Exploring Temporal Communication Through Social Networks

Liaquat Hossain, Kon Shing Kenneth Chung, and Shahriar Tanvir Hasan Murshed

School of Information Technologies, University of Sydney,
Sydney, Australia
{lhossain,ken,tanvir}@it.usyd.edu.au

Abstract. The dissemination of information in social networks and the relative effect of ICT (Information and Communications Technology) use has long been an interesting area of study in the field of sociology, human computer interaction and computer supported cooperative work. To date, a lot of research has been conducted regarding an actor's mobile phone usage behavior while disseminating information within a mobile social network. In this study, we explore the structured network position of individuals using mobile phone and their ability to disseminate information within their social network. Our proposition is that an actor's ability to disseminate information within a social group is affected by their structural network position. In this paper, we determine an actor's structural network position by four different measures of centrality—(i) degree, (ii) closeness, (iii) betweenness, and (iv) eigenvector centrality. We analyse the Reality Mining dataset, which contains mobile phone usage data over a 9 month period for exploring the association between the structural positions of different actors in a temporal communication. We extract relational data to construct a social network of the mobile phone users in order to determine the association between their position in the network and their ability to disseminate information. The following questions form the basis for this study: Does information dissemination capability of an actor reflect their structural position within a social network? How do different measures of centrality associate with the information dissemination capability of an actor? Are highly central actors able to disseminate information more effectively than those who have a lower central position within a social network?

Keywords: Social Networks, Mobile Usage Behaviour, Centrality, Information Dissemination, Temporal Communication.

1 Introduction

Mobile telephone use has proliferated in recent years. They have become embedded in society culture and are no longer unusual to view people using mobile phones in various different contexts. Recent studies regarding mobile phones have primarily focused on the usage behaviour of individuals over a set period of time. Studies have also been conducted into the usability of the mobile communication devices for

C. Baranauskas et al. (Eds.): INTERACT 2007, LNCS 4662, Part I, pp. 19–30, 2007.

understanding dissemination behaviour and innovation patterns [3, 15]. Concrete research into the relationship between a user's position in the mobile social network and their ability to disseminate information in that particular network are relatively few. The objective of this study is to explore this relationship to discover the effect of different network structure and positions on an actor's ability to effective disseminate information within the network. In particular, the study explores how different types of centrality measures give an actor the best ability to disseminate information within a social network. We suggest that the positioning of an actor within a network is closely associated with their ability to disseminate information. An actor's structural position is determined by various measures of centrality: in-degree, out-degree, closeness, betweenness and eigenvector centrality [10]. These measures allow for calculation of an actor's position in the network whilst determining their relationship with surrounding actors in the network. We analysed the reality mining dataset, which contains information collected on 100 mobile users over a nine month period, to elicit the mobile social network for the study. We then conducted preliminary analysis of the mobile usage behavior of the individual subjects contained within the dataset to form the basis for argument. A model representing different network structures affecting information dissemination is constructed using the software application to illustrate the differences between the different measures of centrality. This study may have a profound impact on the way organisations select individuals for projects due to their capability to co-ordinate resources, which directly come from their ability to disseminate information. Furthermore, insights on the association between an individual's social structure and capacity to disseminate information provide us with an awareness of human-interactions in mobile and temporal contexts. This is useful for building contextualized awareness HCI tools.

2 Background

The Reality Mining project represents one of the largest mobile phone experiment attempted in academia to date [8]. The study consists of 100 Nokia 6600 smart phones pre-installed with several pieces of software developed by the Massachusetts Institute of Technology (MIT), as well as a version of Context application from the University of Helsinki. The dataset, which is a MySQL relational database, was collected for the purpose of monitoring mobile phone usage behavior in order to model complex social systems. The information collected includes call logs, Bluetooth devices in proximity, cell tower IDs, application usage and phone status (such as charging and idle). The Reality Mining dataset contains data collected by one hundred human subjects over the course of nine months and represents approximately 350,000 hours of data on users location, communication and device usage behavior over the course of the 2004-2005 academic year. The project is still ongoing and upon completion of the study the dataset will contain 500,000 hours (~60 years) of continuous data on daily human behavior. The subjects contained in the dataset are 75 users who are either students or faculty in the MIT Media Lab, while the remaining 25 users are incoming students at MIT Sloan business school. Of the 75 users at the Media Lab, 20 are incoming masters students and 5 are incoming MIT freshmen.

