small			
- all	0.16 (497K)	0.17 (1.0M)	0.19 (921K)
- same country	0.19 (10K)	0.21 (16K)	0.22 (14K)
- diff. country	0.16 (43K)	0.17 (76K)	0.19 (65K)
large-large			
- all	0.17 (376K)	0.19 (766K)	0.21 (784K)
- same country	0.21 (17K)	0.24 (29K)	0.25 (28K)
- diff. country	0.17 (76K)	0.19 (141K)	0.21 (140K)

Two small, random users are the *least* likely to be congruent, whereas two small friends are the most likely to be congruent.

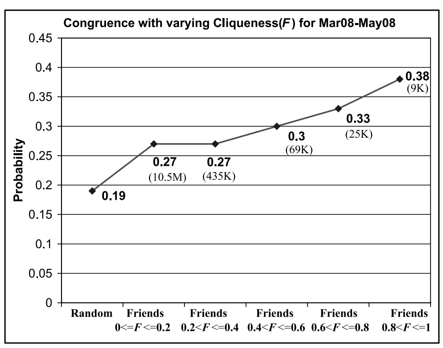

Fig. 4. Congruence probability for pairs of friends with varying degree of cliqueness. Cliqueness turned out to be the strongest camera-type independent factor in our study.

More interestingly, as already observed for the breakdown into small and large friends, the difference between high and low cliqueness friends is getting smaller from 2006 to 2008. As we could only observe the present status of a friendship relation, it would be interesting to look at this more closely by taking the age of friendship links into account.

When the size of the friendship lists is combined with the degree of cliqueness, an interesting reversal takes places. Whereas, globally, pairs of small users are more likely to be congruent than pairs of large users, Table XI show this order is reversed, when we only look at friends with a high degree of cliqueness. This can be understood by observing that sharing more than 50% of friends, when both people involved have

Table X. Breakdown of the Congruence Probabilities between for Friends According to a Common/Different Country and the Degree of Cliqueness for the Two Users

$F_J(X,Y)$	*Mar-May06*	*Mar-May07*	*Mar-May08*
baseline	0.22 (5.9M)	0.25 (11M)	0.27 (11M)
$0 \leq F_J \leq 0.5$			
- all	0.22 (5.9M)	0.24 (10M)	0.27 (11M)
- same country	0.23 (906K)	0.27 (1.4M)	0.29 (1.5M)
- diff. country	0.21 (1.3M)	0.24 (2.1M)	0.28 (2.1M)
$0.5 < F_J \leq 1.0$			
- all	0.32 (27K)	0.32 (51K)	0.33 (53K)
- same country	0.30 (2.0K)	0.34 (2.8K)	0.33 (2.5K)
- diff. country	0.28 (317)	0.22 (383)	0.24 (346)

Table XI. A More Detailed Breakdown of the Congruence Probabilities for 2008 According to (i) the Degree of Cliqueness F_J, (ii) whether Two Friends Share a Common Country, and According to (iii) the Size of the Friendship Lists of the Two Friends

	User type (more or less than five friends)		
$F_J(X,Y)$	all	small-small	large-large
baseline	0.22 (5.9M)	0.25 (11M)	0.27 (11M)
$0 \leq F_J \leq 0.5$			
- all	0.27 (11M)	0.29 (58K)	0.27 (11M)
- same country	0.29 (1.5M)	0.30 (2.7K)	0.29 (1.4M)
- diff. country	0.28 (2.1M)	0.20 (569)	0.28 (2.1M)
$0.5 < F_J \leq 1.0$			
- all	0.33 (53K)	0.32 (26K)	0.35 (22K)
- same country	0.33 (2.5K)	0.33 (945)	0.32 (1.2K)
- diff. country	0.24 (346)	0.24 (176)	0.27 (32)

Note that neither the counts for the same and different countries have to add up to "all" (as only users with country information are used for the first), nor do the numbers for small-small and large-large have to add up to "all" (as there are also other edge types).

long lists of friends, is arguably a stronger sign of a clique than if both sides only have a single friend, but this friend is in common. In fact, if we require the cliqueness to be larger than 0.8 (and not just 0.5), then pairs of large friends have a probability of 48% to share a common brand in 2008 (out of 2,268 pairs). For 2007, this percentage goes up to 51% (out of 2,740 pairs) and it is even 57% (out of 1,365 pairs) in 2006.

