2009 International Conference on Computational Science and Engineering

Finding Cohesive Subgroups and Relevant Members in the Nokia Friend View Mobile Social Network

Alvin Chin
Nokia Research Center
Building 2, No. 5 Donghuan Zhonglu, Economic and Technological Development Area
Beijing, China, 100176
Email: alvin.chin@nokia.com

***Abstract*—The mobile phone can be used as a ubiquitous internet device to help users connect with the relevant people instantaneously to grow their social networks and provide quality friendship. In this paper, we present usage statistics from Nokia Friend View and identify its relevant users, by describing and applying our SCAN method adapted from [1] for finding cohesive subgroups to the friend and interaction networks. Our results show that the connection patterns of the top 10 relevant users in the interaction network reflect the friend relationships of those users, and that members of cohesive subgroups send more messages and have more friends than others in the interaction network. These results can be used to improve friend recommendations, search, and user interaction.**

***Index Terms*—cohesive subgroups; mobile social network; relevant members; social behaviour**

I. Introduction

With mobile phones having Internet and web capability, users can easily access and update their online social networks in real time and with real location. We call the social network that is created from the connection of mobile phones through a shared content medium (normally the web) a mobile social network. Mobile social networks present huge opportunities for businesses and for users to take advantage of real-time interactive and user-generated content for providing recommendations, search, and personalization. However, there has been few research into how to analyze this mass user-generated input and social interactions [2] that arise from the social networking data. It also becomes important for users to connect with the relevant people in order to grow their online and mobile social networks and to provide quality friend relationships, thus preventing friend spam (where users want to become your friend but you do not have anything in common with them). As a result, research is needed into developing recommendation algorithms for finding the relevant people that users should connect with.

Friend recommendations are already implemented in online social networking sites such as Facebook and LinkedIn. These recommendations are typically based on analyzing user profiles for similar content (like attending the same company or university, or having similar interests) and friend-of-a-friend relationships. However, there is no mention as to why users should connect with these people or how they are relevant. Ranking algorithms also suffer from the effects of friend and comment spam. For example, TwitRank and other Twitter ranking engines rank Twitter people based on the number of followers and number of comments, but this fails if the comments are spam. Therefore, there is a need to examine the link connections within interactions in the online social network, as echoed by Chun et al [3], in order to identify relevant people who can be the recommended friends. We can find relevant people by finding cohesive subgroups, that is, subgroups of people that have intense interactions within the group than outside the group [1].

In this paper, we analyze the friend network and interaction network in a mobile social network using the Friend View service created by Nokia Research Center, in order to describe the social behaviour and identify cohesive subgroups using the modified version of the SCAN method in [1]. For convenience and practicality, we identify relevant members as the top 10 ranked members based on highest betweenness, degree and closeness centrality. We make comparisons between the top 10 members of the friend network and the interaction network, and discover that subgroups formed in the interaction network are based on friend relationships, and that subgroup members send more comments and have more friends than others in the interaction network.

Our contributions are the following. First, we describe a novel method based on [1] for finding cohesive subgroups and for finding relevant members in mobile social networks. Second, we demonstrate the viability of our method by applying it to a case study of Friend View. Third, we outline a method that can be used for ranking users and forms the basis for creating friend recommendations.

The remainder of this paper is structured as follows. Section II describes related work and Section III explains the SCAN method used to find cohesive subgroups. We present an overview of Friend View and describe the usage statistics in the Friend View mobile social network from analyzing the friend and message interactions and for characterizing social behaviour in Section IV. Section V describes our method and results for finding cohesive subgroups and the top 10 members for friend and interaction networks in Friend View. We discuss the results of the analysis in Section VI in terms of how the friend network influences the interaction network, and the characteristics of cohesive subgroup and non-subgroup members. Finally, we provide conclusions and recommendations for future work in Section VII.

978-0-7695-3823-5/09 $26.00 © 2009 IEEE
DOI 10.1109/CSE.2009.152

II. Related Work

There has been much work in social network analysis and studying structural properties of online social networks for characterizing and describing social behaviour. However, there has been only few work to date for mobile social networks due to the privacy and security issues that are involved.

