New Review of Hypermedia and Multimedia,
2011, 1–24, iFirst article

Taylor & Francis
Taylor & Francis Group

Comparative analysis of articulated and behavioural social networks in a social news sharing website

ANDREAS KALTENBRUNNER†*, GUSTAVO GONZALEZ‡,
RICARD RUIZ DE QUEROL† and YANA VOLKOVICH†
†Barcelona Media Centre d'Innovació, Barcelona, Spain
‡Mediapro Research, Barcelona, Spain

(Received 11 June 2010; final version received 15 June 2011)

This study analyses and contrasts the explicit (articulated) and implicit (behavioural) social networks on the Spanish Digg-like social news website meneame.net. The explicit network is given in the form of declared but not necessarily bidirectional friendship links; the behavioural network is extracted from conversations through comments to the shared links. These two directed social networks and their intersection are analysed and described in detail, which leads to some important conclusions about user behaviour on link sharing websites and online conversation habits in general. We find that reply interactions are more likely to occur between non-friends and that these interactions are (if bidirectional) also more balanced in the case of non-friends. A *k*-core decomposition of the networks reveals a fundamental difference in the practice of establishing behavioural and articulated links.

Keywords: Social media; Social network analysis; Link sharing; *k*-Core decomposition

1. Introduction

Recently a considerable amount of research effort has been devoted to the study of social networks from the perspective of complex network analysis. Most of these studies rely on large datasets retrieved from imprints of interactions on social media web sites (e.g. Goh *et al.* 2006, Gómez *et al.* 2008, Kunegis *et al.* 2009, Leskovec *et al.* 2010). These interactions either correspond to explicitly declared connections such as friendship links in a social network or implicit ones retrieved from interactions via e-mail, chat clients, forum comments, etc. This issue has been recently addressed by Danah Boyd[1]: she differentiates between articulated (explicit), behavioural (implicit) and personal ("real") social networks.

It seems evident that the connections between users in the above networks often do not correspond to the same type of social relation. In fact, some

*Corresponding author. Email: andreas.kaltenbrunner@barcelonamedia.org

New Review of Hypermedia and Multimedia
ISSN 1361-4568 print/ISSN 1740-7842 online
http://www.tandfonline.com
http://dx.doi.org/10.1080/13614568.2011.598192

connections may not correspond to any relation at all and an absence of a connection between users in the articulated, e.g. friendship, network may help interpretation of the type of relationship in the behavioural network. Repeated interactions in the behavioural network may also enhance the value of an articulated connection. Therefore, if someone wants to understand the nature of the social interactions on a certain website, it is not sufficient to just look exclusively on one of these two networks.

In this work we present a comparative study between the articulated and behavioural social networks extracted from Menéame,[2] a Spanish social news website. The articulated network corresponds to explicitly declared friendship relations between users, and the behavioural network is extracted from reply relations between user comments. For more details on Menéame and the dataset we used we refer to Sections 1.2 and 2.

In this study, we aim to answer the following questions: What are the structural differences between the behavioural and articulated networks? Is it more likely to find behavioural interactions on link sharing websites between friends or are, on the contrary, reply relations preferentially to be found between non-friends? What can we say about the most active users in these two networks?

We address these questions in the following way. First, we briefly review related literature in Section 1.1 and describe the Menéame platform in Section 1.2. In Section 2 we discuss the principal characteristics of the dataset and explain the construction process of the two social networks. We analyse these networks in Section 3 separately and then in Section 4 uncover a fundamental structural difference between them by performing a k-shell decomposition. In Section 5 we perform a similar analysis as in Section 3 for the subset of users who are connected in both networks. Finally, we extract the core users of every network and analyse the co-occurrence between friendship and reply relations. The conclusions are presented in Section 6.

1.1 Related work

To the best of our knowledge this is the first attempt of a combined analysis of overlapping explicit and implicit social networks. However, there have been several studies which have analysed only one of these two networks for socials news websites and similar contexts.

Zhongbao and Changshui (2003), Matsumura *et al.* (2005) and Goh *et al.* (2006), extracted and analysed the behavioural social network from messages in discussion forums. The same methodology was later applied by Gómez *et al.* (2008) on the discussion forums in Slashdot and by Laniado *et al.* (2011) on the discussion pages of the English Wikipedia. Slashdot (like Menéame) possesses an explicit (articulated) social network of friend (and foe) links between the users. This friendship network was analysed by Kunegis *et al.* (2009), and was later used (together with similar data from Epinions and Wikipedia) to derive a predictive algorithm for the sign (friend or foe) of the relation between two users in Leskovec *et al.* (2010).

Menéame was inspired by the social news aggregator Digg. Launched almost one year before Menéame, Digg has been the subject of several research studies. In particular, Wu and Huberman (2007) defined a model for collective attention based on the popularity of Digg stories and Szabó and Huberman (2008) predicted the number of votes (diggs) of a story after a certain time period. A general model for user behaviour in social news websites can be found in Hogg and Lerman (2009). This model incorporates information about the users' friendship links and allows a more detailed prediction of the evolution of the number of votes.

Lerman (2007) investigated how the number of friendship connections influences the promotion of a link to the front page. The author found a significant correlation between the ratio of successful stories of a user and the number of his/her friends. Another study by Lussier *et al.* (2010) went a step further and investigated not only the influence of the direct neighbourhood of a user but also the number of second- and third-degree friends responsible for the success of the user's stories. The results of Lussier *et al.* (2010) indicate that models for voting behaviour should also incorporate the network surrounding of a user at distances greater than one to assess the probability of success of a link in this type of social news aggregators. To go beyond these results we believe that it is engaging to contrast these articulated social networks with behavioural networks to assess in more detail the value of every friendship connection. Our study would be a first step in this direction.

1.2 What is Menéame?

In this work we focus on data obtained from Menéame, the most successful Spanish news aggregator. The website is based on the idea of promoting user-submitted links to news (stories) according to user votes.

The front page of Menéame consists of a sequence of stories recently promoted to the front page, as well as a link to pages containing the most popular, and newly submitted stories. Registered Menéame users can:

1. publish links to relevant news which are retained in a queue until they collect a sufficient number of votes to be promoted to the front page of Menéame. The promotion of links follows a relatively complex set of rules[3];
2. comment on links sent by other users (or themselves);
3. vote (*menear*) comments and links published by other users;
4. establish friendship connections to other users. The connections can be unidirectional and are similar to follower relations on Twitter.

To measure user reputation Menéame follows a set of rules[4] to calculate a user karma. The main idea is that the users' karma is modified according to the votes their stories and comments receive. To ensure participation of the users, their karma diminishes if they do not vote regularly.

Table 1. Principal quantities of the dataset.

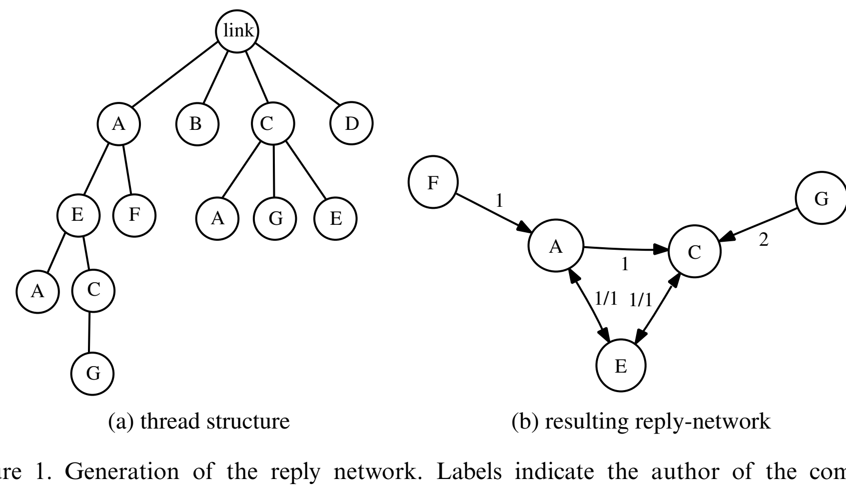
No. of links sent	625,995
No. of links published	59,303
No. of comments	4,528,424
No. of users in database	120,315
No. of users commenting	53,879
No. of users in reply network	33,194
No. of users in friendship network	8,673

2. Dataset and network construction

We start with a brief description of the main characteristics of the dataset. Later we explain how we extracted the social networks used in our analysis.

