Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia - Social and Behavioral Sciences 79 (2013) 18 – 26
ELSEVIER
Procedia
Social and Behavioral Sciences

9th Conference on Applications of Social Network Analysis (ASNA)

Identifying Close Friendships in a Sensed Social Network

Rahman O. Oloritun[a,*], Anmol Madan[b], Alex Pentland[b], Inas Khayal[a,b]

[a]*Masdar Institute of Science and Technology, PO.Box 54224, Abu Dhabi, UAE*
[b]*Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

Abstract

Studies have suggested that propinquity; social, cultural, physical and psychological similarities are major factors in close friendship ties. These studies were subject to human recall of interactions with no details of length or time of interactions. Recently, advancements in mobile technology have enabled the measurement of complex systems of interactions. This study uses social network analysis of data comprising of time-resolved sensed interactions to predict and explain close friendship ties via interactions at different periods, residence (floor) similarity and gender similarity. Results indicate residence (floor) proximity and duration of weekend night interactions have the potential of explaining close friendship ties.

Selection and/or peer-review under responsibility of Dr. Manuel Fischer

Keywords: close friendship; bluetooth; phones; similarities; QAP; social networks

1. Introduction

The factors necessary for the creation of close friendship ties have been the subject of several studies. Past studies have suggested propinquity (Festinger, Back, & Schachter, 1963; Thibaut & Kelly, 1956), social, cultural, physical and psychological similarities (homophily) (McPherson, Smith-Lovin, & Cook, 2001) are the perquisite factors for the creation of close friendship ties. These studies collected data utilizing routine or generally accepted techniques, such as surveys and interviews which are subject to the human ability to recall and are constrained in spatial and time scales by technical difficulties and cost (Isella et al., 2011). These complexities in acquiring human interaction data are the reasons why past studies in social networks have used sampling procedures which are biased towards selecting only friends and relatives (Christakis & Fowler, 2007; Goodreau, Kitts, & Morris, 2009), or limiting respondents to a fixed number of peers (Goodreau et al., 2009).

However, the availability of new data acquisition techniques for logging human face-to-face interaction is opening new avenues for understanding the dynamics of interactions in social networks and providing researchers with access to almost complete social networks. Time-resolved face-to-face interactions by individuals in real-world settings can be captured using embedded sensing techniques. Pervasive and ubiquitous devices such as cell phones make it possible to study and predict patterns of human mobility (Dong, Lepri, & Pentland, 2011; Gonzalez, Hidalgo, & Barabasi, 2008; Song, Qu, Blumm, & Barabasi, 2010) within a dormitory (Dong et al., 2011), city (Chowell, Hyman, Eubank, & Castillo-Chavez, 2003) and between cities (De Montis, Barthelemy, Chessa, & Vespignani, 2005), in countries (Brockmann, Hufnagel, & Geisel, 2006),

* Corresponding author. Tel.: +971-56-753-5283; fax: +971-2-810-9901.
E-mail address: roloritun@masdar.ac.ae

1877-0428 © 2013 The Authors. Published by Elsevier Ltd.
Selection and/or peer-review under responsibility of Dr. Manuel Fischer
doi:10.1016/j.sbspro.2013.05.054

and globally (Isella et al., 2011). Blue-tooth and Wi-Fi technologies have been used to capture proximity patterns (Anmol Madan, 2010), and even face-to-face presence can be resolved with high spatial and temporal resolution (Cattuto et al., 2010).

With recent advancements in mobile technology, mobile phones and wireless devices, which have features such as short-range Bluetooth radios, Cellular-Tower Identifiers, Global Positioning Systems (GPS) and other location technologies, can now be used measure these complex systems of interactions. It is possible to identify who an individual interacts with, where these interactions took place and the frequency of such interactions. Thus, with such fine-grained interaction data the behavior of individuals and the activities in which they are engaged can be inferred (Eagle & Pentland, 2006). These technological advances allow researchers to gather data that have been traditionally scarce in social network analysis. This relatively new approach of using sensor technology to capture real world social interactions avoids the inaccuracies (Anmol Madan, 2010) inherent in self-reported data.