3 Information Dissemination Through Social Network

Disseminating information in social networks is a complex and nuanced process that is the sum of many individual actions. It is difficult to overestimate the importance of social networks in the processes of disseminating and receiving information [13]. Based on the information itself and on contextual factors, a person may choose to share the information as it to selected people in his social network or may modify the information (e.g. remove details) before disseminating it. Goecks and Mynatt [11] state that contact and availability information is often closely guarded and shared only with the people in one’s personal social network. Friends talk about books that they’ve recently read and share photos of their children. Colleagues share ideas, data, and references when collaborating. These are some examples of the types of information disseminated in a social network. Fisher and Dourish [9] argue that the information disseminated can be traced and that there is structure in dissemination. Furthermore, such network structures can be used to build contextualized awareness tools that successfully present an appropriate selection of information.

Borgatti [5] documented how different measures of centrality can be matched to the kinds of information flows that they are appropriate for. Specific simulations were conducted to examine the relationship between type of information flow and the differential importance of nodes with respect to measurements such as speed of reception of traffic and frequency of receiving traffic. It was discovered that traditional centrality measures were fully applicable only for specific flow processes they are designed for, can be regarded as generating expected values for certain kinds of node outcomes (such as speed and frequency of reception) given implicit models of how information disseminates. There are five well known measures of centrality which are commonly used in social network analysis. These are in-degree, out-degree, closeness, betweenness and eigenvector centrality.

Closeness centrality of a node (or actor) as defined by Freeman [10], is the sum of the graph-theoretic distances from all other nodes, where the distance from a node to another is defined as the length (in links) of the shortest path from one to the other. Borgatti [5] argues that in a flow context, ordinarily you would interpret closeness as an index of the expected time until arrival of something flowing through the network. Borgatti also noted that nodes with lower raw closeness scores have short distances from others so will tend to receive flows sooner, assuming that what flows originates from all other nodes with equal probability, and also assuming that whatever is flowing manages to travel along shortest paths.

Betweenness centrality is defined as the ‘share’ of times that a node i needs a node k (whose centrality is being measured) in order to reach a node j via the shortest path (Freeman 1979). Betweenness is conventionally thought to measure the volume of traffic moving from each node to every other node that would pass through a given node [16]. Thus, it measures the amount of network flow that a given node ‘controls’ in the sense of being able to shut it down if necessary.

Eigenvector centrality is defined as the principal eigenvector of the adjacency matrix defining the network [4]. The idea proposed by Bonacich is that even if a node influences just one other node, who subsequently influences many other nodes (who themselves influence still more others), then the first node in that chain is highly influential. At the same time it can be seen that eigenvector centrality is providing a model of nodal risk such as a node's long term equilibrium risk of receiving information traffic is a function of the risk level of its contacts [5].

Finally, degree centrality can be defined as the number of ties incident upon a node [10]. That is, it is the sum of each row in the adjacency matrix representing the network. Degree centrality can also be defined as the number of paths of length one that emanate from a node [14]. As a result, one way to interpret the measure would be in terms of an implicit process that involves no indirect links. For information flow this means that a highly central actor will be more active in disseminating information in that they have more ties to other actors in the network. In light of the social network concepts and measures discussed for making information dissemination structure visible, we ask the research questions: (i) does information dissemination capability of an actor reflect their structural position within a social network? (ii) How do different measures of centrality associate with the information dissemination capability of an actor? (iii) Are highly central actors able to disseminate information more effectively than those who have a lower central position within a social network?