The dataset covers the time span between 7 December 2005 (the date when Menéame was launched) and 14 July 2009. It contains approximately 625,000 links which received about 4,500,000 comments; 59,000 of those links received enough votes to be promoted to the front page. Out of slightly more than 120,000 users in the database, around 54,000 users wrote at least one comment. Approximately 33,000 of them replied to comments or received at least one reply. These users form the behavioural social network (See Table 1 for the exact numbers mentioned in this paragraph).

To construct the behavioural reply network we use the structure of the conversations and the implicit relations between the author of a comment and the users who reply to it. Menéame does not have an interface for nested comments, e.g. explicit discussion threads. However, comments with "#*n*" in the text reply to the *n*-th comment in the comment list. We use this to extract the reply relations. When a user *A* replies to a comment of user *B* we draw a directed link between these two users. This mechanism is illustrated in Figure 1. There the reply network of a small example thread [Figure 1(a)] is



(a) thread structure

(b) resulting reply-network

Figure 1. Generation of the reply network. Labels indicate the author of the comment. The link author is not considered for the network. (a) Thread structure; (b) resulting reply-network.

transformed into the corresponding reply network [Figure 1(b)] with two bidirectional edges between two pairs of users with mutual replies and three edges with a weight of one or two. The latter reflects the two replies of user G to user C. Note that we do not consider the author of the link in the generation of the reply network. We also remove 11 users from the reply network whose only links are self-replies.

The friendship network, on the other hand, consists of approximately 9,000 users who have at least one friendship link (incoming or outgoing). Note that the friendship network of Menéame allows negative edges [similar to the ones studied in Kunegis *et al.* (2009) and Lescovec *et al.* (2010)]. Such edges correspond to ignored users, whose comments are hidden to the ignoring user. However, this feature is used only seldomly. In this study we therefore omit these links. This reduces the above reported numbers by 319 users and 1,578 links.

We have defined the articulated and behavioural networks that we use in this study. Note that both networks are directed: neither friendship relations have to be bidirectional, nor a user who has received a reply to one of his/her comments will necessarily answer the reply with another comment. In the next two sections we will study these networks separately, and in Section 5 we continue with an analysis of various combinations of these networks.

3. Social network analysis

In this section we present the results of the analysis of the above-defined social networks. We calculate several fundamental quantities, which are listed in Table 2. For the corresponding definitions we refer to Newman (2003) unless stated otherwise.

Table 2. Global measures of the two social networks analysed.

Variable	Friendship network	Reply network
No. of nodes with edges, N	8,673	33,194
No. of nodes with in-degree ≥ 1	7,854	30,477
No. of nodes with out-degree ≥ 1	4,458	23,041
No. of edges, M	56,676	964,489
No. of nodes in giant component, G	6,926 (79.9%)	30,405 (91.6%)
Mean distance, l	3.78 (0.97)	3.10 (0.62)
Maximal distance, D	11	8
Clustering coefficient, C	0.085 (0.19)	0.124 (0.19)
Mean in-degree, $\langle d_{\mathrm{in}} \rangle$	6.53(22.76)	29.06 (97.76)
Mean out-degree, $\langle d_{\mathrm{out}} \rangle$	6.53 (28.17)	29.06 (105.65)
Network density, $\hat{a}$	7.53×10^{-4}	8.75×10^{-4}
Reciprocity, ρ	0.67	0.35

Values within parenthesis indicate standard deviation (or percentage of nodes in the giant component).

3.1 Giant component

As in Gómez *et al.* (2008) we consider here *weakly* connected components, i.e. two vertices are connected and in the same component if there exists a directed path between them in at least one of the two possible directions. Both networks are highly connected: in the friendship network the giant component, i.e. the largest connected sub-network, contains 80% of all nodes, and this percentage raises even to 91.6% for the reply network. Some of the measures discussed below are calculated only for the giant component.

3.2 Average and maximal distances

The average distance between any pair of nodes is very short in both networks, 3.8 for the friendship network and 3.1 for the reply network. This indicates that two arbitrary users need on average to pass only over two or three other users (intermediaries) to reach each other. However, in extreme cases it may take up to 10 intermediaries in the friendship network (or eight in the reply network). This quantity plus one is referred to as the maximum distance D (i.e. the diameter of the network) in Table 2. These results are in agreement with the results on other traditional social networks, see for example Newman (2003).

3.3 Degree distributions and density

The in- and out-degree distributions of both networks show typical heavy-tailed behaviour (Sigman 1999, Resnick 2007). We plot these distributions in Figure 2 together with corresponding fits with truncated log-normal (LN) distributions. Kolmogorov-Smirnov tests accept the fits with high p-values (shown in the figure legends) for the in- and out-degree distribution of the friendship network. However, in the case of the reply network these fits only give a good visual approximation but are rejected by the statistical test as an explanation for the data. We also tried to adjust power law distributions (with cut offs) to the data, but the results were worse than the ones presented here with LN-distributions.

We find a large difference in the average number of in- and out-degrees ($\langle d_{\text{in}} \rangle$ and $\langle d_{\text{out}} \rangle$). It is larger in the reply network (where it takes a value of 29) than in the explicit friendship network (where the users on average have only 6.5 declared friends). Clearly, the interaction via comments is more common in Menéame than the practice of establishing friendship links. It is also interesting to note that in both cases the standard deviation of the out-degrees is larger than the corresponding statistic of the in-degrees. From the cumulative distributions of both networks (see the right column of Figure 3) we also observe that the proportion of users with only one in-link is larger than the corresponding proportion of users with only one out-link. Obviously, a user is less likely to receive more than one link (be it friendship or reply links) than to establish more than one. This difference is the main cause of the larger standard deviations of the out-degrees in both networks. It is also

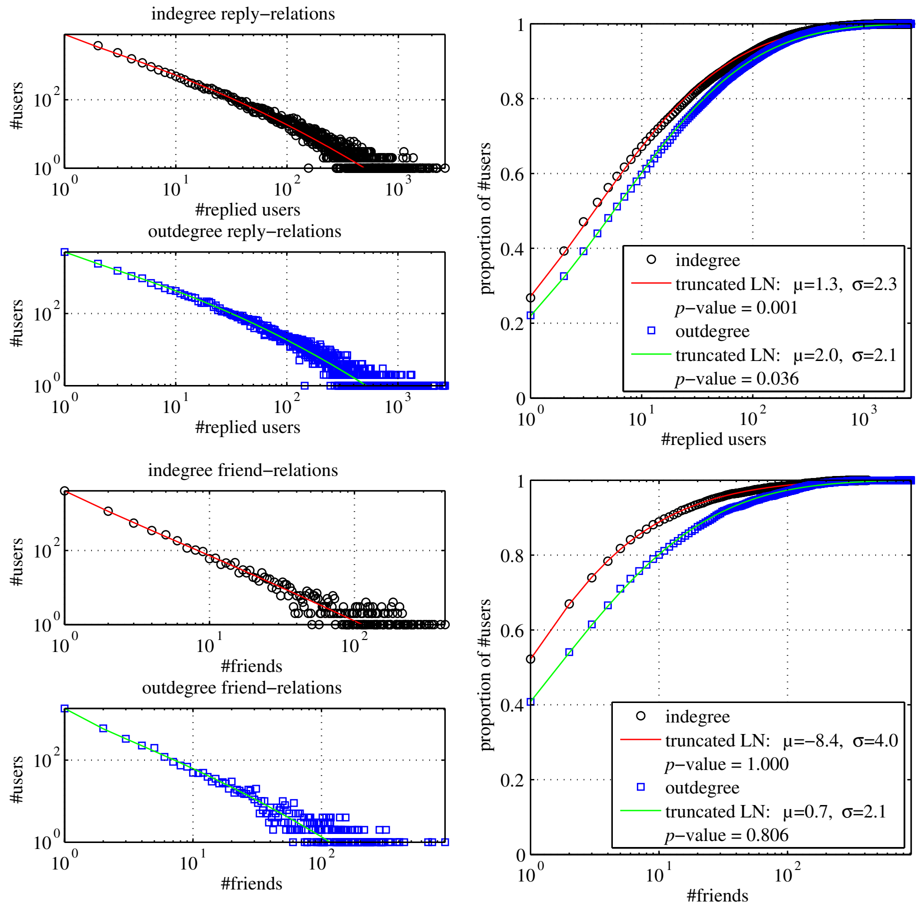


Figure 2. In- (black circles) and out-degree (blue squares) distributions of the reply (top sub-figures) and friendship (bottom sub-figures) network. The solid lines represent truncated LN fits of the data.