Two Important User Types: P&S vs. DSLR. Apart from the social aspects studied above, we also looked at a break-down of "low-end vs. high-end" users. Concretely, we focused on those users who either used a point-and-shoot camera (P&S), which is usually a cheaper (by comparison) camera for non-expert users and on those with a digital single-lens reflex camera (DSLR), which is usually a more expensive camera for more advanced users. The results for friends are presented in Table XII and in Table XIII for random users. Note that for this analysis other camera types, such as camera phones, were not used.

Table XII. Breakdown of Brand Congruence Probabilities for Pairs of Friends According to the Camera Quality Used (Point-and-Shoot vs. Digital-Single-Lens-Reflex) and Further Conditioning on the Country

Type of pair	*Mar-May06*	*Mar-May07*	*Mar-May08*
P&S - P&S			
- all	0.23 (1.5M)	0.25 (2.1M)	0.26 (1.7M)
- same country	0.26 (232K)	0.28 (294K)	0.29 (226K)
- diff. country	0.18 (280K)	0.19 (351K)	0.21 (240K)
P&S - DSLR			
- all	0.18 (663K)	0.19 (1.3M)	0.20 (1.3M)
- same country	0.22 (109K)	0.23 (183K)	0.24 (180K)
- diff. country	0.16 (157K)	0.17 (28K)	0.18 (266K)
DSLR - P&S			
- all	0.18 (580K)	0.19 (1.1M)	0.20 (1.1M)
- same country	0.23 (96K)	0.24 (165K)	0.25 (168K)
- diff. country	0.15 (131K)	0.16 (234K)	0.18 (225K)
DSLR - DSLR			
- all	0.49 (447K)	0.48 (1.3M)	0.47 (2.1M)
- same country	0.52 (87K)	0.50 (231K)	0.49 (338K)
- diff. country	0.47 (110K)	0.45 (316K)	0.46 (470K)

Table XIII. Breakdown of Brand Congruence Probabilities for Pairs of Random Users According to the Camera Quality Used (Point-and-Shoot vs. Digital-Single-Lens-Reflex) and Further Conditioning on the Country

Type of pair	*Mar-May06*	*Mar-May07*	*Mar-May08*
P&S - P&S			
- all	0.19 (652K)	0.19 (1.2M)	0.20 (799K)
- same country	0.25 (14K)	0.26 (19K)	0.28 (13K)
- diff. country	0.18 (55K)	0.18 (82K)	0.18 (52K)
P&S - DSLR			
- all	0.17 (167K)	0.17 (434K)	0.19 (454K)
- same country	0.25 (5K)	0.26 (10K)	0.27 (10K)
- diff. country	0.15 (16K)	0.18 (82K)	0.17 (38K)
DSLR - P&S			
- all	0.17 (167K)	0.17 (437K)	0.19 (456K)
- same country	0.24 (5K)	0.26 (11K)	0.27 (10K)
- diff. country	0.15 (17K)	0.16 (37K)	0.17 (38K)
DSLR - DSLR			
- all	0.44 (43K)	0.42 (158K)	0.42 (259K)
- same country	0.45 (1.8K)	0.45 (6K)	0.45 (8K)
- diff. country	0.43 (4.9K)	0.41 (16K)	0.41 (27K)

There are some interesting observations to make. First, for pairs of P&S users the effect of being in the same country is striking. For two friends in this group, the difference for 2008 is between 29% (for the same country) and 21% (for different countries).

Table XIV. Breakdown of Brand Congruence Probabilities for Pairs of friends for the Period March-May 2008 According to the Camera Quality Used and the Degree of Cliqueness ($F_J \leq 0.5$ or $F_J > 0.5$)

	Cliqueness		
Type of pair	ignored	low	high
all	0.27 (11M)	0.27 (11M)	0.33 (53K)
P&S - P&S	0.26 (1.7M)	0.26 (1.7M)	0.39 (15K)
P&S - DSLR	0.20 (1.3M)	0.20 (1.3M)	0.23 (4.1K)
DSLR - P&S	0.20 (1.1M)	0.20 (1.1M)	0.23 (4.1K)
DSLR - DSLR	0.47 (2.1M)	0.47 (2.1M)	0.66 (5.8K)

For two random users the corresponding numbers are 28% vs. 18%. This even goes so far that the friendship information seems nearly irrelevant in this setting and it shows that the local dominance of “cheap” camera brands differs around the globe. It also seems to imply that users of such cheap P&S cameras are not strongly affected by the influence of their friends. However, this later claim will be refuted, once we additionally take cliqueness into account (Table XIV). Second, pairs of DSLR users are far more likely to use cameras of the same brand than other users. Third, for such pairs of “expert” users, the country influence is much weaker and the global differences seem to be washed away for high-end models. These users also seem to show a higher influence to peer pressure, as the degree of congruence, though already at a high level for random users, increases further when conditioned on friendship (42% vs. 47% for 2008, 45% vs. 49% for 2008 and a common country).