A. Characterizing and describing social behaviour

Researchers have used statistical analysis [3]–[5], and social network analysis [3], [6] for characterizing social behaviour in online social networks. Java et al [6] used social network analytic measures such as degree, clustering coefficient and reciprocity to study the topological properties of Twitter and why people use Twitter. Leskovec et al [5] analyzed an instant messaging network and analyzed communication statistics such as number of friends, duration of conversations, and time interval between each conversation. However, none of them analyzed whether the interaction between users in the social network followed the declaration of friend relationship.

Chun et al [3] demonstrated that the friend network is similar in structure to the interaction network (which they call the activity network) and they conducted a detailed social network analysis into the structural characteristics of the friend and activity networks of CyWorld and their correlations. Even though we also compare the friend and interaction networks, our work is different in that we do not examine the structural characteristics of friend and interaction networks, but rather examine the problem of identifying relevant members in the friend and interaction networks by finding cohesive subgroups.

There has been little work in performing the above analysis for mobile social networks [2] such as mobile phone networks and mobile online networks due to the security and privacy issues in obtaining that data. However, with access to mobile phone logs and call records, researchers are making use of social network analysis to mine these datasets for characterizing the structure [7] and addressing the role and strength of social ties in the formation and growth of groups, or communities [8].

B. Finding subgroups and relevant people

Huberman et al [4] showed that the most relevant active people in your social network are actually your friends, from their Twitter study. Much work has looked into finding relevant people within communities and subgroups using social network analysis [6], [9], network centrality [10] and clustering [11]. Social network analysis is now being used to mine call graphs derived from call logs obtained from telecom networks to find dense communities [7], where each user communicates with a large subset of the remaining people and the subgraph has high average degree.

However, the problem becomes which measures are appropriate for identifying subgroups, are computationally efficient, and are accurate for finding subgroups. Our previous work [1] demonstrates that our method is superior over clique and related methods, is simple to implement, and provides accurate subgroups.

The next section will explain about our method for finding cohesive subgroups.

III. Method for Finding Cohesive Subgroups: The SCAN Method

The SCAN method is a three-step process for identifying cohesive subgroups on the basis of social networks inferred from online interactions [1]. The SCAN method assumes that the social network has been previously inferred using one of a number of data mining and crawling techniques that are available. In this paper, we deal with static networks with one period of time, so the SCAN method can be reduced to two steps which we explain below.

A. Select

In the first step, the possible members of cohesive subgroups are identified from the social network. We set a cutoff value on a measure that is assumed to be correlated with likelihood of being a subgroup member, and then filter out people who fail to reach the cutoff value on that measure. We can first use betweenness centrality as this cutoff measure as it is a good measure for finding subgroups from previous research reviewed in [1], followed by other centrality measures such as degree and closeness centrality because they have been shown to be strongly correlated with a strong sense of community, which is a behavioural trait of cohesive subgroups [12]. The cutoff centrality can be obtained by creating a frequency distribution graph of centralities of all members in the social network and then selecting the cutoff as a break in the distribution that separates a smaller group of higher centrality members from the larger group of people with lower centralities. From this cutoff centrality measure, we obtain a subgraph of the original social graph where all members that have a centrality below the cutoff centrality measure are removed, due to their overall low connectivity to, and cohesion with, others in the social network. This results in a list of potential active members of subgroups.

B. Collect

In the second step, the objective is to recognize active subgroups from the subset of network members identified in the Select step. This is accomplished by forming subgroups of the selected members using cluster analysis, specifically weighted average hierarchical clustering because this is more computationally efficient than k-plex analysis [12]. The output of hierarchical clustering is a set of nested, non-overlapping clusters (or tree) in a dendrogram. The extraction of the hierarchy shows potential cohesive subgroups, but it does not actually partition the people into a particular set of non-nested subgroups.

