Using Cohesive Subgroups for Analyzing the Evolution of the Friend View Mobile Social Network

Alvin Chin[1], Hao Wang[1]

[1] Nokia Research Center
Building 2, No. 5 Donghuan Zhonglu, Economic and Technological Development Area
Beijing, China, 100176
{alvin.chin, ext-hao.10.wang}@nokia.com

Abstract. The mobility of users and the ubiquity of the mobile phone and Internet are leading to the development of mobile social networks. Much work has been done on modeling the evolution of online social networks using mathematical, social network analysis, and graph theoretic methods, however few using cohesive subgroups and similarity. In this paper, we present a study of the evolution of the Nokia Friend View mobile social network using network and usage statistics, and use the DISSECT method [7] for characterizing this evolution through the movement of cohesive subgroups. We discover that the friend network becomes less dense and less clustered (with fewer subgroups) over time, and the DISSECT method [7] helped to identify these cohesive subgroups and accurately predicted its most active users. We visualized these cohesive subgroups and modeled the evolution using persistence of subgroups. These results point the way towards an analytical framework for comparing mobile social networks which may help facilitate development of new recommender applications.

Keywords: Mobile social network, social network evolution, social network analysis, subgroup identification, cohesive subgroups, centrality, similarity modeling.

1 Introduction

Online social networks such as Facebook and LinkedIn are increasingly being used for sharing content and keeping in contact with friends, colleagues and family. With the ubiquity of the mobile phone and wireless technologies, location can be added as a context to localize the content such as location of photos taken and status updates, thus creating mobile social networks. Mobile social networking applications such as Foursquare and BuddyCloud use location of users to provide services such as finding people and places nearby, providing relevant content, providing search and updated points of interest, and creating specific topic channels from which other people can subscribe to.

Previous work has studied the structure and properties of online social networks [9, 17, 18, 22] and their evolution [1, 3, 17, 22, 24]. Most use social network properties over time or create models of evolution using group formation, clustering and partitioning, or mathematical modeling and graph theory, but fail to enumerate the cohesive subgroups and their persistence over time. There have been few studies of analyzing cohesive subgroups in the evolution of a mobile social network because the data is difficult to obtain. Therefore, our objective is to use cohesive subgroups as the method for describing the evolution of a mobile social network. Our research questions are the following. First, what are the cohesive subgroups of people that exist in a mobile social network over time? Second, which subgroups are ephemeral and which subgroups persist over time, and third, how do the persistent subgroups evolve and change over time?

In this paper, we analyze Friend View, a mobile social network created by Nokia Research Center, by quantifying its network properties, and then describe the evolution of Friend View in terms of its cohesive subgroups and how they change over time using the DISSECT method of Chin and Chignell [7]. Friend View is a service that allows mobile users to find others, update their status and location, add friends, and keep updates with others. We obtained the dataset of Friend View with the user names being anonymized. We discover that the DISSECT method helped to find cohesive subgroups in the interaction network, their persistence over time, and accurately predicted the most active users.

Our contributions are the following. First, we study the evolution of a mobile social network from the first day that it started to its last day. Although it can be argued that the dataset for Friend View is short (only 11 months), it provides us with the complete data for the entire network, as opposed to others who study only a subset of the entire social network. Second, we identify cohesive subgroups in different time periods during the evolution, and classify their persistence based on how the members within the subgroups move from one subgroup to another, split into several subgroups, or stay within the same subgroup.

The paper is organized as follows. Section 2 describes related work on identifying subgroups and modeling the properties and evolution of online social networks. In Section 3, we introduce and describe the user interface and usage statistics of Friend View along with the number of users and friend pairs over time. In Section 4, we analyze the comment interaction network of Friend View by using the DISSECT method of Chin and Chignell [7] to track the cohesive subgroups and their persistence and movement over time. In Section 5, we discuss the implications of the Friend View analysis and applications of our work. Finally, we conclude the paper in Section 6 and provide areas for future work.

2 Related work

There has been much work in studying the properties and structure in online social networks such as Twitter [18], Wealink [17], Yahoo! 360 and Flickr [22], and Cyworld [9], however there have been only a few works studying the evolution of social networks by analyzing cohesive subgroups over time. In this section, we review methods for finding subgroups and modeling evolution.

2.1 Finding Subgroups

Most available methods for identifying subgroups are based on some combination of the following measures and techniques: centrality, cohesiveness, and clustering and partitioning. Centrality [16] identifies the most important active people that are well connected in the network. Centrality is a useful predictor of subgroup membership because those who are actively involved in one or more subgroups will generally score higher with respect to centrality scores within the surrounding network. Betweenness centrality has been used to find and measure subgroup and community membership [31, 38], whereas degree [14, 29, 41] and closeness centrality [23, 28] have been used for characterizing influential members.