worth mentioning that the number of link-receptors is larger than the number of link initiators in both networks (by about 75% in the friendship network and 32% in the reply network) as can be seen from the values in Table 2.

3.4 Clustering coefficients

The clustering coefficient C allows to measure the level of cohesiveness of the network. Here we use its directed version, which is defined as the probability that two nodes with a common neighbour are connected. It is equal to the average over the individual clustering coefficients C_i, which we calculate by extending the definition of Watts and Strogatz (1998) to directed networks. We use: $C_i = \frac{|E_i|}{k_i(k_i-1)}$, where k_i is the sum of in- and out-degrees of node i and E_i the set of edges between the direct neighbours of node i. The clustering coefficients are similar in magnitude in the two networks, being slightly higher in the reply network (0.085 in the friendship network vs. 0.124 in the reply network) but with higher standard deviation in the friendship network. Moreover, the C's are similar in magnitude in other social networks and much larger than in comparable random networks (Newman 2003). We notice, however, a more clustered community in Menéame than in Slashdot, for which smaller clustering coefficients are reported (0.056 for the friendship network with negative edges (Kunegis *et al.* 2009) and 0.027 for the reply network (Gómez *et al.* 2008)).

3.5 Link reciprocity

To study the proportion of bidirectional edges in the network we calculate the link reciprocity ρ. Instead of using the simple ratio between bi- and uni-directional edges which would depend on network size and link density, we follow the definition of Garlaschelli and Loffredo (2004). The definition uses the adjacency matrix ($a_{ij}=1$ if there exists a connection from user i to user j) and states:

$$\rho = \frac{\sum_{i \neq j} (a_{ij} - \hat{a})(a_{ji} - \hat{a})}{\sum_{i \neq j} (a_{ij} - \hat{a})^2},$$

where $\hat{a}$ is the network density,

$$\hat{a} = \frac{\sum_{i \neq j} a_{ij}}{N(N-1)}, \tag{1}$$

and N the number of nodes in the network.

We find that the friendship network has a reciprocity which is almost twice as large (0.67 vs. 0.35) than the one of the reply network. Apparently, the social pressure of responding to an incoming friendship link is much stronger than the one of replying to a comment. The reciprocity of the Menéame reply network is still slightly larger than the Slashdot equivalent, where a value of 0.28 was obtained (Gómez *et al.* 2008).

3.6 Mixing coefficient

The mixing coefficient, or degree correlation, r allows to detect whether highly connected nodes preferentially link to other highly connected nodes (r is positive in this case). Such a phenomenon is known as assortative mixing, and is present in many social networks (Newman and Park 2003). If, on the contrary, the highly connected nodes link to less connected ones, we speak of dissortative mixing ($r<0$). This can be found, for example, in food webs or in the Internet (Newman 2002). The intermediate case with absolute values of r close to 0 is referred to as neutral assortativity.

The above-cited studies use undirected networks to calculate r. Here we follow a novel approach for computing assortativity in direct networks which has recently been proposed by Foster *et al.* (2010). To account for the edge direction the authors introduce a notation in which $\alpha, \beta \in \{\text{in}, \text{out}\}$ are used to index the degree type. For each edge e, i_e^α and j_e^β are the α- and β-degree of the source node and the target node. A set of four assortativity measures is defined using the Pearson correlation:

$$r(\alpha, \beta) = \frac{M^{-1} \sum_{e \in E} [(i_e^\alpha - \bar{i}^\alpha) * (j_e^\beta - \bar{j}^\beta)]}{\sigma^\alpha \sigma^\beta}$$

where E is the set of edges in the network and M its cardinality, $\bar{i}^\alpha = M^{-1} \sum_{e \in E} i_e^\alpha$, and

$$\sigma^\alpha = \sqrt{M^{-1} \sum (i_e^\alpha - \bar{i}^\alpha)^2}$$

$\bar{j}^\beta$ and σ^β are analogously defined.

The absolute values of the correlation coefficients are dependent on the network degree distribution; to compute statistical significance, the degree–degree correlations are compared with those of an ensemble of 100 randomised networks with the same in- and out-degree sequence as the original network. These networks are obtained using the configuration model (Bender and Canfield 1978, Molloy and Reed 1998, Newman 2003). The statistical significance of each correlation $r\ (\alpha, \beta)$ is computed as the difference between the value observed in the original network and its average in the randomised ensemble $\bar{r}_{\text{rand}}(\alpha, \beta)$ in units of the standard deviation $\sigma_{\text{rand}}(\alpha, \beta)$:

$$Z(\alpha, \beta) = \frac{r(\alpha, \beta) - \bar{r}_{\text{rand}}(\alpha, \beta)}{\sigma_{\text{rand}}(\alpha, \beta)}.$$

Values of $|Z| > 2$ can be considered statistically significant. As a last step, as Z scores are dependent on the network size, they are normalised by defining an assortative significance profile (ASP) for each network:

$$\text{ASP}(\alpha, \beta) = \frac{Z(\alpha, \beta)}{\sqrt{\sum_{\alpha,\beta} Z(\alpha, \beta)^2}}.$$

The sign of Z or ASP indicates whether the networks are more (Z-assortative) or less (Z-dissortative) assortative than expected given their degree distribution.

The results of this analysis for the friendship and the reply networks are shown in Table 3 and Figure 3 together with those of variants of these networks which will be analysed in subsequent sections. We find that both the friendship (blue line with pentagrams) and the reply networks (green line with diamonds) are Z-assortative, showing the largest ASP for r (in, in). The ASP profile is more balanced in the case of the friendship network while the Z-assortativity in the reply network is larger for (in, out) and (in, in) than for (out, in) and (out, out). This implies that users show a preference to reply to other users with in- or out-degrees similar to their own in-degree, while their out-degree is less determining in this behaviour.

Table 3. *Z*-scores of directed assortativity and ASP profile of the six networks analysed.

