



HAL
open science

Are Friends Overrated? A Study for the Social Aggregator Digg.com

Christian Doerr, Siyu Tang, Norbert Blenn, Piet Van Mieghem

► To cite this version:

Christian Doerr, Siyu Tang, Norbert Blenn, Piet Van Mieghem. Are Friends Overrated? A Study for the Social Aggregator Digg.com. 10th IFIP Networking Conference (NETWORKING), May 2011, Valencia, Spain. pp.314-327, 10.1007/978-3-642-20798-3_24 . hal-01597982

HAL Id: hal-01597982

<https://inria.hal.science/hal-01597982v1>

Submitted on 29 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Are Friends Overrated?

A Study for the Social Aggregator Digg.com

Christian Doerr, Siyu Tang, Norbert Blenn, Piet Van Mieghem

Department of Telecommunication
TU Delft, Mekelweg 4, 2628CD Delft, The Netherlands
{C.Doerr, S.Tang, N.Blenn, P.F.A.VanMieghem}@tudelft.nl

Abstract. The key feature of online social networks is the ability of users to become active, make friends and interact with those around them. Such interaction is typically perceived as critical to these platforms; therefore, a significant share of research has investigated the characteristics of social links, friendship relations, community structure, searching for the role and importance of individual members.

In this paper, we present results from a multi-year study of the online social network Digg.com, indicating that the importance of friends and the friend network in the propagation of information is less than originally perceived. While we note that users form and maintain social structure, the importance of these links and their contribution is very low: Even nearly identically interested friends are only activated with a probability of 2% and only in 50% of stories that became popular we find evidence that the social ties were critical to the spread.

1 Introduction

The recent explosive growth of online social network (OSN) platforms such as Facebook, Twitter, or Digg has sparked a significant interest. As several hundred millions users now regularly frequent these sites to gather and exchange ideas, researchers have begun to use this comprehensive record to analyze how these networks grow by friendship relations, how information is propagated, and who are the most influential users. A good understanding of these principles would have many applications, such as effective viral marketing [1], targeted advertising [2], the prediction of trends [3] or the discovery of opinion leaders [4].

A fundamental assumption of previous research is that friendship relations are a critical component for the proper functioning of social networks [5], i.e., they assume that information, opinions and influences are sourced by single individuals and then propagated and passed on along the social links between members. The extent, density, layout and quality of the social links and the network of links will therefore determine how information can be spread effectively.

In this paper, we report on results from a multi-year empirical study of the OSN Digg.com, a so-called social news aggregator, that indicate that the criticality and importance of individual friendship relations and the friendship network is less than previously perceived. In these social news aggregators, users

submit news items (referred to as “stories”), communicate with peers through direct messages and comments, and collaboratively select and rate submitted stories to get to a real-time compilation of what is currently perceived as “hot” and popular on the Internet. Yet, despite the many means to communicate, interact and spread information, an analysis of eleven million stories and the commenting and voting patterns of two million users revealed that the impact of the friendship relations on the overall functioning of the social network is actually surprisingly low. In particular, we find that, while users indeed form friendship relations according to common interests and physical proximity, these friendship links are only activated in 2% of the information propagation. Furthermore, in 50% of all stories that became “hot”, there was no sufficient prior contribution by the friend network to trigger the popularity of the story. Instead, we find that a critical mass was reached through participation through random spectators.

The remainder of this paper is structured as follows: Section 2 discusses related work and prior findings on the role and characteristics of friendship links and the friendship network in OSNs. Section 3 describes background information about the social network used in our experimentation and our data collection methodology. Sections 4 and 5 discuss the role of friendships and selected individuals to the successful information propagation. Section 6 present an outlook on the issue of time dependencies in Digg. Section 7 summarizes our findings.

2 Related Work

Ever since the publication of Katz and Lazarsfeld’s argument for the origin and spread of influence through communities [6], researchers have investigated the mechanisms by which ideas and opinions are passed along social relationships. The wide-spread popularity of OSNs now provides an easily accessible, machine-readable data source for broad-scale analysis. As it is commonly assumed that friendship interactions are the backbone of social networking platforms [5] along which a “web of influence” [7] is extended and maintained, the investigation of how links are formed and used has received significant attention.

