# Groupthink and Peer Pressure: Social Influence in Online Social Network Groups

Pan Hui, Sonja Buchegger
Deutsche Telekom Laboratories, TU-Berlin
Ernst-Reuter-Platz 7, D-10587 Berlin, Germany
phone: +49/30/835358517, fax: +49/30/835358409
firstname.lastname@telekom.de

***Abstract*—In this paper, we present a horizontal view of social influence, more specifically a quantitative study of the influence of neighbours on the probability of a particular node to join a group, on four popular Online Social Networks (OSNs), namely Orkut, YouTube, LiveJournal, and Flickr. Neighbours in OSNs have a mutually acknowledged relation, most often defined as friendship, and they are directly connected on a graph of a social network. Users in OSNs can also join groups of users. These groups represent common areas of interest. We present a simple social influence model to describe and explain the group joining process of users on OSNs. To this end, we extract the social influence from data sets of OSNs of a million sample nodes. One of our findings is that a set of neighbours in the OSN is about 100 times more powerful in influencing a user to join a group than the same number of strangers.**

***Keywords*-social networks; social influence;**

## I. Introduction

Social influence occurs when an individual's thoughts or actions are affected by other people. The process of interpersonal influence that affects actors' attitudes and opinions is an important foundation of their socialization, identity, and decisions. [1].

Online social networks (OSNs), such as Facebook, Orkut, Flickr, and LiveJournal, have become very popular and continue to grow their user base. Users on these OSNs usually have explicitly labeled friends, which we consider to be proximity relationships on the networks. Users can choose to create or join a group or community based on common interests, such as affiliations, hobbies, political stands, or religions. They can invite other users to join the same groups. It is commonly believed that the close-proximity peers, such as friends and friends of friends, have social influence on the joining of a certain group of that node. As in any social networks, online or offline, the effect can also be the other way around, when people get to know others by virtue of their common interests and memberships in the same groups. These acquaintances then become new social relations. In OSNs, these new relations may result in links between users and thus new neighbours in the network topology. In this paper, we analyse snapshots of network topologies. There is therefore no information on what came first, the friendship link or the group membership. At least for teenage users, it has been found that *the dominant usage pattern is to connect with friends, family, and acquaintances, thereby reinforcing the structural dimensions of peer social worlds that exist in schools. It is less common for teens to seek out new friends online* [2]. We therefore think it is reasonable to assume that most users that are friends with others who joined the same group, are friends first and then join OSN groups.

Group memberships have been studied in social network analysis as two-mode social networks [3], variously referred to as *affiliation networks, membership networks* or *hyper-networks*, with relations being termed *affiliation relation* or *involvement relation*. Affiliation networks exhibit a duality of social relations and affiliations. They are two-mode networks consisting of subsets of actors, and connections among members of one of the modes are based on linkages established through the second mode. More precisely, the first mode is a set of actors as usual in social networks, the second, additional mode is a set of *events*, which can be a wide range of specific kinds of social occasions: e.g., social clubs, boards of directors of corporations, university committees. In this paper we use the terms *groups* and *memberships* when referring to the common-interest (second mode) part of the affiliation network, and *social networks* and *neighbours* for the friendship (first mode) part.

An interesting feature of affiliation networks is that, thanks to their dual nature, one can look at either part of the network and derive predictions for the other. We are interested in how friendships in OSNs, or, more generally, neighbors in social networks, influence the choice of groups to join. Conversely, when we look at group memberships, the question becomes what we can predict about a user's friends given the groups she has joined.

Affiliation networks are most commonly represented as two-mode sociomatrices, bipartite graphs and hypergraphs. While the latter provide a more immediate intuition thanks to their visualization, they are not scalable to large data sets as used in this paper.

The rest of the paper is organized as follows. We discuss related work in II. We explain the social influence model in Section III and the data sets used in this paper in Section IV. A description of the analysis and results can be found in Section V, followed by an outline of future work in

Section VI and conclusions in Section VII.

## II. Related Work

Online social networks are a fairly recent phenomenon. One of the first projects to study the impact on people's behaviour and social life was the Digital Youth project. A qualitative study [2] found that there is strong direct peer pressure to join OSNs among American teenagers in addition to their own feeling of being left out if they do not join. Reasons for not joining vary, but the dominant categories of non-participators seem to be disenfranchised teens with little access to OSNs, conscientious objectors, and former users.

