Social Network Generation and Friend Ranking Based on Mobile Phone Data

Mustafa İlhan Akbaş, Raghu Nandan Avula*, Mostafa A. Bassiouni, Damla Turgut
Department of Electrical Engineering and Computer Science
University of Central Florida, Orlando, FL
Email: {miakbas,bassi,turgut}@eecs.ucf.edu, *raghu.avula@ucf.edu

***Abstract*—Social networking websites have been increasingly popular in the recent years. The users create and maintain their social networks by themselves in these websites by establishing or removing the connections to friends and sites of interests. The smart phones not only create a high availability for social network applications, but also serve for all forms of digital communication such as voice or video calls, e-mails and texts, which are also the ways to form or maintain our social network.**

In this paper, we deal with the problem of automatically generating and organizing social networks by analyzing and assessing mobile phone usage and interaction data. We assign weights to the different types of interactions. The interactions among users are then evaluated based on these weight values for certain periods of time. We use these values to rank the friends of users by a sports ranking algorithm, which recognizes the changes in the collected data over time.

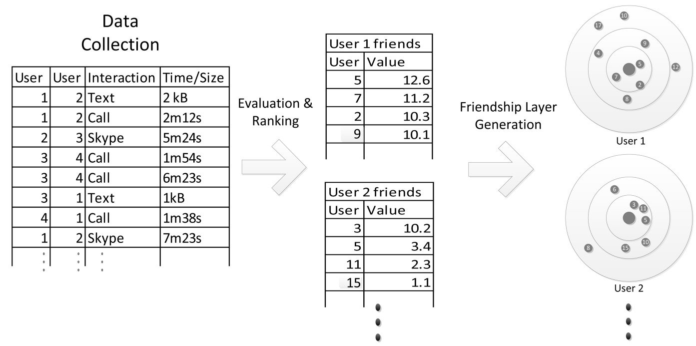


Figure 1: Overview of social network generation for users.

I. INTRODUCTION

Over a billion people frequently use social network web sites. These social networks are successful in providing digital social networks of friends and acquaintances to individual users. As a result of this process, each user creates a digital version of his or her real-life social network. However, the creation and maintenance of the virtual social network requires a large number of interactions, as it is the user's task to establish or remove the connections to friends and sites of interests.

The explicit interactions in virtual social network give users the impression of full control over their ego network. However this process is time-consuming and the users are confronted with the need to organize the internal structure of their personal virtual social network. Some online social networking sites such as Facebook offer a partial support for creating default lists for family members and spouses. Other social networks try to offer a full support by enabling the creation of social circles as desired (Google+). In either case, the user assigns the friends and acquaintances to an individual structure.

In this paper, we deal with the problem of automatically generating the user's social network by using different sources of available interaction data such as physical proximity, text messages, phone calls and video chats. We start our approach to this problem by evaluating the interactions. The interactions are assigned with weights according to their types. The total value of the interactions between two users is evaluated by using the number of communications and the weights of the interaction types. Since the interactions are quantified by our approach, these values are used to rate and rank the friends of users and to find the friendship levels in the social network of each user.

The rating and friendship level of a user's friend is determined by our approach as shown in Fig. 1. We propose using sports ranking methods to determine the ratings of friends. In this approach, the friends of a user are considered as the sports teams competing against each other. Each friend's performance is assessed for a certain period of time as in sports competitions. One of the major sports ranking methods is chosen according to its suitability for social networks to assign ratings for the users' friends. The pattern of the user's personal network formed by the calculated ratings is analyzed to find the different levels of intimacy in this network.

The approach proposed in the paper is applied to a real-life dataset, which is collected during the Nodobo project [1]. Nodobo is an experiment to gather communications metadata from a group of high school students. For this purpose, each student is provided with an Android cell phone. The set of relations, which we use to infer the presence of social ties between the study participants, are explained in the Nodobo project. The social graph of the total group of users, which denotes the ties among the participants, is also drawn. Unlike Nodobo project, our approach concentrates not on the social network among the participants of the study but on the social network of each individual and the organization of these personal networks.

The rest of the paper is organized as follows. We provide a detailed description of our approach in Section II. The simulation results are given in Section III and the related work is summarized in Section IV. We conclude in Section V.