Cohesive subgroups within social networks can indicate the most active members within a community [15, 40]. Cliques and k-plexes have been used to characterize groupings in social networks [8, 13], but are not suited to large networks because their computational complexity scales exponentially with the number of nodes in the network and their discovery is an NP-complete problem [2].

Clustering and techniques such as link analysis [4, 19] and co-citation analysis [20, 21] can be used to detect subgroups. Hierarchical clustering is often used to quantify the structure of community in web networks (e.g., [10, 12]) where the cluster orderings in the dendrogram form the subgroups. In contrast, groups formed in partitioning methods are not nested, but partitioning techniques are computationally efficient. Criteria and methods aimed at identifying optimal partitions include modularity [34], vector partitioning [39] or normalized cut metrics [24] for finding subgroups.

Hierarchical clustering has been shown to produce similar subgroupings as k-plex analysis for some data examples and is less computationally intensive [6]. Modularity has been proposed as an optimizing method for partitioning dendrograms [30]. Sometimes clustering and partitioning algorithms are combined in order to identify subgroups (e.g. [26]). However, little evaluative research has been carried out for determining which methods of unsupervised subgroup formation work well in subgroup analysis of social networks, and under what conditions.

2.2 Modeling Evolution

Online social networks evolve over time and much research has looked into the temporal aspects of social networks changing over time such as Leydesdorff et al [25] and Snijders et al [35]. Within social networks, subgroups of people may be found that vary in cohesiveness [33].

Using social network analysis, Kumar et al [22] analyzed the structure of Yahoo! 360 and Flickr networks and Barabási et al [3] analyzed scientific collaborations over time, to create a model of evolution and used simulation to test the model. Hu and Wang [17] studied the evolutions of degree, network density, clustering coefficient, number of users, modularity, and degree assortativity, in order to reveal the properties and evolutionary patterns of the Wealink online social network.

Other researchers have created models of evolution using group formation, clustering, and partitioning methods. For example, Backstrom et al [1] developed a method for measuring the movement of individuals between communities, examined the properties of membership of how groups formed, and identified which communities grew over time. Tang et al [37] adopted a spectral clustering framework using the temporal information to detect, identify and model community evolution in dynamic multi-mode networks. Cortes et al [11] proposed a bottom-up data structure to represent all small subgroups based on the "Communities of Interests" concept on each user in the dynamic network, and then updating all communities. Palla et al [32] quantified the social group evolution using cliques.

Mathematical modeling and graph theory can be used for modeling evolution of communities. Lin et al [27] detected communities using a non-matrix factorization and iterative algorithm for solving the evolutionary clustering problem in Chakrabarti et al. [5]. Sun et al [36] proposed a tool named GraphScope, based on information theory, to monitor communities and their membership changes in a stream of graphs efficiently. Leskovec et al [24] created a mathematical theoretical model for characterizing and describing the densification and shrinking diameter phenomenon in social networks over time.

For modeling evolution, few researchers have looked into the visualization and activity of the most influential members in the network over time, concentrating instead on the statistics of the network. Chin and Chignell [6, 7] proposed a similarity modeling approach to quantify changes in subgroup structure over time. In this approach cohesiveness over time is quantified in terms of the similarity of the subgroupings that are identified in different time periods. We adopt this approach for studying the evolution of a mobile social network.

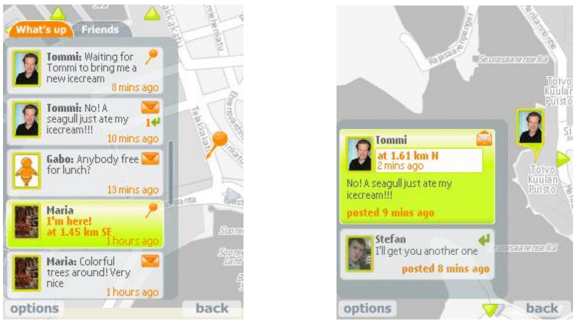
3 Nokia Friend View

In this section, we introduce Nokia Friend View and its user interface. We then report on usage statistics and network properties of Friend View.

3.1 User Interface

Friend View is a location-based microblogging service from Nokia Research Center that was launched in the beginning of November 2008, and was discontinued at the end of September 2009 since it was an experimental service. It allowed 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.

Friend requests can be sent manually by entering the person's name or selecting the person who commented on the original status message. If the recipient accepts the friend request, the two users become friends and can see each other's message updates and threads. Users can post an optional status message with annotated GPS location. In the Friend View mobile phone interface shown in Figure 1, users can see all their friends' locations and status messages on a map in the What's up tab (a), and view and make comments to other conversation threads started from status messages (b).