The effect of friendship becomes much more pronounced, when we look at the results for high-cliqueness friends in Table XIV. Now a pair of cliqued P&S friends has a probability of 39% of being congruent, compared to 28% for two random P&S users in the same country. Similarly, the congruence goes up to 66% for pairs of highly cliqued DSLR users, compared to 45% for two such random users in the same country and compared to 49% for two such friends (regardless of cliqueness) in the same country. The same table also shows, in agreement with the results in Tables XII and XIII, that for pairs of different user types, any friendship connection is essentially irrelevant. We were expecting the “inexperienced” P&S users to be influenced by their “expert” DSLR user friends, which does not seem to be the case.

Evolution Over Time. So far, our analysis has mostly focused on a single time slot at a time. In this and the next paragraphs, we will focus on the evolution of brand congruence. The histogram in Figure 5 shows how pairs of users have become more and more likely to share a brand. For pairs of friends this effect is at least partly due to the fact that the current friendship graph of 2008 is only an approximation of the friendship network for 2007 or 2006. This means that pairs of “friends” could actually be random pairs when the link was only recently established. Hence one would expect a lower degree of congruence among friends for 2007 and 2006. Interestingly, the same trend towards a higher brand congruence in 2008 exists even for pairs of random users. This is due to the fact that the variance in terms of brand distribution has gone down. More concretely, the two strongest brands (Canon and Sony) combined, increased their share of users from 48% in 2006, to 49% in 2007 and finally to 54% in 2008 and the entropy of the global brand distribution has decreased from 3.1 bits in 2006 to 2.9 bits in 2008.

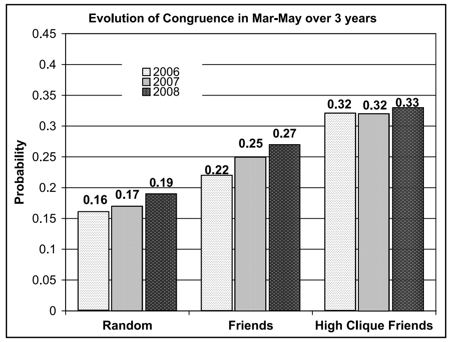

Fig. 5. Changes in congruence probabilities over time. Recall that all friendship links are with respect to the network in 2008.

Table XV. Users of DSLR Cameras Are, Compared to Users of Point-and-Shoot Cameras, (i) Less Likely to Change Their Camera Model over the Course of a Year and (ii) Less Likely to Change Their Brand, Even if They Do Change Their Camera Model

	Camera type in 2007	
	P & S	DSLR
Number of users considered	142K	66K
Changes in brand	34%	15%
Changes in model	52%	36%
Brand change, given model change	67%	40%

The last statement remains true when the fact that there is generally less variability among DSLR users is taken into account (Tables XII and XIII). Unrelated to the brand loyalty aspect, it is also noteworthy that more than half of the P&S users changed their model between 2007 and 2008.

Loyalty for Low-Budget vs. High-End Users. Intuitively, one would expect photographers to "evolve" in terms of brand loyalty. That is, in their early days as photographers they would be using comparably cheap point-and-shoot cameras and they would not feel particularly attached to their current brand. For these users, the buying decision was probably more influenced by offers at major electronic discount stores, than it was by consciously considering all possible options in terms of their quality. Eventually, they might decide to take photography more seriously and purchase a single lens reflex camera. At this point, one would expect their decision to be more conscious and the user feeling more attached to the particular model she purchased. This would then lead to a higher brand loyalty when it comes to future purchases. Furthermore, for high end cameras accessories such as lenses might be reusable between different models of the same brand.