The mechanisms by which these social ties are formed is still subject to investigation. Fono and Raynes-Goldie [8] for example studied the semantics of friendship relations in the OSN LiveJournal and find that a number of overlapping processes drive the formation of links: ties may be formed as measures of trust, reciprocity, an extension of offline acquaintance or as a facilitator for information transmission. This function of friends as content providers [9] is a strong force, as it even drives the friend selection of users: in a usage survey on Facebook, subscribers named their interest to obtain information from friends to form friendship relations [10]. According to network theory, such content dissemination should work best across “weak ties” [11] which are linking tightly connected clusters, as these should transmit information with the least amount of redundancy. Therefore, the predominant share of OSN research has been conducted to investigate the topological properties of OSN and to understand how users behave and exchange content and information across their social links.

Along these lines, Mislove et al. [12] studied the topological properties of four OSNs at large-scale: Flickr, YouTube, Live-Journal, and Orkut, for which they evaluated different network metrics, e.g. link symmetry, node degree, assortativity and clustering coefficient. The OSNs were characterized by a high fraction of symmetric links, and contained a large number of highly connected clusters, indicating tight and close (transitive) relationships. The degree distributions in the OSNs followed a power-law with similar coefficients for in- and out-degree, showing the mixed importance of nodes in the network - there are few well connected and important hubs to which the majority of users reach to.

In [13], Leskovec et al. presented an extensive analysis about communication behaviors of the Microsoft Messenger instant-messaging users. The results of 30 billion conversations among 240 million people showed that people with similar characteristics (e.g., age, language, and geographical location) tend to communicate more. Structurally, in terms of node degree, cluster coefficient, and shortest path length, it was shown that the communication graph is well connected, robust against node removal, and exhibits the small-world property.

Another line of research aims to discover factors for content popularity and propagation in OSNs. For instance, it is shown in [14] that there are different photo propagation patterns in Flickr and photo popularity may increase steadily over years. A similar study was performed with YouTube and Daum in [15]. Here, video popularity is mostly determined at the early stage after content submission. Such observations are also voiced in the viral marketing literature where it is assumed that “few important trends reach the mainstream without passing the Influentials in the early stages, ... they give the thumbs-up that propel a trend“. [16, p. 22]. As it is however quite difficult to evaluate concretely and at a population scale how content is disseminated, there exists to this date no population-level (on the order of millions of users thereby also considering possible emergent properties) evaluation of these hypotheses.

This paper aims to address this void, and from this current state of research, we therefore formulate two hypotheses that will be addressed and investigated for the scope of an entire social network:

- H_1 There exist critical members inside the community who have better or earlier access to important information.
- H_2 Inter-personal relations and the overall network of friendships are the key component to the successful spread of information.

3 The “Digg” data collection

The news portal digg.com is a social content website founded in 2004. According to the Alexa rating (alexa.com), Digg belongs to the top 120 most popular websites. At the time of this work 2.2 million users are registered at the webpage, submitting about 15,000 to 26,000 stories per day. Out of those, approximately 180 stories are voted to be popular per day. Users are able to submit and vote on content they like, an activity called “digging”. A submission consists of the link to a webpage, where a news story, an image or a video is stored, and a short

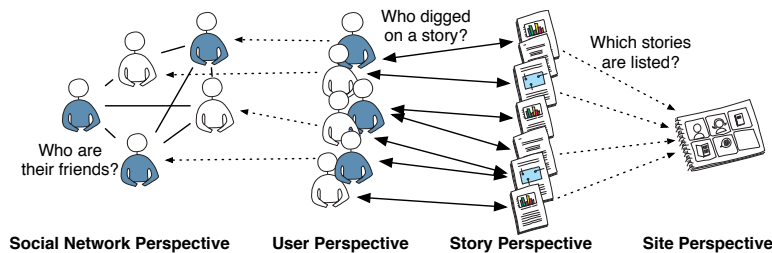


Fig. 1. Different components of the Digg crawling process.

description of the content. New submissions gets enqueued to the upcoming pages where they are staying for a maximum of 24h. Users may explore the upcoming section by topic or a recommendation system, which displays stories that already gained attention and votes. A secret algorithm by digg.com chooses stories from the upcoming section to promote them to the front pages which is the default page when entering the Digg website. After this promotion a story will therefore obtain a lot of attention.

Within the Digg network, it is possible for users to create friendship connections to other users. One may either be a fan or a mutual friend to another person. Fans and friends are notified by the friends interface of Digg if their friend has “digged” or submitted a story.¹ We studied different aspects of Digg, such as the friendships, user characteristics and activities, and the properties and dynamics of the published content. While most social network traces are crawled using friendship relations, e.g. [12] and [17], the Digg dataset was obtained by a simultaneous exploration from four different perspectives, as shown in Fig. 1:

Site perspective: The Digg website lists popular and upcoming stories in different topic areas. Periodically all front pages with all popular stories and all upcoming stories were collected. All discovered stories are added to an “all-known story” list maintained by us.