This study evaluates correlations between close friendship ties and duration of interactions measured via embedded sensing techniques at different periods of time while controlling for propinquity and similarity of personal attributes. This study uses time-resolved face-to-face interactions data of individuals from a real-world setting, captured using embedded sensing techniques, to predict close friendship ties. The data for this study consists of duration of interactions in four different periods, namely, weekday daytime, weekday night, weekend daytime and weekend night. The data for this study is from a multi-faceted research that studied social evolution, effects of exposure on lifestyle and political opinions, (Anmol Madan, 2010) and recently, the effect of exposure on change in BMI (Body Mass Index) in face to face network.

This multi-faceted research is a precursor to an ongoing study (Aharony, Pan, Ip, Khayal, & Pentland, 2011) that uses more advanced embedded sensing techniques to gather fine grained social data on a larger scale. As the collection of real life interaction data improves and becomes widespread, the complexity of data acquired will increase. There will be the need to infer relationship types from automatically captured interaction data. Here, we used network visualization, social network analysis and Multi-Regression Quadratic Assignment Procedure to predict and explain close friendship ties via interactions networks at different periods in time, spatial (floor of residence similarity) proximity and gender similarity.

2. Background

One of the very first studies that captured and analyzed face-to-face interactions in social networks using sensor technology was the study by Choudhury and Pentland (2004), where a device called the Sociometer was developed to define the structure of the network, and record when and if people were conversing. In, another study (Choudhury, Philipose, Wyatt, & Lester, 2006) turn-taking in face-to-face conversations was analyzed. The study reports that the influence of each participant in joint turn-taking, and their “betweenness” centrality, estimated from the network, are correlated.

There are over five billion (BBC, 2010) mobile phones globally, and it is expected to exceed the global population in 2012 (Perez, 2012), which can be used as sensors of location, proximity and communication. Thus, the term ‘Reality Mining’ was used in studies (Eagle & Pentland, 2006; A. Madan & Pentland, 2009; A. Pentland, Lazer, Brewer, & Heibeck, 2009) at MIT that used mobile phones as sensors to learn the structure of social networks. Another study (Gonzalez et al., 2008) showed that call detail records (CDR) can be used to characterize temporal and spatial behaviors in human mobility pattern. Examples of other studies that used mobile phones to map human interaction in a social network includes Mobiscopes for Human Spaces by Abdelzaher et al. (2007) and the Darwin phone projects at Dartmouth (Avancha, Baxi, & Kotz, 2009; Miluzzo et al., 2010).

Past studies (Eagle, Pentland, & Lazer, 2009; Anmol Madan, 2010), have suggested that different types of relationships are expressed in different periods in time. Hence, these periods in time can provide insight into the identification of the relationship types. Therefore, interaction properties at different periods in time can help identify close friendship ties.

3. Methodology

This section presents the social network analysis approach used in this paper to evaluate how duration of interactions and homophily explain close friendship ties. The approach consists of data collection, network/graph construction, visualization of the networks, description of the networks and logistic regression of reported networks on the interaction networks at the different periods in time.

3.1. Data Collection

The data for this paper is from a study performed in a real-world setting of an American university undergraduate dormitory in 2009. The experiment comprised of 42 participants composed of 50% males and 50% females. The study was approved by the Institutional Review Board (IRB) of the university and conducted under strict protocol guidelines. Participants in the study were given socially-aware mobile phones. The phones had proximity detecting software installed with Bluetooth sensors or transceivers to detect other proximate phones in the study. The class 2 Bluetooth transceivers on the mobile phones have a realistic indoor sensing range of approximately 10 feet. Once a phone detects another phone in proximity, it captures the other phone's identifier and records the length of time the phones were in proximity. A detailed description of the data collection platform and the technologies used is available elsewhere (Anmol Madan, 2010).