IV. The Friend View Mobile Social Network

In this section, we introduce an overview of Nokia Friend View explaining the user interface and its operation, followed by a summary of usage statistics from anonymized Friend

View data from the first 80 days that the Friend View service started. For privacy considerations, GPS location is not included in the dataset.

A. Overview of Friend View

Friend View is a location-enhanced microblogging application and service from Nokia Research Center that was launched in Nokia Beta Labs in November 2008, and is available for public use at friendview.nokia.com. It allows users to post messages about their status and activity with other friends in their social network from GPS-enabled Nokia S60 phones or from the web. A friend request is made by specifying an e-mail, finding a user in the Friend View directory, or adding a user who has commented on a message originated by a user's friend, and then a friend request is sent. The recipient to be added as a friend, can decide to accept or ignore the friend request. If the recipient accepts the friend request, the two users become friends and can see each other's message updates and threads. Users can send two types of messages in Friend View, an "I am here" message which just posts the current location coordinates of where the user is from the phone without writing a message, or a message which posts the location or not. In the Friend View mobile phone interface, users can see all their friends' locations on a map which is navigable and zoomable, along with their messages, and comments from others.

B. Usage Statistics in Friend View

Friend View has two different types of networks, a *friend network* which consists of all users and their friends (from acceptance of friend invitations), and an *interaction network* which consists of all the posted messages by Friend View users and their associated comments. From the gathered data of the first 80 days when Friend View started, we obtain the usage statistics of approximately 7000 users of Friend View and divide the results into three types: friend statistics, message statistics, and comment statistics.

1) Friend statistics: We discover that most people (around 70 percent) have small social networks of less than 10 friends, with few people having large numbers of friends (50 or more). The majority of users accept all friend invitations with very few users that ignored 2 or more friend invitations.

2) Message statistics: Most people post "I am here" messages during the morning getting into work, with heavy usage in the early morning and late evening. People generally write messages less than 200 characters in length, similar to Twitter that has a limit of 140 characters. Most Friend View users post at most 25 messages within the 80 days, while the rest post less than 200 messages. Most messages have a very short lifetime of less than 3 days (the time between when the message is posted until the last comment to the post), indicating messages are discussed almost instantaneously and have a short attention span.

3) Comment statistics: Most users post less than 50 comments, with few users that send more than 300 comments. The majority of Friend View users write very short comments of less than 50 characters, with an average of less than 5 comments to each message. On the other hand, a small number of users write long comments averaging at least 100 characters, but on average, these users receive more than 10 comments to their messages.

V. Finding Subgroups and Relevant Members in Friend View

We apply our method presented in Section III to find cohesive subgroups and relevant members in the Friend View friend and interaction networks, and to answer the following research questions:

1) Does the friend network influence the interaction network? That is, are the connections in the interaction network representative of the friend connections in the friend network?
2) Do subgroup members send more comments than others in the interaction network?
3) Do subgroup members have more friends than others in the interaction network?

A. Data Collection and Procedure

We obtained the friend network data from the Friend View team which consisted of pairs of anonymized ids which indicate that the two ids are friends of each other. The interaction network data was also obtained and consisted of pairs of anonymized ids indicating that the first member made a comment to the second member's comment or original message. Due to privacy concerns, we did not obtain the location of the Friend View member that posted the message. We then apply the SCAN method presented in Section III to the friend and interaction networks. In order to better assess and evaluate the cohesive subgroups formed and the relevant members, we decide to take the top 10 ranked members that have the highest centrality ordered by betweenness, degree then closeness centrality from all the members identified in the Select step.