Network	(α, β)	$r\ (\alpha, \beta)$	$\bar{r}_{\text{rand}}$	σ_{rand}	$Z\ (\alpha, \beta)$	ASP
Friendship	(out, in)	−0.088	−0.148	1.4×10^{-3}	41.0	0.48
	(in, out)	0.001	−0.102	2.7×10^{-3}	38.3	0.45
	(out, out)	−0.084	−0.118	1.6×10^{-3}	21.9	0.26
	(in, in)	0.038	−0.130	2.8×10^{-3}	60.7	0.71
Replies	(out, in)	−0.117	−0.120	3.6×10^{-4}	7.2	0.21
	(in, out)	−0.099	−0.106	3.8×10^{-4}	18.2	0.53
	(out, out)	−0.110	−0.112	3.6×10^{-4}	4.9	0.14
	(in, in)	−0.102	−0.113	3.8×10^{-4}	27.92	0.81
Friendship reduced	(out, in)	−0.085	−0.145	1.7×10^{-3}	35.9	0.42
	(in, out)	0.005	−0.102	2.6×10^{-3}	40.5	0.47
	(out, out)	−0.083	−0.118	1.9×10^{-3}	18.4	0.22
	(in, in)	0.038	−0.128	2.6×10^{-3}	63.9	0.75
Replies reduced	(out, in)	−0.132	−0.141	5.6×10^{-4}	16.4	0.38
	(in, out)	−0.108	−0.121	6.3×10^{-4}	20.6	0.48
	(out, out)	−0.122	−0.131	6.2×10^{-4}	13.9	0.33
	(in, in)	−0.114	−0.131	5.7×10^{-4}	30.6	0.72
Intersection	(out, in)	−0.054	−0.141	4.7×10^{-3}	18.3	0.45
	(in, out)	−0.009	−0.118	5.3×10^{-3}	20.6	0.52
	(out, out)	−0.051	−0.129	5.2×10^{-3}	15.1	0.38
	(in, in)	−0.007	−0.130	5.0×10^{-3}	24.7	0.62
Difference	(out, in)	−0.140	−0.129	7.1×10^{-4}	−15.3	−0.64
	(in, out)	−0.117	−0.110	6.7×10^{-4}	−10.7	−0.45
	(out, out)	−0.128	−0.119	6.9×10^{-4}	−13.86	−0.57
	(in, in)	−0.124	−0.120	6.9×10^{-4}	−6.3	−0.26

Note: All values are significant ($|Z > 2|$).

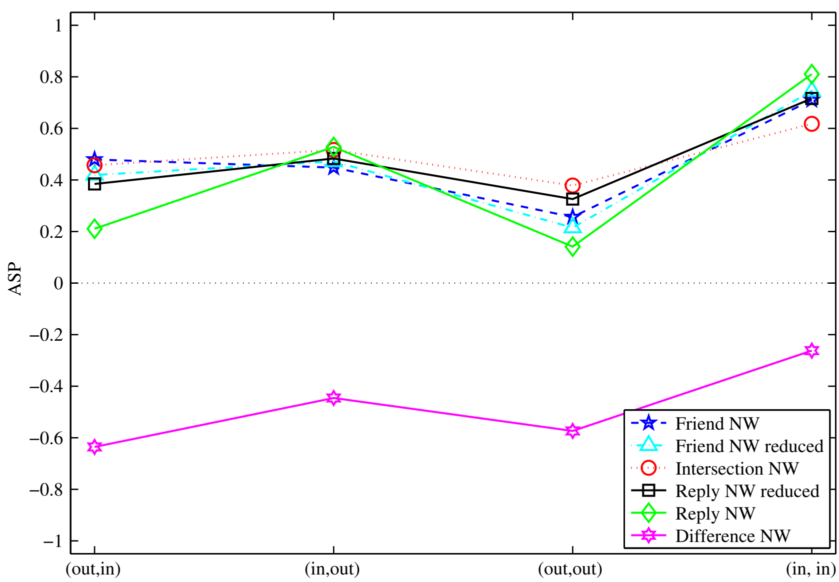


Figure 3. Comparison of the ASPs of the six networks analysed. All values are significant ($|Z| > 2$).

4. *k*-Core decomposition

In this section we further analyse the structure of the reply and friendship networks via k-core decomposition. This allows us to detect significant structural differences between the networks.

The notion of k-core decomposition of a network was introduced by Seidman (1983) and has been used to study various topics in biology or computer science (see Alvarez-Hamelin *et al.* (2008) and the references therein). Recently, Kitsak *et al.* (2010) showed that the k-core decomposition is more suitable for detection of the most efficient information spreaders in a social network than the degree or the centrality of the users.

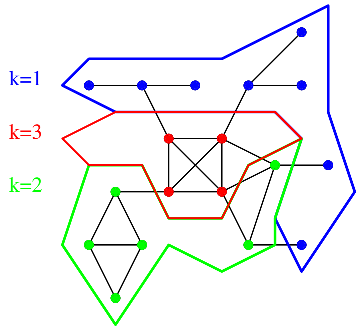
In simple words, the *k-core* of a graph is the maximal sub-graph in which each vertex is adjacent to at least k other nodes of the sub-graph. Formally, for a given graph $G = (V, E)$ we define a sub-graph $H = (C, E|C)$ induced by the set $C \subseteq V$. H is a k-core of G if and only if for every vertex $v \in C : \deg_H(v) \geq k$, and H is the maximum sub-graph which fulfils this condition. With $\deg_H(v)$ we denote the degree[5] of the vertex v in the sub-graph H. A vertex v has shell index k if it belongs to the k-core but not to the $(k+1)$-core. The set of all vertices with shell indexes equal k is called the k-shell.

An example for the k-core decomposition of a small undirected graph is presented in Figure 4(a). We start with $k = 1$ and peel away all nodes with degree 1. We repeat this removing procedure iteratively until no vertices with degree 1 remain. Next, we assign to all removed nodes shell index 1. We continue with the same procedure for $k = 2$ and obtain vertices with shell indexes 2, and so on. We stop when we have removed the last node from the network at the $k_{\max}$-th step. The variable $k_{\max}$ is then the maximum shell index of the graph ($k_{\max} = 3$ in our example).

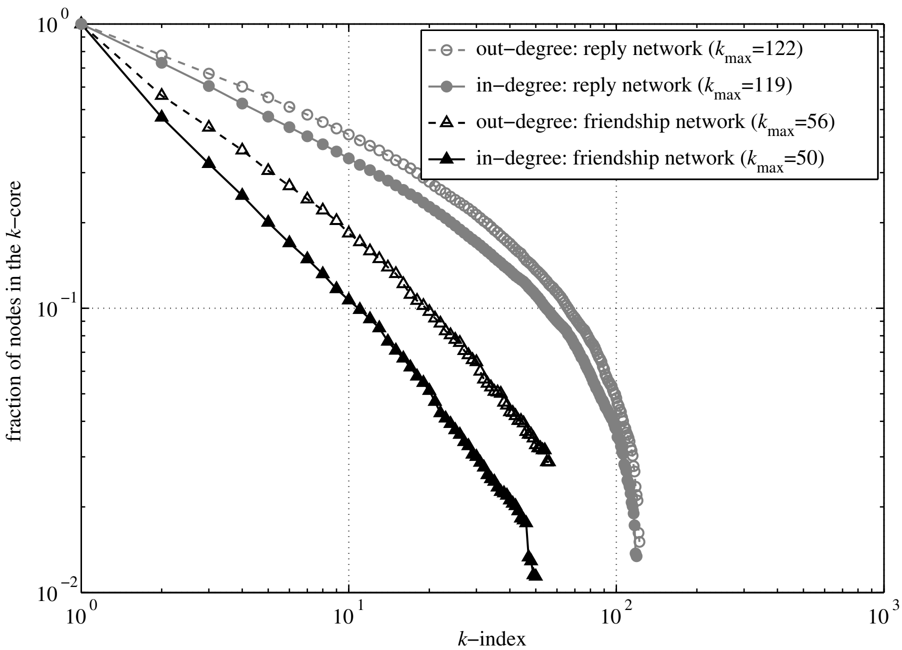
We will calculate the in- and out-degree k-core decomposition of the friendship and reply networks. In Figure 4(b), we present the proportion of users in the k-cores for these networks and the in- and out-degree variant of the k-core decomposition. All these curves expose heavy-tailed behaviour.

If we compare the in- and out-degree k-shell indexes of every user with its in- and out-degrees in the corresponding networks we obtain the characteristic scatter-plots (black and grey circles or squares) as shown in Figure 5. We compare them with the equivalent plots of randomised versions of these networks, obtained using again the configuration model. The randomisation preserves the in- and out-degrees of every user but otherwise shuffles the connections between the users.

