

Inferring Individual Social Capital Automatically via Phone Logs

VIVEK K. SINGH, Rutgers University and Massachusetts Institute of Technology
ISHA GHOSH, Rutgers University

Social capital is one of the most fundamental concepts in social computing. Individual social capital is often connected with one's happiness levels, well-being, and propensity to cooperate with others. The dominant approach for quantifying individual social capital remains self-reported surveys and generator-methods, which are costly, attention-consuming, and fraught with biases. Given the important role played by mobile phones in mediating human social lives, this study explores the use of phone metadata (call and SMS logs) to automatically infer an individual's social capital. Based on Williams' Social Capital survey as ground truth and ten-week phone data collection for 55 participants, we report that (1) multiple phone-based social features are intrinsically associated with social capital; and (2) analytics algorithms utilizing phone data can achieve high accuracy at automatically inferring an individual's bridging, bonding, and overall social capital scores. Results pave way for studying social capital and its temporal dynamics at an unprecedented scale.

CCS Concepts: • **User Machine Systems** → **Human Factors**; *Computer Applications*; Social and Behavioral Sciences; **Human-centered computing** → **Ubiquitous and mobile computing** → **Empirical studies in ubiquitous and mobile computing**

KEYWORDS

Social Capital; Mobile Phones; Social Behavior; Measurement; Computer-Mediated-Communication; Ubiquitous Computing

ACM Reference format:

Vivek K. Singh, and Isha Ghosh. 1997. Inferring Individual Social Capital Automatically via Phone Logs. *ACM PACMHCI*, 1, CSCW, Article 95 (September 2017), 12 pages. <https://doi.org/10.1145/3134730>

INTRODUCTION

Social capital describes the ability of individuals or groups to access resources embedded in their social network [8, 12]. Its presence in a network has been found to have significant effects at community as well as individual level. At a societal level, multiple studies have connected an increase in social capital to better public health, lower crime rates, and more efficient financial markets [2]. Putnam (2000) links a decline in social capital to increased social disorder, reduced participation in civic activities, and potentially more distrust among community members [40]. On an individual level, social capital has been connected with higher levels of life satisfaction, trust, and mental health [26].

The use of social capital in various disciplines (politics, sociology, public policy, communication, social computing) has underscored the challenges in identifying a reliable measure for it.

Author's addresses: Vivek K. Singh, 4 Huntington St., New Brunswick, NJ 08901, USA; Isha Ghosh, Vivek K. Singh, 4 Huntington St., New Brunswick, NJ 08901, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org. 2573-0142/2017/November - 95 \$15.00
Copyright is held by the owner/author(s). Publication rights licensed to ACM. <https://doi.org/10.1145/3134730>

An early study of social capital focused on societal-level social capital and used levels of social trust, participation in voluntary associations, and other forms of political and civic engagement as indicators of social capital [40]. However, this method was critiqued as measuring the cause and consequences of social capital rather than its constructs [20, 30].

Similarly, multiple approaches have also been proposed to measure individual level social capital (the focus of this work). Traditional approaches for quantifying individual social capital have focused on aspects that could be simply observed (e.g., gender, race, age) or elicited in a small period of time in lab settings (e.g., via surveys and generator methods) [24, 32, 33, 21, 50]. However, each of these approaches relies on active human reporting and must contend with numerous hurdles such as subjectivity in re-ports and observations, social and cognitive biases, and narrow observation chances, while dealing with pressures such as budget, time, and the effort required [22].

Social capital has also been formulated as a function of an individual's interactions with their strong and weak ties [31]. Social network sites in particular allow for a characterization of an individual's network and quantify their interactions with bridging and bonding ties within the network. Social media usage, therefore, has been studied in an attempt to provide a standardizable measure of social capital [16, 17, 9]. More recently, smartphones have created convenient interfaces to access social network apps and remain "constantly connected". Research has explored the impact of this affordance on an individual's social capital and found a positive relationship between the intensity of access to social network sites via smartphones and social capital [28, 39]. These studies however, still rely on the use of self-reported data for measuring social behavior, and in turn, quantifying social capital. This again limits the scalability of these approaches due to the costs of human time and attention. Similarly, recent studies have also tested the relationship between communication frequency and tie strength [53]. While this study operationalizes communication using phone metadata, it attempts to quantify tie strengths rather than looking at the overall social capital that an individual has access to. *As such, there are no current approaches that infer an individual's social capital levels without the need for active human effort.*

As the use of smartphones continues to grow, the volume of metadata generated by individuals using these phones is also increasing. Emerging research has explored the potential of using information passively (i.e. without active human effort) gathered from these phone-mediated interactions to make predictions about an individual's economic status [7], personality traits [13], and well-being [43].

Building upon this line of work, this exploratory study examines the possibility of using phone metadata to automatically infer an individual's social capital by investigating the following research questions:

RQ1: Do long-term phone-use patterns have some intrinsic associations with an individual's social capital?

RQ2: Can a data analytics algorithm be used to automatically infer individual social capital based on phone metadata?