II. Ranking and Grouping Friends

There are various applications on smart phones, which allow us to trace multiple types of interactions between a user and the members of the user's social network. For instance, text messages, call logs and e-mail conversations are stored as history on the device or in the cloud. Moreover, smartphones are equipped with many sensors, which sense, evaluate and record even more information about the user, the environment and contacts. For instance, the proximity of two users is detected by acoustic sensors or by applications using Bluetooth. Most of the mobile phones are also equipped with GPS receivers. The location information can also be collected through these GPS receivers.

Usage of all available data on a mobile phone enables the interpretation of the social network. However, one of the critical problems is that the generation of the network cannot be measured in terms of accuracy or other measures that would allow us to state that the collected data sufficiently describes the corresponding social network. The answers of the questions related to social network dynamics are visible influences on models such as ours.

The collected interaction data of Nodobo project is analyzed in terms of basic interaction information such as the identifications of the parties in the communication, the time of the communication and type of the communication method. Instead of plainly reading and transforming the interaction data, these structural artifacts are used as a structural framework. Additional to the information derived from the interaction data, this framework gives information on how the nodes are socially located in the network and how the links are created.

A. Interaction evaluation

The interaction data derived from mobile devices and sensors need to be evaluated in terms of their importance. The type of an interaction is an important factor to determine its importance. Intuitively, an e-mail sent to a person appears to be more distant than calling the same person. A phone call is shown to be less effective for personal relationships than meeting with the other party in person [2]. Considering these differences, we suggest assigning specific weights to different types of interactions. Weight assignment results in the ability to change the weight according to experience or context and to include additional interaction types as needed. The interaction values are defined in our approach as follows:

$$i_{A,B} = \alpha \cdot F(T) + \beta \cdot V(T) + \gamma \cdot nP(T) + \delta \cdot E(S)$$

where $P(T)$, $V(T)$ and $F(T)$ denote the number of times a phone call, a video conference and a face-to-face interaction occurred respectively for a particular amount of time, T. $E(S)$ denotes the number of e-mails or text messages with size S. The number and types of interactions can be increased or decreased based on the capabilities of the mobile phone or the application to collect the data.

Each interaction type has a different constant $(\alpha, \beta, \gamma, \delta)$, which reflects the variety in effects of different interaction types on personal relations. Formulation of this equation and finding the exact values of constants is one of the next steps of our work. Due to the nature of social sciences, these values may change depending on various factors such as the social group under investigation. Our approach provides the means to utilize results of works in social sciences such as the study of Okdie et al. [3].

B. Ranking Friends

The interactions between friends correlate in number with the strength of friendship [2]. Therefore, after determining the friends of a person and evaluating the interactions, our system also ranks the friends to find different levels of friendship in the social network of users.

We define a sports competition style relationship among users. The friend with a larger interaction value in a defined period of time has a win against the friend with lower interaction value for the same time period. Therefore the sports ranking methods, in which teams win or loose against each other, can be utilized to rank friends.

The most common sports ranking method utilizes the winning percentage to rank the teams or individuals participating in the competition. The winning percentage is the ratio of the games a team won to the total number of games played by that team. Hence this method can be employed to find the rating of a friend i of a user u by using the following formula:

$$r_i = \frac{w_i}{(w_i + l_i)}$$

where w_i and l_i are the number of wins and losses of the friend i of user u.

When the Winning Percentage method is used to calculate the ratings of friends, the rating of a friend depends only on the number of times that friend has a better interaction value than the other friends and the number of comparisons made. Win percentage generally does not satisfy the particularities of sports organizations. Therefore more complicated ranking methods such as Colley [4] and Massey [5] are used in sports.

Colley and Massey ranking methods take the schedule of the competition and the current ranking into consideration. Chartier et al. [6] conducted a sensitivity analysis of these methods and concluded that the Colley and Massey methods are insensitive to small changes. The insensitivity is a desirable property for a ranking method in the social network analysis. For instance when a person moves to another state or starts to work in a new company, the social network of the user may be extended by new friends. Then the user may spend most of the time with one of these friends. If a sensitive ranking algorithm is used, the new friend immediately becomes a strong tie of that person, which rarely happens in real social networks.

Colley and Massey methods differ in terms of their inputs. Colley's ranking method is based only on the results from the comparisons and each result is either a win or a loss. On the other hand, Massey method utilizes the actual game scores and homefield advantage, which may improve the ranking in sports but have no correspondence in social networks. Therefore, we choose the Colley method as the basis for ranking friends of a person.