B. Subgroups and Relevant Members in the Friend Network

We construct the Friend View friend network as a directed graph *G(V,E)*, where *V* represents the set of Friend View members and *E* represents the set of friend relations similar to Java et al [6] where they constructed a friend network in Twitter. This results in the Friend View friend network consisting of 3374 Friend View users that have 1 or more friends and altogether having 9110 ties, with each Friend View member having an average of 3 friends. Using the frequency distribution for betweenness, closeness and degree centrality as in [1], the cutoffs that we obtain are 0.000207 for betweenness, 0.05 for closeness, and 0.0008 for degree for selecting the members. From the list of members that meet this criteria, we rank order the members in descending order according to betweenness, then degree and closeness centrality. Next, we select the top 10 users from the ranking and their friend networks as shown in Figure 1, where the large node indicates

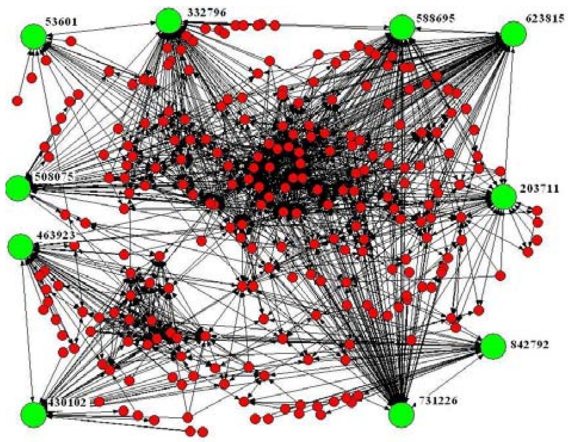

Fig. 1. Top 10 Friend View members and their friend networks (from the Select step)

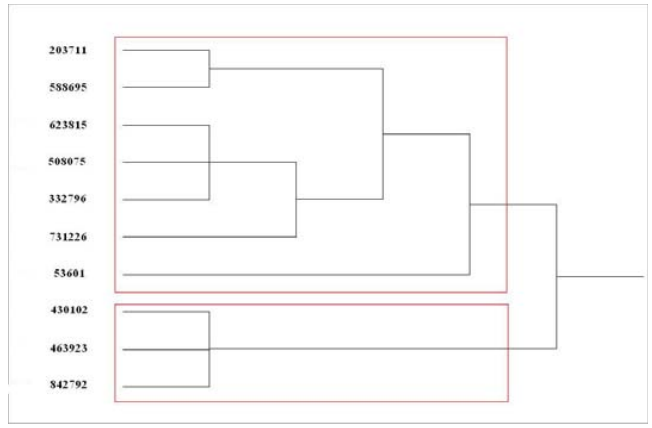

Fig. 3. Cohesive subgroups and dendrogram of the top 10 Friend View members in the friend network (from the Collect step)

one of the top 10 users and the small nodes are the friends of the users.

We see that the top 10 Friend View members' friend networks are highly connected with a total of 264 friends and 2314 friend ties. To determine whether the top 10 Friend View members which we select according to our criteria also are friends of each other and if they form a cohesive subgroup, we extract just the large nodes and their associated ties from Figure 1 and perform the Collect step. This results in Figure 2 which shows that the top 10 Friend View members from the Select step indeed form a cohesive subgroup.

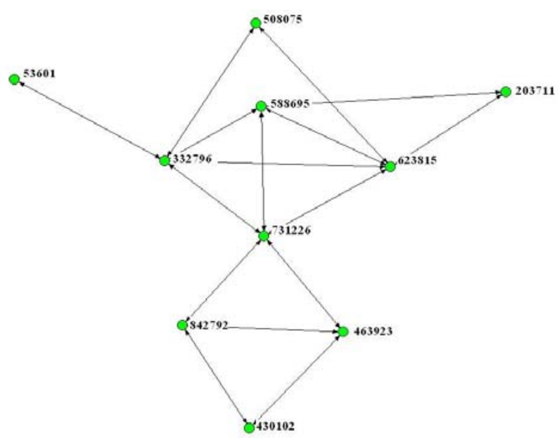

Fig. 2. Cohesive subgroup formed from the top 10 Friend View members from the friend network (large nodes of Figure 1)

For the Collect step, the results of performing weighted average hierarchical clustering on the top 10 Friend View members, is the dendrogram in Figure 3 which illustrates two cohesive subgroups of friends indicated by rectangles. These subgroups are then joined together by member 731226 to form the top 10 members' cohesive subgroup.