Story perspective: For all stories, a complete list of all diggs performed by users is collected. Newly discovered users will be added for future exploration.

User Perspective: For each discovered user, a list of their previous activities is obtained. If new stories are found the entire activities is retrieved.

Social Network Perspective: A list of friends is downloaded for every user. If a user is newly found, he is added to the data discovery process, and a list of all friends and user profile information are fetched.

¹ It should be noted at this point that the semantics of a friend in Digg (obtaining information) is certainly different from a friendship in Facebook (personal acquaintance) or LinkedIn (business contact) [9], as also the main function differs between these social networks. As this paper investigates information propagation and social news aggregators such as digg.com focus on the exchange of information, these results are only immediately applicable to this type of OSN. To what extent these findings can be extended towards other types needs more investigation.

The above procedure is continued until no new user and story can be found and periodically repeated afterwards to discover new activity. By using the above crawling methodology, we are able to collect (nearly) the entire information about friendships, activities of users and the published content in the Digg network. This is important as traditional crawling techniques will only discover those users which are linked within the social network and will overlook for example all users without any friendship relations who are still otherwise active members. In our case our crawling technique discovered nearly twice as many users than could have been identified by a pure crawl of the social network alone. This outcome might (partially) explain some of the contrary findings in our paper. Our data covers the entire span since the beginning of Digg in 2004 until the end of our study in July 2010 and contains a volume of more than 600 GB, covering the history of 2.2 million registered users and 11 million stories.

4 Information Spread through the Network of Friends

As discussed in the introduction, it is commonly assumed that the friendship relations within a social network are a critical components to the successful spread of information. This section will dissect this process and investigate for the case of the Digg OSN, whether the propagation of news is indeed the results of the activation of users' ties.

4.1 Self-Organization of the Friendship Network

According to sociological theory, friendship relations in an OSN grow directed by common interests and tastes [9]. Within Digg, all news stories are classified within eight major topic areas, subdivided by 50 special interests. When matching the users' concrete digging behavior with the topic area into which a story was classified, we find that the subscribers exhibit quite strong and distinct preferences and tastes for individual topic areas. As shown in Fig. 2, if a particular user reads, diggs and is therefore interested in two distinct topic areas, say for example "Science" and "Technology", more than 70% of all consumed stories fall within the most preferred genre. For three topic areas, the favorite one draws 65%, and even for users interested in all eight categories the top two will on average still account for 60% of read stories. These rankings of user interest provide a direct measure of how similar the tastes and preferences of users in their information acquisition are. When comparing two users and their ranking of topics, we use the number and distance of permutation steps required to transform one list into the other (the Wasserstein rank distance [18]) as a measure of user similarity. A network-wide analysis of the similarities between friends shows that users directly connected to each other have a very high alignment of their preferences and tastes: 36% of rank lists are identical, 20% require one transformation, and within three transformation steps 80% of all friendship relations are aligned.

While there exists a perfect overlap between the interests and tastes of friends, there is a surprisingly low amount of common activity among friends

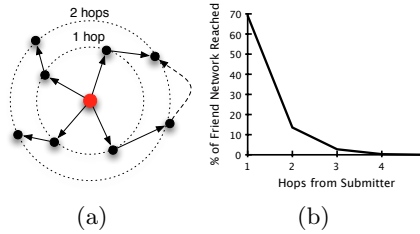
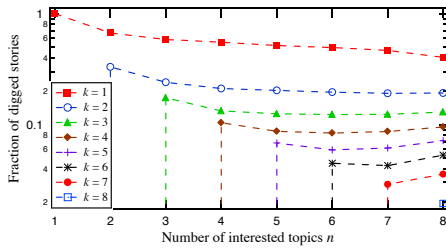


Fig. 2. The average ratio of the k -th favorite topic of users that are interested in work spreading out from the submitter is n topics after ranking. **Fig. 3.** Most of the entire friendship network actually already covered by the first hop.

and only 2% of all friend pairs actually do dig on the same story. The hypothesis that common interests result in the formation of friendships in order to gain information from neighboring peers [8] would also predict that the more similar the tastes between friends are, the closer the alignment of clicking patterns would be. In practice however, we found this not to be the case.

