Teasing Apart Behavioral Protocols in Longitudinal Self-reported Friendship Networks

Tanmay Sinha[1] **Wenjun Wang**[1] **Xuechen Lei**[2]

[1]**School of Computer Science**
Carnegie Mellon University
Pittsburgh, PA 15213
{tanmays,wenjunw}@cs.cmu.edu

[2]**Carnegie Institute of Technology**
Carnegie Mellon University
Pittsburgh, PA 15213
xuechenl@andrew.cmu.edu

Abstract

How can we effectively model the network-mediated process of social selection in longitudinal friendship? Self-reported measures have long been used as major information sources for tapping people's behavior. Recent years have witnessed highly cited research pushing this idea further to additionally utilize sensor-based information as means of obtaining fine-grained access to social behavior. While such approaches have been shown to be promising, there is lack of work that comprehensively evaluates the contribution of these two different information sources in decoupling group behavioral dynamics. In an attempt to unleash the combined potential of self-reports and sensor data in understanding social interaction, we utilized multidimensional self-reported survey data along with yearly call, SMS and bluetooth logs from 131 participants. By leveraging a statistical representation capable of encoding simultaneous social processes that mobilize and constrain individuals in a network, our results provide rich insights into behavioral rules that govern the dynamics of friendship among these participants.

1 Introduction

The human social system is a complex entity. We are intentional beings having multiple motivations and multiple expressions for social action. This is greatly reflected in the social choices we make and their impact on our behavior. Human social networks arising from such phenomena have been extensively studied in mathematical graph theory and social sciences with a strong focus on how, why and when actors connect to each other. Traditionally, formation of these social networks has been heavily dependent on self-reported measures [19]. Self-reports provide an explanatory lens into explicit perceptions of people and often serve as ground-truth data in experimental studies [22]. However, foundational studies suggest weak correlations between behavioral data and self-reported measures, highlighting systematic biases ranging from features of the research instrument [23], to cognitive influence that affects people's long-term and short-term behavior reporting [9]. Therefore researchers in the last decade have begun to acknowledge the added value in tapping rich fine-grained behavioral information from ubiquitous devices such as smartphones [17] to better understand people's interaction [5,7,8,30].

In this work, our objective is to utilize both self-reported as well as directly observable sensor information captured by smartphones (in form of calls, SMS and bluetooth proximity information) to empirically investigate the network-mediated process of *social selection*. Social selection occurs when individuals select or form social relationships on the basis of certain characteristics possessed by themselves and others. While a wealth of work in social network analysis has looked into non-network metrics as well as aggregated centrality based features to better understand the structure of human communication networks and their impact on people's behavior [15,21,25,26,27], problems of developing or utilizing principled approaches that can *explicitly* model dependency between network ties have emerged as an important theme in social computing and human factors research -

dependency, which makes social networks fundamentally different from other forms of relational data. Non-parametric bootstrapping approaches employed in Quadratic Assignment Procedures (QAP)[16] offer limited flexibility, since they create random network permutations with solely the same number of arcs, but not necessarily with other similar graph distributions (for instance same number of arcs and same number of reciprocated ties as the observed network). Moreover, bootstrapping approaches consider each effect, one at a time, completely ignoring the nesting of configurations. For instance, when there are significantly more transitive triads, it is impossible to know whether that arises because of an increased number of stars or because of the triangulation process itself.

At the heart of this dialog is thus the consideration of tie-dependence that is crucial to model selection based network effects. This also motivates an increased need to move beyond the assumption of temporal homogeneity that underlies methods using frequency counts of specific behaviors, and understand network self-organization from different perspective. In our work, we therefore seek to investigate how network ties depend on each other as a function of 3 effects: first, internal local network processes, which represent patterns formed from the network ties (from the mere virtue of people being connected to each other) and provide evidence for ongoing structural processes; second, factors that are external to the network, which mainly include attributes possessed by actors, and third, dyadic covariates that look simultaneously into the presence of multiple types of relations between actors.