The information used for this study is a subset of data available online (Alex Pentland, Madan, Dong, & Reid, 2009), with an appraisal of the data quality, noise levels, and sensor characteristics. Participants' demographic data, gender, floor of residence (pseudonyms was used to represent the floors), race and school year, were part of the data collected. The participants lived on a floor of their choosing subject to the agreement of other floor residents in accordance to the residency policy of the dormitory. Participants also listed their close friends, who are also participants in the study. A close friend in this study refers to a person with whom an individual is comfortable discussing very personal issues/topics or from whom an individual seeks or gets emotional support. Non-close friendships are defined as individuals that you interact with that were not specifically labeled as close friends. The average degree of the interaction networks is approximately 33 and there are 42 nodes. This implies that an average node connects to 80% of the nodes.

3.2. Creating the Close friendship and Interaction Networks

Reported Close friendship: The reported close friend data collected is a directed network of the 42 (nodes) participants in the study where participants *i* names participant *j* as a close friend. By design, participant can only name other study participants as close friends. It is however not usual in studies of networks that a participant *i* indicates a participant *j* as a close friend and participant *j* does not indicate participant *i* as a close friend. Relationships where only one of participants *i* and *j* names the other as a close friend (i.e. relationships that are not reciprocal) are not indicative of close friendship and were taken out of the network. The result of removing the unreciprocated relationships is a symmetric network where for every participant *i* that names participant *j* as a close friend; participant *j* also names participant *i* a close friend. These reciprocal connections in a fully symmetric network can approximate to an undirected network as illustrated in figure 1.

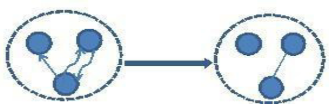

Fig 1. An illustration showing the conversion of reported close friendship to un-directed graph of relations.

Interaction Networks: The sensor measured or automatically captured interactions featured the entire interactions in the community studied. A key assumption is that interaction occurs whenever the socially aware phones are within their Bluetooth transceivers range even if the participants do not engage in any form of observable interactions. These interactions comprised close friends and non-close friend's interactions. The automatically captured networks comprised four networks for four time spaces, *t*, daytime (a period between 8a.m to 8p.m) and nights (a period between 8p.m to 8a.m) for weekday and weekends respectively. The interaction networks are such that if a participant *i* that spent some length of time interacting with participant *j* then participant *j* must also interact with participant *i* for the same duration. The duration (hours) of the interaction is weight of the connection between participants *i* and *j*. Hence, the interaction networks are inherently symmetrical and therefore can be taken as undirected weighted networks, where the weights are the duration of interaction.

3.3. Analysis

We graphed the reported close friendship and interaction networks using the Fruchterman-Reingold Algorithm (Fruchterman & Reingold, 1991) in the igraph (Csardi & Nepusz, 2006) library for R-2.14.2 (R Development Core Team, 2012). We report our observations from the visual depictions of the networks and drew inferences. The shape of the nodes/vertices indicates the participant's gender while the color implies the location (floor) of the participant's residence in the dorm. In our depiction of the interaction networks, we highlighted the sub-network that represents the close friend's interactions. To understand how participants divide their time across close friends and non-close friends, we introduced a network property that we referred to as the mean duration (hours) per dyad (MDD) in close friends, non-close friends and the community networks for each participant and we visualized the MDDs for close friends, non-close friends networks for the comparison. The mean duration per dyad (MDD) for each participant is defined as the weighted degree of the interaction network divided by the un-weighted degree of the interaction network. The MDD is mean duration per dyad (hours per dyadic interaction) by each participant was calculated as follows.

Let ***A*** be the adjacency matrix for a weighted network ***G***, with n nodes, such that the elements A_{ij} in **A** are the weights (duration) of the network and ***B*** is the un-weighted adjacency matrix for an equivalent of network ***G*** such that element B_{ij} in ***B*** is one if a weight in G is greater than zero and zero, otherwise. There weighted degree w_i (total hours spent interacting by each participant), can be written as (equation 1)

$$w_i = \sum_{j=1}^{n} A_{ij} \tag{1}$$

The un-weighted degree k_i, which equivalent to the total number of nodes connected to a node, can be written as in (equation 2)

$$k_i = \sum_{j=1}^{n} B_{ij} \tag{2}$$

The Mean duration per dyad, $\boldsymbol{MDD_i}$ (equation 3), for participant i is therefore obtained by