Yet, pairs of friends do not exist in isolation, but are embedded within a larger network of the friends of the friends. This in OSNs very dense network of friends [12] may be a powerful promoter, as theoretically a large group can be reached if information can be passed on from friend to friend over several steps.

Our analysis shows that information can indeed travel over multiple hops from the original submitter (see Fig. 3(a)) and on average does reach 3.7 hops from the source until the propagation dies down. The actual contribution of the multi-hop network, i.e., the number of friends of friends that can be activated, is however rather limited. As shown in Fig. 3(b) nearly 70% of the ultimately participating *network* of friends consists of the submitter’s direct friends, while the benefit of the additional hops decreases super-exponentially. This result is not astonishing given the generally low activation ratios of friends and possible redundancies in the spread as indicated by the dashed line in figure 3(a).

4.2 Reaching Critical Momentum

All news stories submitted to Digg are initially collected in the “upcoming” list, which with more than 20 000 submissions a day has a very high turnover rate (more than 800/h). In order to become promoted to the front pages, a story has to attract sufficient interest, i.e., a large enough number of diggs, within 24 hours, which as shown in Fig. 4 the majority reaches within 16 hours. We experimentally determined that about 7 diggs/h are necessary to qualify for the promotion, thereby stories should gather on average around 112 diggs.

A story can rally this support initially from random spectators or friends of the submitter, who were notified about the newly placed story. Figure 5 shows the probability that a given number of friends is active on the website on the same day, and we compute the likelihood that at least 112 friends are online within 24

Table 1. Ratio of friends/non-friends among the number of diggs for popular stories

Average ratio	Before popular		After popular	
	Friends	Non-friends	Friends	Non-friends
a) 63484 stories	0.72	0.28	0.25	0.75
b) 51679 stories	0.23	0.77	0.14	0.86

hours to be about 0.01. This expectation corresponds with the actually observed promotion success ratio. The fine line between failure and success, whether a story will be forgotten or becomes popular, therefore strongly depends on the underlying stochastic process, whether at a certain time a sufficient number of friends are online to support the story. In the remaining 99% of the cases when not enough friends are online, additional support needs to be rallied from users outside the friend network to reach the promotion threshold.

4.3 Promotion without Friendships

As the likelihood that a story becomes popular solely through the submitter’s friendship network is rather slim (given the low activation ratio of friends, the limited contribution of the network of friends and the low probability of a critical mass of friends active on the same day), in most cases the contribution of non-friends is necessary to push a story up to the promotion threshold.

Table 1 shows the ratio of friends and non-friends active on a story both before and after the promotion for all stories that became popular within the Digg network. Here, two distinct groups emerge. In about 54% of all cases, a story was marketed predominantly by friends, although a contribution of non-friends (28%) was necessary for the story to reach critical mass. Figure 6(a) shows this aggregated pattern for an example story in this class. In the remaining cases (46%), stories were spread and digged predominantly by users outside the submitter’s friendship network. Figure 6(b) shows a typical example for this pattern. Once the promotion threshold is crossed, both types of stories are

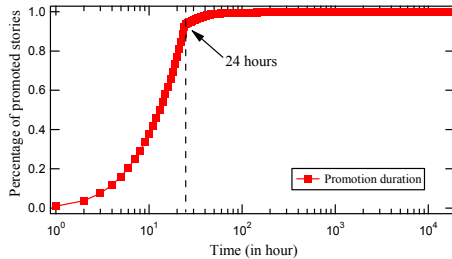


Fig. 4. Stories have to gain enough momentum within 24h to be promoted to the front pages.

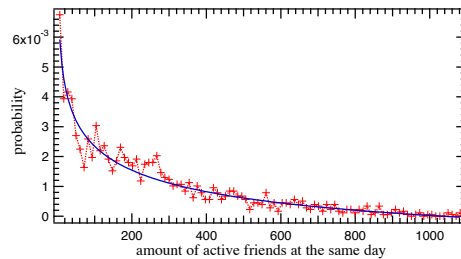


Fig. 5. The probability for a given number of friends to be active within 24h decreases rapidly.