Our *research question* in this study is the following: What insights can we get into the processes of network tie formation, by fusing two sources of data that are of fundamentally different nature: longitudinal sensor data (gathered through mobile phones) and surveys collected at different time granularities? More specifically, we are interested in knowing what social rules govern friendship dynamics (popularity, activity, reciprocity etc), and whether these dynamics are totally explained by such local mechanisms, or are there some external mitigating factors too? We answer this question by first performing theoretically grounded explanatory analyses that motivate presence of potential effects in our network (section 4), and then utilizing the methodology of Simulation Investigation for Empirical Network Analysis (SIENA) to quantify the strength of similarity selection in the dynamics of peoples behavior (section 5). Consistent with theory, our results (section 6) confirm that there is significant presence of reciprocity and tendency for dense triangulation patterns in the friendship network. We also found significant effects for males self-reporting more friends than females, while people with similar race, number of children and level of community involvement self-reporting each other as friends. However, results of big 5 personality traits revealed social selection effects somewhat consistent with prior research and somewhat inconsistent, thus painting a slightly different picture than when looking at each personality trait in isolation from others.

2 Study Context

We utilized the *Friends and Family* data [1] collected from a young-family residential living community in North America over a period of 1 year. The dataset comprised of a)sensor data collected from subjects' android based mobile phones via an external application that polled the device every 6 minutes to record call logs, SMS and bluetooth proximity information, b)individual demographic information such as race, gender, number of children and big-5 personality traits, c)social activities (leisure reading, watching TV, video, movies, serious discussion or arguments, work or school related tasks etc), level of community involvement, moods (happy, stressed, productive etc) and eating preferences (healthy eating and consumption of meat, seafood, grains etc) from weekly and monthly surveys. Thus this dataset was replete with rich information about people's daily lives, which provided an interesting opportunity to explore the effects of different kinds of behavioral variables on tie formation processes.

3 Related Work

3.1 Theoretical Framing

We drew on prior work as a theoretical lens to understand tie formation (why ties might be present in a network) and tie patterning (how ties might come to form local network patterns or configurations). For instance, reciprocity or exchange is one of the fundamental human activities [3], and so generally we would expect ties in a social network to be reciprocated. Similarly, triangulation patterns (which reflect the human propensity to operate in group like structures) in social networks have been traditionally studied as path closure and network closure [4]. In contrast to closure, theo-

ries of prominence in networks also suggest that people who are well connected may be advantaged or have a distinctive status [10]. Furthermore, the homophily principal implies that it is easier or even more rewarding for an actor to connect and interact with a similar other than a dissimilar other [20]. Processes of network tie formation such as these often lead to interesting phenomena such as preferential attachment [2] where popularity may induce further popularity. This fundamental theoretical grounding motivated empirical research that relates to our current work.

We situated our work in the *reality mining* framework proposed by [8]. The term reality mining refers to a the potential functionality of a ubiquitous and distributed infrastructure of mobile devices to unveil regular rules and structure in the behavior of both individuals and organizations. Smartphones allow for unobtrusive and cost-effective access to previously inaccessible sources of data related to daily social behavior. By a)continually logging and time-stamping information about a user's activity, location, proximity to other users and, b)leveraging such fine-grained features indicative of dyadic relationships to infer daily or weekly routines of people, the large-scale dynamics of collective human behavior can be analyzed. In recent years, this recognition has led to a growing chorus of social computing work that argues in favor of leveraging mobile phone as wearable sensors. Motivated by this rationale, we incorporated covariate networks built on SMS, call log and bluetooth data in our work, to study how they entrain and co-evolve with the social network of friendship among users.

In social psychology it is assumed that people's behavior can be explained to some extent in terms of underlying personality traits, which are seen as enduring dispositions relatively stable over time. Big-5 is a well known example of a multi-factorial model that describes an individual's personality through a number of fundamental dimensions known as traits (Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness), derived through factorial studies [12]. Prior work [13] has experimentally examined the effect of individual personality differences and psychological predispositions on their ego networks. There is also evidence to support the evidence that personality can be reliably predicted from standard mobile phone logs. Using summative features reflective of regularity (average inter-event time for calls and SMS), diversity (number and entropy of contacts), spatial and active behavior, the work of [21] was able to predict Big-5 personality traits using machine learning classifiers. However, no network metrics were used to assess characteristics of the underlying interaction network.