$$MDD_i = w_i / k_i \tag{3}$$

QAP Logistic Regression: The quadratic assignment procedure (QAP) is an approach for statistical significance testing using social network data. One assumption of parametric statistical techniques, which determine statistical significance by comparing observed values to appropriate theoretical distributions, is that the observations being analyzed are independent of one another. This assumption is not accurate and does not hold in social network analysis. QAP is a non-parametric technique, meaning it does not rely on assumptions of independence; it is also a general procedure that is frequently used for both correlation and multiple regression analysis. The QAP approach (Krackhardt, 1988) estimates regression model coefficients and then uses random permutations of the network data to generate a distribution of coefficient estimates from random networks with the same structure. The actual estimates are then compared with this generated distribution to test for significance. Improvements in this procedure have been made to ensure conservative estimation of standard errors across less-than-ideally structured data (Dekker, Krackhardt & Snijders, 2007).

The procedure for the improved QAP regression starts with a standard multiple regressions across corresponding elements of the dependent and independent matrices and the beta weights and R^2 values are obtained. The rows and columns of the dependent matrix are randomly permutated a large number of times – in the case of this analysis, the matrix was permutated 1000 times. After each permutation, a new regression is conducted. The beta weights and R^2 values from each regression are stored to form a distribution against which the original beta weights and R^2 values are compared. A t-statistic is calculated to provide an estimation of the probability that the original beta weights and R^2 value were obtained by chance. The effects of other variables in the analysis, Z, are partialled out from X and the resulting residuals were entered into a regression of Y on both the residuals and Z. The technique is called "double" semi partialling (Dekker et al., 2007) because Z enters the regression twice. The effect of any collinearity amongst the dependent variables is partialled out by performing a double regression (Dekker et al., 2007) first on the residuals and the independent variables and then on all the variables and residuals.

We developed a QAP model to predict likelihood of close friendship ties. Our QAP model includes reported close friendship network as the independent variable and network of interactions on weekday daytime, weekday nights, weekend daytime, weekend nights, and dyadic variables, (floor similarities, gender similarities and race similarities) respectively as independent variables.

The QAP model is illustrated in equation 4.

Close friendship = $\beta_0 + \beta_1$ • *weekday daytime* + β_2 • *weekday night* + β_3 • *weekend daytime* + β_4 • *weekend night* + β_5 • *same floor* + β_6 • *same gender*+ β_7 • *same race*, *Where* β_i *is the coefficient of regression.* (4)

Weekday daytime, weekday night, weekend daytime, and *weekend night* are lengths of interactions. *Same floor* is a dyadic variable which is 1 if participant i and participant j both resides on the same floor and 0 otherwise, same gender is a dyadic variable which is 1 if participant *i* and participant *j* are of the same gender and 0 otherwise and same race is a dyadic variable which is 1 if participant *i* and participant *j* are of the same race and 0 otherwise. The model was evaluated using the "double" semi partialling QAP implementation in the "sna" library (Butts, 2008, 2010) for R (R Development Core Team, 2012).

4. Results

This section presents the result of visualization and statistical analyses performed on the close friendship network and the other parameters.

4.1. Reported Close Friendship and Interaction Networks

The reported close friendship network is depicted in figure 2. The circles indicate female participants while the squares are the male participants with the colors representing the floor of residence. Nodes with the same color all live on the same floor. In close friendship network diagram, we observe clusters of close friends that live on the same floor. In the reported close friendship network, female participants appear to make up the largest percentage of isolates, that is nodes without any connections (female isolates = 4 and male isolates = 1). There are quite a number of mix gender (e.g. female-male or vice-versa) and more male-male edges than female-female and female-male edges.

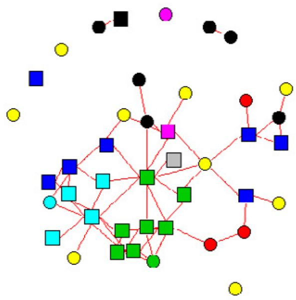

Fig. 2. Graph of reported close friendships. The round nodes represent female participants while the square nodes represent male participants. The color of nodes represents the location of their domicile (i.e. floor location in the dormitory), therefore node with the same color are said to live at the same floor. Clusters of close friends living on the same floor can be seen in the network plots. We observe that female participants tend to be at the edge of this network and have fewer close friends within the study community.