The Colley method of sports ranking can be defined by a linear system [7], $C\vec{r} = \vec{b}$, where $\vec{r}_{n\times 1}$ is a column-vector of all the ratings, $\vec{r}_i$. The right-hand-side vector, $\vec{b}$, is defined with the components as follows:

$$b_i = 1 + (w_i - l_i)/2$$

$C_{n\times n}$ is called the Colley coefficient matrix and defined as follows:

$$C_{ij} = \begin{cases} 2 + (w_i + l_i) & i = j \\ -n_{ij} & i \neq j \end{cases}$$

The scalar n_{ij} is the number of times friends i and j are compared to each other. The Colley system $C\vec{r} = \vec{b}$ always has a unique solution since $C_{n\times n}$ is invertible. Then the rating of a friend of a user is defined as follows [6]:

$$r_i = \frac{1 + \frac{(w_i - l_i)}{2} + \sum\limits_{k \in F_u} r_k}{2 + (w_i + l_i)}$$

where F_u is the set of all contacts that user u communicated.

With this calculation method, the rating of a friend depends on the ratings of the other friends of user u, which has an important reflection in the social networks. Hence interacting with someone more than that person's best (top ranked) friend has a high impact on the friend rating. In contrast to traditional methods, the initial rating of any friend with no changes is equal to $\frac{1}{2}$, which is the median value between 0 and 1. Depending on the comparisons, a win increases and a loss reduces the value of r. This approach results in a system less sensitive to changes. Therefore the communication between a user and a friend needs to retain a high level for several ranking periods until it has a remarkable effect on the friend's rating. In other words, a friend of a user is not assigned with a high intimacy role by Colley method by just having a high interaction value for a short amount time.

After their ratings are calculated, the friends of a user can be sorted and assigned with ranks. Generation of a complete list of friends for a user and their rankings depend on the level of variety in the collected data, the weights assigned to different interaction types and the length of the interval for the data collection. The weights and the length of the interval must be chosen according to the sociological characteristics of the group under consideration. The ratings calculated by the ranking method used in our approach depend on the ratings of other users and the ratings are updated periodically. Therefore the resulting friend network for a user will have a pattern of discrete groups of ratings according to Dunbar et al. [2]. Then the levels of the intimacy or the circles of friends are decided according to the ratings.

Table I: Simulation parameters

Total time	4 months
Game period	1 month
Number of users	27
Call weight (w_c)	1.25
SMS weight (w_s)	1
Number of Calls	1309
Number of SMS	25,982

III. Simulation Study

In this section, we evaluate our approach by measuring how the patterns in the estimated friendship networks vary with the ranking algorithm and interaction characteristics. The Nodobo study includes call, SMS and proximity records. The direction of the calls and messages are in the record along with the associated phone number and the duration of the call or length of the message. The participants of the study used provided mobile phones to communicate with their personal contacts as well as the other participants of the study. In our approach, we aim to generate the social network with all contacted friends for each participant. Therefore, the data must be analyzed and used for all friends of each user. Since the proximity data exist only for the current participants, it cannot be used in the same analysis with the call and SMS records. Additionally, there are false positive ties formed in the social graph when the proximities of the Nodobo dataset are used [1].

We use different weight values for calls and texts depending on the studies of Okdie et al. [3] and Boucher et al. [8]. According to these studies, users declare a higher satisfaction level in vocal communication compared to the communication based on texts. Therefore we assigned a higher weight to the calls, $w_c = 1.25$, than SMS, $w_s = 1$, in our simulations. The data collection period is divided into intervals of one month and the interactions of the users are compared for each month to calculate the wins and losses of the friends. Table I summarizes the parameters used in the study.

Fig. 2 shows the resulting ratings of friends with Colley method for nine of 27 users. The friends of the users are numbered and sorted according to their ratings in the figures. The pattern of the rating distribution reflects the discrete groups of friends for users. The resulting rating graphs have the same concavity characteristics, which is an effect of the unsensitive Colley ranking method. This pattern also exists in the rating results of all students which are not included in Fig. 2. For 20 users ($\tilde{7}$4%), friends can be grouped into three discrete groups such that friend with maximum rating in an intimacy group has a rating at least 15% lower than the friend with the minimum rating in the higher intimacy group.