3.2 Empirical Research

In the process of making inferences on the patterns of relationships between users, prior work such as [25] has explored how mobile behavioral data collected via SMS, call log and bluetooth records can be used to a)build communication and proximity networks, b)extract basic network properties such as centrality (e.g. degree, betweenness, closeness and eigenvector), small world measures such as efficiency (how close to a small-world a given ego-network is) and aggregate triadic measures, c) predict Big-5 personality traits of users. Many of these insights have emerged from foundational work in social network analysis around modeling relationships among social entities, and on their patterns and implications [27]. In line with these studies, we investigated the hypothesis of an individual's personality traits affecting the formation of network (friendship) ties. However, in our work we exploited three sources of data to study the formation and impact of friendship ties, specifically combining a)fine-grained mobile sensor data, b)fixed actor attributes collected via demographic surveys, and c)variable actor attributes collected via monthly and weekly surveys, with structural configurations representative of network self-organization process. The hope was that addition of these local structural network patterns would help us understand the complex combination of social processes by which network ties are formed, which is hard to tap by just using aggregated centrality measures as features.

While [26] have studied a natural interdependence between network structure and individual characteristics of the network actors via classic network autocorrelation models and proved to some extent that a)under the homophily principle, network ties tend to form according to similarity on certain actor attributes, b)network autocorrelation emerges as a consequence of tie selection over time, their conclusion could only be applied to one specific personal behavior (music taste) and it is not very clear how well the results would generalize to other personal attributes. Moreover, they didn't take any static personal traits or covariate networks into consideration, which will be explored in our current study. Finally, while the work of [28] has supported the claim that communication is an indicator of social network tie strength, in the hindsight it has also questioned the assumption of

using communication as a direct proxy for tie strength. The authors demonstrated that using simple communication features such as communication frequency and duration to predict tie strength is not a good idea, especially when dealing with sparse data. Thus, in an attempt to utilize more communication channels that can serve as a proxy for face-to-face communication, we leveraged bluetooth proximity data in our work.

4 Explanatory Data Analysis

4.1 Social Network Creation

We constructed four longitudinal one-mode networks from our data for the months of September 2010, December 2010, March 2011 and May 2011: a)Longitudinal friendship network (extracted using friendship surveys in which people were asked to self report friends on a scale of 1 to 7; we recoded links and binarized this directed network by empirically choosing 4 as the threshold for establishment of a strong friendship tie), b)Longitudinal call log network (originally extracted at a temporal resolution of 6 minutes, but, aggregated monthly; the link weight in this $directed$ network represented the number of times person A called B and vice versa), c)Longitudinal SMS network (originally extracted at a temporal resolution of 6 minutes, but, aggregated monthly; the link weight in this $directed$ network represented the number of times person A sent SMS to B and vice versa), d)Longitudinal bluetooth proximity network (originally extracted at a temporal resolution of 6 minutes, but, aggregated monthly; the link weight in this $undirected$ network represented the number of times A and B were recorded to be in physical proximity to each other). Next, we analyzed what systematic differences were likely to underlie forms of conditional dependence among tie variables of same and distinct types, in order to understand what effects were likely to play a big or small role in social selection process. In the analysis below, we therefore performed comparison of the chosen effects with an Erdos Renyi random graph having same number of nodes and similar density.

4.2 Presence of Network Based Effects

Among structural effects, we discovered substantially higher counts in our operationalized networks (than the random graph) for reciprocity, transitivity, transitive path closure (that indicates tendency for structural holes to close when there is another path between the nodes - in terms of friendship, i tends to choose j and vice versa who are friends of their friends h) and complete subgraph (that is indicative of people tending to cluster together in denser regions of a network). On the other hand, counts of popularity (incoming star-like configuration) and activity (outgoing star-like configuration) were much lower than what would be present in a random graph. At a higher level, these analyses also suggest presence of multiple social processes that might operate in our friendship social network structure, with certain structural configurations being more likely and the others being less likely than in a random graph. However, since endogenous structural network effects are based on the assumption of homogeneity in the ways people are connected in the friendship network, we also utilized exogenous actor attributes to relax this homogeneity. In the literature there is ample, though not always converging, evidence of a relationship between social network measures such as centrality and Big-5 personality traits. For instance, according to [14], all the Big Five personality traits, with the exception of agreeableness, correlate closely with degree, and more precisely with in-degree. Gender, interestingly has been shown to be an important determinant in social selection, with prior research providing evidence for gender-based homophily effects in friendship networks. Interestingly, females are more likely to form intimate social networks than males [20]. Therefore, in order to assess the impact of Big-5 traits and some standard demographic variables on friendship tie formation, we binarized each attribute value is shown in table 1. We counted number of ties sent or received by actors based on their attribute values, to study the sender, receiver and homophily effects, which measure respectively the propensity of people posessing certain attributes to send ties, receive ties and bi-directionally self-report friendship connections with similar others.