4 Method

The initial dataset contained voice and text message data regarding interactions made by the 100 participants in the study to members, both internal (eg. other subjects) and external (eg. friends, family) to the study. As a first step in data analysis, we concentrated on the interactions made by the participants (in the study) to other members who were also involved in the study. Such interactions were considered to be internal interactions. Interactions between members of the study and outsiders were deemed as external interactions.

There are a total of 897,921 interactions contained in the Reality Mining Dataset. We decided the best way of separating the data was to create a new table containing only the voice records which contained internal interactions. We then used certain thresholds to gather only the data we thought would be suitable to determining valid relationships between members of the study. We applied the bootstrapping mechanism by applying a threshold limit of a minimum of 5 interactions to take place before a valid relationship could be deemed to have substantial information disseminated. This approach is similar to the email study conducted by Adamic and Adar [1]. Calls of duration 0, which surprising there were many of, were also deemed as invalid data for our study. It could only be assumed that such calls were made to retrieve voice messages as the number called by the sender was the same as the number of the sender.

From this threshold we were able to gain a table of internal interactions for our data of 30,620 voice call records. The respective strengths, which were the number of interactions between each participant, were then calculated and exported into a text file, later which was read into the UCINET program [7] for social network analysis. We were then able to alter the view of the network diagram according to different variables based on the strengths of the relationships, for example only displaying relationships with tie strength greater than 10.

In order to calculate the information spread, we used an information dissemination index adapted from Bae [3]. According to Bae, the information dissemination index was a numerical value based on data collected that determined which members of a network had been more involved in communicating with other members. Our own information dissemination index to determine which participants contributed more to the information flow in the network was as follows:

Information Dissemination Index:

$$\frac{\#\text{voice calls sent} - \#\text{voice calls received}}{\#\text{voice calls sent} + \#\text{voice calls received}}$$

Regardless of range, the information dissemination index is +1 if somebody only makes calls and doesn't receive any calls. The index is -1, if somebody only receives calls and does not make any calls. The index is 0, if somebody has totally balanced communication behaviour, sending and receiving the same number of calls.

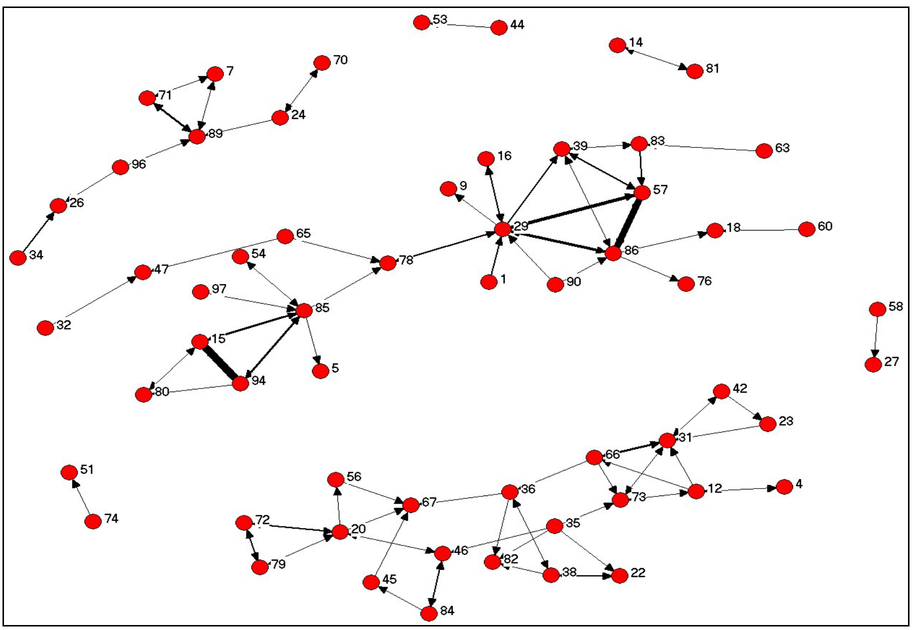

Fig. 1. Visualisation of the mobile phone interaction pattern between participants with tie strength

Results and analysis were gathered by mining the modified dataset using PHP scripts with embedded SQL commands. Using these commands we were able to discover call activity information which could then be combined with centrality information from UCINET to report the findings seen in the next section.