In the second set of the analyses, the ratings are calculated by using the Winning Percentage method. The experiment interval is divided into periods of one month. Fig. 3 shows a comparison of ratings by Colley and Winning Percentage method for two users. Both methods rank friends close to each other. However one of the main disadvantages of Winning Percentage method is that some of the users are not included in the friend list. Although the user communicated with these

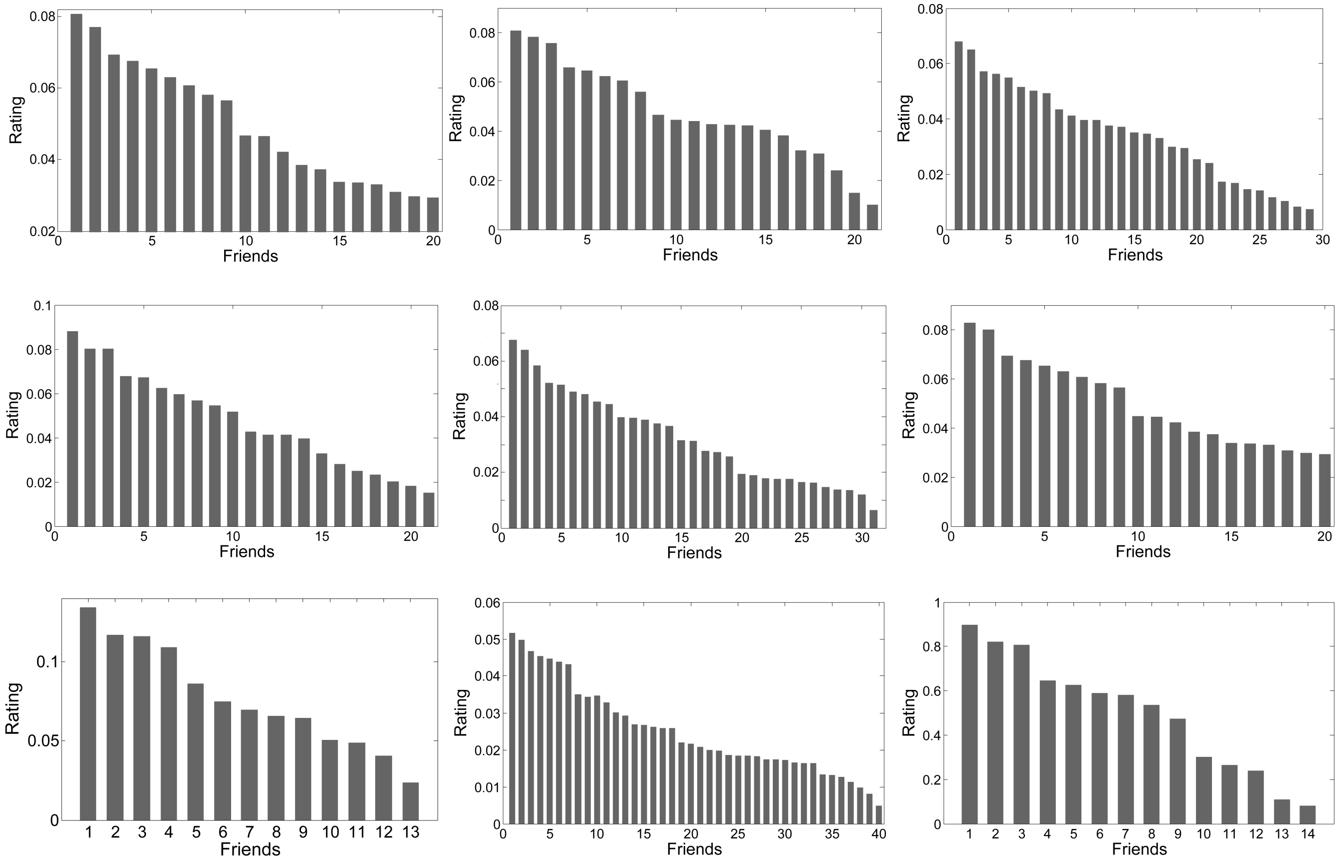
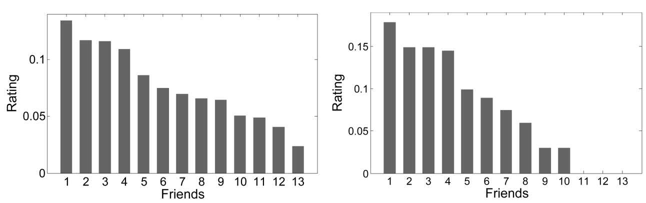
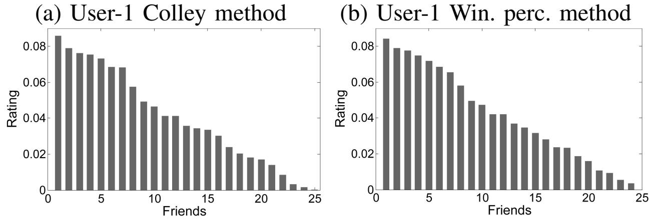


Figure 2: Ratings of friends with Colley method for nine users.