For sender effects, we discovered that a)males self-reported sending more friendship connections than females, b)people with greater than 1 children sent more friendship connections, c)people who were very involved in the community had substantially more activity in terms of sending friendship connections. With regard to Big 5 Personality traits, we could infer that: a)people who were rated highly on conscientious and extrovert personality dimensions sent substantially more friendship connections. Prior work using Facebook data [11] has also found a positive correlation between the number of friends (taken as a measure of degree centrality) and extraversion, b)people who were rated highly on neuroticism dimension seemed to have more activity in terms of sending friend-

Table 1: Discretization of Actor Attributes

Big 5 Personality Scores Extraversion (1) versus Intraversion (0): *(8 questions: discretized into 0 (score <=25), 1(score>25))* Agreeableness (1) versus Antagonism (0): *(9 questions : discretized into 0 (score <=30), 1(score>30))* Conscientiousness (1) versus Lack of Direction (0): *(9 questions : discretized into 0 (score <=30), 1(score>30))* Neuroticism (1) versus Emotional Stability (0): *(8 questions : discretized into 0 (score <=25), 1(score>25))* Openness (1) versus Closeness to Experience (0): *(10 questions : discretized into 0 (score <=35),1(score>35))*
Gender: Male (1), Female (0)
Number of children: marked on a scale of 1-4 *(discretized into 0 (<=0 children) and 1(>0 children))*
Feel involved: marked on a scale of 1-5 points *(How involved do you feel in the community?), (discretized into 0 (<=2) and 1(>2))*
Race: Asian (1), Black (2), Hispanic (3), Middle Eastern (4), Mixed-Race (5), Native American (6), White (7), Other (8)

ship ties, c)people who were rated low on openness personality dimension also seemed to have the attribute based sender effect.

For receiver effects, we found that a)females received more friendship connections than males, b)people with greater than 1 children received more friendship connections, c)people who were very involved in the community had substantially more popularity in terms of receiving friendship connections. With regard to Big-5 personality traits, we could infer that: a)people who were rated highly on neuroticism and extrovert personality dimensions received substantially more (roughly twice) friendship connections than other personality types. This finding is in sync with prior work [6] that operationalized actor-based features (e.g. number and duration of calls, BT hits, etc) from recordings of real-life smartphone usage and personality surveys in order to automatically classify personality traits, and found that extroverts are more likely to receive calls and spend more time on them. The work of [15], however has found that people low in neuroticism tend to have high indegree centrality scores in advice and friendship networks, which is contradictory to our findings about receiver effects for people rated highly on neuroticism, b)people who were rated highly on conscientiousness dimension seemed to have more popularity in terms of receiving friendship ties, c)people who were rated low on openness personality dimension also seemed to posess the attribute based receiver effect. This finding is in sync with the study of [15], who have found negative correlation between in-degree centrality and openness, d)Receiver effects for agreeableness were less prominent.

For the homophily effects, we found that: a)homophily based effects were not prominent for the gender attribute, since count of ties within same and different gender were very similar. Intuitively, this also made sense because the study population comprised of couples, b)people with similar number of children formed substantially more friendship connections with each other than with different number of children, c)people who were very involved in the community formed substantially more friendship connections with each other than with people with different levels of involvement in the community (*rich get richer*). With regard to Big-5 personality traits, we could infer that: a)There seemed to be a very small homophily based effect for the personality traits of agreeableness, conscientiousness, extraversion and neuroticism, b)Homophily based effect seemed to diminish over the 4 network timestamps for people rated highly on openness, meaning that there was a smaller difference between total number of homophilous ties and total number of heterophilous ties over time.