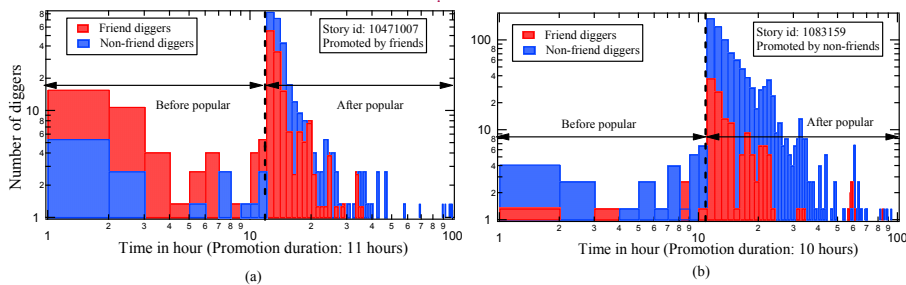


Fig. 6. Propagation pattern of a popular story since publication. (a) The story is promoted by friends. (b) The story is promoted by non-friends. (log-log scale)

read more by non-friends, as the quantity is usually significantly larger and the contribution of the submitter’s friendship network may already be exhausted.

5 The Criticality of Individuals

The successful spread of information cannot be explained directly from the social ties inside our investigated online social network, neither through the relationships among individual friends nor from the usage and outreach of users into their friendship network. This naturally raises the question whether all users are equal inside the network, or whether there are some individuals in the social community (a) who themselves have better (or earlier) access to important content and are therefore able to get a high number of popular stories, (b) can use the friendship network more efficiently, act as motivators able to overproportionally recruit friends, or (c) able to early on spot content that will later resonate with the masses and become a hit. These questions will be the focus of this section.

There exist a number of ways to define the importance or criticality of individuals in networks. In complex network theory and social network analysis, importance is typically defined from a structural perspective, using topological metrics such as node degree or betweenness [19], which measure how well a particular node is connected to its surrounding peers and how many possible communication paths between nodes in the network will traverse this node. Using this definition of importance, most studies of online social networks find a small number of topologically critical nodes [12, 20, 13], resulting from the power-law degree distribution of these complex networks: there exist a few well-connected nodes with whom a large number of users are friends. In our analysis, we confirm these findings and will (for now) use this definition of critical individuals.

Contrary to other OSNs however, we do not only observe a skewed distribution in the degree and connectivity of nodes, but also in the symmetry of relationships among users. While most OSNs show high levels of link symmetry, for example 74% of links in LiveJournal and 79% of links in YouTube are found to be bi-directional [12], the relationships in Digg are less reciprocative (38% on

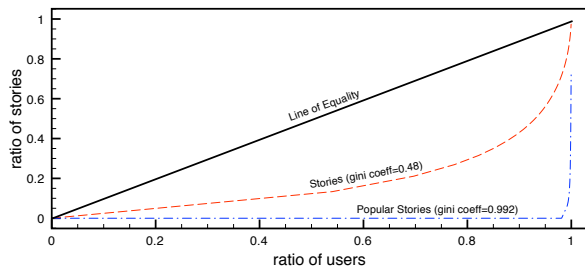


Fig. 7. While a significant share of users submit stories into the system, only the stories of a selected few are reaching critical mass to become popular.

Degree	# Users	Sym. link
[0, 10)	282536	0.53
[10, 100)	49416	0.42
[100, 1000)	13993	0.39
1000	111	0.31

Fig. 8. Fraction of symmetric links in the Digg network

average) and also vary with the degree of the node: the more connections an individual B already has, the less likely it is to match an incoming new friendship request from A . In Digg, A thus becomes a “fan” of B , thereby receiving notifications about the activities of B , but not vice versa.

This observation of decreasing reciprocity is consistent with sociological theory and ethnographic studies of social networks which showed that friendship requests in OSNs are often driven by users’ interests to become passively informed by means of these social ties [8, 10]. The fact that the average symmetry is significantly lower and also dependent on the degrees of remote nodes, underlines (a) that users are engaging in friendships in Digg with the intention of information delivery and (b) the existence of individuals which act (or view themselves) as sources and broadcasters of knowledge, which according to [16] would embody the critical influentials in the network.

5.1 Submitting Successful Stories

When looking at all stories submitted in the past 4 years, we indeed find that users are very different when it comes to generating and sourcing important information. While the content published on Digg is followed about 25 million visitors a month, only a limited number of registered users are actively submitting content. The activity patterns of these users is furthermore biased, as shown in the Lorenz curve [21] in Fig. 7: the 80% least active users of the network are together submitting only about 20% of the entire content as indicated by the dashed red line. While this is far from an equitable system, the same skew – commonly referred to as the 80-20 or Pareto rule – has been found repeatedly in economics and sociology. Here it is however more drastic when only considering those stories that gained enough support and were promoted to popular. As the figure shows, these successful stories can be attributed to a selected minority of 2% of the community, which is able to find and submit 98% of all stories.