5 Results and Analyses

From the preliminary analysis that was conducted, it was revealed that an actor who was centrally based in the mobile communication network was more likely to have a greater ability to disseminate information throughout the network. Figure 1 depicts a visual analysis (sociogram) of the communication of the participant network - that is, calls made to and from the participants with the threshold applied are represented in the sociogram. Clearly, there are three major components in the sociogram which represents a clique group or cluster. It is possible that such cliques arise as a result of participants belonging to the same school or group within MIT (eg. Sloan School of Management). There are also a couple of periphery actors who do not belong to any of the clusters. Their communication frequency is low. The number of interactions made is visually represented as strength of the tie, as indicated by thickness of the lines. Hence, actors 57, 86, 15, and 94 are interesting nodes to take interest of as communication pertaining to these actors is particularly high. Further analysis shows that these actors represent the top four participants who communicated the most within the network of the participants themselves. Figure 2 below shows the top ten participants in the study when their interactions with other members were analysed.

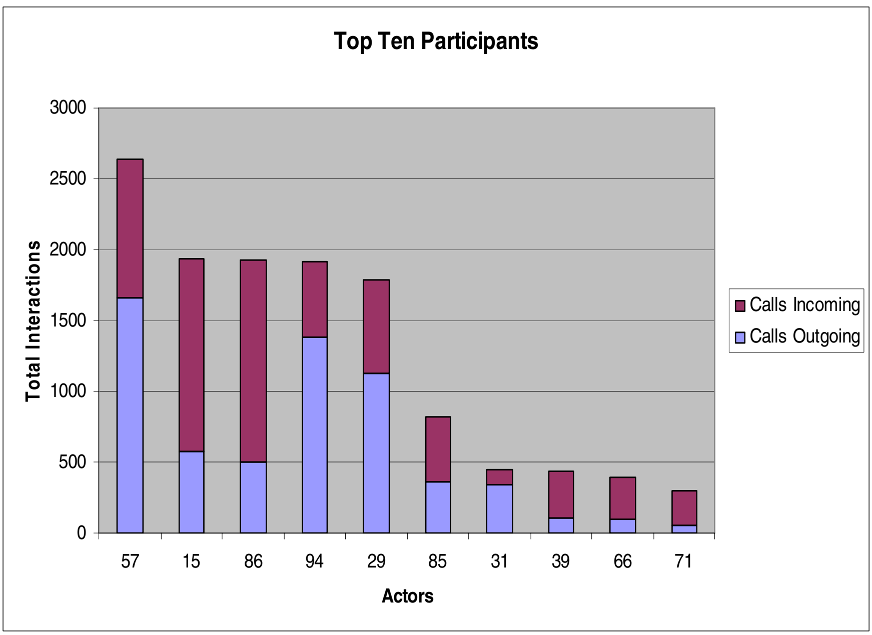

Fig. 2. Top ten participants showing call activity

Table 1 below highlights particular actor's in and out degree statistics as well as their information dissemination index calculations.

Table 1. Information dissemination index and degree values of the top ten participants of the study

Actor	OutDegree	InDegree	Total Interactions	Information Dissemination Index
57	1660	974	2634	0.26
15	578	1362	1940	-0.404
86	503	1425	1928	-0.478
94	1381	537	1918	0.44
29	1126	663	1789	0.258
85	367	448	815	0.042
31	339	111	450	0.504
39	108	325	433	-0.501
66	99	297	396	-0.5
71	56	237	293	-0.618

Hanneman [12] states that degree centrality is actually the greatest measure of actors in positions to disseminate information. He suggests that actors who have more ties to other actors may be advantaged to an extent because they have many ties, they may access to, and be able to call on more resources of the network as a whole. The evidence and the results from our study seem to agree with this notion, that an actor with a high degree of centrality is in a better position than others to disseminate information to other members of the network.