Finally, we looked at dyadic covariate effects by including call, SMS and bluetooth networks to understand alignment in their pattern of ties with the friendship network. The motivating intuition was that people were more likely to be connected to others they felt close to, via multiple channels of communication. We investigated entrainment effects using Quadratic Assignment Procedure (QAP), which fundamentally compares the degree of observed association between networks (friendship and dyadic covariates) to that which would be expected to arise from a process in which individuals were randomly assigned to positions within the two networks, holding the structure constant. Overall, we found that the QAP regression estimates were significant at $p < 0.001$ and more preditive for call log (0.111, 0.198, 0.161, 0.155 over 4 timestamps) and bluetooth (0.147, 0.155, 0.143, 0.149 over 4 timestamps) networks, when compared to estimates for the SMS network (0.064, 0.038, 0.043, 0.016 over 4 timestamps), which although significant at $p < 0.001$, provided the least information in predicting friendship ties. Moreover, the SMS network got less predictive as we moved from Sept 2010 to May 2011. Overall, these results suggested significant presence of entrainment-based effect between friendship and the covariate networks, reflective of how ties in dyadic covariate networks predicted presence of ties in the friendship network.

5 Method

Oue explanatory analysis suggested interesting differences in network and attribute statistics in the friendship network in comparison to a random graph. As a next step, we tested significance of these results by considering potentially competing theoretical reasons why social ties in the observed friendship network might have arisen. We also focused on modeling the distribution of other stochastic actor attributes measured at an individual level across the friendship network of relational ties. We utilized the statistical methodology of Simulation Investigation for Empirical Network Analysis (SIENA) [24] for our longitudinal network data, the intuition being to find a distribution of graphs where the observed effects were central (we fitted the model by estimating parameters). This would allow us to discover which configurations were important in the network, which effects had independent explanatory value and which among others, could be explained by other effects. Essentially, SIENA comprises of 3 steps: simulation, estimation and heuristic goodness of fit (GOF). In the simulation step, the overarching goal is to generate a sequence of m graphs successively updated through small changes via Metropolis Hastings algorithm. After the algorithm converges, we choose the last graph in the sequence as our sample point. In the simulation step, each successive graph will be accepted with probability shown in Equation (1)

$$\min_{\theta}\{1, \frac{P_\theta(x^*)}{P_\theta(x^{(m-1)})}\} \tag{1}$$

$x^{(m-1)}$ is the previous state and x^* is the current state that may or may not be accepted. P is calculated according to Equation (2), $\theta_1, \theta_2 ... \theta_p$ are parameters and $z_i(G)$ are count of configurations that we selected based on our analysis in section 4.

$$P_\theta(G) = ce^{\theta_1 z_1(G) + \theta_2 z_2(G) + ... + \theta_p z_p(G)} \tag{2}$$

In the estimation step, maximum likelihood principle is used as guidance. Formally, we want the expected value of statistics of the sampling graph sequence to be equal to the observed statistics (as shown in Equation (3) and (4))

$$z_\theta - z(x_{obs}) = 0 \tag{3}$$

$$z_\theta = \frac{1}{M}(z(x^{(1)} + x^{(2)} + ... + x^{(M)})) \tag{4}$$

Robbins-Monro algorithm is applied here to perform a heuristic approximation on optimal parameters, with an objective to find a solution to equation (3). Step size decreases as the algorithm progresses. The update rule for parameters is shown in Equation (5)

$$\theta^{(m+1)} = \theta^{(m)} - \alpha_r D_0^{-1}(z(x^{(m)}) - z_{obs}) \tag{5}$$

D_0, the diagonal value of the covariance matrix D of configurations, is calculated as shown in Equation (6)

$$D = \frac{1}{M_1}\sum_m z(x^{(m)}) z(x^{(m)})^T - z_{\bar{\theta}} z_{\bar{\theta}}^T \tag{6}$$

Finally, in assessing GOF, we see how well the model captures features that are not explicitly modeled. The t-ratio of a feature X, which represents the standardized difference [Mean value of X (in observed graph) - Mean value of X (in simulated graphs) divided by standard deviation of X (in simulated graphs)], is calculated according to Equation (7)

$$\frac{\tilde{z}_\theta - z(x_{obs})}{SD_\theta(\{z_k(x^{(m)})\}_{m=1,2,...,M_3})} \tag{7}$$