The scatter-plots of the randomised friendship networks (coloured dots in the top plots of Figure 5) have a significantly lower $k_{\max}$ value. This implies that the structure of the friendship network is significantly different from a randomly generated network. In the randomised version the correlation between k-index and degree is larger, therefore users with larger degree are more likely to end up with larger k-index as well. We suspect that these observations may be caused by the fact that the users in the friendship network do not connect randomly to other users, rather they choose the



(a) An example of the k-core decomposition

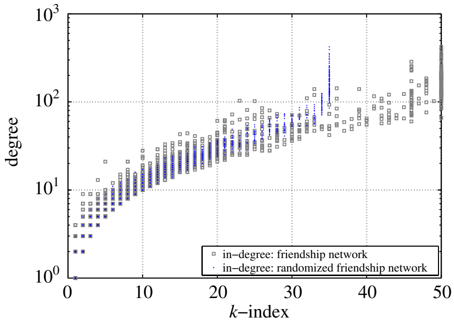


(b) The fraction of nodes in the k-shells versus the value of k.

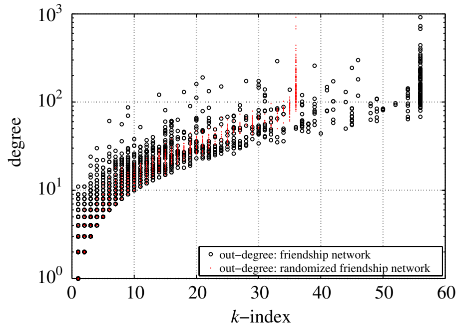
Figure 4. k-shell decomposition. (a) An example of the k-core decomposition; (b) The fraction of nodes in the k-shells vs. the value of k.

connections with some criteria and end up in a larger cluster where nearly everybody is connected to everybody (see also Figure 7(a)). Many users try to be connected to the top users which explains why the out-degree of the friendship network shows an even greater difference between the randomised and the actual k-cores.

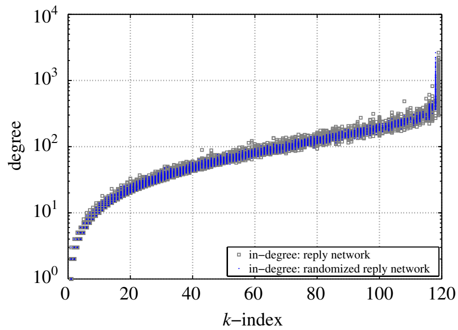
We also observe many upper outliers in the original friendship network (black dots in Figure 5), nodes whose degree is larger than expected given their k-index. They lie above the region marked by the coloured dots. We suspect that such users are outside hubs, i.e. they are not friends with many of



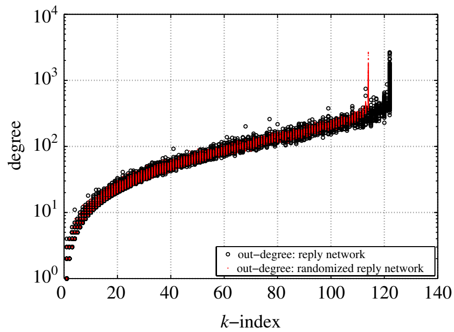
(a) in-degree friendship network



(b) out-degree friendship network



(c) in-degree reply network



(d) out-degree reply network

Figure 5. Scatter plots for k-index vs. degree for real and randomised networks. (a) In-degree friendship network; (b) out-degree friendship network; (c) in-degree reply network; (d) out-degree reply network.

the most connected users and they have their own "friendship domain". For example, when somebody prefers to connect to real-life friends rather than to Menéame gurus.

The picture in the reply network is very different (bottom plots in Figure 5). The scatter-plots of the randomised networks (coloured dots) are more similar to the ones of the original reply network. The users reply less distinctively to the comments of other users, which creates a higher correlation between the k-index and the degrees in these networks and more similar values of $k_{\max}$ between original and randomised networks.

Similar results to the ones described in this section are also observed if undirected equivalents of the networks are used to calculate the k-index (data not shown).

5. Combined analysis of explicit and implicit networks

In the previous sections we have seen that the behavioural and articulated networks are very different in size. This happens due to the fact that many Menéame users do not use the feature which allows to establish friendship connections. To overcome this limitation of our dataset we continue with an analysis of the subsets of the reply and friendship network generated only by the users which are active in both networks. More precisely, we focus on the subset of users with at least one link in both networks and the connections between these users. Such a joint network contains 7,196 users.[6]

5.1 Social network analysis of the reduced networks

We start with an analysis similar to the one performed in Section 3 using the subset of the friendship and reply network containing only these 7,196 users. We also analyse the intersection and the difference of these reduced networks. The intersection is the (directed) network where a user A is connected to another user B if and only if they are connected in both networks, i.e. there exists a friendship and a reply connection from user A to user B. If we remove all these connections from the reduced reply network, we obtain the difference network. This network contains thus only the reply connections which do not have an equivalent in the friendship network.

In Table 4 we present the results of the social network analysis of these four networks. Since the original friendship network (compare with Table 2) is nearly identical to its reduced version, we do not observe many significant changes in the considered quantities. The largest difference can be found in the average degree which increased by about 0.9. This indicates that in the construction of the reduced friendship network users connected less than

Table 4. Global measures of the reduction of reply and friendship network, their intersection and their difference.

Variables	Friend NW	Reply NW	Intersection NW	Difference NW
No. of nodes with edges N	6,899	7,011	2,371	6,972
No. of nodes with in-degree ≥ 1	6,440	6,820	1,985	6,776
No. of nodes with out-degree ≥ 1	3,613	6,087	1,517	6,068
No. of edges, M	51,348	417,973	13,504	404,469
No. of nodes in giant component, G	5,977 (86.6%)	6,816 (97.2%)	1,756 (74.1%)	6,773 (97.2%)
Mean distance, l	3.71 (0.95)	2.64 (0.60)	3.58 (0.97)	2.65 (0.60)
Maximal distance, D	11	8	10	6
Clustering coefficient, C	0.083 (0.19)	0.164 (0.15)	0.027 (0.11)	0.143 (0.13)
Mean in-degree, $\langle d_{\mathrm{in}} \rangle$	7.44 (24.05)	59.62 (117.46)	5.70 (16.47)	58.01 (111.4)
Mean out-degree, $\langle d_{\mathrm{out}} \rangle$	7.44 (28.76)	59.62 (120.90)	5.70 (15.87)	58.01 (115.8)
Network density, $\hat{a}$	1.1×10^{-3}	8.5×10^{-3}	2.4×10^{-3}	8.3×10^{-3}
Reciprocity, ρ	0.679	0.40	0.56	0.38

Note: Values within parenthesis indicate standard deviation (or percentage of nodes in the giant component).

average have been removed in greater proportion than others. The picture changes a little in the case of the reply network. Its reduced version is significantly smaller. Nearly 80% of the users are removed, but roughly 43% of the connections persist. Then the average degree doubles the one of the original network, the average distance becomes shorter (2.64 vs. 3.2 in the original network) and the clustering coefficient increases (0.164 vs. 0.124). The reciprocity also slightly increases (0.40 vs. 0.35).

If we compare the network densities $\hat{a}$ we observe that the reduced versions of friendship and reply network do not have a similar density as is the case for their unreduced counterparts. Both densities increase, but while the reduced reply network is about 10 times "denser", $\hat{a}$ increases in the friendship network only by approximately 70%.

The ASP-profile of the reduced networks is again very similar to the original networks (see Figure 3 and Table 3).

The intersection of these reduced two networks is a smaller network of only 2,371 users. The proportion of users within the giant component is smaller (74.1%) than in the other networks analysed here. Other interesting differences are the lower clustering coefficient, an intermediate reciprocity (between reply and friendship network) and a slightly more balanced ASP-profile (red dashed line with circles in Figure 3) as in the case of the friendship network.

The difference network on the other hand is very similar to the reply network: nearly all the values are identical. Only the maximal distance D is smaller, which is caused by eliminating some nodes from the periphery of the giant component.