The interaction networks for periods in time (weekday daytime, weekday nights, weekend daytime and weekend nights) are shown in figure 3. The cluster of close friends living on the same floor is also noticeable in network diagram of interactions at the different periods in time. The table 1 shows the average durations and average MDDs of the reported close friendship network and the interaction networks. The female participants appear to have more connections in the non-close friend interaction networks than they appear to have in close friend interaction networks, the sub-network highlighted in red in figure 3. Participant are members of close friends and non-close friends interaction networks at the time periods depicted in figure 3.

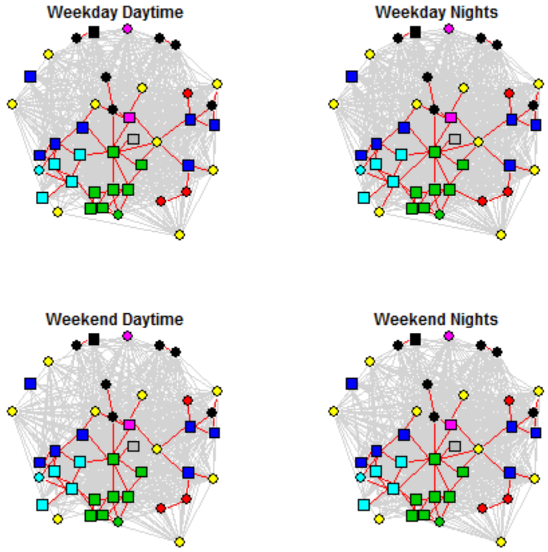

Fig. 3. The graphs of the sensor measured social interaction networks. The social interactions between close friends are colored red (the red edges). The round nodes represent female participants while the square nodes represent male participants. The color of the nodes represents the location of their domicile (i.e. floor location), therefore nodes with the same color are said to live at the same floor. Clusters of interacting close friends living on the same floor can be seen in the network plots for different time periods with subtle edge differences.

Figure 4 suggests that participants spend more time interacting with a close friend than a non-close friend. However, participants interact with more non-close friends than close friends, but for shorter lengths of time. In this study these multiple interaction networks are close friends interaction network and non-close friends' interaction network.

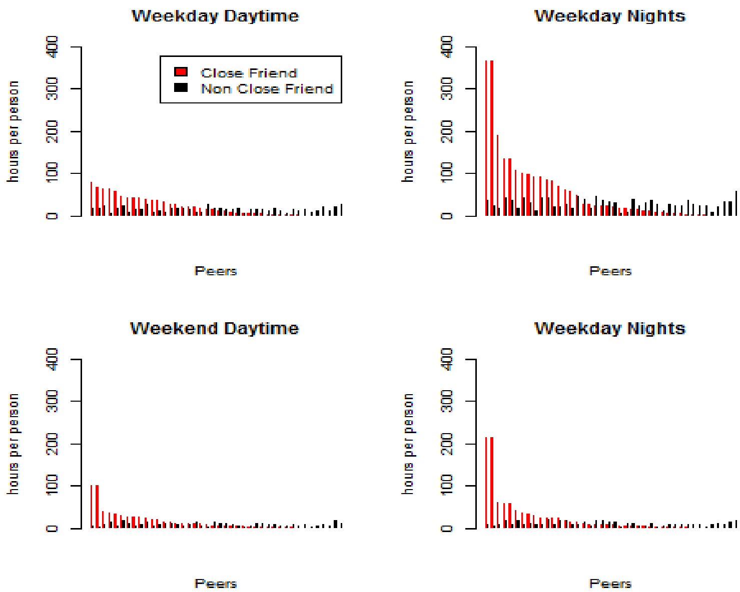

Fig. 4. Bar plots of Mean duration per dyad mean duration of interactions (MDD) between peers in close and non-close friends interaction networks. The plots suggest that there are longer interactions at nights (weekdays and weekends) than daytime (weekdays and weekends).