There have been theoretical and empirical studies on group formation and preferential behavior in online groups. Backstrom et at. [4] studied the membership, growth and evolution of large social networks. They observed that the tendency of an individual to join a community is influenced by both the number of friends he has within the community, and more crucially how those friends are connected to one another. Another work also by Backstrom et al. [5] examined the preferential behavior of Yahoo! Groups. They found that different types of groups produce varying degrees of engagement. Members of a smaller, private group usually have higher engagement than members of a large, public one. The more groups a person belongs to, the less likely that they would be heavily involved in all of them. Anagnostopoulos et al. [6] studied the influence and correlation in social networks. They define several general models that replicate the aforementioned sources of social correlation and propose simple tests that can identify influence as source of social correlation.

While these studies have shown the existence of correlation between user actions and social influence, there is no quantitative study of a large range of OSNs and characterisation of the social influence of each network. Most of the previous works are trying to show the evolution of a certain type of OSN. Instead, we try to do a broad study of several OSNs with a certain static snapshots of time. Our aim is to produce a quantitative big picture of the social influence of different types of OSN and to try to explain the difference based on the nature of the service and the network topology of the OSN. More specifically, we look at how the proximity friendship will influence the joining of groups/communities on the OSN.

Prior to OSNs, affiliation networks have been studied in different contexts [3], such as of memberships on a corporate board of directors, club memberships of a set of community decision makers or elites, memberships in voluntary organizations, researchers' affiliations with academic institutions, committees of faculty members, trade partners of major oil exporting nations, high-school clubs, and eremonial events attended by members of a village. It has been found that not only are pairwise ties more likely between people who share a focus, or an affiliation, but these ties are likely to form specific kinds of network patterns, such as clusters.

## III. Social Influence Model

We propose a simple social influence model based on social influence network theory [1]. The theory tells us how a network of interpersonal influence enters into the process of opinion formation. In social influence network theory, the final outcome of an idea in a group of actors is the result of the interpersonal influences between these actors, and also actors' susceptibilities to interpersonal influence.

Here we model the action of joining an online social group as the result of the node's original opinion about the group, the influence of other nodes in the network, and its susceptibilities to interpersonal influence on joining the group. In OSNs, the influence can be in the form of direct invitation from another node in the system (e.g., a friend) or just the indirect observations of the activities of another node related to the group (e.g., the posting of a picture in a social group by a friend). In this model, we assume that the social structure of the group of actors is fixed during the process of opinion formation. This is realistic in an OSN in the way that the network would not evolve too fast for the small amount of time a user needs to decide whether to click or not to click on a button to join a group.

Setting aside the possibility of a set of users that become neighbours on a social graph because they met virtually in an OSN group and then became friends, we assume that a user's decision of joining a particular group on the OSN is influenced by (N-1) other nodes in the whole network. We then define the group-joining process in the OSN in a set of N actors to be:

$$y^{(t)} = AWy^{t-1} + (I - A)y^{(1)} \tag{1}$$

for $t$ = 2,3,..., where $y^{(1)}$ is an $Nx1$ vector of actors' initial opinions about joining a group, $y^{(t)}$ is an $Nx1$ vector of actors' opinions at time $t$, $W = [w_{ij}]$ is an $NxN$ matrix of interpersonal influences ($0 \leq w_{ij} \leq 1, \sum_j^N w_{ij} = 1$) and A = $diag(a_{11}$,$a_{22}$,...,$a_{NN})$ is an $NxN$ diagonal matrix of actors' susceptibilities to interpersonal influence on joining a group ($0 \leq a_{ii} \leq 1$).

If we apply the equation (1) iteratively, we obtain

$$y^{(t)} = V^{(t-1)}y^{(1)} \tag{2}$$

where,

$$V^{(t-1)} = (AW)^{t-1} + [\sum_{k=0}^{(t-2)} (AW)^{(k)}](I - A) \tag{3}$$

for t = 2, 3...

When we consider the process in an equilibrium state (assuming convergence), equation (1) becomes

$$y^{(\infty)} = AWy^{(\infty)} + (I - A)y^{(1)} \tag{4}$$

If we assume $I - AW$ to be non-singular, then

$$y^{(\infty)} = Vy^{(1)} \tag{5}$$

where,

$$V = (I - AW)^{(-1)}(I - A) \tag{6}$$

$V$ is a matrix of reduced-form coefficients describing the total interpersonal effects that transform a user's initial opinion about joining a group into final opinions. Considering that in an OSN, an invitation to join a group usually can only be sent to friends (i.e., direct neighbours) or users can usually only observe the activities of their direct friends, we can simplify the number $N$ to be the target actor and her direct neighbours. We will show later in the result section that actually strangers have very little similarity in joining the same groups as an actor.