(a) User-1 Colley method (b) User-1 Win. perc. method



(c) User-2 Colley method (d) User-2 Win. perc. method

Figure 3: Friend ratings for two users.

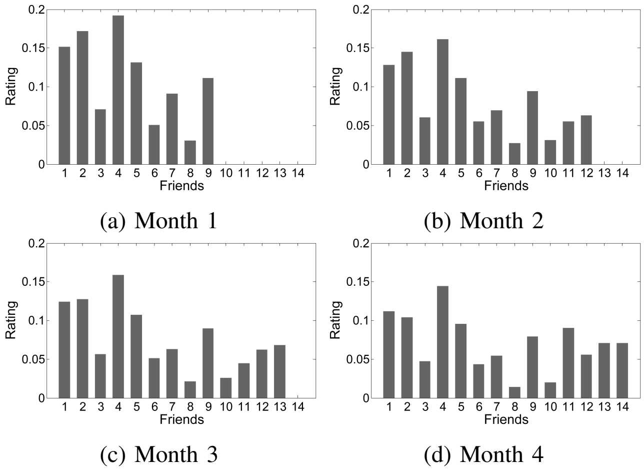
friends, they didn't have an interaction value considered as a win to be included in the friends list. Additionally, the pattern of discrete groups of friends can not be observed in Winning Percentage method. The ratings of the friends are distributed in a more continuous pattern compared to the ratings of Colley method.

In Fig. 4, the change in the ratings of a user's friends is demonstrated for each month during the four month interval. Since the number of communicated contacts increases over time and the rankings change accordingly, the ratings are not sorted in Fig. 4 to keep the order of friends. The user communicates at most with newly added friends after each month, which can be observed in the results. However the ratings do not change drastically and Colley method generates a consistent friend network through four months.

IV. Related Work

There are approaches in the literature, which use online communication data to analyze social network structures. Stumbl [9] is a Facebook application that collects the information from users about their daily face-to-face meetings with their Facebook friends. This information is combined with the interaction data to create the social graph of users. Ilyas and Radha [10] utilized principal component centrality approach to identify the nodes with the most influential roles in the network. Xu et al. [11] studied a similar problem and adopted random walk to find the experts for relevant information in an online social network.

The vast usage of the social networking applications on mobile phones leads to an important research approach for social network analysis. The applications and the operating systems of the mobile phones become increasingly personalized with



(a) Month 1 (b) Month 2

(c) Month 3 (d) Month 4

Figure 4: The change in friend ratings in four months.

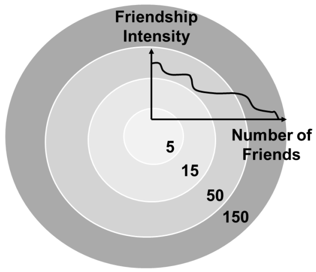


Figure 5: Discrete friend groups in a social network.

the specific communication software and the supplementary data. The social network studies on mobile networks are conducted conducted either by utilizing the real life data gathered by experiments [12], or by the simulations of mobile phone users (eg. [13]) and their behaviors. Eagle et al. [14] proposed an approach to compare the communication, location and proximity information from mobile phones to self-reported relationships. Their analyses showed that the self-reported data and the behavioral data have both overlapping and distinct features. Gaito et al. [15] proposed a framework to capture and model the human behavior in the dimensions of space, sociality and time by using smartphone data.

Dunbar et al. [2] showed that the number of friends, which a person can have, has a limit and the personal social networks of people have common patterns. The types of social groups formed according to ties in our social networks have clear boundaries between each other [16]. Social networks are hierarchically organized in discretely sized groups, which are also refered as circles as shown in Fig. 5. The inner most circle includes up to five people and the sizes of circles increase by a factor of three. In this paper, we improve our friendship layer determination concept [17] and implement it for the Nodobo dataset.