As one would expect, the intersection of the two networks is more similar to the smaller one, i.e. the friendship network. It is interesting to observe that removing the intersection of friendship and reply network hardly alters the statistics of the remaining reply network. However, the ASP profile changes dramatically. The elimination of interaction with friends leads to a Z-dissortative network (line in magenta with hexagram markers in Figure 4). This implies that reply interactions between non-friends follow a different behavioural pattern. Users show a preference to reply to users which are of different degree. This tendency is strongest in the relation between the out-degree of the replier and the in-degree of the user being replied to. In other words, when considering only replies between non-friends, a user who writes a large number of replies has a tendency to do so to users who do not receive many replies and vice versa.

In the following section we will investigate these phenomena further. In particular, we will analyse the coincidence of reply and friendship links.

5.2 Analysis of link coincidence

In this section we perform a simple probabilistic analysis: we count for every possible directed link whether there exists a connection in both networks or in only one of them. We break these counts down according to the different relationship types (uni- or bi-directional) in the two networks and present them in Table 5.

Note that these values coincide with the number of edges M from Table 4. We will use them to calculate the following conditional probabilities:

1. p (**reply**|**friends**) is the probability of a user A to reply to at least one comment of a user B given that A is a friend of B.
2. p (**friends**|**replies**) is the probability that a user A who replies to at least one comment of a user B establishes also a friendship relation to user B.

Using Table 5 we find that p (reply|friends) = 0.26 and p (friends|replies) = 0.03. Thus, in general, the replying activities do not lead to friendship connections. However, at least in one out of four cases a friend of a user will reply to one of the user's comments.

Next, we investigate if this picture changes if we filter the network and consider only the most active users or connections. In Figure 6 we present results for the cases when we consider only

1. users with more than a certain number of incoming or outgoing friendship links (their in- or out-degrees in the friendship network are larger than f_θ), or
2. reply relations which occur more often than a certain threshold w_θ.

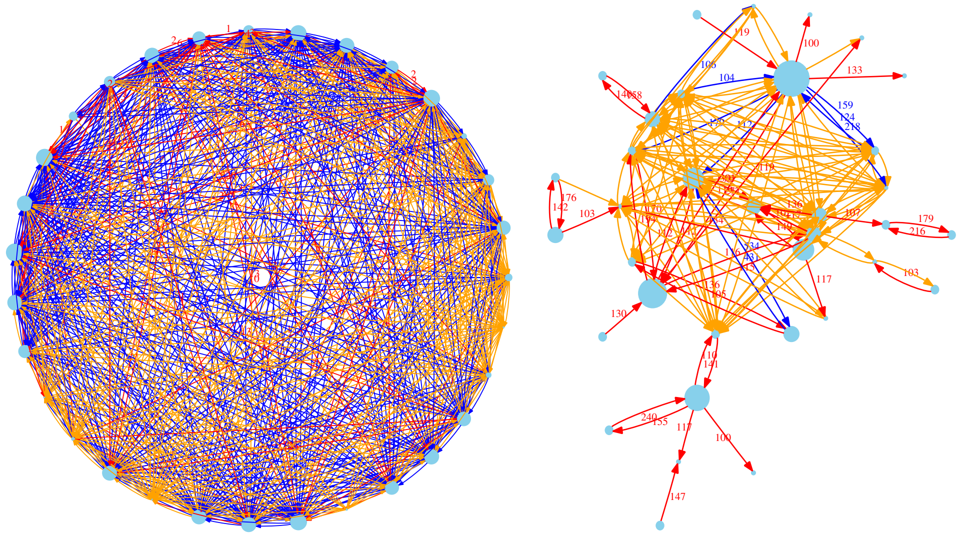
The two bottom plots of Figure 6 depict the evolution of the number of users who remain in the network if the threshold criteria are applied. In both cases we start with the 7,196 users with connections in both, reply and friendship, networks. The number of users decays then as the value of the threshold is increased.

The top left sub-plot in Figure 6 shows that both probabilities defined above increase nearly steadily and finally stabilise if we concentrate only on the users with larger degrees in the friendship network. For a user with an in- or out-degree >120 in the friendship network (slightly more than 100 users) there is a 50% chance that a user who received a reply is also a friend of this user. The inverse conditional probability even reaches 80% for users with a degree higher than 140, indicating that four out of five users who receive a reply are also declared as friends.

An example of the topology of such a filtered friendship network is given in Figure 7(a)) for a threshold of $f_\theta = 220$. The friendship-links without reply

Table 5. Co-occurrence of the different link types in the combination of the two social networks.

Relations	A replies to B	A does not reply to B	Total
A is friend of B	13,504	37,844	51,348
A is not friend of B	404,469	25,431,793	25,836,262
Total	417,973	25,469,637	25,887,610



(a) friendship links: $f_\theta = 220$ (b) reply interactions: $w_\theta = 100$

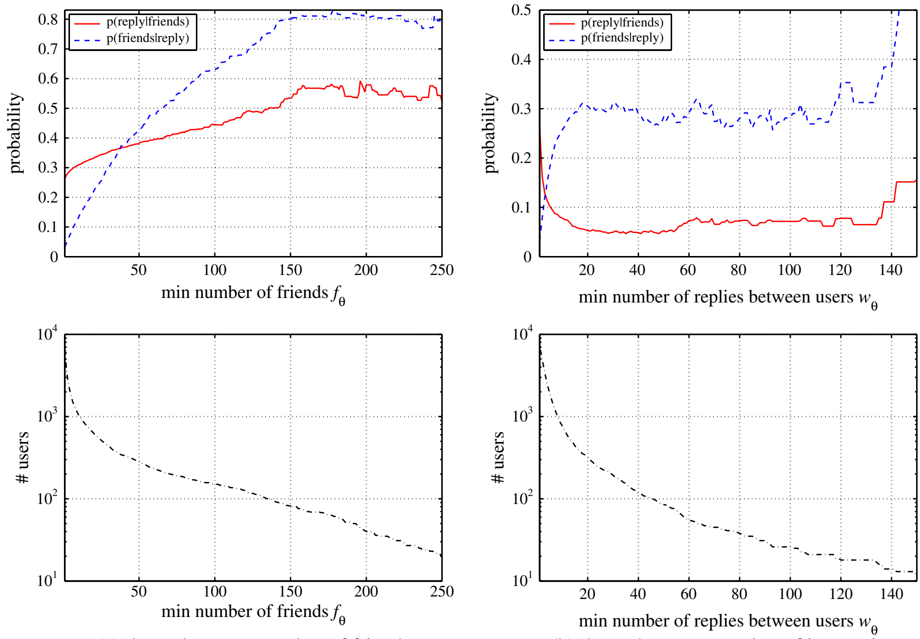
Figure 6. Filtered friendship (a) and reply network (c). Different link colours indicate different types of relationships in the combination of both networks; blue: (directed) interaction in both networks, orange: only friendship link is present, red: only reply interaction (with a weight above threshold). The node size corresponds to the degree in the reply network. (a) Friendship links $f_\theta = 220$; (b) reply interactions: $w_\theta = 100$.

relations are drawn in orange while connections which exists in both networks are drawn in blue. Clearly, the combined relations dominate, and the number of friendship links is significantly larger than the number of links that is absent.

A quite different picture can be seen in the top right sub-plot of Figure 6 for case where we put a threshold on the number of replies between the users. We start from the same point as in the top left sub-plot but here the probability p (reply|friends) increases while p (friends|replies) decays if we increase the threshold. Both probabilities more or less stabilise around a threshold of 20 reply interactions. Given that a user A replied more than 20 times to comments of another user B we get a 30% probability of a directed friendship link from user A to user B, but have on the other hand a very low probability of around 6% in the inverse case. The values start to fluctuate a lot for thresholds larger than 120 where the sample sizes are too small and only <20 users remain in the network.