(a) (b)

Fig. 1. Mobile phone user interface of Friend View showing (a) status messages from friends showing location on a map, and (b) comments to a friend's status message.

3.2 Usage Statistics

We obtained a complete dataset of Friend View interactions for the entire period of its operation that consisted of all users that posted status messages and comments. Consistent with ethics requirements, all data was anonymized prior to analysis, conformed to standard personal data protection laws, and people were labeled with unique, but not identifying, codes. The usage statistics, broken down into user and friend statistics, along with status message and comment statistics are summarized below.

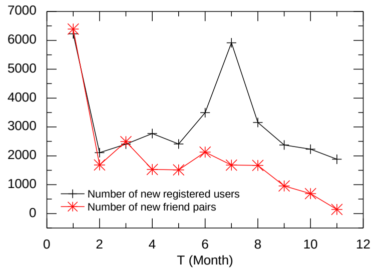
A total of 62736 status messages were posted by 16176 users, providing an average of 3.88 status messages per user. Out of the 62736 status messages, 9363 status messages had associated comments that were posted by 2395 users, and 22251 comments were posted from 2283 users, producing a mean of 3.91 status messages with comments per user, and a mean of 2.38 comments per status message. Overall, users posted a small number of status messages (around 20, and generally a lot fewer), and a small number of comments (around 10 or less), but there were a few users in the long tail of the distribution that had an excessive number of status messages and comments that amounted in the hundreds and even thousands.

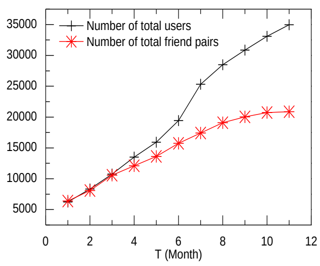
3.3 Network Statistics

We studied the network statistics of Friend View with the number of users and friend pairs over its entire duration. We extracted 11 snapshots of the dataset with an interval of one month from November 1, 2008 to September 30, 2009, like that in [17]. Figure 2 (a) shows the number of new users and new friend pairs for every month over the lifetime of Friend View, and Figure 2 (b) shows the total number of users and friend pairs.

The growth patterns for the number of new users and new friend pairs are similar in shape, with the highest number occurring at the beginning when many users joined Friend View and were eager to try out the service, and then decreases until the end of the service, with the exception of a few spikes like the seventh month. It is interesting that in the 7th month (June 2009), the number of new users encounters a huge spike and almost reaches the level from the beginning. We believe that this behaviour may have been caused as a result of a touch version of Friend View that was released during that time. However, the number of new pairs of friends in June 2009 decreases albeit very little. We speculate that if the huge spike of new users did not happen, then the number of new pairs of friends would decrease even more still and that value in August 2009 would not remain at the same level as the previous month. In Figure 2 (b), the number of users and friend pairs start small, rapidly accelerate, and then stabilize. In the case of the number of users, it still continues to grow while the number of friend pairs appears to plateau during the 10th and 11th months.

In the next section, we explain a method for modeling evolution by discovering cohesive subgroups over time.





(a) **(b)**

Fig. 2. Time evolution of (a) the number of new users and friend pairs and (b) the total number of users and friend pairs.

4 Modeling Evolution With Cohesive Subgroups Using the DISSECT Method

To model the evolution of a social network, rather than create a simulation, theoretical, or statistical model, we seek to understand the behavior of its relevant members. We determine who are the relevant members over time, how they are grouped with other members, and how relevant members move within or between subgroups, using the DISSECT (Data-Intensive Socially Similar Evolving Community Tracker) method of Chin and Chignell [7]. For Friend View, we apply the DISSECT method to the interaction network. The interaction network consists of users with status messages and associated comments which form the conversations among users and from which relevant members can be added as friends. The detailed description about the interaction network is given in Section 3.2.

The DISSECT method tracks evolution of multiple known subgroups in terms of similarity-based cohesiveness over time, which is an improvement over the original SCAN (Social Cohesion Analysis of Networks) method also by the same authors [30].

The main steps in the framework of DISSECT consist of: 1) finding the time periods for cohesive analysis, 2) selecting the possible members of latent cohesive subgroups based on a certain network centrality cutoff, 3) performing hierarchical cluster analysis of each network snapshot in each of the time periods to discover the latent subgroups, and 4) calculating similarity of the known subgroups and finding the best network centrality that results in the largest similarity between subgroups in successive time periods. For more detailed implementation and discussion of the DISSECT and SCAN methods, please refer to [7] and [30] respectively.

4.1 Steps in the DISSECT Method

4.1.1 Find the initial time periods for analysis

The dataset is divided into time periods for tracking subgroups in the network over time. Time periods should be long enough so that there is enough data to distinguish potential subgroups, and there should be a sufficient number of them to estimate cohesion over time.