From our analysis, we discover that the top 10 relevant members in the Friend View friend network have a minimum of 1 friend, a maximum of 5 friends, and an average of 3.2 friends.

C. Subgroups and Relevant Members in the Interaction Network

We construct the Friend View interaction network as a directed graph *G(V,E)*, where *V* represents the set of Friend View members and *E* represents the set of comments to messages by members, where a directed edge or tie exists between node A and node B if node B posted a comment in response to node A's message and the edge weight *w* indicates the number of comments that B made in response to all of A's messages. This results in the interaction network consisting of 1150 Friend View users and 1947 ties, which is significantly less in number compared to the friend network, and each user has an average of 8 users that they converse with.

For the Select step for finding cohesive subgroups, we remove edges with weights less than 2 in order to focus on the active members that comment on messages, similar to the approach of Huberman et al [4]. We remove all unconnected nodes from the network which results in 620 unique members. We obtain the following centrality cutoffs based on betweenness, degree, and closeness centrality distributions: betweenness of 0.001, degree of 0.002, and closeness of 0.05. We select the members whose centrality values are above the cutoff, rank order the members in decreasing order ordered by betweenness, then degree and closeness centrality. The top 10 members from this ranking is chosen along with their egonetworks, which forms the network graph of Figure 4 consisting of 262 members and 835 ties.

We can see that the top 10 members (the large nodes) from Figure 4 are highly connected compared to the other members (the small nodes). To determine if the top 10 members from the interaction network of Figure 4 form cohesive subgroups, we extract just the top 10 members and their ties with each other and present the resulting subgraph in Figure 5.

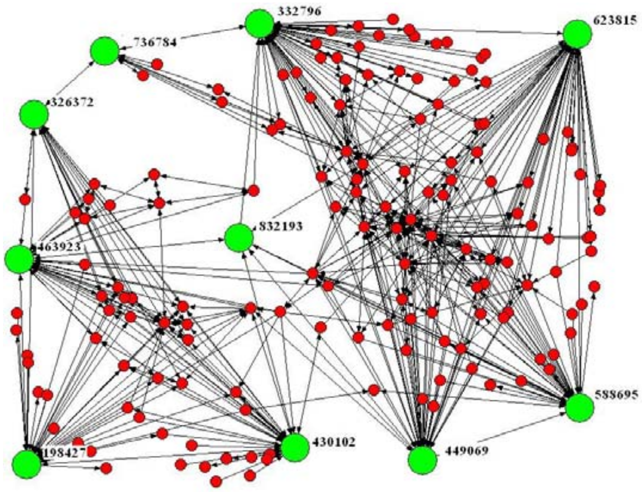

Fig. 4. Social network of selected members from Friend View interaction network with highlighted top 10 members

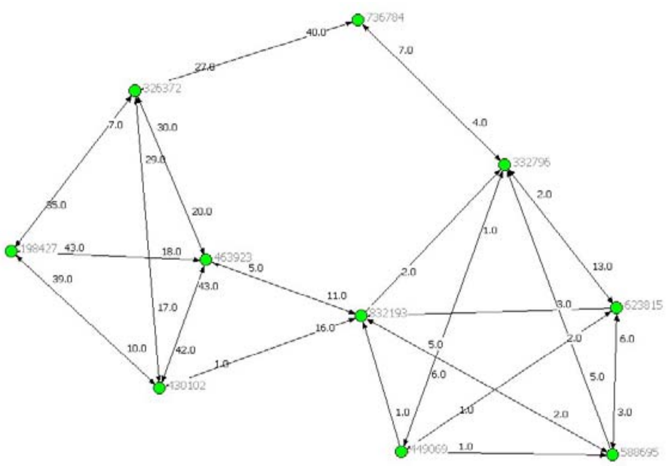

Fig. 5. Cohesive subgroup formed from the top 10 Friend View members in the interaction network (large nodes of Figure 4)

Figure 5 shows that the top 10 Friend View members from the interaction network of the Select step, do indeed form a cohesive subgroup and comment on messages with other members in the subgroup. The edge weights indicate the number of comments made. Members 832193, 332796, 449069, 588695, and 623815 form a completely connected graph. The top 10 members comment an average of 212.9 times throughout the first 80 days of Friend View use, with a minimum of 39 and maximum of 612.