This is indeed what we observed. Namely, users of point-and-shoot cameras are far more likely to switch to a different brand when they buy a new camera, then users of single-lens-reflex cameras. See Table XV for details. However, this has to be put in relation to the observation that DSLR users generally have less variability in terms of their brands used and tend to stick to fewer brands. Still, comparing the numbers of Table XV with Table XIII one observes that a random DSLR user in the same country has a probability of “only” 45% of sharing my brand in 2008, but if I change my model from 2007 to 2008, and if I have a DSLR camera in 2007, then I have a probability of 64% of ending up with the same brand again. The corresponding numbers are 29% and 48% for P&S users. Interestingly, the probability of staying with the same brand is in both cases lower than the probability of sharing the same brand with a high cliqueness friend of the same user type (DSLR or P&S). This could be seen as an indication that users are more influenced by “viral marketing” through their close friends than by “brand loyalty” resulting from their own past experience. It could also be an indication that the event of a user changing her model is already a small indication of dissatisfaction, which might encourage a brand change.

Triggering of Events by Friends. We also tried to investigate, whether the fact that a user changes her brand or at least acquires a new camera model, has some measurable influence on the probability that her friends do likewise, especially the high-cliqueness friends, with a cliqueness higher than 0.5. Again, such effects are indeed observable.

Given a user changes her camera model between 2007 and 2008, on average 54% of her high cliqueness friends and 51% of her low cliqueness friends also change their model. However, a random user only has a probability of 48% of doing so. Also, out of all the low-cliqueness friends who change their model, 29% change to a camera of the same brand as the user of attention, and this percentage increases to 38% for the cliqueness friends. This percentage is higher than the 33% probability of two high-cliqueness friends in 2008 to share a common brand, which is a further indication that they indeed changed together. On the other hand, random users who change their model during the same period, only have a probability of 20% of changing to the same brand in 2008. This percentage is then just one percent above the 19% probability for two random user to share a brand.

Similarly, given that a user changes her brand (and hence the model) in the same period, on average 38% of her low-cliqueness friends do the same (out of which 18% change to the same brand), 43% of the high-cliqueness friends (out of which 27% change to the same brand) and 38% of random users (out of which 13% change to the same brand). In all of these cases, taking a common country into account changed very little, generally adding 1-2% to all numbers.

4.4 Propagation of Models Through Graph

The analysis in the previous sections focused (i) on the static case and (ii) on brands. In this section, we take a different approach and try to describe how new models propagate through the friendship graph, that is, how users become “converted”. Our experiments focus (i) on the Nikon D80, a DSLR camera which became very popular between 2007 and 2008 (from 2.1% to 3.7% of users with known model), (ii) on the Canon SD1000, a P&S camera with a similar gain in popularity (from 0.2% to 1.5% of users with known model), and (iii) on the Canon EOS 20D, a DSLR model which sees its spread shrink from 2007 to 2008 (from 2% to 1.5% for users with known model). In the first part of this section, we describe how the graph and in particular the connected components of converted users change as a model spreads/shrinks in the network. In the second part, we analyze factors which influence the probability that a particular user will convert to the model under consideration.

Behavior of Converted, Connected Components. Following a similar approach to Leskovec et al. [2006], where the (graph) patterns of recommendation are examined, we looked

Table XVI. Distribution of the Weakly Connected Components of Nikon D80 Users

Weakly CC	Nikon D80 (DSLR), Growth					
	2007		2008		Random-2008	
(nodes,edges)	#CC	% users	#CC	% users	#CC	% users
(1, 0)	7424	66.0%	11079	59.2%	13704	73.2%
(2, 1)	292	5.2%	515	5.5%	404	4.3%
(3, 2)	36	1.0%	68	1.1%	51	0.8%
(3052, 7280)	1	27.1%	0	0%	0	0%
(3982, 7955)	0	0%	0	0%	1	21.3%
(6214, 17297)	0	0%	1	33.2%	0	0%

Weakly connected components accounting for less than 1% of the users in all three settings (2007, 2008, Random-2008) are not listed. The distributions of the real growth and the random growth from 2007 to 2008 for the Nikon D80 differ considerably.

at how the distribution (in terms of counts) of weakly[10] connected components changes from 2007 to 2008 as more nodes become "converted". Concretely we did the following.