In this paper, we look at how the social influence of the neighbours affects a node's opinion about joining a group. Since we use snapshot data of the OSNs, which do not show the dynamics of the evolution of the network, we are more interested in looking at the similarity of the group memberships of a node with its neighbours, and what the difference is compared to random nodes in the system. This would be useful for us to validate our claim above that we can limit the number of actors $N$ to be the node and its direct neighbours. Additionally, it can be useful for us to estimate the interpersonal influence parameters for each OSN.

## IV. Online Social Networks Data Sets

The OSNs we study in this paper include Flickr, Orkut, LiveJournal, and YouTube. The data were collected by crawling in late 2006 and 2007 [7]. The networks may have evolved in the past two years, which is the same for any real social network, and it is true for all dynamic systems. Studying a human social network at a certain point in time can still give us some general insight and knowledge into human behavior, in the sense of anthropology and sociology. Furthermore, this is the biggest available OSN data, and it usually takes a lot of effort and resources to do large-scale data collection, making it extremely difficult to obtain data that shows changes over a meaningful period of time or that even just exceeds the extent of the one used here. Our interest is more on the steady state of the membership distribution instead of the evolution. We believe it is justified for us to look into these data sets, as has traditionally been done for social network analysis of static graphs, in order to better understand correlations of group memberships and friendships in OSNs. The data sets were collected by automated crawling scripts on a cluster of 58 machines. Statistics of the data sets [7] are summarised in Table I.

### A. Flickr Data Set

Flickr (www.flickr.com) is an image and video hosting website, web services suite, and online community platform. It was launched in February 2004, and as of November 2008, it claims to host more than 3 billion photos. On a Flickr user's profile page, there are contacts (friends) and public groups this user is belonging to. Groups are used in Flickr to share content and for conversation. A group can be public or private. Each group has a pool for photos, which are shared by the members of this group, and a discussion board for talking. The data set we used in this paper contains over 1.8 million users and 22 million links. The neighbour and membership information are obtained via the APIs exported by Flickr for third-party developers. This data set covers a large fraction of the large weakly connected component (WCC), which is the set of nodes in a directed graph where each node has a path to every other node in the same set if all links are viewed as undirected.

### B. Orkut Data Set

Orkut (www.orkut.com) is a social network website run by Google to help users meet new friends and maintain existing relationships. It is a more explicitly social network website compared to the other three networks we study. Brazil and India constitute the major user base of Orkut. As of May 2008, 53.86% users are from Brazil, and 16.97% from India. In Orkut, a group is called a community. Orkut communities allow members to connect over a shared interest or hobby. Users can join a community or create their own community.

In Orkut, links are indirectly created by trust relationships and require consent from the target. The data was collected using HTML screen-scrapping between October 3rd and November 11th, 2006. The crawled subset consists of 3,072,441 users, corresponding to 11.3% of Orkut's user popultion of about 27 millions at the time of the crawl. This data collection was limited because Orkut requires a logged-in account to browse the network and limits the rate at which a single IP address can download information.

### C. LiveJournal Data Set

LiveJournal (www.livejournal.com) is an online social network for bloggers. Users can share their blog, journal, and diary. LiveJournal was started in March 1999. In LiveJournal, a group is also called community. A community is a journal run by a member for people with common interests. The data set we used covers around 95.4% of the users of the whole LiveJournal community at that time. It contains over 5.2 million users and 72 million links. The LiveJournal data in this paer was obtain from a crawl from December 9-11, 2006, with the APIs provided by the website.