Figure 6 illustrates the dendrogram from weighted average hierarchical clustering in the Collect step, showing two independent cohesive subgroups highlighted by rectangles which are joined together to form the larger cohesive subgroup.

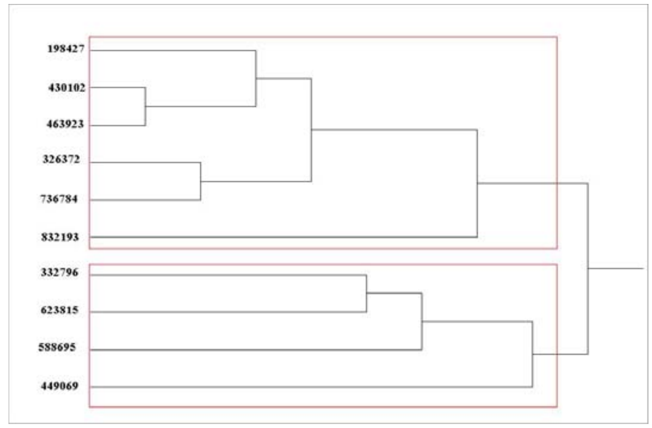

Fig. 6. Dendrogram of the top 10 Friend View members from the interaction network (from the Collect step)

VI. Discussion

In this section, we discuss the influence that the friend network has on the interaction network, and whether subgroup members send more comments and have more friends than others in the interaction network.

A. Does the friend network influence the interaction network?

We discover that Friend View members that are declared as friends, also make comments to each other's messages, or members that comment on each other then become friends. We cannot clearly distinguish which is the more prominent case, because the time that each friend invitation is accepted and the time when the comment is made, are not recorded in the dataset. Suffice to say, the most relevant members in the friend network, as discovered from our method, are also the most relevant members in the interaction network, as shown in Table I in bold.

TABLE I
Top 10 members in the Friend View friends network and the interaction network ranked in descending order of betweenness, degree and closeness centrality (from the Select step)

Top 10 members in friends network	Top 10 members in interaction network
508075	**332796**
731226	**463923**
332796	**588695**
842792	832193
203711	**623815**
623815	736784
463923	326372
588695	449069
430102	198427
53601	**430102**

We discover that 9 out of the top 10 members from the friend network are also in the interaction network, and that 5 out of the top 10 members from the friend network are in the top 10 members from the interaction network as indicated by the bolded members in Table I. Therefore, the friend network does influence the interaction network.

B. Do subgroup members send more comments and have more friends than others in the interaction network?

Members of cohesive subgroups send significantly more comments than others throughout the entire 80 days of study. Specifically, subgroup members post an average of 212.9 comments compared to 7.775 for the entire interaction network. Subgroup members also have more friends than others with average of 57.8 friends compared to 3.075 friends for the entire network. By comparing the cohesive subgroups formed by the top 10 members in the interaction network to the cohesive subgroups formed by the top 10 members in the friend network, we discover that the cluster orderings exactly follow the declaration of friend relationships between the members. That is, members that are friends of each other, are clustered together from their conversations to form subgroups as shown in Figure 6. Therefore, members of cohesive subgroups in Friend View interact with each other significantly more than non-subgroup members and are most likely to be friends with each other.

C. Friend recommendations

The cohesive subgroup members that are common in the friend and interaction networks can then be used as the basis for friend recommendations, because these are the members that are active and influential to the social network. For large networks, the top X of the members in rank order from the Select step in the friend network that are also in the interaction network, can be recommended as friends to a particular user if they are in the user's ego-centric interaction network and not in the user's ego-centric friend network. If the friends have a profile, then the profile can be compared with the user's profile to determine any similarities which can be used to even further refine and personalize the friend recommendations.