An example for the filtered reply network with threshold 100 on the number of replies is given in Figure 7(b). It illustrates the lack of interaction between friends at this level of intensity.

Combining these two analyses we can conclude that, although users with a large degree in the friendship network will reply with high probability to users with similar degrees among their friends, the number reply interaction between them is likely to be small. On the other hand, a user who interacts



(a) dependence on number of friends (b) dependence on number of interactions

Figure 7. Conditional probabilities of interactions between (a) the top connected users (more than f_θ friends) in the friendship network and (b) the pairs of users who frequently (more than w_θ times) reply to each other.

repeatedly (e.g. more than 20 times) with another one via comments declares this user a friend only in less than one of 10 cases.

5.3 Reply reciprocity

Previously, we have mainly analysed the relations in the reply network without taking into account the number of actual interactions between the users (besides of using it as threshold in Section 5.2). In this section we will analyse how balanced the numbers of these directed interactions are. For this purpose we will introduce a measure which we will call the reply reciprocity ρ_{reply}. The same measure has been used by Lussier *et al.* (2010) to measure the reciprocity of user votes in Digg. It is inspired by the norm of reciprocity (Gouldner 1960): the social expectation that people respond to each other in the same way. That is returning positive actions for the positive actions and acting with indifference or hostility to negative actions.

The combined analysis of reply and friendship networks allows to analyse whether this norm is stronger if there exists a friendship relation between the involved users. We will use the following interpretation of the norm of reciprocity in the context of replies to comments. Although, a priori we cannot assess whether a reply will be seen as a positive or negative action, it seems plausible to assume that, on average, a reply between users with an explicit friendship relation is seen as a positive action and should also invoke a positive reaction. However, we can then also assume that, on the contrary, the absence of a friendship link would indicate the existence of indifference or even hostility between the users and, in consequence, a reply between such two users will then be more likely to have negative connotations. It would cause, according to the norm of reciprocity, either negative responses or indifference.

Formally, we define the reply–reciprocity between two users A and B as:

$$\rho_{\text{reply}}(\text{user } A,\ \text{user } B) = \log \frac{|\text{replies of } A \text{ to } B| + 1}{|\text{replies of } B \text{ to } A| + 1} \tag{2}$$

Using the logarithm we achieve a symmetrical measure around 0. A positive value indicates that the user A replies more often to comments from user B than vice versa and a negative value indicates the contrary. If we do not allow indifference in the definition of the norm of reciprocity then ρ_{reply} should be 0 if the norm is fulfilled. The larger ρ_{reply} becomes, the lesser the two users fulfil the norm.

We will break the results down according to the type of relationship between the users (mutual friends, unidirectional and non-friends). Note that we only consider pairs of users with a bidirectional connection in the reply network since we want to ignore indifference in the reply interactions.

We present the results[7] in Table 6. Surprisingly, we observe that the average reciprocity is much lower (0.21) among users who are non-friends than for pairs of user with a unidirectional (0.32) or bidirectional friendship link (0.38). This indicates that the norm of reciprocity is less satisfied in the case of a friendship relation between the commentators. Given that a pair of users has mutually replied to each other, it is more likely that this relation is balanced in the case of non-friends than in the case of friends. We suspect that this happens because the content of a friend's reply usually does not represent a challenge to the original statement, rather it supports the statement's points and can therefore be left unanswered.[8]

5.4 Extraction of the k-cores of the networks

In this section we will analyse the size of the innermost k-shell and the corresponding k-index ($k_{\max}$) for the eight possible combinations between

Table 6. Average values of the reply–reciprocity between users with mutual replies for different types of relations in the friendship network.

Friend–relation	Reply–reciprocity
No friendship link	0.21 (0.28)
Unidirectional friendship	0.32 (0.34)
Mutual friendship	0.38 (0.35)
Total	0.22 (0.29)

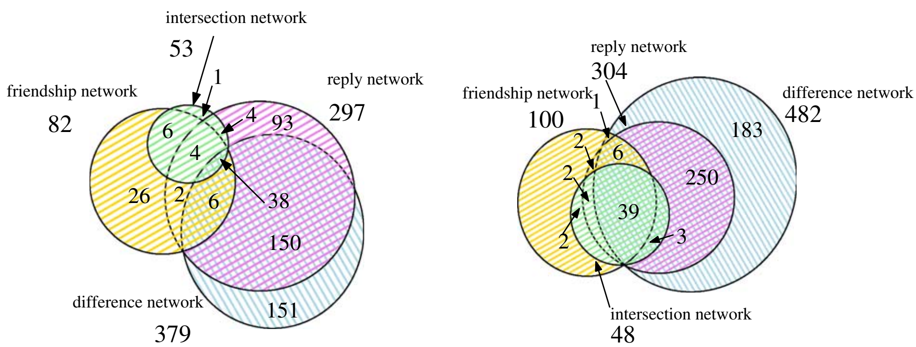
Table 7. Size and k-index of the k_{max}-shells of the different network types.

	In-degree		Out-degree	
	k_{max}	Size of k_{max}-shell	k_{max}	Size of k_{max}-shell
Friendship network	50	83	56	108
Reply network	119	408	122	345
Intersection network	25	56	25	67
Difference network	96	458	95	374

network type and the use of in- or out-degrees for the decomposition. The corresponding numbers are given in Table 7. Note that we consider here again the entire networks and not their reduced versions as in the previous sub-sections.

We recall that k_{max}-shell represents the most connected component of the entire network. To be member of the k_{max}-shell a user needs to have at least k_{max} connections to other users with the same shell. Therefore, the ratio between k_{max} and the size of the shell is a lower bound for the proportion of connections among the members of this shell. This ratio is especially small in the friendship networks where all members of the k_{max}-shell (calculated using the in-degree) are connected to more than 60% of all other members.

In Figure 8 we also plot Venn diagrams that illustrate the numbers of users in the networks innermost k-cores who are shared by the different types of networks. Interestingly, there is a considerable fraction of the users who are in the intersection of all four k_{max}-cores. We find 38 such users if we use the in-degree for calculation and 39 of them in the case of the out-degree (28 of them belong to both intersections). We claim that they are the most active users who are not only friends to the large number of users but also frequently reply independently to friends and non-friends. According to Kitsak *et al.* (2010) such users would be the most efficient information spreaders in



(a) using in-degree (b) using out-degree

Figure 8. Non-proportional Venn diagrams of the maximal k-cores of the different network types. The numbers indicate the cardinality of the different k-cores and their intersections. (a) Using in-degree; (b) using out-degree.

Menéame. They would be the best candidates to be studied in a more detailed analysis about individual behaviour patterns in these articulated and behavioural networks and on the social news website in general. We leave this analysis for future research.

6. Conclusions

In this study we have compared and analysed the (behavioural) reply network and the (articulated) friendship network of Menéame, a Spanish Digg clone. We find significant differences between the two networks in the reciprocity of the directed links. It is almost twice as large in the friendship network which can be explained by the higher social pressure to respond an incoming friendship link. At the individual level the larger reply network translates into a larger average degree, indicating that the reply network is clearly more important than the friendship network on Menéame.

We found that the reply network is Z assortative (Foster *et al.* 2010), which is similar to what has been reported for the reply network of Slashdot by Laniado *et al.* (2011). However, if only replies between non-friends are considered the network becomes Z dissortative. This uncommon structure of the reply network is also confirmed by a k-core decomposition which finds that rewiring the reply connections randomly (without changing the degree of the users) does not lead to much smaller maximal k-indexes, as for example in the case of the friendship network. This is quite exceptional, if we compare to the results in Kitsak *et al.* (2010) and may indicate that the practice of replying to comments is very much content-based and there is no preference to reply specifically to the comments of a specific user. Moreover, from the dissortative structure of the network of replies between non-friends, we conclude that users who reply a lot are less likely to reply to comments of users who themselves reply a lot. We suspect that it is to avoid an extended and time-consuming discussion involving several mutual replies or to ignore "trolls", i.e. users who post-inflammatory comments which seek to provoke emotional responses.