For the DSLR models under consideration (Nikon D80 and Canon EOS 20D) we obtained a distribution of sizes of the weakly connected components of converted users in 2007. This distribution is shown in the first and second columns in Table XVI and Table XVII respectively for Nikon and Canon Models. Then, for 2008, we again computed this distribution, both for the actual graph (shown in third and fourth columns) and also for a random version, (shown in fifth and sixth columns) in the same tables. In the random version random nodes are converted irrespective of their neighbors. The number of random nodes converted was chosen in order to maintain (i) the number of users who convert *away* from the model under consideration between 2007 and 2008, and (ii) the number of users who convert to the model under consideration. Differences or similarities between the distribution in clique sizes are hence an indication how far or close to random growth/decline the actual growth/decline of the model was.

Table XVI shows that for the actual growth of the Nikon D80 the percentage of isolated nodes in 2008 out of all converted nodes is only 59%, compared to 73% for a random growth. Furthermore, the giant component encompasses 33% of nodes in the real graph, compared to only 21% in the random graph. This is a clear indication that the nodes were not converted independently of each other, but that conversions tended to be locally boosted.

Interestingly, Table XVII also shows that the decline of the model Canon EOS 20D appears to follow the random decline quite closely. Both in terms of (remaining) isolated nodes and in terms of the (remaining) size of the giant component no big differences can be observed. This is an indication that people convert to a target model in a somewhat co-ordinated manner, but that they convert from a target model in an essentially random fashion. This nicely coincides with the intuition: You tell your friends about the new camera you just bought, but you do not tell them about the old camera you just got rid of.

[10]Recall that friendship connections are directed in Flickr, so weakly connected components will usually differ from connected components. However, as shown in Section 4.3 ("Mutual vs. one-way links"), there does not appear to be a noticeable difference between the influence of unidirectional and bidirectional links. For this reason, we worked with the larger weakly connected component, effectively treating all edges as bidirectional.

Table XVII. Distribution of the Weakly Connected Components of Canon EOS 20 Users

Weakly CC	Canon EOS 20D (DSLR), Decline					
	2007		2008		Random-2008	
(nodes,edges)	#CC	% users	#CC	% users	#CC	% users
(1, 0)	6909	64.8%	5123	69.6%	5183	70.4%
(2, 1)	237	4.4%	222	6.0%	177	4.8%
(3, 2)	36	1.0%	22	0.9%	25	1.0%
(1671, 3248)	0	0%	1	22.7%	0	0%
(1682, 3627)	0	0%	0	0%	1	22.9%
(3119, 8441)	1	29.2%	0	0%	0	0%

Weakly connected components accounting for less than 1% of the users in all three settings (2007, 2008, Random-2008) are not listed. The distributions of the real decline and the random decline for the Canon EOS 20D almost coincide.

Table XVIII. The Table Shows the Percentage of Users Converted from 2007 to 2008 in Relation to the Number of Their Converted Friends in 2007

		Canon SD1000 (P&S)		Nikon D80 (DSLR)	
cliqueness	# conv. fr. '07	# users	% conv. '08	# users	% conv. '08
all	0	223054	1.1	174856	1.6
	1	4799	2.6	24329	3.0
	2	418	2.6	8789	3.3
	[3, 4]	88	2.3	7068	4.2
	[5, 8]	20	0	4546	5.0
	[9, ∞)	1	0	3282	5.0
high	0	227366	1.1	215974	2.0
	1	962	4.7	5570	3.9
	[2, ∞)	52	1.9	1326	5.4

The more of a user's friends had the (new) model under consideration in 2007, the more likely was the user to convert to this model in 2008. Converted high cliqueness friends ($F_J > 0.1$) were a strong indication toward a conversion of the user herself. Out of all users with a known camera model in 2007 and 2008, given that 2007 model is distinct from the model under consideration, 1.1% converted to a Canon SD1000 and 2.0% to a Nikon D80. These values serve as baseline conversion rates.

Factors Influencing the Probability of Conversion. Rather than a global view of the conversions (in terms of the size distribution of connected components), here we take a local view and analyze which factors influence the conversion probability of a single user.

First, we wanted to see if the number of converted friends of a particular user in 2007 has an influence on her conversion probability and, if yes, whether there are "diminishing returns" where having many converted friends does not enhance the conversion probability a lot compared to having at least a few converted friends. Not surprisingly, the number of infected friends did indeed have an influence (see top part of Table XVIII). So we looked further and tried to understand if there are certain types of friends, which have the biggest influence. For example, friends with a high cliqueness (see bottom part of Table XVIII), or friends from the same country (see Table XIX).