### D. YouTube Data Set

YouTube (www.youtube.com) is a video-sharing website where users can upload, view, and share videos, and it includes a social network. YouTube was created in February 2005. Besides video sharing, YouTube also allows users to create groups. Groups allow multiple people to discuss

| OSN | Orkut | Flickr | LiveJournal | YouTube |
|---|---|---|---|---|
| Number of users | 3,072,441 | 1,846,198 | 5,284,457 | 1,157,827 |
| Estimated Crawled Fraction | 11.3% | 26.9% | 95.4% | unknown |
| Number of links | 223,534,301 | 22,613,981 | 77,402,652 | 4,945,382 |
| Mean number of friends per user | 106.1 | 12.24 | 16.97 | 4.29 |
| Fraction of links symmetric | 100.0% | 62.0% | 73.5% | 79.1% |
| Number of groups | 8,730,859 | 103,648 | 7,489,073 | 30,087 |
| Mean group membership per user | 106.44 | 4.62 | 21.25 | 0.25 |

Table I
HIGH-LEVEL STATISTICS OF THE CRAWLED OSN DATA SET

things publicly and to post videos that apply to the discussion. A creator or member of a group can add video, invite other members, start a conversation, and offer comments to videos and topics that other members have added. The YouTube data set we studied in this paper was obtained on January 15th, 2007, it consists of over 1.1 million users and 4.9 million links, and is believed to cover a large fraction of the whole network.

## V. RESULTS AND ANALYSIS

### A. Sample Properties

Since the network is too big for membership analysis, we decided to sample a portion of each network with sample points chosen uniformly distributed. Because different networks have different link densities and the computer memory can only handle a limited amount of operations, we sample a different number of nodes from different networks. For example, we sample 5,000 nodes from Flickr, LiveJournal, and YouTube, but can only handle 350 nodes from Orkut [1]

Figure 1 shows the degree distributions of the sampled nodes for all these four OSNs [2], and the descriptive statistics are shown in Table II. We can see that Flickr, LiveJournal, and YouTube all have low degree distributions compared to the Orkut data set, which focuses on friendships. Flickr has an extreme degree distribution with an average degree of 5.6547 and the maximum can be as large as 3612. The table also shows the types of the distributions for each data set and the corresponding Kolmogorov-Smirnov value measuring the fitness of the fitting.

| OSN | Distribution | Kol-Smir | Mean | Median | Max |
|---|---|---|---|---|---|
| Orkut | Dagum | 0.028 | 74.301 | 46 | 857 |
| Flickr | Frechet | 0.278 | 5.6547 | 0 | 3612 |
| LiveJournal | Gen. Gama(4P) | 0.096 | 16.965 | 6 | 733 |
| YouTube | Lognormal | 0.081 | 9.068 | 3 | 656 |

Table II
DEGREE STATISTICS OF OSN SAMPLES.

Figure 2 shows the distribution of the number of groups each node has, and Table III summarises the descriptive statistics of the graphs. We can see that the mean numbers of groups of our sample are very similar to the mean of

[1]There are around 45,000 neighbours of these 350 sample nodes and the program thus needs to process 45,350 nodes instead of 350.

[2]For Flick, LiveJournal, and YouTube, we only shows degree up to 100 for better resolution, and the same case for Figure 2

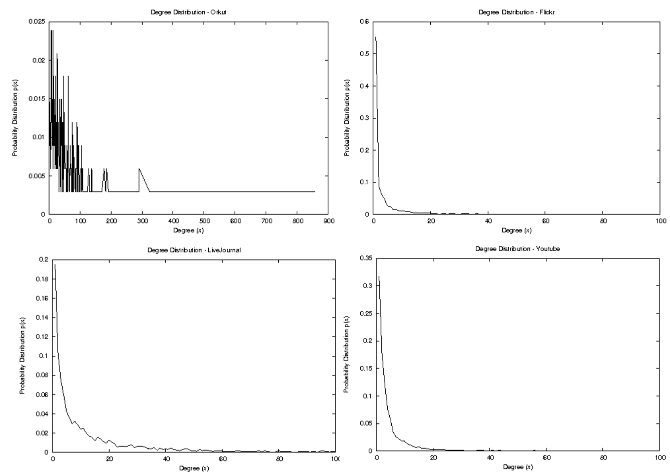


Figure 1. Degree distribution of Orkut(T-L), Flickr(T-R),LiveJournal(B-L), and YouTube(B-R) .

| OSN | Distribution | Kol-Smir | Mean | Median | Max |
|---|---|---|---|---|---|
| Orkut | Pareto2 | 0.072 | 108.2 | 27 | 996 |
| Flickr | Logistic | 0.072 | 5.6547 | 0 | 656 |
| LiveJournal | Gen. Extr. value | 0.181 | 23.335 | 6 | 254 |
| YouTube | Logistic | 0.345 | 0.478 | 0 | 66 |

Table III
MEMBERSHIP STATISTICS OF THE OSN SAMPLES

the whole network (Table I). Orkut has an average of as high as 108 groups per node, but each YouTube user tends to have less than half the number of group memberships. This is easily explainable by the fact that the main purpose of YouTube is not for socialising. LiveJournal has a mean group number of 23 and for the Flickr case, the average is 5.6.