In the second part of the study we performed an analysis of the combination between behavioural and articulated networks, which we restrict to the subset of users with connections in both networks. We also analyse the intersection and the difference of both networks. The first one is found to be very similar to the friendship network and the latter has similar characteristics as the reply network.

Moreover, we analyse the link coincidence between the two networks where we find that although there exists a relatively high probability of replies from a user to his/her friends, the number of reply interactions between these users is relatively small and users who interact repeatedly with others do this preferentially with non-friends.

If we analyse the reciprocity of the reply interactions we find that non-friends who reply to each other are more likely to have similar number of replies between each other than friends. According to the social norm of reciprocity (Gouldner 1960) this could be explained by a negative connotation of replies between non-friends. Replies between friends may be seen more neutral and thus do not provoke replies to the same extend.

In summary, the actual personal network of Menéame users is somehow divided: users have reply relations, which in many cases do not coincide with their friends. The interaction between users is thus more motivated by the user generated content than by established friendship relations. So, new approaches should be emerging in social media using social currencies to promote interaction between users who are not explicitly friends. This would have implications in the design of a community sustaining a social network. We hypothesise that stimulating the users to interact around content rather than only declaring friendships is very likely to generate a higher level of interaction in the network.

Acknowledgements

The authors wish to thank Vicenç Gómez and David Laniado for helping in the evaluation of this study and Meneame.net for allowing access to an anonymised dump of their database. This work has been developed inside the i3media project (CDTI 2007–2010) and partially funded by MediaPro, the Centro para el Desarrollo Tecnológico Industrial (CDTI) within the Ingenio 2010 initiative. Yana Volkovich acknowledges support from the Torres Quevedo Program from the Spanish Ministry of Science and Innovation, co-funded by the European Social Fund.

Notes

[1] http://www.zephoria.org/thoughts/archives/2009/07/28/would_the_real.html (accessed 22 April 2010).
[2] http://www.meneame.net.
[3] See http://www. blog.meneame.net/2009/05/07/explicacion-simple-del-algoritmo-de-promocion-de-noticias-promote/ (accessed 22 April 2010) for details.
[4] Described in http://www.meneame.wikispaces.com/Karma (accessed 22 April 2010).
[5] If the network is directed, then one can use either the in- or the out-degree of the node v. There exist therefore two different k-core decompositions for directed networks. Such an extension of the original definition of Seidman (1983) has been first proposed by Batagelj and Zaversnik (2002).
[6] Note that, as can be seen from the top row of Table 4, some of these users do not have any links in the reduced networks, since all of their link targets are users who do not fulfil the above condition.
[7] Since the measure is symmetric we calculate the statistics only for positive values of ρ_{reply} and use only half of the cases when $\rho_{\mathrm{reply}} = 0$.
[8] Another possibility to explain missing back-replies between friends would be the existence of other communication channels (besides those of Meneame) between friends to interact. However, given the nature of public discussion it seems to be unlikely that the discussion switches exclusively to such hidden channels if a friend's reply would have had negative connotations.

References

J.I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat and A. Vespignani, "K-core decomposition of Internet graphs: Hierarchies, self-similarity and measurement biases", *Networks and Heterogeneous Media*, 3, pp. 395–411, 2008.

V. Batagelj and M. Zaversnik, "Generalized cores", *CoRR*, cs.DS/0202039, 2002.

E.A. Bender and E.R. Canfield, "The asymptotic number of labeled graphs with given degree sequences", *Journal of Combinatorial Theory, Series A*, 24(3), pp. 296–307, 1978.

J.G. Foster, D.V. Foster, P. Grassberger and M. Paczuski, "Edge direction and the structure of networks", *Proceedings of the National Academy of Sciences*, 107(24), pp. 10815–10820, 2010.

D. Garlaschelli and M.I. Loffredo, "Patterns of link reciprocity in directed networks", *Physical Review Letter*, 93(26), p. 268701, 2004.

K.I. Goh, Y.H. Eom, H. Jeong, B. Kahng and D. Kim, "Structure and evolution of online social relationships: Heterogeneity in unrestricted discussions", *Physical Review E*, 73(6), p. 066123, 2006.

V. Gómez, A. Kaltenbrunner and V. López, "Statistical analysis of the social network and discussion threads in Slashdot", *WWW '08: Proceedings of the 17th international conference on World Wide Web,* New York, NY, USA: ACM, pp. 645–654, 2008.

A. Gouldner, "The norm of reciprocity: A preliminary statement", *American Sociological Review*, 25(2), pp. 161–178, 1960.

T. Hogg and K. Lerman, "Stochastic models of user-contributory web sites", *Proceedings of 3rd International Conference on Weblogs and Social Media*, MenloPark, CA: The AAAI Press, pp. 50–57, 2009.

M. Kitsak, L. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. Stanley and H. Makse, "Identifying influential spreaders in complex networks", *Arxiv preprint arXiv:1001.5285*, 2010.

J. Kunegis, A. Lommatzsch and C. Bauckhage, "The slashdot zoo: Mining a social network with negative edges", *WWW' 09: Proceedings of the 18th inter-national conference on World wide web*, New York, NY, USA: ACM, pp. 741–750, 2009.

D. Laniado, R. Tasso, Y. Volkovich and A. Kaltenbrunner, "When the Wikipedians talk: Network and tree structure of Wikipedia discussion pages", *ICWSM-11: Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*, Menlo Park, CA: The AAAI Press, pp. 177–184,, 2011.

K. Lerman, "Social information processing in social news aggregation", *IEEE Internet Computing: Special issue on Social Search*, 11(6), pp. 16–28, 2007.

J. Leskovec, D. Huttenlocher and J. Kleinberg, "Predicting positive and negative links in online social networks", *WWW' 10: Proceedings of the 19th international conference on World wide web*, New York, NY, USA: ACM, pp. 641–650, 2010.

J. Lussier, T. Raeder and N.V. Chawla, "User generated content consumption and social networking in knowledge-sharing OSNs", *SBP 2010: Proceedings of Third International Conference on Social Computing, Behavioral Modeling, and Prediction*, Berlin, Germany: Springer, pp. 228–238, 00202010.

N. Matsumura, D.E. Goldberg and X. Llorà, "Mining directed social network from message board", *WWW' 05: Proceedings of the 14th international conference on World Wide Web*, New York, NY, USA: ACM, pp. 1092–1093, 2005.

M. Molloy and B. Reed, "The size of the giant component of a random graph with a given degree sequence", *Combinatorics, Probability, and Computing*, 7, pp. 295–306, 1998.

M.E.J. Newman, "Assortative mixing in networks", *Physical Review Letters*, 9(20), pp. 208–701, 2002.

M.E.J. Newman, "The structure and function of complex networks", *SIAM Review*, 45, pp. 167–256, 2003.

M.E.J. Newman and J. Park, "Why social networks are different from other types of networks", *Physical Review E*, 68(3), p. 036122, 2003.

S. Resnick, *Heavy-tail phenomena: probabilistic and statistical modeling*, Berlin, Germany: Springer Verlag, 2007.

S.B. Seidman, "Network structure and minimum degree", *Social Networks*, 5(3), pp. 269–287, 1983.

K. Sigman, "Appendix: A primer on heavy-tailed distributions", *Queueing Systems*, 33, pp. 261–275, 1999.

G. Szabó and B.A. Huberman, "Predicting the popularity of online content", *CoRR*, abs/0811.0405, 2008.

D.J. Watts and S.H. Strogatz, "Collective dynamics of small-world networks", *Nature*, 393(6684), pp. 440–442, 1998.

F. Wu and B.A. Huberman, "Novelty and collective attention", *Proceedings of the National Academy of Sciences*, 104(45), pp. 17599–17601, 2007.
K. Zhongbao and Z. Changshui, "Reply networks on a bulletin board system", *Physical Review E*, 67(3), p. 036117, 2003.