Table 1. Average of the total time spent interacting by participants and average MDD (Mean duration per dyad for participants) in the sensor measured interaction networks. The hours per close friend (i.e. MDD in close friend networks) is obtained by dividing the number of close friends that a participant interacts with by the total number of hours that participant spent with close friends. The hours per non-close friend (i.e. *MDD* in non-close friend networks) is obtained by dividing the number of non-close friend that a participant interacts with by the total number of hours that participant spent with non-close friends. The same procedure was repeated for all people within the community that the participants interacted with. The average of the hour per person, (close friend/ non-close friend/everybody) across all the participants is presented here. The fourth column of the table below indicates that participants spend more time on average with close friends than non-close friends, even though there are more non-close friends.

		Average Duration of Interactions	Average MDD
Weekday daytime	All	656.39	15.05
	Close Friends	62.33	23.99
	Non close Friends	594.07	14.34
Weekday nights	All	1578.18	29.61
	Close Friends	363.25	62.98
	Non close Friends	1214.93	27.80
Weekend daytime	All	220.25	8.49
	Close Friends	99.31	16.72
	Non close Friends	120.94	8.01
Weekend nights	All	452.79	10.61
	Close Friends	214.89	26.90
	Non close Friends	237.90	15.05

4.2. Results of QAP Logistic Regression

Burris (2005) argues that when interpreting QAP regression results, the focus should be on the comparative magnitude of the coefficients, rather than on the overall model R^2 or the level of statistical significance for each coefficient. In Table 2, we report the coefficients for each independent variable, their odd ratios and their significance level. Discussion will focus on the comparative magnitude of those coefficients which are significant at $p < 0.1$. The regression results suggest that living on the same floor in the residence significantly increases the likelihood of a close friendship tie and the interaction network on weekend night and that close friendship tie are correlated.

Table 2. The results of the QAP Logistic Regression of reported close friendship on measured social interactions and similarities in residence and gender. Interactions at weekend nights and floor (in dorm) similarity are significant ($p < 0.1$). N = 861

Parameters	Estimates	Odds Ratio (e^{β})	p-values
Intercept	-2.600	0.078	0.000
Social Interactions			
Weekday Daytime	0.004	1.004	0.710
Weekday Nights	-0.006	0.993	0.269
Weekend Daytime	-0.008	0.991	0.692
Weekend Nights	0.029	1.029	0.060
Same			
Floor of Residence	1.882	6.569	0.000
Gender	-0.335	0.715	0.168
Race	-0.113	0.893	0.717

5. Discussion

In this study, a close friend is defined as a person with whom an individual is comfortable discussing very personal issues/topics or from whom an individual seeks or gets emotional support. This study found that participants spend more time interacting with close friends than they do with non-close friends even though they have more non-close friends. The duration of close friend interactions is greater at night rather than during daytime, when participants maybe engaged in activities outside the residence. Reported close friendship network is similar in structure to the close friends' interaction networks at the time spaces, weekday daytime, weekday night, weekend daytime and weekend night. Reported close friendship network plot shows clusters of close friends living on the same floor. The close friend interaction networks plots also show clusters of close friends' interactions living on the same floor.

When performing a QAP regression with the reported close friendship network as the dependent variable, the four (4) interaction networks at different periods in time (weekday daytime, weekday night, weekend daytime and weekend night), same floors in the residence (a dyadic variable which is 1 if participant *i* and participant *j* both resides on the same floor and 0 otherwise), same gender (a dyadic variable which is 1 if participant *i* and participant *j* are of the same gender and 0 otherwise) and same race (a dyadic variable which is 1 if participant *i* and participant *j* are of the same race and 0 otherwise) as independent variables, we found the interaction network on weekend nights and floor (in dormitory) to be significant at $p<0.1$. The effect of weekend nights' interaction network on close friendship ties is confounding. The interaction network on weekend nights could be significant because close friends tend to interact more on weekend nights or more interactions on weekend night make participants become friends.