### B. Metrics

The metrics we are looking for from the data sets are the similarities of a node's memberships with its neighbours. Here we definite the result of mutual influence, or we can also call it similarity, of a node $i$ by its $k$ neighbours as:

$$S(i) = \frac{\sum_{j=1}^{k}\left(\frac{|\bigcap(G(i)G(j)|}{|\bigcup(G(i)G(j))|}\right)}{k} \quad (7)$$

where, $G(i)$ is the groups joined by node $i$, and $k$ is the degree of node $i$.

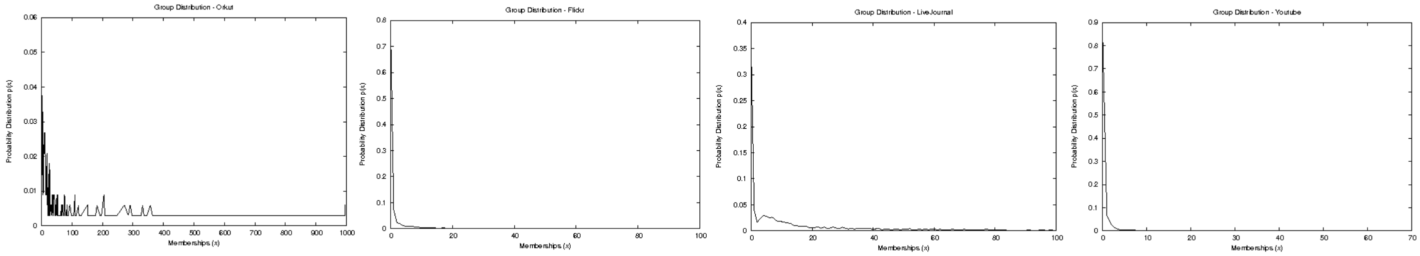


Figure 2. Group distribution of Orkut, Flickr, LiveJournal, and YouTube.

This definition is based on the classic Jaccard index [8] which was proposed by Jaccard over a hundred years ago to evaluate the similarity of two communities. This metric basically measures how strong the mutual influence of a node and each of its neighbours is on joining the same group.

We can also define the influence power of node $i$ on its neighbours as:

$$I(i) = \frac{\sum_{j=1}^{k}(\frac{|\bigcap(G(i)G(j)|}{|G(j)|})}{k} \tag{8}$$

The difference between $S(i)$ and $I(i)$ is that for S(i), the result is normalised with the union of both the groups of the sample node and its neighbour, but for I(i), the result is only normalised with the group membership of the neighbour node $j$. This shows how many groups joined by node $j$ are correlated with the influence of node $i$.

### C. Social Influences

In order to compare the difference of influence from the neighbour nodes and from random nodes in the network, we take all the neighbour nodes for the sample nodes and randomly assign the same number of nodes as control neighbour nodes to each sample node.

From Table IV, we can see that the neighbours have a much higher similarity value compared to the control nodes, and the difference is as high as 105 times for the YouTube data set. Since we select the control nodes from the common neighbour sets of all the sample nodes, this difference is expected to be large if we do random sampling from the whole network.

### D. Similarity by Degree and Clustering

According to Backstrom et al. [5], the probability that a node will join a group depends on how many friends of the node already are in the group, but more importantly also on how well these friends know each other. We try to investigate this by looking at the correlation between the similarity values and the clustering coefficients [9] of the sample nodes. The clustering coefficient of a node roughly measures how many triangles are formed among its neighbours, the higher the clustering coefficient the better the neighbours know each other. It can be calculated by dividing the number of links between the neighbours by the maximum number of links that can be formed among these neighbours. We show the correlation of these two metrics in Figure 3. Also from Table V, we can see that the maximum S-C correlation is 0.4745 in the Orkut case, and for the YouTube case, it is as low as 0.0670. We can say that from the data sets available, we cannot observe a strong correlation between similarity and the clustering coefficient.