4.1.2 Select the possible members of known subgroups to be tracked (from previous step) using Select from the SCAN method

While betweenness centrality appears to be a useful filter for screening potential subgroup members, other centrality measures such as degree and closeness centrality may also be used [6]. Degree centrality may be a good default measure with which to screen potential subgroup members because it deals with direct interactions where the ties having stronger bonds indicate stronger cohesion, and also because it has the lowest computational complexity compared to the other centrality measures [6].

4.1.3 Perform clustering of snapshots in time of known subgroups of people using the Collect step from the SCAN method

This step is identical to the Collect step from the original SCAN method, with the provision that other cluster methods may be used in addition to, or instead of, weighted average hierarchical clustering.

4.1.4 Repeat previous two steps for different values of centrality

Since there is as yet no known "best" or most appropriate centrality cutoff value for selecting potential subgroup members, a search process may be used to identify particular cutoff values that lead to identification of the most cohesive groups. Different types of search strategies may be employed but they would involve repetition of steps 4.1.2 and 4.1.3 over a range of values of centrality, with the goal of maximizing the cohesiveness (self-similarity over time) of the obtained subgroupings. In this case, the similarity measures suggested by [6] are recommended, although other similarity measures may also be used.

4.1.5 Select and characterize the obtained subgroupings in terms of their cohesiveness and their behaviour over time

The search process briefly described above may also be expanded by search over different definitions of time periods, as well as different centrality measures and centrality cutoff values. For instance, both the starting points and durations of time periods could be varied. It seems likely that strongly cohesive subgroups that remain intact over a sustained period of time should be "easy to find" with a range of time period definitions and centrality measurement and filtering strategies. In contrast, the search process envisioned above might be useful in finding more ephemeral subgroups that exist for only short periods of time and for tracking, in more detail, evolution in subgroupings.

4.2 Applying the DISSECT Method

We now apply the DISSECT method to the Friend View interaction network. The interaction network is a directed graph *G(V,E)*, where *V* represents the set of Friend View users and *E* represents the set of comments to status messages posted by users, where a directed edge exists between user A and user B if user A posted a comment in response to user B's status message and the edge weight *w* indicates the number of comments that A made in response to all of B's status messages. We then remove all nodes that have no edges.

We performed a link analysis (commenter and poster names) using the DISSECT method [7] to determine how well the method performs in the absence of content analysis (which is subjective, error prone, and time consuming). Due to the anonymized nature of the data, no labeling information was attached to the people in the network. The DISSECT method was applied as follows.

4.2.1 Define the time periods to partition the network into time snapshots

We define the time periods used for the evolution in Table 1 using Figure 2 to help guide us for the different phases of evolution based on the user growth.

Table 1. Time periods chosen for analysis in the evolution of Nokia Friend View.

Time period	Time range	Growth phase
T1	Nov. 1 to 30, 2008	Beginning, initial growth
T2	Dec. 1 2008 to Feb. 28, 2009	Early growth
T3	Mar. 1 to May 31, 2009	Rapid growth
T4	Jun. 1 to Jul. 31, 2009	Rapid slowing of growth
T5	Aug. 1 to Sept. 30, 2009	Growth decline

In the first month of Nokia Friend View (T1), many early technology adopters tried out the service, which continued into the first couple of months in T2. At the beginning of March through June (T3), due to the adoption by internal employees, Nokia fanatics and friend recommendations, the service spread to many social networking news sites, which led to Friend View's rapid growth. The service then reached its peak during June and July (T4), as most people already were using the service. Finally, with the announcement that Friend View was going to be discontinued during the beginning of September, this naturally led to the decline in growth in T5.

4.2.2 Selecting possible members

For each time period, we first selected the cutoff points for normalized betweenness and degree centrality (as suggested by [7]) in order to select the possible subgroup members whose centrality was above this cutoff value. To determine the betweenness and degree centrality cutoff points, their frequency distributions were inspected for each time period. Figure 3 shows the betweenness centrality frequency distribution for the first time period T1. This tends to result in a cutoff value that focuses subsequent analysis on a relatively small set of active people. The degree centrality frequency distribution is similar, and therefore omitted in this paper.

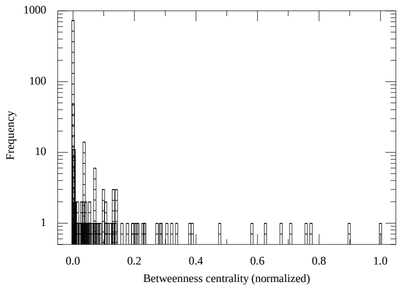


Fig. 3. Betweenness centrality distribution for the first time period (T1) in Nokia Friend View.