The visualization of reported close friendship network shows most isolates are female participants (nodes), i.e. nodes with no connection, and the connected female participants are frequently found at the edges. In the interaction networks however all the node that is isolates in close friends interaction network are connected in non-close friends interactions. It could be that the close friends of these isolates are either outside the community understudy or they have no close friends. The network plots suggests floor of residence and gender may play some role in the explaining close friendship.

The significant impact of living on the same floor in the dormitory on close friendship may be explained by propinquity. Propinquity has been suggested by psychologists to be one of the major factors in the formation of friendship (Festinger et al., 1963; Thibaut & Kelly, 1956). Festinger et al. (1963) followed friendships in a small two-floor apartment building and found that neighbors were mostly likely to be friends. They also, found that people on separate floors are least likely to be friends and people who live near ground-floor staircases and mailboxes had friends on both floors.

Despite propinquity explanation, the effect of living on the same floor on close friendship is confounding. Cluster of close friends on floors could be a result of participants who are close friends deciding to live together on the same floor or living on the same floor made the participants become close friends.

However, the lack of a wide range in the ages of the study population limits the generality of the study although the study population includes subjects from different races that are with varying levels of income, and from different cultures. Another limitation of this study is the assumption that interaction occurs whenever the socially aware phones (devices used to capture the interaction data analyzed in this study) are within their Bluetooth transceivers range even if the participants do not engage in any form of observable interactions. It should be noted that the range of the sensor devices is generally visible to participants, i.e., participants will most of the time see the other participants if their sensors can connect.

6. Conclusion

This study suggests that close friendships can be explained via the duration (hours) of interactions, the period of interactions (e.g. weekday daytime, weekday night, weekend daytime and weekend night), floors similarities (residence) and gender similarities.

This study highlighted the importance of social interactions measured using embedded sensing techniques in the real world, presenting a new avenue for understanding individual behavior and explanation of various social ties. These results paint a bright picture for future studies of social networks fueled by the latest advances in wireless communication and embedded sensing technologies. These technologies allow us to collect fine-grained data on a larger scale that would not have been possible earlier. Future directions include continued research into understanding and showing mathematically the factors that engender close friendship by applying these analyses and more on a larger dataset from an ongoing project (Aharony et al., 2011)

Acknowledgements

The authors would like to thank Dupe Adelaja and Camile Pinnard for their useful comments during the writing of this paper. This research was supported by the MI-MIT Joint Funding Program. The authors also extend their gratitude to the reviewers for their very constructive comments.

References

Abdelzaher, T., Anokwa, Y., Boda, P., Burke, J., Estrin, D., Guibas, L., et al. (2007). Mobiscopes for Human Spaces. *IEEE Pervasive Computing, 6*, 20-29.