Table 2. Betweeness compared to degree centrality – scores and rankings

Actor	Betweeness	Ranking	InDegree	Ranking	OutDegree	Ranking
29	9.585	1	663	4	1126	3
15	8.914	2	1362	2	578	4
86	8.741	3	1425	1	503	5
85	7.663	4	448	6	367	6
94	7.614	5	537	5	1381	2
57	6.987	6	974	3	1660	1
35	5.195	7	10	71	65	20
20	5.309	8	109	21	191	10
66	4.24	9	297	8	99	17
71	3.713	10	237	9	56	23

It can be seen from table 2 that the actors with high betweenness centrality scores generally agree with their high scores for degree centrality. Actors 29, 15, 86, 85, 94 and 57 have high betweenness centrality scores which are well above average for the network. There is also a low level of variance in the network relative to the mean when betweenness is concerned (table 3); this could mean that the actors with high

Table 3. Summary statistics for all measures

	Degree	**Closeness**	**Betweeness**	**Eigenvector**
Mean	4.068	2.332	1.034	7.814
Std Dev	2.593	0.347	1.952	16.501
Sum	244.068	139.939	62.069	468.868
Variance	6.722	0.120	3.809	272.267
Min	1.695	1.695	0.000	-0.000
Max	13.559	2.670	9.585	72.480

betweenness scores can be deemed as quite centralised. The top six actors in betweenness scores also feature prominently in the rankings for out and in degree. These members are considered to have a lot of influence in the network as a lot of other participants depend on them to make connections with other people. In other words, as they are "between" a lot of other interactions, their level of call activity is relatively high. Interesting though are some of the participants that scored highly with regards to betweenness but have a quite a low ranking with in degree or out degree compared. A prime example of this is actor 35 who has the 7th highest betweenness centrality score but fares quite lowly when in degree is concerned. This can be seen as an indication that actor 35 may have a structural network position of a peripheral facilitator to separated sub networks within the entire network. This position may not be apparent when just viewing the network using in and out degree measurements. This is also the case for actor 20, whose betweenness centrality score ranking is much higher than their degree calculations. In both cases they have a higher out degree score than in degree, which agree with a position of passing on information to other groups in the network. It can thus be proposed that different types of centrality measures affect a participant's network position. Clearly, some members of the network have a higher position when a different centrality measure is used and this can distinguish their position in the network for different purposes, where it may become apparent of different positions a member may hold.

Closeness centrality and eigenvector centrality measures were also calculated to discover if it has any impact on the findings about actors who have demonstrated to have a better position to disseminate information. Table 4 shows the top ten eigenvector centrality scores achieved by participants in the study and is compared to their ranking in the closeness centrality calculations.

The eigenvector approach is an important measurement as it is an effort to discover the most central actors (i.e. those with the smallest farness from others) in terms of "global" or "overall" structure of the network, and to pay lesser attention to patterns that are more "local". In other words, we want to find nodes that have an influence on the way other nodes may disseminate information. Again it can be seen from table 4 that actors 29, 86, 57, 39 and 94 have high eigenvector centrality scores and this indicates they are more central to the main pattern of distances among all of the actors. There is high variance in the eigenvector calculation relative to the mean (table 3). This may add to the fact that these actors have a lot of influence in the network which aids the results already seen that they are able to disseminate more information due to their structural position in the social network.

Table 4. Eigenvector versus closeness centrality

Actor	Eigenvector	Ranking	Closeness	Ranking
29	72.48	1	2.67	1
86	63.661	2	2.656	3
57	56.145	3	2.654	4
39	56.145	4	2.654	4
94	50.76	5	2.633	12
90	34.347	6	2.648	7
85	30.256	7	2.655	6
78	22.13	8	2.667	2
16	18.286	9	2.643	8
1	18.286	10	1.881	31

Closeness centrality approach emphasises the distance of an actor to all others in the network by focusing on the distance from each actor to all others. Again, even though the numbers of results of the closeness calculations vary very little, the same highly central actors which were mentioned previously are again ranked very quite highly but with a very low standard deviation. It seems the actors are quite close to one another which facilitates dissemination of information for the central actors.