6 Results

We conducted 3 sets of experiments following the methodology of SIENA, which mainly focused on: a)how network substructures shape the evolution of the friendship network, b)how static actor attributes affect evolution of the friendship network, c)how dyadic covariates influence evolution of the friendship network. As a sanity check, we looked at tie changes between subsequent observations of the friendship network and found the jaccard distance (number of ties maintained from 1 network wave to the other, divided by the sum of number of ties created, destroyed and maintained from 1 network wave to the other) to be greater than 0.5, in turn signaling that our friendship networks could indeed be considered as evolving. For the next set of results described below, we considered an estimate significant if it was 2 times the standard error [18, p. 177]. Moreover, a graph statistic with t-ratio less than 2.0 (in absolute value) was considered to reasonably capture the corresponding effect in the observed friendship network [18, p. 182]. For each of the significant effects, we outline the SIENA *(estimate$\pm$standard error, t-ratio)* below.

6.1 Modeling Endogenous Structural Effects

We included structural effects described in our explanatory analysis in section 4. The only difference is that we included *k-triangle* configurations for transitivity-closure and complete subgraph to allow for measurement of highly clustered regions, which essentially comprised of two connected nodes that were also jointly connected to k other nodes. We found that there was high propensity for ties to be reciprocal (2.41$\pm$0.09, -0.04). In addition, a significant positive estimate for transitive triplets (0.35$\pm$0.03, -0.03) suggested high tendency for group formation in the friendship network. However, on looking at estimates for transitive triplets and k-triangle transitivity closure (1.64$\pm$0.08, 0.38) in combination, we could infer that there was significantly stronger propensity for triangles to cluster together and occur in denser graph regions, rather than being homogenously distributed. Moreover, we saw a significant negative estimate for the a)outdegree (arc) effect (-2.99$\pm$0.07, 0.16), which was like an intercept in a linear regression, indicating baseline propensity for tie formation (it was expected because of the sparsity of the network), b)activity based effects (-0.05$\pm$0.009, 0.12), suggesting most actors had similar levels of activity (the network was not centralized on out-degree), c)transitive reciprocated triplets (-0.21$\pm$0.02, -0.21). Looking at estimates of transitive triplets and its reciprocated version together, we could further infer that the way closure happened was usually uni-directional, meaning that if i was a friend of j and j was a friend of h, the more likely form of closure was that only i reported h as a friend and not vice versa (closure was not bi-directional).

6.2 Modeling Exogenous Actor Attribute Effects

We included actor attribute based sender, receiver and homophily effects described in our explanatory analysis in section 4. However, we divided our experiments into two parts. First, we looked at the effect of Big-5 personality traits on social selection in the friendship network. We then investigated the effects of demographic attributes. The results for Big-5 personality traits clearly suggested that there was a significant sender (0.25$\pm$0.10, -0.09) and receiver (0.52$\pm$0.13, -0.02) effect for people who were extroverts. Extroverts self-reported themselves as more socially active. Sender effects for conscientiousness and openness, receiver effects for neuroticism, and similarity effects for conscientiousness and extraversion were also positive and in sync with intuitions formulated on the basis of our explanatory analysis, although not significant. More careful investigation of these subtle variations in Big-5 effects will be required in future work. Second, we looked at the effects for demographic variables. These results suggested that there was a significant sender effect for gender (0.12$\pm$0.06, 0.01). Moreover, there was a receiver (0.03$\pm$0.01, 0.04) and homophily (1.04$\pm$0.08, -0.09) effect associated with race, signaling that people with certain race categories tended to receive significantly more ties, and there was also a tendency of friendship tie formation within the same race. We also found that the level of community involvement had significant presence of sender (0.32$\pm$0.07, 0.05), receiver (0.16$\pm$0.05, -0.04) and homophily (0.25$\pm$0.07, 0.05) effects. Finally, a similiar pattern existed for the number of children attribute with respect to sender (0.19$\pm$0.06, 0.02), receiver (0.16$\pm$0.05, -0.04) and homphily (0.20$\pm$0.05, 0.07) effects.