Table XVIII shows two things. First, at least for the DSLR model considered the conversion probability increases with the number of converted friends, though there is

Table XIX. The Table Shows the Percentage of Users Converted from 2007 to 2008 in Relation to the Number of Their Converted Friends in 2007

		Canon SD1000 (P&S)		Nikon D80 (DSLR)	
cliqueness	# conv. fr. '07	# users	% conv. '08	# users	% conv. '08
same country	0	227226	1.1	209598	1.9
	1	1046	2.2	8659	3.3
	2	79	2.5	2476	3.9
	[3, 4]	28	3.6	1472	4.8
	[5, 8]	1	0	502	4.6
	[9, ∞)	0	0	163	5.5
differ. country	0	227738	1.1	209113	1.9
	1	608	1.3	7700	3.4
	2	25	0	2530	4.7
	[3, 4]	7	0	1825	5.4
	[5, 8]	2	0	1045	5.1
	[9, ∞)	0	0	657	4.0

Friends are broken up into "same country" or "different country". The baseline conversion probabilities, irrespective of any friendship information, were again 1.1% and 2.0% for the Canon SD1000 and the Nikon D80 respectively.

the typical leveling off effect ("diminishing returns"). Second, high cliqueness friends[11] are "more valuable" in terms of their conversion power than arbitrary friends.

Earlier in Section 4.3 ("Two important user types: P&S vs. DSLR"), we already observed that the probability of brand congruence of P&S users depended a lot more on the effect of the country than for DSLR users. Table XIX confirms this observation from the angle of conversion probabilities. Though DSLR users are a lot more likely to convert if they have many converted friends, the location of the friend seems to play less of a role than for P&S users. Possible explanation are that P&S cameras are marketed more locally or that high-end users invest more time in product research and are also effected by remote friends.

4.5 Who's a Friend?

All of our analysis so far indicates that friendship, and especially high cliqueness friendship has a big influence on brand congruence and on model propagation. But what constitutes a "friend" or rather a "contact" on Flickr?[12] Are such virtual friends also friends in real life? Are they online acquaintances? Or family members?

Though we have no way to give a comprehensive answer to the complex issue of "friendship," we did look at certain observable aspects that help to shed some light on this complex question. A complete analysis of Flickr's social graph would be well beyond the scope of this work. Still, in this section, we answer the following questions by looking at our data. Are friends more likely to share a common last name, that is, most probably be family members? Are they more likely to share common interests? Are they more likely to live in the same country? And how do the answers differ between high-cliqueness and low-cliqueness friends?

Even though our analysis of the social graph is partly fragmentary and several explanatory factors are not part of our dataset, all the findings in the following indicate

[11]We chose a threshold of $F_J > 0.1$ and not 0.5 as, for example, in Table X as otherwise the number of matching users would have been too small to draw any conclusions.
[12]Recall that we use the term "friend" to refer to "contacts" in Flickr. We chose the terminology "friend" as it is more standard in the general context of social networks.

Table XX. The Fraction of Friend-Pairs Sharing the Same Last Name Increases with Increasing Cliqueness

cliqueness	# edges	%-age same
$0 \le F_j < 0.2$	9.5M	.48%
$0.2 \le F_j < 0.4$	0.2M	4.1%
$0.4 \le F_j < 0.6$	60K	9.3%
$0.6 \le F_j < 0.8$	24K	14.1%
$0.8 \le F_j \le 1.0$	9K	19.2%

Two random friends, independent of cliqueness, share the same last name in 0.7% of cases and two random users, irrespective of friendship, share the same last name in 0.03% of cases. Only users with an extractable last name were considered for this analysis.

Table XXI. The Overlap of Groups Joined by Two Friends as a Function of Their Cliqueness

cliqueness	# edges	F_G
$0 \le F_j < 0.2$	40.6M/37M	.032/.035
$0.2 \le F_j < 0.4$	1.8M/1.2M	.075/.110
$0.4 \le F_j < 0.6$	0.39M/0.14M	.048/.126
$0.6 \le F_j < 0.8$	0.19M/44K	.041/.156
$0.8 \le F_j \le 1.0$	90K/15K	.039/.198

The first numbers in columns 2 and 3 refer to the setting where a user without any joined groups has 0 overlap with other users. The second numbers refer to the setting where users without any joined groups are dropped from the analysis.

that “high cliqueness” for a pair of Flickr contacts does indeed correlate with “close friendship” in real life as high cliqueness increases the probability for pairs of users to be (i) family members, (ii) share interests, (iii) live in the same country, and (iv) they reciprocate their friendship links.