VII. Conclusions and Future Work

In this paper, we presented preliminary results on Nokia Friend View usage on anonymized data from the early months (first 80 days) of the Friend View service, for characterizing social behaviour and for finding cohesive subgroups and their relevant members. Our contributions are the following. First, we described a novel method called the SCAN method and modified it for finding cohesive subgroups and for finding relevant members in mobile social networks. Second, we compared the friend and interaction networks for a mobile social network to determine if friend relationships affect interaction networks. Third, we outlined a method that can be used for ranking users and forms the basis for friend recommendations. We discovered that the interaction network for the top 10 members derived from the SCAN method, resembled the friend network of those members in Friend View. That is, the top 10 members in the interaction network are friends of each other.

Our work attempts to address the gap in information analysis in mobile social networks [2], but there are many areas for future work. First, our analysis was based on link analysis of users without looking into the content of the messages and the comments which could be irrelevant. Second, our method for finding cohesive subgroups looked at a static dataset in time, which was the first 80 days. It is recommended to analyze the Friend View dataset in different time periods in order to observe how subgroups evolve and if relevant members change over time. A third area for future work is to create a recommendation and ranking algorithm based on our method and provide it to users to see if they would accept these recommended friends. Finally, our analysis can be applied to other social networks such as Facebook to compare with our results.

Acknowledgment

The authors would like to thank the Nokia Friend View team for their input, and Davin Wong for providing the Friend View dataset and usage statistics for this paper.

References

[1] A. Chin and M. Chignell, "Automatic detection of cohesive subgroups within social hypertext: A heuristic approach," *New Review of Hypermedia and Multimedia*, vol. 14, no. 1, pp. 121–143, 2008.

[2] C. Zigkolis, Y. Kompatsiaris, and A. Vakali. (2009, Jan.) Information analysis in mobile social networks for added-value services in w3c workshop on the future of social networking. [Online]. Available: http://www.w3.org/2008/09/msnws/papers/mobile-social-data-analysis.pdf

[3] H. Chun, H. Kwak, Y.-H. Eom, Y.-Y. Ahn, S. Moon, and H. Jeong, "Comparison of online social relations in volume vs interaction: a case study of cyworld," in *IMC '08: Proceedings of the 8th ACM SIGCOMM conference on Internet measurement*. New York, NY, USA: ACM, 2008, pp. 57–70.

[4] B. Huberman, D. Romero, and F. Wu, "Social networks that matter: Twitter under the microscope," Social Computing Lab, HP Laboratories, Tech. Rep., 2008.

[5] J. Leskovec and E. Horvitz. (2008) Planetary-scale views on an instant-messaging network. [Online]. Available: http://arxiv.org/abs/0803.0939

[6] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter: understanding microblogging usage and communities," in *WebKDD/SNA-KDD '07: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*. New York, NY, USA: ACM, 2007, pp. 56–65.

[7] V. Pandit, N. Modani, S. Mukherjea, A. Nanavati, S. Roy, and A. Agarwal, "Extracting dense communities from telecom call graphs," Jan. 2008, pp. 82–89.

[8] K. Dasgupta, R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjea, A. A. Nanavati, and A. Joshi, "Social ties and their relevance to churn in mobile telecom networks," in *EDBT '08: Proceedings of the 11th international conference on Extending database technology*. New York, NY, USA: ACM, 2008, pp. 668–677.

[9] N. Du, B. Wu, X. Pei, B. Wang, and L. Xu, "Community detection in large-scale social networks," in *WebKDD/SNA-KDD '07: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*. New York, NY, USA: ACM, 2007, pp. 16–25.

[10] J. R. Tyler, D. M. Wilkinson, and B. A. Huberman, "E-mail as spectroscopy: Automated discovery of community structure within organizations," *The Information Society*, vol. 21, no. 2, pp. 143–153, 2005.

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

[12] A. Chin and M. Chignell, "Identifying subcommunities using cohesive subgroups in social hypertext," in *HT '07: Proceedings of the eighteenth conference on Hypertext and hypermedia*. New York, NY, USA: ACM, 2007, pp. 175–178.