From this distribution, the cutoff was chosen for each centrality measure to be the largest value in the distribution that separates a smaller group of higher centrality members from the larger group of people with lower centralities [7]. In this case, a betweenness centrality cutoff of 0.2 was chosen for T1. Based on inspection of the corresponding distributions for T2 through T5, a betweenness centrality cutoff of 0.2 was also chosen for those time periods. For degree centrality, the same process was also used and a cutoff of 0.005 was chosen. Other values of betweenness and degree centrality selected for the search process were a betweenness centrality cutoff of 0.1 and a degree centrality cutoff of 0.03. Higher cutoff values were not considered in this case as they would have resulted in a very small number of people being considered for possible subgroup membership.

4.2.3 Forming possible cohesive subgroups

Once we selected the betweenness and degree centrality cutoff values, then for each cutoff value, all members in the network with betweenness centrality and degree centrality higher than the selected cutoff values are chosen for weighted average hierarchical clustering (from the DISSECT method), in order to find cohesive subgroups.

4.2.4 Similarity analysis and visualization for evolution of cohesive subgroups

We use modularity [30] to select the partition to cut the dendrogram (resulting from hierarchical clustering) and find the optimum possible cohesive subgroups. Next, to determine which of these optimum possible cohesive subgroups (computed for each betweenness and degree centrality cutoff) are the actual cohesive subgroups for that time period, we perform a similarity analysis to find the most similar overlap in cohesive subgroup members using the method of Chin and Chignell [7]. The selected betweenness and degree centrality cutoff values are determined from the highest similarity as computed using the method in [7], and are summarized in Table 2.

From the visualization of optimum possible cohesive subgroups (not shown here due to space limit), subgroup members tend to be tightly connected with each other, suggesting strong cohesion. As well, the number of cohesive subgroups decreases over time as the subgroup becomes larger, indicating that people are connected to new friends and their conversations.

Table 2. Centrality cutoffs for each time period based from the DISSECT method [7].

Time period	Betweenness centrality cutoff	Degree centrality cutoff
T1	0.2	0.005
T2	0.1	0.003
T3	0.1	0.003
T4	0.2	0.005
T5	0.1	0.003

Figure 4 shows the movement of people between the Friend View subgroups based on the similarity analysis and quantifies the subgroups similar to [32]. The members have been grouped together based on the subgroups found from the DISSECT method, with each member having a shape that corresponds to the time period where that member is first found in other time periods as indicated in the legend. The arrows indicate how the member or subgroup moved from one subgroup in the previous period to the next period. From this figure, it can be seen that there are two subgroups at T1 (3, 32, 37, 7, 9, 34, 16) and (5, 21, 17, 24, 28, 11, 32, 33, 13). The second of these groups started splitting up in T2 and had completely dispersed by T3. Across the entire trial only three people stayed together across all the time periods (7, 9, and 37) and a further two people paired up in T2 and then stayed together for the remainder of the trial. Other people moved around between the time periods, sometimes moving to a different subgroup and sometimes appearing to drop out and become singletons. Since the data was anonymized, it is not possible to identify who participated in the cohesive subgroups. However, ID numbers were assigned in the order of adoption (beginning with the number 1) so the low ID numbers shown in Figure 4 (all below 50) indicated that the people who participated in subgroups were the early adopters in the trial.

The subgroups shown in Figure 4 can be characterized in terms of how persistent they are over the time periods. Persistence is based on how long a user belongs to and stays in a subgroup. The most *persistent group* (37, 7, 9) remains together through all the time periods. *Semi-persistent groups* may be defined as groups that stayed together for some portion of the trial. They include the following groupings (6, 26) that persist from T2 through to T5 and (5, 21) from T1 to T2. *Temporal groups* have