This effect is however not the result of the pure quantity. In other words, there exists no statistically significant relationship between the number of stories a person has submitted and the ratio of stories that will become popular

(Pearson’s correlation coefficient $r^2=-0.01$). While the presence of such a highly skewed distribution seems to suggest the existence of a few “chosen ones”, a closer inspection reveals that these highly successful submitters are also not those users responsible for the effective spread of information. First of all, the average ratio of popular to submitted stories of the top 2% successful submitters is only 0.23. Second, the group of users who rank among the top successful members of the community is highly volatile, and the set of successful users changes substantially between studied time intervals. As we do not find a significant number of members who are able to continuously repeat their previous successes, we may conclude that there exists no conceptual difference or strategic advantage with those who do score successful stories. It appears that they were simply in the right place at the right time.

In conclusion, it is not predominantly the well-connected nodes that are the originator of wide-spreading content, as there is no significant relationship between a user’s success ratio and its degree with those around it (p-value>0.5).

5.2 Activation of the Social Network

While there do not exist any particular nodes that are overly successful in injecting content, it may be possible that there exist users highly successful in activating their friendship network, and therefore would be a key component in helping stories reach widespread popularity.

It turns out that the activation ratio of a node’s direct friends is surprisingly low. On average, a particular node is only able to generate 0.0069 clicks per friendship link. This recruitment is furthermore quite stable with the structural properties of the network nodes. While the literature predicts that nodes in a social network achieve an exponentially increasing influence compared to their importance [16, p. 124], we find a linear relationship ($r^2=0.76$) between the size of a nodes’ friendship network and the amount of users it can recruit to click on a story. As the slope of the linear regression is low ($a=0.102$), there is no overproportional impact of higher-degree nodes: 1 activated user with 100 friends is on average about as effective as 10 activated users with 10 friends.

While we find no *quantitative* difference in the friendship network around the important nodes, there may be a *qualitative* difference in terms of structural characteristics and the information propagation along links. As complex networks evolve, certain growth processes such as preferential attachment [22] create sets of highly connected clusters, which are interconnected by fewer links. According to social network theory [11, 23], these links among clusters, commonly referred to as “weak ties”, act as a critical backbone for information propagation. Information within a cluster is communicated and replicated between nodes thereby creating high redundancy, while the weak ties transport other, previously unknown information between groups of nodes (see Fig. 9(a)).

To evaluate this hypothesis, we classified the network into weak and strong ties according to their edge betweenness and compared their theoretical importance to the actual amount of content that was propagated between each pair of nodes. Figure 9(b) shows a Lorentz curve of the betweenness and the actual

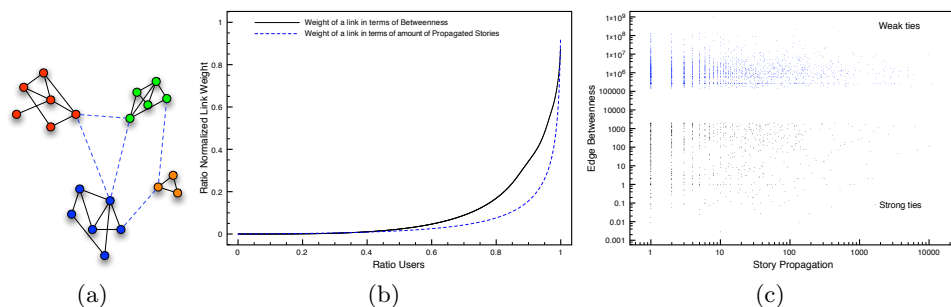


Fig. 9. While according to the weak ties hypothesis, (fig. a) the links connecting different clusters and communities (high edge betweenness) are critical to the spread of information and the topology and usage patterns of the social network showed similar network characteristics (e.g. the weight/degree distribution in fig. b), there existed no relationship between the strength of the tie and the amount of information propagated, neither for the entire network nor for the subclasses of strong and weak links (fig. c).

information conductivity, demonstrating that the distributions are in general comparable. As there is no hard threshold for what characterizes a weak or strong tie, we compared the top and bottom 20% of the distribution as weak and strong ties respectively to the amount of stories propagated along a certain link. As shown in Fig. 9(c), there exists no relationship ($r^2 = 0.00006$), thus information is not propagated more effectively along weak ties.