Finally, we conducted a multi-mode analysis taking into account all forms of centrality measurements and actors call activity, which can be deemed as their information dissemination ability. The basis of the multi-mode analysis is based on evidence that out-degree centrality is the most important determinant of an actor's position in a social network because it indicates the level to which an actor is able to communicate [12]. With this in mind we have taken the top 25 actors as per their out degree centrality score and compared these ranks with their ranking in all other measurements of centrality to determine each actor's position in the network and how this compares to their call activity.

Each actor is given a rank for each respective centrality measure out of 25, if an actor scored 0 for that particular centrality measure then they are given a rank of 25 for that measure. Once all the ranks have been determined then are then added up across each row to determine an actor's total score. The lower the score the more centrally positioned an actor is deemed.

Once again, familiar actors which have been previously reported to have the highest call activity are deemed to be the most centrally positioned and are therefore proposed to be in a better position to disseminate information. Below is a diagram which depicts these results in graphical form with the more central actor's closer to the centre of the circle.

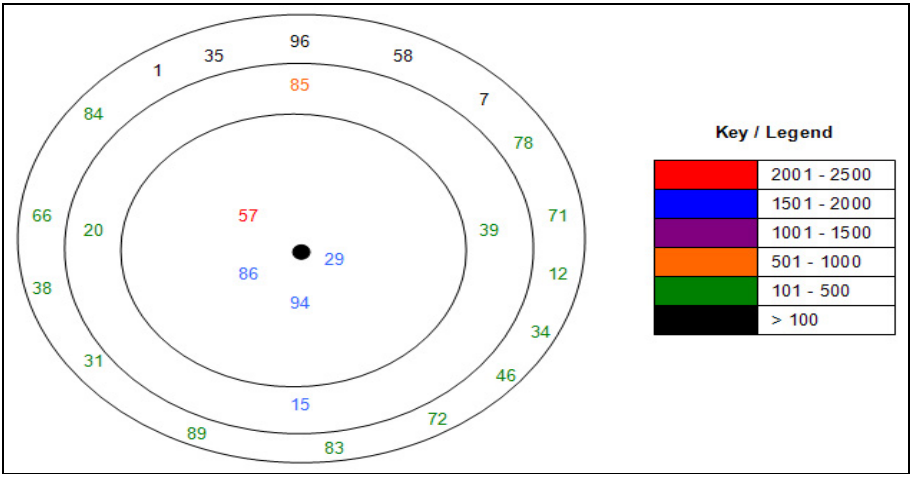

Fig. 3. Actors' central position with call interactions colour-coded

From figure 3, the closer the actor is to the centre the better their overall centrality score is, which leaves them in a better position to disseminate more information than members who are less centrally located. This correlates to the particular actor's call activity, which is also the amount of information disseminated. As seen from figure 3, actor 29 is most centrally positioned in the network and also has the 5th highest call activity with a total of 1789 interactions. Actor 57 who has the highest number of interactions in the group is the 4th best centrally positioned member of the network. Overall, it can clearly be seen that all of the actor's with high call activity (> 1500) are in the two inner most circles which gives them the ability to disseminate more information because they are more central in the network than the other members of the study. In general, the results and findings from this study seem to converge to the argument that an actor's structural position in a social network affects their ability to disseminate information. The pattern indicates that actors who have high centrality scores figure prominently in outgoing and incoming calls which corresponds to their increased ability to disseminate more information. The findings bear implications for researchers and scholars in the human computer interaction field and those involved in computer-supported and cooperative work. First, structure and position of individuals dictate information dissemination and flow. Second, traces from the network structure generate insights which are invaluable for the design and development of communication tools in mobile and dispersed personalised contexts.