Usually, a node with a high degree is expected to have high influence power (i.e., higher degree centrality) [10]. We want to look at how the degree of a node will affect the similarity values, $S(i)$, and also the influence $I(i)$. This can be done by correlating the similarity/influence with the node degree. We show in Figure 4 the correlation for Flickr and YouTube. We can see that there is little correlation between the degree and the similarity of nodes. Similar tendencies are observed in the Orkut and LiveJournal cases, so we omit them for redundancy. Similar results also occur in the node influence case, and we also summarise the descriptive statistics in Table V. Usually, however, node influence has a higher correlation with degree compared to similarity.

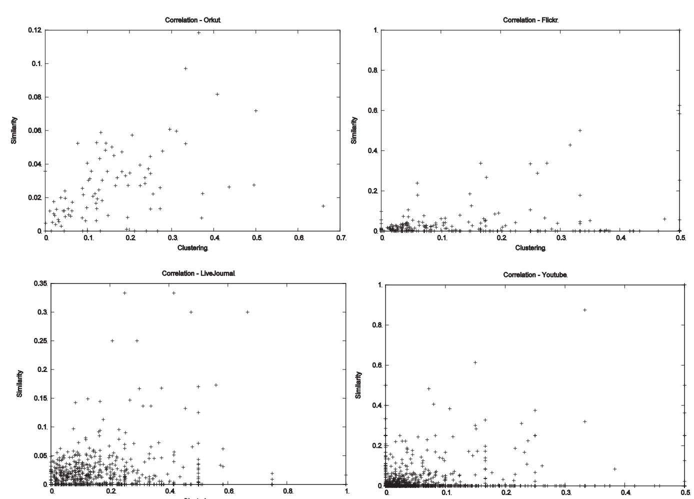


Figure 3. Correlation of similarity and clustering coefficient in OSN samples.

| Statistics | N-Orkut | C-Orkut | N-Flickr | C-Flickr | N-Live | C-Live | N-tube | C-tube |
|---|---|---|---|---|---|---|---|---|
| Sample Size | 350 | 350 | 5,000 | 5,000 | 5,000 | 5,000 | 5,000 | 5,000 |
| Mean | 0.02545 | 0.00117 | 0.01575 | 3.8510E-4 | 0.01985 | 0.00472 | 0.00968 | 9.1566E-5 |
| 5% | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 25% (Q1) | 0.00777 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 50% (Median) | 0.0165 | 6.3500E-4 | 0 | 0 | 0.00212 | 0 | 0 | 0 |
| 75% (Q3) | 0.03543 | 0.00176 | 0 | 0 | 0.02221 | 0.00472 | 0 | 0 |
| Max | 0.22071 | 0.00969 | 1 | 0.01973 | 1 | 0.5 | 1 | 0.08333 |

Table IV
CHARACTERISTICS OF THE MUTUAL INFLUENCE IN THE EXPERIMENTAL DATA SETS

| OSN | Orkut | Flickr | LiveJournal | YouTube |
|---|---|---|---|---|
| Correlation (s-c) | 0.475 | 0.185 | 0.082 | 0.067 |
| Correlation (s-d) | -0.088 | 0.040 | -0.026 | 0.002 |
| Correlation (i-s) | -0.013 | 0.153 | 0.016 | 0.027 |

Table V
CORRELATIONS OF OSN SAMPLES.

Figure 3 shows the correlation between the similarity and the clustering coefficient for these four OSNs.

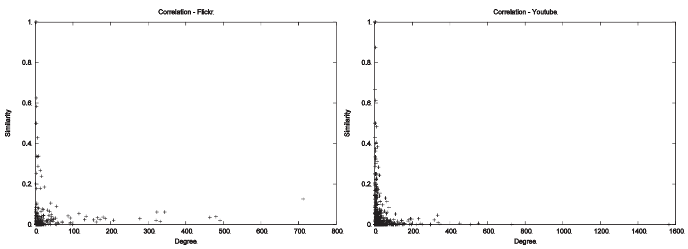


Figure 4. Correlation of similarity and node degree of Flickr(L) and YouTube(R).

Similarity and influence show a very high correlation (Figure 5), but are not completely correlated. Nodes that have low similarity values can, however, have a high influence value sometimes.