There are various ranking methods used in sport organizations. Billingsley [18] developed an algorithm for updating a football team's rating based on its most recent result and the information from the previous seasons. Colley [4] and Massey [5] are important sports ranking methods, which take the schedule of the competition and the current ranking into consideration when calculating the rankings of the teams.

V. Conclusion

The approach outlined in this paper aims to infer the social networks of individuals by assessing their smartphone data. We propose an interaction evaluation function, a ranking and a grouping method in accordance with the research findings in social networks. Our future work includes testing our approach by using multiple real life smartphone data with collection times spread over longer periods. Since the friends of a single individual and their intimacy levels are considered, another possible future work is implementing a smartphone application, collecting and integrating all available interaction data to extend our aproach.

References

[1] J. Banford and J. Irvine, "Estimating Social Graphs in an Education Environment," *IEEE Vehicular Technology Magazine*, vol. 7, pp. 31–37, Mar. 2012.

[2] R. Dunbar, *How Many Friends Does One Person Need?: Dunbar's Number and Other Evolutionary Quirks*. Faber and Faber, Feb. 2010.

[3] B. Okdie, R. Guadagno, F. Bernieri, A. Geers, and A. Mclarney-Vesotski, "Getting to know you: Face-to-face versus online interactions," *Computers in Human Behavior*, vol. 27, pp. 153–159, Jan. 2011.

[4] W. N. Colley, "Colley's bias free college football ranking method: The Colley Matrix explained." http://www.colleyrankings.com/matrate.pdf, Sept. 2003.

[5] K. Massey, "Statistical Models Applied to the Rating of Sports Teams, Master's thesis, Bluefield College, Bluefield," 1997.

[6] T. P. Chartier, E. Kreutzer, A. N. Langville, and K. E. Pedings, "Sensitivity and stability of ranking vectors," *SIAM J. on Scientific Computing*, vol. 33, pp. 1077–1102, May 2011.

[7] T. P. Chartier, E. Kreutzer, A. N. Langville, and K. E. Pedings, "Bracketology: How can math help? Mathematics and sports," *Dolcani Mathematical Expositions, Joseph Gallian, ed. Mathematical Association of America*, vol. 43, pp. 55–70, July 2010.

[8] E. M. Boucher, J. T. Hancock, and P. J. Dunham, "Interpersonal sensitivity in Computer-Mediated and Face-to-Face conversations," *Media Psychology*, vol. 11, pp. 235–258, June 2008.

[9] T. Hossmann, F. Legendre, G. Nomikos, and T. Spyropoulos, "Stumbl: Using Facebook to collect rich datasets for opportunistic networking research," in *IEEE WoWMoM*, pp. 1–6, June 2011.

[10] M. Ilyas and H. Radha, "Identifying Influential Nodes in Online Social Networks Using Principal Component Centrality," in *IEEE ICC*, pp. 1–5, June 2011.

[11] K. Xu, J. Xie, and V. Li, "Locating experts via online social networks," in *IEEE ICC*, pp. 1–5, May 2010.

[12] H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan, and D. Estrin, "Diversity in smartphone usage," in *MobiSys*, pp. 179–194, June 2010.

[13] F. Nazir, H. Prendinger, and A. Seneviratne, "Participatory mobile social network simulation environment," in *IEEE ICC*, pp. 1–6, May 2010.

[14] N. Eagle and A. Pentland and D. Lazer, "Inferring friendship network structure by using mobile phone data," in *Proc. of the National Academy of Sciences*, vol. 106, pp. 15274–15278, Sept. 2009.

[15] S. Gaito, E. Pagani, G. P. Rossi, and M. Zignani, "Sensing multi-dimensional human behavior in opportunistic networks," in *ACM MobiOpp*, pp. 89–90, Mar. 2012.

[16] W. X. Zhou, D. Sornette, R. A. Hill, and R. I. M. Dunbar, "Discrete hierarchical organization of social group sizes," *Proc. of the Royal Society B: Biological Sciences*, vol. 272, no. 1561, pp. 439–444, 2005.

[17] M. İ. Akbaş, M. R. Brust, and D. Turgut, "Social network generation and role determination based on smartphone data," in *IEEE INFOCOM Student Workshop*, Mar. 2012.

[18] R. Billingsley, "College Football Research Center," 2004.