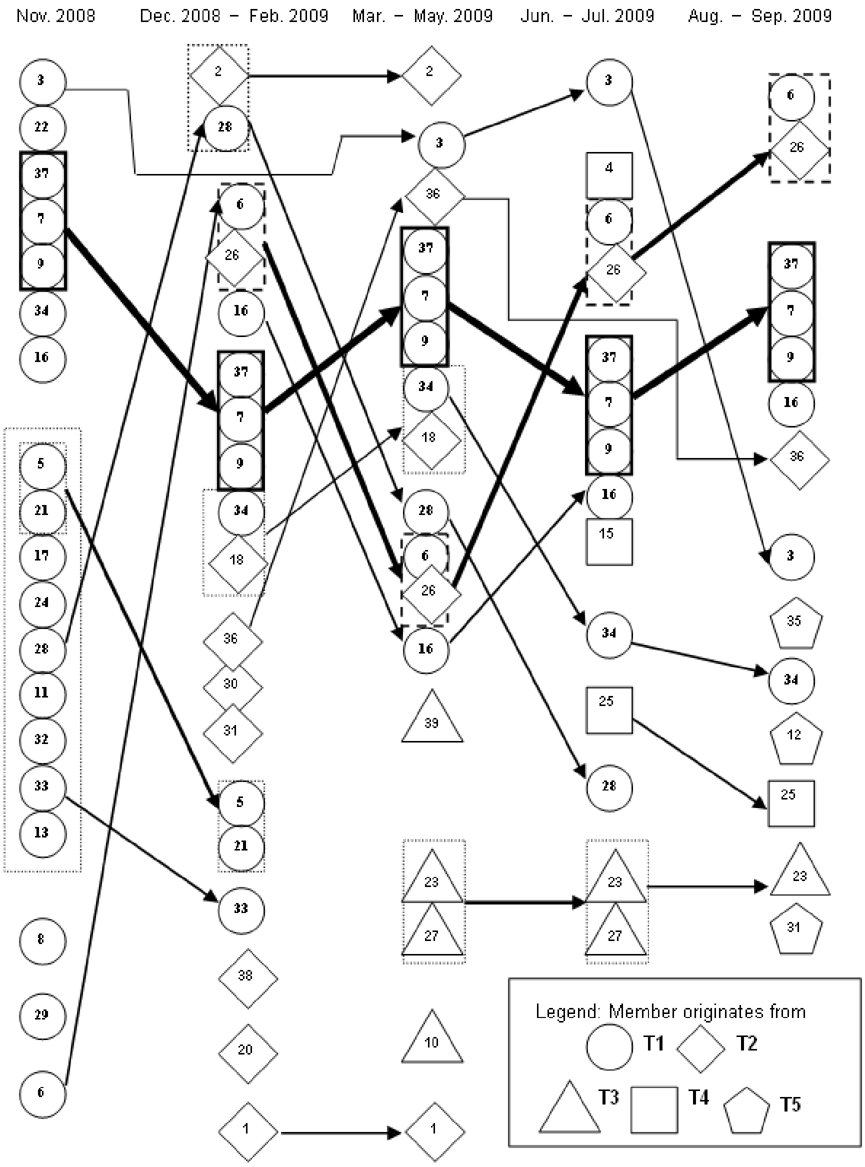


Fig. 4. Visualization of the evolution of members and subgroups in different time periods in Friend View.

persistent members in one period and then divide into different groups in the next time period. An example temporal group is (5, 21, 17, 24, 28, 11, 32, 33, 13) which formed in T1 but then divided into different groups in later time periods. *Ephemeral groups* may be defined as having members that are together in only one period such as (3, 22) in T1, and (30, 31) in T2. Based on this analysis, the different types of subgroups observed in the Friend View trial are listed in Table 3 according to their persistence. Note that since this analysis extends over five time periods, a particular person may participate in more than one type of subgroup at non-overlapping time durations.

We discovered very little evidence of new subgroups (involving new members) being formed after T2. Thus cohesive subgrouping activity was driven by the early adopters, and with most of the cohesive subgroup formation being done in the early part of the trial (although with some movement of members between and out of subgroups in later stages of the trial).

Table 3. Enumeration of all the different type of Friend View subgroup members discovered from the DISSECT method according to persistence

Type of subgroup	Subgroup members	Persistence
Persistent core subgroup	(37, 7, 9)	T1 to T5
Semi-persistent group	(6,26)	T2 to T5
	(5, 21)	T1 to T2
Temporal group	(5,21,17,24,28,11,32,33,13) (2,28)	T1, T2
Ephemeral group	(3,22) (30,31)	T1, T2

4.3 Cohesion and Message Activity

The DISSECT method for finding cohesive members in Friend View only takes into account the social network properties, but not the actual number of status messages posted by each member. Therefore, we wanted to see if members of cohesive subgroups tend to post more status messages, have a higher degree and betweenness centrality, and higher indegree and outdegree (with respective to unique members that make comments) than others that are non-members of cohesive subgroups. Since cohesive subgroup members tend to have more message activity and conversation than non-members [8], we expected that relationship to also hold in the Friend View data.

We discovered the number of unique people that a member makes comments to another member's status message or comment for that period, is the highest correlated with the number of status messages (posted at that period), followed by the number of unique individuals that make comments to a member's status message. Of those people selected using the centrality cutoff, people in the identified subgroups sent more messages as expected. In this sample, the correlations between betweenness centrality and message activity were of the same order as the correlations between degree centrality and message activity.

5 Discussion

Overall, Friend View had persistent core subgroups that stayed within all the time periods, as well as semi-persistent, temporal and ephemeral subgroups. Since the

Friend View dataset did not contain the actual content of the status messages and the comments, we could not perform a content analysis in order to more accurately filter out Friend View users that would not be part of cohesive subgroups. Since we did not perform any analysis concerning different sized durations of time periods, future work will involve developing similarity measures that take into account variable time windows and all possible combinations of time windows (rather than pairwise consecutive time windows). Some possible methods for selecting time periods could be probabilistic stochastic models [35], sliding time windows [30], and time graphs and burst analysis [1].