Friends and Family. Intuitively, family members are more likely to influence each other’s purchase decisions than non-family friends. Therefore knowing roughly what fraction of friends is made up by a user’s set of family members is relevant information for marketing campaigns. Though we have no way of determining the actual family relation between two given users, Table XX shows the fraction of befriended user pairs who share a common last name. To derive the last name we looked at all user’s who had provided at least two sequences of non-whitespaces as part of their “real name” on Flickr. Whenever there were at least two such tokens, we used the last one after casting it to lower case and removing punctuation characters. In this way, a last name could be extracted for 720K out of 2.1M users in our study.

Friends and Common Interests. Apart from family membership, another possible factor influencing (the closeness of) friendship is the set of common interests. Though the general interests of a user could not be directly observed by us, we used the set of groups joined by a user as an indication. One would expect that high cliqueness friends are more likely to share common Flickr groups than other types of friends. Table XXI shows that this is indeed the case but only after users who have not joined a single group are dropped from the analysis.

Friends Around the Corner. If "high cliqueness" does indeed correspond to, or at least correlate with "close friendship" then high cliqueness should also correlate with local proximity as close friends can be assumed to be more likely to be, well, close also in the geographical sense. Overall, any pair of friends with valid country information shares the same country in 40% of the cases, compared to 17% for a pair of random users. But for a pair of users with $F_J \geq 0.2$, this percentage has already increased to 76% and highly cliqued users with $F_J \geq 0.8$ share a common country in 89% of the cases.

Friendship is Mutual. On Flickr, the act of adding a user as a "contact" does not need to be authorized which means that friendship links can be unidirectional. See Section 4.3 ("Mutual vs. one-way links"). However, for friends with a cliqueness of at least $F_J \geq 0.2$ there is a chance of at least 90% for the link to be mutual and this increases to 99% for users with a cliqueness of $F_J \geq 0.8$. This means that high cliqueness is not caused by "devotees" who simply copy all of their idol's conctacts.

5. SUMMARY

The main results of our work can be roughly broken up into two groups: results concerning brand congruence and results concerning the dissemination of models through the friendship graph. As for brand congruence, we tested (i) whether friendship on Flickr has a noticeable impact on brand congruence ("Yes."), (ii) whether this impact can be explained by geographical factors alone ("No."), (iii) which factors have the biggest influence on the "strength" of a friendship link regarding brand congruence ("Cliqueness and size of friendship lists."), (iv) whether certain user types need to be considered separately ("DSLR and P&S."), and (v) whether there are noticeable effects of a local change of models or brands being related to changes in the neighborhood ("Yes."). Concerning the propagation of particular models, we analyzed (i) if the connected components of the corresponding users grow randomly ("No."), (ii) if they shrink randomly ("Yes."), (iii) if the model conversion probability of a user depends on the number of converted friends ("Yes."), (iv) if this probability increases for more high cliqueness converted friends ("Yes."), and (v) whether there was a strong dependency on friends being in the same/different country ("Only for P&S users."). Finally, we looked at what constitutes a "high-cliqueness" friend pair on Flickr and found several indications that they correspond to close friends in "real" life as such contacts (i) are more likely to be family members, (ii) are more likely to share common interests, (iii) are more likely to be in the same country, and (iv) are more likely to pertain to mutual friendship links.

Building on these insights, we plan to investigate other aspects apart from cliqueness which govern the strength of a friendship link. Here, possible factors are common cities (rather than merely common countries) but also generalized notions of "friendship link" using, for example, (i) the number of comments left for a user's pictures by another user, and (ii) the number of times these pictures are bookmarked by others. Similarly, it would be interesting to predict users who will be "early adopters" for new models, in particular for models who end up being a major success later.

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 *Proceedings of the 16th International Conference on the World Wide Web (WWW)*. 835–844.

ALGERSHEIMER, R., DHOLAKIA, U. M., AND HERRMANN, A. 2005. The social influence of brand community: Evidence from european car clubs. *J. Market. 69*, 19–34.

BACKSTROM, L., HUTTENLOCHER, D., KLEINBERG, J., AND LAN, X. 2006. Group formation in large social networks: membership, growth, and evolution. In *Proceedings of the 12th International Conference on Knowledge Discovery and Data Mining (KDD)*. 44–54.