5.3 Early Predictors

Finally, we investigated if there exist certain individuals who might be called important and influential in the sense that they are able to early-on identify content that will later on become popular (see for example [16]). In the months of April-May 2009, we followed the voting patterns of all registered users on all stories to determine how successful users were in finding and clicking on content that within the next hours or days would become popular. Of all activity within this two month time period, users identified and reacted on average only to 11.9% of content before it got promoted. With the absence of any high performers, there are thus no specific individuals who are able to consistently and repeatedly find emergent trends. This observation did not change either for the case of the high degree individuals or the users with a high success ratio of submitting successful stories; there exists no statistically significant difference in their ability to find content in the social network before it actually reaches widespread popularity.

6 Beyond Pure Friend Relations

The discussions in section 4 and 5 show that neither the importance of individual users nor the dynamics of the individual friendship relations or the network of

friends can solely explain if a certain story will become a success. Furthermore, as in nearly 50% of all stories the promotion process took place without any dominant contribution by the friendship network, we further investigated how the low participation values of the friendship network may be explained and which features are the dividing force between those stories pushed by friends and those promoted by the general public.

6.1 Spread Without Friends - A Matter of Timely Relevance

To investigate why one story is propagated by friends while another one is pushed by random users, we conducted a controlled experiment and presented the 158 most successful stories in the last year to a group of non-experts. In the experiment, stories were displayed to participants using the same user interface as on digg.com, except that only one story was displayed at a time to eliminate distractions. As we could in retrospect classify these stories as promoted by friends or non-friends, the stories in the experiment were balanced in terms of topic areas and to mimic a similar distribution as on Digg. Looking at the title, description, link, digg count and if available preview image, we asked participants to rate each story in terms of general appeal, their own personal interest and the general importance of a particular story.

From the experiment it became evident that the difference between friends and non-friends promotion was a direct result how important and relevant the participants rated a particular story. Whenever participants marked a story as being “of general interest to the public”, in other words it is likely that one would hear it in the evening news, or attributed it with a high level of timely relevance, the same story has also reached popularity on Digg by non-friends. Thus, whether stories will become friend or non-friend promoted seems to be a function of a stories’ content and appeal (both stat. significant at $p=0.05$).

6.2 Explaining Critical Mass Through Temporal Alignment

As a large number of factors previously assumed to be of importance to information spread turned out in our study of Digg to be rather insignificant and highly volatile over time, we further investigated the influence of time on the story propagation process. We found that some of the unexpected low or highly fluctuating factors are to some extent dependent upon the temporal alignment of users, i.e., whether users in general (and friends in particular) are visiting the site within the same narrow time window or not.

Figure 10 visualizes this idea of temporal alignment on a snapshot of the front pages from April 2009, which shows the position of all popular stories with at least 100 diggs over a 48 hour time interval on the first 50 front pages. There exists a high flux in the amount of stories passing through as within on average 3 hours the entire content on the first front page has been replaced by newer items. From a combined analysis of voting patterns and front page traces, we are able to determine the usual search strategy and search depth of users on Digg. Stories accumulate 80% of their attention received after promotion on the first

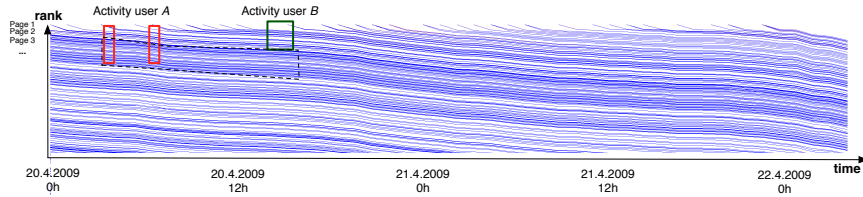


Fig. 10. The high turnover rate of even the popular stories and the limited attention span and activity period of users can offer an explanation of the low importance of friendships. The intensity of a line indicates the rate of diggs a story is accumulating.

and second page only, while the ratio of users who are scanning more than the first 4 front pages is practically zero.

Considering the case of two users active on 20.4.2009, this can explain the surprisingly low amount of common friendship activations, as nearly 70% of the stories visible to user *A* during the two morning visits are already outside of user *B*'s attention window as the user visits the site just six hours later. Unless *B* actively looks for and follows up on *A* activity, the abundance of content and high turnover rate of information combined with limited attention span will therefore largely limit the potential for commonality. This demonstrates that whether a story reaches critical mass depends to a significant amount upon who and how many people are currently active on the site within a short time window.