6 Conclusion

Drawing on the results and findings from this study, it can be seen that an actor in a more centrally favored position does have an increased ability to disseminate more information to other members of a network than someone with a more peripheral position. That said, it can also be seen from the results that the importance of a particular actor may not become apparent until different centrality measures are taken into account. This can dramatically change the structural position of a particular actor and increase their prominence in a social network, without the actor necessarily being a central figure.

The results seem to agree with Hanneman's [12] view, in suggesting out-degree as an influential centrality measure when determining a central figure in an information flow. Although this does not mean they automatically are in the best information disseminating position in the network, eigenvector and betweenness measures also play a critical part in discovering which actors have the most influence and connections to distribute information to key divisions of the network. These measures become just as an important as out and in degree centrality measures for understanding an actors' communication structure [2, 6]. Further study needs to be carried out to explore centrality and information dissemination abilities in terms of correlation within different and similar group scenarios (eg. groups with high information dissemination indices as compared to low ones) rather than a large network to discover if these patterns stay consistent. A multiplex relational analysis (such as short message sent and received, and video calls sent and received) also needs to be conducted within the group and outside the group of the participants' network to gain a further understanding of the complexities and dynamics of information dissemination behaviour, in relation to the actor's network. Such understanding helps us to model social exchange behaviour from a social networks perspective. Furthermore, the interactions of these same actors at different time-periods (quarterly) can be studied to understand changes in individual and group communication structures longitudinally.

References

[1] Adamic, L., Adar, E.: How to Search a Social Network. Social Networks 27, 185–203 (2005)

[2] Ahuja, M.K., Galletta, D.F., Carley, K.M.: Individual Centrality and Performance in R&D Groups: An Empirical Study. Management Science 49(1), 21–38 (2003)

[3] Bae, S.J., Gloor, P., Schnorf, S.: Detection of Power User Patterns among High School Students in a Mobile Communication Network. In: Applications in Social Network Analysis 2005 (University of Zurich, Switzerland 2005)

[4] Bonacich, O.: Power and Centrality: A Family of Measures. American Journal of Scoiology 92, 1170–1182 (1987)

[5] Borgatti, S.: Centrality and Network Flow. Social Networks 27, 55–71 (2005)

[6] Borgatti, S.P., Everett, M.G.: Notions of Position in Social Network Analysis. Sociological Methodology 22, 1–35 (1992)

[7] Borgatti, S.P., Everett, M.G., Freeman, L.C.: Ucinet for Windows: Software for Social Network Analysis, Analytic Technologies, Harvard, MA (2002)

[8] Eagle, N., Pentland, A.: Reality Mining: Sensing Complex Social Systems. Pervasive Ubiquious Computing 10, 255–268 (2006)

[9] Fisher, D., Dourish, P.: Social and Temporal Structures in Everyday Collaboration. In: CHI 2004 Conference. Vienna, Austria, pp. 551–558 (2004)

[10] Freeman, L.C.: Centrality in Social Networks: Conceptual Clarification. Social Networks 1(3), 215–239 (1978)

[11] Goecks, J., Mynatt, E.D.: Leveraging Social Networks for Information Sharing. In: Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work, Chicago, Illinois, pp. 328–331. ACM Press, New York (2004)

[12] Hanneman, R.A.: Introduction to Social Network Methods. 2004 (August 8, 2001), http://faculty.ucr.edu/~hanneman/SOC157/NETTEXT.PDF
[13] Matsuo, Y., Mori, J., Sugiyama, T.: Real-world Oriented Information Sharing using Social Networks. In: Proceedings of the ACM-SIG GROUP 2005, pp. 81–84. ACM Press, New York (2005)
[14] Scott, J.: Social Network Analysis: A Handbook. SAGE Publications, London (2000)
[15] Song, X., Lin, C.-Y., Tseng, B.L., Sun, M.-T.: Modeling and Predicting Personal Information Dissemination Behavior. In: Knowledge Discover from Data 05, Chicago, Illinois, USA (2005)
[16] Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press, New York (1994)