6.3 Modeling Variable Dyadic Covariate Effects

We included tie entrainment based effect for modeling dyadic covariate based effects, as discussed in the explanatory anaysis in section 4. In addition, we also incorporated exchange-based effects, meaning that if i and j were connected in a dyadic covariate network W, how did it affect the probability of j self-reporting i in the friendship network X. For bluetooth network, we found a small and significant positive estimate for entrainment (0.0002$\pm$0.00, 0.005), but significant negative estimate for exchange (-0.0002$\pm$0.0001, -0.01), suggesting that the more node i and j stayed in physical proximity to each other, the more likely was there a strong friendship tie from i to j self-reported by i, but it was less likely that j reported a strong friendship tie too. We discovered a similar pattern for entrainment (0.015$\pm$0.005, -0.08) and exchange (-0.01$\pm$0.006, 0.04) for the call log network, while the SMS network was not very predictive in telling us about the perceived friendship network.

7 Discussion and Conclusion

In essence, we studied three categories of effects to understand social interaction and structure in the social behaviors of individuals in the community. We applied statistical frameworks to give an expression to propositions about the outcomes of dynamic, interactive and local processes that drive friendship network formation. This in turn allowed us to quantify and assess the observable regularities in the social network structure implied by these propositions. The models we applied

are well aligned with many contemporary theoretical views about the evolution of social networks and provided an opportunity to demonstrate potential links between hypothesized network processes and observable network regularities. The SIENA representation allows us to include endogeneity in tie formation, and satisfactorily reproduce different aspects of the network structure that explain the nature and extent of local clustering, distribution of node to node connectivity etc in human communication networks. Intuitively, this is done by centering the distribution of statistics over those of the observed network. Moreover, with application of these models to longitudinal data (such as the friendship network), we stand to learn a great deal about the applicability and robustness of the stronger assumptions we make in cross-sectional observations. The dual interpretability made affordable by using SIENA models mirrors both theoretical considerations - the way individuals form ties but at the same time are also constrained and affected by structure. Finally, since we not only separately characterize possible endogenous processes but also explore them in the presence of unobserved exogenous variables, we do not make any homogeneity assumptions about whether structural effects (such as transitivity, reciprocity etc.) hold for all people or only for people having certain attributes.

However, there are a couple of limitations that are important to consider while interpreting our results: First, we relied on Big 5 personality traits for our analysis, since it was available for our chosen corpus. However, there is a good amount of criticism and opposition to using the big 5 personally traits to quantify different personalities [29], and one could argue that it is not universally accepted as valid. Furthermore, since the Big 5 personality traits were collected only once in our dataset, we could not look at the fleeting personality states that people have been shown to go through in their daily lives (via experience sampling) [31]. Second, in considering the results of the present study, we must note that the dataset is small and may not be representative of a general population, and while some insights that we present could be attained from such a dataset, deeper insights seem to require further probing into the results, for example by interviewing the participants.

There are some potentially exciting course of research for the future. First, when dealing with dyadic covariate data such as the call log, SMS and bluetooth networks, there is an interesting possibility to investigate certain cross network clustering effects. For instance, if i sends an SMS to h, h sends an SMS to j, how does it affect the probability of a friendship tie between i and j (closure of a covariate)? Or, if h calls i and j, how does it affect the probability of a friendship tie between i and j (shared incoming ties)? Alternately, if i and j are in close bluetooth proximity to h, how does it affect the probability of a friendship tie between i and j (shared outgoing ties)? Second, knowing more about robustness of results obtained by fitting SIENA models is an important part of the analysis loop that needs more work. We seek to better understand the extent to which our tests compromised if one or more of the structural, actor attribute or dyadic covariate based effects are omitted from model specification. Is it true that this leads to network dependencies being represented only incompletely, and hence can potentially lead to wrong inferences being drawn?

Acknowledgments

The authors thank Kathleen Carley and Alex Smola for their invaluable feedback.

References

[1] 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(6), 643-659.

[2] Barabsi, A. L., & Albert, R. (1999). *Emergence of scaling in random networks. Science*, 286(5439), 509-512.

[3] Blau, P. M. (1964). *Exchange and power in social life*. Transaction Publishers.