A combination of this temporal perspective with interest and friendship data can go a long way and provide a much more detailed understanding of user behavior, as we were able to improve our analysis accuracy of the activation ratio of friendship links by a factor of 15. Note however that while a temporal view is currently able to reveal in retrospect why certain users clicked on a particular story, it is not yet possible to predict how users will interact on a story in the future for a variety of reasons. Most importantly, an accurate prediction will require a good model of users' future activity periods at a fine enough resolution to minimize the prediction error of which stories users will see. Furthermore, it will be necessary to further understand the concrete decision process that will lead to a users actively clicking on a story.

7 Conclusion

In this paper, we have evaluated the assumption made in OSNs that friendship relations are the critical factor for information propagation. While we find evidence that friendships are formed based on common interests, the actual effectiveness is surprising low and does not confirm the high importance attributed to them – at least in digg.com. We furthermore notice that although there exists a significant skew in the characteristics of network nodes from a topological perspective, we do not find any evidence that some users are more effective in terms of spreading information. They have no better access to information, are not more efficient in triggering their friends nor do predict trends better. Vari-

ous outcomes of our analysis point to a factor that in the past has not received sufficient attention: time period overlaps. We find that when incorporating this factor, the conductivity of friendships and our ability to explain the spread of information improves manyfold. This will be the focus of future research.

References

1. M. Richardson and P. Domingos, "Mining knowledge-sharing sites for viral marketing," in *ACM SIGKDD*, 2002.
2. W.-S. Yang, J.-B. Dia, H.-C. Cheng, and H.-T. Lin, "Mining social networks for targeted advertising," in *39th Hawaii International Conf.*, 2006.
3. J. Surowieck, *The Wisdom of Crowds*. Anchor, 2005.
4. J. Davitz, J. Yu, S. Basu, D. Gutelius, and A. Harris, "ilink: search and routing in social networks," in *ACM SIGKDD*, 2007.
5. D. M. Boyd and N. B. Ellison, "Social network sites: Definition, history, and scholarship," *J. of Computer-Mediated Communication*, vol. 13, no. 1, p. 210, 2007.
6. E. Katz and P. F. Lazarsfeld, *Personal Influence*. Free Press, 1955.
7. D. Koller, "Representation, reasoning and learning," 2001.
8. D. Fono and K. Raynes-Goldie, "Hyperfriends and beyond: Friendship and social norms on livejournal.," *Internet Research Annual*, vol. 4, 2006.
9. K. Raynes-Goldie, "Pulling sense out of today's informational chaos: Livejournal as a site of knowledge creation and sharing," *First Monday*, vol. 8, no. 12, 2004.
10. N. B. Ellison, C. Steinfield, and C. Lampe, "The benefits of facebook "friends:" social capital and college students' use of online social network sites," *Journal of Computer-Mediated Communication*, vol. 12, pp. 1143–1168, July 2007.
11. M. Granovetter, "The strength of weak ties," *Am. J. of Sociology*, vol. 78, 1973.
12. A. Mislove, M. Marcon, K. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *IMC Conference*, 2007.
13. J. Leskovec and E. Horvitz, "Planetary-scale views on a large instant-messaging network," in *WWW2008*, 2008.
14. M. Cha, A. Mislove, and K. P. Gummadi, "A measurement-driven analysis of information propagation in the flickr social network," in *WWW2009*, 2009.
15. M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon, "Analyzing the video popularity characteristics of large-scale user generated content systems," *IEEE/ACM Transactions on Networking*, 2009.
16. E. B. Keller and J. Berry, *The Influentials: One American in Ten Tells the Other Nine How to Vote, Where to Eat, and what to Buy*. The Free Press, 2003.
17. Y. Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong, "Analysis of topological characteristics of huge online social networking services," in *WWW2007*, 2007.
18. L. Rüschemdorf, "The wasserstein distance and approximation theorems," 1985.
19. J. P. Scott, *Social Network Analysis: A Handbook*. Sage, 2000.
20. F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida, "Characterizing user behavior in online social networks," in *IMC2009*, 2009.
21. M. O. Lorentz, "Methods of measuring the concentration of wealth," *Publications of the American Statistical Association*, vol. 9, no. 70, pp. 209–219, 1905.
22. H. A. Simon, "On a class of skew distribution functions," *Biometrika*, 1955.
23. P. Cserehely, *Weak Links: Stabilizers of Complex Systems from Proteins to Social Networks*. Springer Berlin, 2006.