We discovered relatively few subgroups in the Friend View data, therefore based on our experience, large datasets should use degree centrality instead of betweenness centrality because betweenness centrality seems to be a better indicator of activity in a relatively small and densely connected network [7]. In very large networks (such as the one formed in the Friend View trial) where there are only isolated pockets of people who know each other, degree centrality (with its emphasis on whom the people are directly connected to) may be a better reflection of activity and a better filter for selecting people who are likely to belong to subgroups that might be cohesive. In addition, the few relatively small subgroups are probably as a result of selecting a relatively stringent centrality cutoff criterion, therefore, it is recommended to have algorithmic methods of centrality cutoff selection that supplement the visual inspection of the frequency distribution used in the research reported here.

As this study primarily dealt with investigating the detailed evolution using cohesive subgroups, we did not compare our results with other work, however this is reserved for future work.

6 Conclusion

In this paper, we studied the evolution of a mobile social network called Nokia Friend View using network properties. We then identified cohesive subgroups and relevant members using the DISSECT framework [7]. We discovered that the DISSECT method found different types of cohesive subgroups based on persistence in time: persistent, semi-persistent, temporal and ephemeral groups. We compared the persistent groups with the original post and comment statistics, and discovered that the most active members were the ones who were a part of the persistent groups.

Since Nokia Friend View has been discontinued, we do not have an opportunity to obtain a larger dataset. For future work, we plan to use other social network datasets to examine the value of the DISSECT approach. Relevant research issues include the choice of time periods using various statistical and time-based models, developing other measures for similarity assessment, and using content analysis to determine semantic properties that govern or explain subgroup membership. The DISSECT framework may also be applied to other online social networks to determine whether the similar cohesive subgroups along with network and message statistics, and the social graph of the friend and interaction network, can be used to improve friend recommendations.

Acknowledgements

We would like to thank James Reilly and Matti Sillanpaa from the Nokia Friend View team for providing us with the anonymized data set from Friend View.

References

1. Backstrom, L., D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: membership, growth, and evolution. . In ACM SIGKDD. 2006. pp. 44 - 54 ACM
2. Balasundaram, B., S. Butenko, I. Hicks, and S. Sachdeva, Clique relaxations in social network analysis: The maximum k-plex problem. Tech. rep., Texas A and M Engineering, 2008:
3. Barabasi, A., H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek, Evolution of the social network of scientific collaborations. Physica A: Statistical Mechanics and its Applications, 2002. 311(3-4): 590-614
4. Brin, S. and L. Page. The anatomy of a large-scale hypertextual web search engine. In WWW 1998. pp. 107 - 117
5. Chakrabarti, D., R. Kumar, and A. Tomkins. Evolutionary clustering. In Proceedings of the 12th ACM SIGKDD. 2006. pp. 554 - 560
6. Chin, A. and M. Chignell, Automatic detection of cohesive subgroups within social hypertext: A heuristic approach. New Review of Hypermedia and Multimedia, 2008. 14(1): 121-143
7. Chin, A. and M. Chignell, DISSECT: Data-Intensive Socially Similar Evolving Community Tracker. Computational Social Network Analysis. 81-105
8. Chin, A. and M. Chignell. Identifying subcommunities using cohesive subgroups in social hypertext. In HT 2007. 2007. pp. 175 - 178 ACM
9. Chun, H., H. Kwak, Y. Eom, Y. Ahn, S. Moon, and H. Jeong. Comparison of online social relations in volume vs interaction: a case study of cyworld. In Proc, of the 8th ACM SIGCOMM IMC Conference. 2008. pp. 57-70. ACM New York, NY, USA
10. Clauset, A., Finding local community structure in networks. Physical review E, 2005. 72(2): 26132
11. Cortes, C., D. Pregibon, and C. Volinsky, Communities of interest. Intelligent Data Analysis, 2002. 6(3): 211-219
12. Donetti, L. and M. Munoz, Detecting network communities: a new systematic and efficient algorithm. Journal of Statistical Mechanics: Theory and Experiment, 2004. 2004: P10012
13. Du, N., B. Wu, X. Pei, B. Wang, and L. Xu. Community detection in large-scale social networks. In 1st SNA-KDD. 2007. pp. 16-25. ACM
14. Fisher, D., Using egocentric networks to understand communication. IEEE Internet Computing, 2005. 9(5): 20-28
15. Fortunato, S., V. Latora, and M. Marchiori, Method to find community structures based on information centrality. Physical review E, 2004. 70(5): 56104
16. Freeman, L., Centrality in social networks conceptual clarification. Social networks, 1979. 1(3): 215-239
17. Hu, H. and X. Wang, Evolution of a large online social network. Physics Letters A, 2009. 373(12-13): 1105-1110
18. Java, A., X. Song, T. Finin, and B. Tseng. Why we twitter: understanding microblogging usage and communities. . In 1st SNA-KDD. 2007. pp. 56-65. ACM
19. Kleinberg, J., Authoritative sources in a hyperlinked environment. Journal of the ACM (JACM), 1999. 46(5): 604-632