Aharony, N., Pan, W., Ip, C., Khayal, I., & Pentland, A. (2011). Social fMRI: Investigating and shaping social mechanisms in the real world. *Pervasive and Mobile Computing, 7*, 643-659.
Avancha, S., Baxi, A., & Kotz, D. (2009). Privacy in mobile technology for personal healthcare. *Submitted to ACM Computing Surveys*.
BBC. (2010). Over 5 billion mobile phone connections worldwide. Retrieved 12 July 2012, from http://www.bbc.co.uk/news/10569081
Brockmann, D., Hufnagel, L., & Geisel, T. (2006). The scaling laws of human travel. *Nature, 439*, 462-465.
Burris, V. (2005). Interlocking Directorates and Political Cohesion among Corporate Elites. *American Journal of Sociology, 111*, 249-283.
Butts, C. T. (2008). Social network analysis with sna. *Journal of Statistical Software, 24*, 1-51.
Butts, C. T. (2010). SNA: tools for social network analysis. *R package version, 2*.
Cattuto, C., Van den Broeck, W., Barrat, A., Colizza, V., Pinton, J. F., & Vespignani, A. (2010). Dynamics of person-to-person interactions from distributed RFID sensor networks. *PLoS ONE, 5*, e11596.
Choudhury, T., & Pentland, A. (2004). Characterizing Social Networks using the Sociometer. *Proceedings of the North American Association of Computational Social and Organizational Science (NAACSOS)*.
Choudhury, T., Philipose, M., Wyatt, D., & Lester, J. (2006). Towards activity databases: Using sensors and statistical models to summarize people's lives. *IEEE Data Eng Bull, 29*, 1-8.
Chowell, G., Hyman, J., Eubank, S., & Castillo-Chavez, C. (2003). Scaling laws for the movement of people between locations in a large city. *Physical Review E, 68*, 066102.
Christakis, N. A., & Fowler, J. H. (2007). The Spread of Obesity in a Large Social Network over 32 Years. *New England Journal of Medicine, 357*, 370-379.
Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695.
De Montis, A., Barthelemy, M., Chessa, A., & Vespignani, A. (2005). The structure of Inter-Urban traffic: A weighted network analysis. *Geographical, 34*, 12.
Dekker, D., Krackhardt, D., & Snijders, T. A. B. (2007). Sensitivity of MRQAP tests to collinearity and autocorrelation conditions. *Psychometrika, 72*, 563-581.
Dong, W., Lepri, B., & Pentland, A. (2011). Modeling the co-evolution of behaviors and social relationships using mobile phone data. *Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia*, 134-143.
Eagle, N., & Pentland, A. (2006). Reality mining: sensing complex social systems. *Personal Ubiquitous Comput., 10*, 255-268.
Eagle, N., Pentland, A. S., & Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 15274-15278.
Festinger, L., Back, K. W., & Schachter, S. (1963). *Social Pressures in Informal Groups: A Study of Human Factors in Housing*: Harper.
Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force directed placement. *Software: Practice and experience, 21*, 1129-1164.
Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. (2008). Understanding individual human mobility patterns. *Nature, 453*, 779-782.
Goodreau, S. M., Kitts, J. A., & Morris, M. (2009). Birds of a Feather, or Friend of a Friend? Using Exponential Random Graph Models to Investigate Adolescent Social Networks. *Demography, 46*, 103-125.
Isella, L., Stehle, J., Barrat, A., Cattuto, C., Pinton, J.-F., & den Broeck, W. V. (2011). What's in a crowd? Analysis of face-to-face behavioral networks. *Journal of Theoretical Biology, 271*, 166-180.
Krackhardt, D. (1988). Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social Networks, 10*, 359-381.
Madan, A. (2010). *Social Evolution: Opinions and Behaviors in Face-to-Face Networks.* Unpublished Doctoral thesis, Massachusetts Institute of Technology, Cambridge, MA, USA.
Madan, A., & Pentland, A. (2009). Modeling social diffusion phenomena using reality mining. *AAAI Spring Symposium on Human Behavior Modeling*.
McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology, 27*, 415-444.
Miluzzo, E., Cornelius, C. T., Ramaswamy, A., Choudhury, T., Liu, Z., & Campbell, A. T. (2010). Darwin phones: the evolution of sensing and inference on mobile phones. In (pp. 5-20): ACM.
Pentland, A., Lazer, D., Brewer, D., & Heibeck, T. (2009). Using reality mining to improve public health and medicine. *Stud Health Technol Inform, 149*, 93-102.
Pentland, A., Madan, A., Dong, W., & Reid, T. (2009). Social Evolution Dataset - Data Collected. In. http://realitycommons.media.mit.edu.
Perez, S. (2012). The Number Of Mobile Devices Will Exceed World's Population By 2012 (& Other Shocking Figures). Retrieved 29 June 2012, from http://techcrunch.com/2012/02/14/the-number-of-mobile-devices-will-exceed-worlds-population-by-2012-other-shocking-figures/
R Development Core Team. (2012). R: A Language and Environment for Statistical Computing. In.
Song, C., Qu, Z., Blumm, N., & Barabasi, A. L. (2010). Limits of predictability in human mobility. *Science, 327*, 1018-1021.
Thibaut, J. W., & Kelly, H. H. (1956). The Social Psychology of Groups. In *Social Science Classics Series* (pp. 39-43): Transaction Publishers.