CHA, M., MISLOVE, A., AND GUMMADI, K. P. 2009. A measurement-driven analysis of information propagation in the flickr social network. In *Proceedings of the 18th International Conference on World Wide Web (WWW)*. 721–730.

CHAUDHURI, A. AND HOLBROOK, M. B. 2001. The chain of effects from brand trust and brand affect to brand performance: The role of brand loyalty. *J. Market. 65*, 81 – 93.

CRANDALL, D. J., BACKSTROM, L., HUTTENLOCHER, D., AND KLEINBERG, J. 2009. Mapping the world's photos. In *Proceedings of the 18th International Conference on World Wide Web (WWW)*. 761–770.

GARG, N. AND WEBER, I. 2008. Personalized, interactive tag recommendation for flickr. In *Proceedings of the Conference on Recommender Systems (RecSys)*. 67–74.

KIRK, R. E. 2007. *Statistics: An Introduction* 5th Ed. Wadsworth Publishing, Beverly, MA.

KLEINBERG, J. 2007. Cascading behavior in networks: Algorithmic and economic issues. In *Algorithmic Game Theory*, N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani Eds., Cambridge University Press, Chapter 24, 613–632.

LESKOVEC, J., ADAMIC, L. A., AND HUBERMAN, B. A. 2007. The dynamics of viral marketing. *ACM Trans. Web 1,* 1, 5.

LESKOVEC, J. AND HORVITZ, E. 2008. Planetary-scale views on a large instant-messaging network. In *Proceedings of the 17th International Conference on the World Wide Web (WWW)*. 915–924.

LESKOVEC, J., SINGH, A., AND KLEINBERG, J. M. 2006. Patterns of influence in a recommendation network. In *Proceedings of the 10th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD)*. 380–389.

MARLOW, C., NAAMAN, M., BOYD, D., AND DAVIS, M. 2006. Ht06, tagging paper, taxonomy, Flickr, academic article, to read. In *Proceedings of the 7th Conference on Hypertext and Hypermedia (HT)*. 31–40.

MISLOVE, A., KOPPULA, H. S., GUMMADI, K. P., DRUSCHEL, P., AND BHATTACHARJEE, B. 2008. Growth of the Flickr social network. In *Proceedings of the 1st Workshop on Online Social Networks (WOSP)*. 25–30.

MISLOVE, A., MARCON, M., GUMMADI, K. P., DRUSCHEL, P., AND BHATTACHARJEE, B. 2007. Measurement and analysis of online social networks. In *Proceedings of the 7th SIGCOMM Conference on Internet Measurement (IMC)*. 29–42.

POPESCU, A. AND GREFENSTETTE, G. 2010. Mining user home location, travel patterns and gender from Flickr tags. In *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media (ICWSM)*. 307–310.

RATTENBURY, T., GOOD, N., AND NAAMAN, M. 2007. Towards automatic extraction of event and place semantics from Flickr tags. In *Proceedings of the 30th International SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*. 103–110.

REINGEN, P., FOSTER, B., BROWN, J. J., AND SEIDMAN, S. 1984. Brand congruence in interpersonal relations: A social network analysis. *J. Consum. Res. 11,* 3, 771–783.

RICHARDSON, M. AND DOMINGOS, P. 2002. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the 8th International SIGKDD Conference on Knowledge Discovery and Data Mining (SIGKDD)*. 61–70.

SIGURBJORNSSON, B. AND VAN ZWOL, R. 2008. Flickr tag recommendation based on collective knowledge. In *Proceedings of the 17th International Conference on the World Wide Web (WWW)*. 327–336.

SINGLA, A. AND WEBER, I. 2009. Camera brand congruence in the Flickr social graph. In *Proceedings of the 2nd International Conference on Web Search and Data Mining (WSDM)*. 252–261.

SINGLA, P. AND RICHARDSON, M. 2008. Yes, there is a correlation - from social networks to personal behavior on the web. In *Proceedings of the 17th International Conference on the World Wide Web (WWW)*. 655–664.

YANG, W.-S., DIA, J.-B., CHENG, H.-C., AND LIN, H.-T. 2006. Mining social networks for targeted advertising. In *Proceedings of the Hawaii International Conference on System Sciences (HICSS) 6*, 137a.

Received February 2010; revised August 2010, December 2010; accepted February 2011