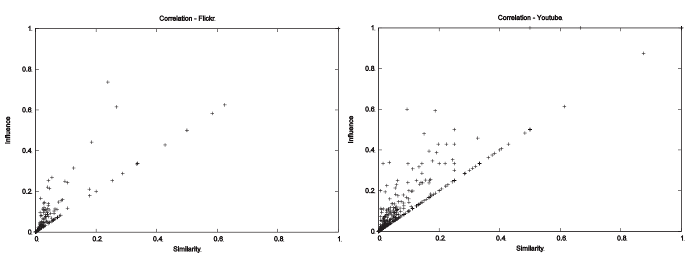


Figure 5. Correlation of node influence and mutual influence of Flickr(L) and YouTube(R).

## VI. Future Work

This paper presents a first step toward analysing the relationship between social ties and group memberships in online social networks.

Some open questions include: What part of social relations are created by virtue of being members of the same group first? Is there a correlation between the number of friends an OSN user has and how many groups she has joined, are joiners more social? How has group-joining behavior changed over the years since the phenomenon of OSNs started? By now, OSNs have become quite mainstream, predominantly with younger people, has group joining become more wide-spread as users get used to OSN features and explore more?

Since we only have data on the online social networking part and not on the history of how the links between nodes were made in relation to the social ties in the real world (the *meat space*, as it is sometimes called to distinguish from the *virtual space* on the Web), we have no information on how many friendships in an OSN stem from real-world acquaintances from real-world groups or other interest affiliations. Also, links between users in one OSN can be the result of getting to know each other in a different virtual space, such as another OSN or other interactive websites, such as blogs with commenting features.

In Section III we presented a social influence model that, in the next step, we will use to further analyse the data sets listed in Section IV.

## VII. Conclusions

In this paper, we analyse the influence of friendship on group membership and vice versa using data sets of four popular online social networks, with the data sets consisting of millions of sample points. We calibrate the influence using similarity measurements, and characterise the differences of these four systems.

We note that most users of OSNs so far do not join groups. When they do join, most limit themselves to very few groups, but there is a long tail in the distribution and a few nodes join a large number of groups. We can say that from the data sets available, we cannot observe a strong correlation between similarity and clustering coefficients.

We found that basically there is little correlation between the degree and the similarity of nodes in our data sets. Similarity and influence show quite a high correlation, but they are not completely correlated. Nodes that have low similarity values can, however, have a high influence value sometimes.

The correlations we found do not indicate causation, as we analyse a snapshot of the topologies. Causation in one way would mean social influence by friends leading to joining groups, in the other way it would mean that group membership leads to social relations with other group members. In both interpretations, it is about the correlation between social ties and personal interests.

Comparing the influence of neighbours in the social network on group joining behavior to the influence of random nodes in the network, we find that the neighbors, i.e., friends (in varying senses of the word, depending on the nature of the online social network the samples are taken from) have a much higher impact than random nodes, i.e., strangers.

Additionally, we proposed a simple social influence model based on social influence network theory. Future work includes the application of this model to the analysis of online social network data sets.

## VIII. Acknowledgements

The authors would like to thank Alan Mislove for providing us with the OSNs data. This work is in collaboration with the EC IST SOCIALNETS - Grant agreement number 217141.

## References

[1] N. Friedkin and E. Johnsen, "Social influence networks and opinion change," *Advances in Group Processes*, vol. 16, pp. 1–29, 1999.

[2] danah boyd, "Taken out of context: American teen sociality in networked publics," Ph.D. dissertation, University of California, Berkeley, January 2009.

[3] S. Wasserman and K. Faust, *Social Network Analysis Methods and Applications*. Cambridge University Press, 1994.

[4] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan, "Group formation in large social networks: membership, growth, and evolution," in *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM, 2006, pp. 44–54.

[5] L. Backstrom, R. Kumar, C. Marlow, J. Novak, and A. Tomkins, "Preferential behavior in online groups," in *WSDM '08: Proceedings of the international conference on Web search and web data mining*. New York, NY, USA: ACM, 2008, pp. 117–128.

[6] A. Anagnostopoulos, R. Kumar, and M. Mahdian, "Influence and correlation in social networks," in *KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM, 2008, pp. 7–15.

[7] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proceedings of the 5th ACM/USENIX Internet Measurement Conference (IMC'07)*, October 2007.

[8] P. Jaccard, "Etude comparative de la distribution florale dans une portion des alpes et des jura." *Bulletin de la Societe Vaudoise des Sciences Naturelles*, vol. 37, p. 547, 1901.

[9] M. E. J. Newman and J. Park, "Why social networks are different from other types of networks," *Phys. Rev. Lett. E*, vol. 68, p. 036122, May 2003.

[10] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, pp. 35–41, 1977.