[4] Cartwright, D., & Harary, F. (1956). Structural balance: a generalization of Heider's theory. *Psychological review*, 63(5), 277.

[5] 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(7), e11596.

[6] Chittaranjan, G., Blom, J., & Gatica-Perez, D.: Whos who with Big-Five: Analyzing and Classifying Personality Traits with Smartphones. *ISWC* (2011)

[7] Eagle, N., Pentland, A. S., & Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, 106(36), 15274-15278.

[8] Eagle, N., & Pentland, A. (2006). Reality mining: sensing complex social systems. *Personal and ubiquitous computing*, 10(4), 255-268.

[9] Freeman, L. C., Romney, A. K., & Freeman, S. C. (1987). Cognitive structure and informant accuracy. *American anthropologist*, 89(2), 310-325.

[10] Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social networks*, 1(3), 215-239.

[11] Golbeck, J., Robles, C., Edmondson, M., & Turner, K.: Predicting Personality from Twitter. *SocialCom/PASSAT*, (2011)

[12] John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research*, Vol. 2, pp. 102138. New York: Guilford Press

[13] Kalish, Y. & Robins, G.L.: Psychological predispositions and network structure: The relationship between individual predispositions, structural holes and network closure. *Social Networks*, (2006)

[14] Kanfer, A. & Tanaka, J.S.: Unraveling the Web of Personality Judgments: The Influence of Social Networks on Personality Assessment. *Journal of Personality*, 61(4) 711-738 (1993)

[15] Klein, K. J., Lim, B. C., Saltz, J. L., & Mayer, D. M.: How do they get there? An examination of the antecedents of network centrality in team networks. *Academy of Management Journal*, 47, 952-963 (2004)

[16] Krackhardt, D. (1988). Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social networks*, 10(4), 359-381.

[17] Lane, N. D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T.,& Campbell, A. T. (2010). A survey of mobile phone sensing. *Communications Magazine*, IEEE, 48(9), 140-150.

[18] Lusher, D., Koskinen, J., & Robins, G. (Eds.). (2012). *Exponential random graph models for social networks: Theory, methods, and applications*. Cambridge University Press.

[19] Marsden, P. V. (1990). Network data and measurement. *Annual review of sociology*, 435-463.

[20] McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 415-444.

[21] de Montjoye, Y. A., Quoidbach, J., Robic, F., & Pentland, A. S. (2013). Predicting personality using novel mobile phone-based metrics. *In Social Computing, Behavioral-Cultural Modeling and Prediction*, pp. 48-55. Springer Berlin Heidelberg.

[22] Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of management*, 12(4), 531-544.

[23] Schwarz, N. (1999). Self-reports: how the questions shape the answers. *American psychologist*, 54(2), 93.

[24] Snijders, T. A. (2001). The statistical evaluation of social network dynamics. *Sociological methodology*, 31(1), 361-395.

[25] Staiano, J., Lepri, B., Aharony, N., Pianesi, F., Sebe, N., & Pentland, A. (2012, September). Friends don't lie: inferring personality traits from social network structure. In Proceedings of *the 2012 ACM conference on ubiquitous computing*, pp. 321-330. ACM.

[26] Steglich, C., Snijders, T. A., & West, P. (2006). Applying SIENA: An Illustrative Analysis of the Coevolution of Adolescents' Friendship Networks, Taste in Music, and Alcohol Consumption. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 2(1), 48.

[27] Wasserman, S. (1994). *Social network analysis: Methods and applications*, Vol. 8. Cambridge university press.

[28] Wiese, J., Min, J. K., Hong, J. I., & Zimmerman, J. (2015). You Never Call, You Never Write: Call and SMS Logs Do Not Always Indicate Tie Strength. In Proceedings of *the 2015 conference on Computer supported cooperative work-CSCW15*.

[29] Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological bulletin*, 117(2), 187.

[30] Stopczynski, A., Sekara, V., Sapiezynski, P., Cuttone, A., Madsen, M. M., Larsen, J. E., & Lehmann, S. (2014). Measuring large-scale social networks with high resolution. *PloS one*, 9(4), e95978.

[31] Fleeson, W. (2001). Toward a structure-and process-integrated view of personality: Traits as density distributions of states. *Journal of personality and social psychology*, 80(6), 1011.