20. Kleinberg, J., Bursty and hierarchical structure in streams. Data Mining and Knowledge Discovery, 2003. 7(4): 373-397
21. Kumar, R., J. Novak, P. Raghavan, and A. Tomkins, Structure and evolution of blogspace. Communications of the ACM, 2004. 47(12): 35-39
22. Kumar, R., J. Novak, and A. Tomkins. Structure and evolution of online social networks. In Proceedings of the 12th ACM SIGKDD. 2006. pp. 611 - 617. ACM
23. Kurdia, A., O. Daescu, L. Ammann, D. Kakhniashvili, and S. Goodman. Centrality measures for the human red blood cell interactome. Engineering in Medicine and Biology Workshop. 2007. pp. 98-101. IEEE
24. Leskovec, J., J. Kleinberg, and C. Faloutsos, Graph evolution: Densification and shrinking diameters. ACM Transactions on Knowledge Discovery from Data (TKDD), 2007. 1(1): 2
25. Leydesdorff, L., T. Schank, A. Scharnhorst, and W. De Nooy, Animating the development of Social Networks over time using a dynamic extension of multidimensional scaling. http://arxiv.org/pdf/0809.4655, 2008:
26. Li, N. and G. Chen. Analysis of a Location-Based Social Network. In Proceedings of the Intern. Confer. on Computational Science and Engineering. 2009. pp. 263-270. IEEE
27. Lin, Y., Y. Chi, S. Zhu, H. Sundaram, and B. Tseng. Facetnet: a framework for analyzing communities and their evolutions in dynamic networks. In WWW 2008. pp. 685-694. ACM
28. Ma, H. and A. Zeng, The connectivity structure, giant strong component and centrality of metabolic networks. Bioinformatics, 2003. 19(11): 1423 - 1430
29. Memon, N., H. Larsen, D. Hicks, and N. Harkiolakis, Detecting hidden hierarchy in terrorist networks: Some case studies. LNCS in Intelligence and Security Informatics: 477-489
30. Moody, J., D. McFarland, and S. Bender deMoll, Dynamic network visualization1. American journal of sociology, 2005. 110(4): 1206-1208
31. Newman, M. and M. Girvan, Finding and evaluating community structure in networks. Physical review E, 2004. 69(2): 26113
32. Palla, G., A. Barabási, and T. Vicsek, Quantifying social group evolution. Nature, 2007. 446(7136): 664-667
33. Piper, W., M. Marrache, R. Lacroix, A. Richardsen, and B. Jones, Cohesion as a basic bond in groups. Human Relations, 1983. 36(2): 93
34. Ruan, J. and W. Zhang. An efficient spectral algorithm for network community discovery and its applications to biological and social networks. Proceedings of the 2007 Seventh IEEE International Conference on Data Mining 2007. pp. 643-648
35. Snijders, T., C. Steglich, and M. Schweinberger, Modeling the co-evolution of networks and behavior. Longitudinal models in the behavioral and related sciences, 2007: 41-71
36. Sun, J., C. Faloutsos, S. Papadimitriou, and P. Yu. Graphscope: parameter-free mining of large time-evolving graphs. In Proceedings of the 13th ACM SIGKDD. 2007. pp. 687 - 696. ACM
37. Tang, L., H. Liu, J. Zhang, and Z. Nazeri. Community evolution in dynamic multi-mode networks. In 14th ACM SIGKDD. 2008. pp. 677-685
38. Tyler, J., D. Wilkinson, and B. Huberman, E-mail as spectroscopy: Automated discovery of community structure within organizations. The Information Society, 2005. 21(2): 143-153
39. Wang, G., Y. Shen, and M. Ouyang, A vector partitioning approach to detecting community structure in complex networks. Computers & Mathematics with Applications, 2008. 55(12): 2746-2752
40. Wellman, B., Structural analysis: From method and metaphor to theory and substance. Contemporary Studies in Sociology, 1997. 15: 19-61
41. Welser, H., E. Gleave, D. Fisher, and M. Smith, Visualizing the signatures of social roles in online discussion groups. Journal of Social Structure, 2007. 8(2): http://www.cmu.edu/joss/content/articles/volume8/Welser/