The ability to assess a user's social capital through automatically generated data could allow for individuals to keep track of their own social capital scores (just like fitness trackers and credit scores) easily over time and have options to receive suggestions and recommendations.

As a first step towards this vision, this work analyzes data from a 10-week long field study involving phone metadata collection via a mobile app and uses Williams Social Capital Survey as the "ground truth" [54]. We report that a Lasso regression analysis of phone metadata can be used to automatically infer an individual's social capital with high accuracy.

2 STUDY

The data used in this paper was obtained as part of the Rutgers Well-being Study. The participants for the study were recruited using flyers, email announcements, and social media posts in the area surrounding Rutgers University. The participants needed to: (1) be between 18-75 years of age; (2) comfortable with written and oral English; (3) use an Android smartphone; (4) carry their phone on them most of the time; and (5) be willing and able to travel to the study site for three in-person sessions. This can hence be considered a convenience sample, which was considered acceptable, given the exploratory nature of this work.

Participants in this study were asked to attend three in-person sessions for surveys and install a mobile application onto their smartphone. A screenshot of the app is presented in Figure 1. The app was developed using the “Funf in a box” framework [3] and was released via a URL shared with the study participants.

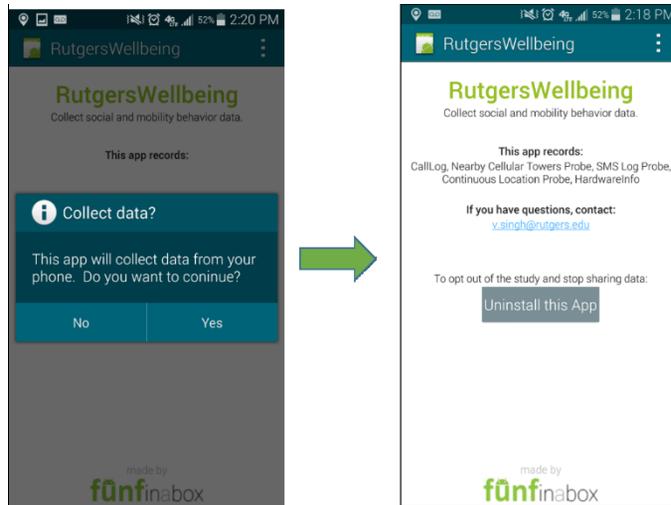


Fig. 1. Screenshot of the app used for this study.

The app recorded anonymized call and SMS metadata (calls/sms initiated or received - number, times and anonymized id but no actual audio or text). The app also recorded location metadata, which is not relevant to the current discussion, though.

The study included 59 participants. However, some of the participants did not complete all the surveys, and some did not enter their unique identifying code consistently across different surveys, resulting in a set of 55 participants for whom we have the mobile-based data as well as the scores for the two surveys of interest (more details on surveys presented later). The survey order was randomized for different participants. Participation in this study was voluntary to the study was incentivized monetarily. The participants were compensated up to a sum of \$100 on successful completion of the study over the ten-week period.

The sample consisted of 36 male and 19 female participants (university population is 50% male, 50% female). The most common age group for participants was 18-21 years, the most common education level was “some college,” the median annual family income was in the range \$35,000-\$49,000 and 96% of them were single. All personnel involved with the study underwent human subject training and IRB certification. All data were anonymized before analysis.

A summary of the data collected over the ten-week period in Spring 2015 is shown in Table 1. The participants on average made 511 phone calls (median = 312; minimum = 43; maximum = 2,711) and exchanged 3,413 sms messages (median = 2,423; minimum = 21; maximum = 17,625) over this period.

Table 1. Summary of data considered in this study.

Data type	Number of participants	Data points
Calls	55	28,132 calls
SMS	55	187,720 messages
Surveys	55	2 surveys per participant.

3 MEASURES

3.1 Social Capital

This survey was based on the Williams (2006) Internet Social Capital Scales (ISCS) [54]. This survey has been actively used by multiple studies in the literature (>750 citations as per Google Scholar) including recently by one to study associations between Facebook usage and social capital [17]. While the original survey was intended to measure ‘online’ social capital and contrast it with the ‘offline’ social capital, we were interested in a general purpose interpretation of social capital. Hence, we removed the words ‘online’ and ‘offline’ in the survey to get a general purpose understanding of an individual’s social capital. For example, the question “when I feel lonely, there are several people online (respectively offline) I can talk to”, was simply replaced by “when I feel lonely, there are several people I can talk to.”

This survey consisted of 20 questions (10 related to bonding social capital and 10 related to bridging social capital). While bridging social capital focuses on access to newer resources and information obtained from diverse ties, bonding social capital quantifies the access to emotional and substantive support from close ties. The participants were asked to answer the question on a 5-point scale ranging from Strongly Disagree (1) to Strongly Agree (5). The sum of bridging and bonding scores gave the overall social capital score.

A descriptive summary of the scores is shown in Table 2.

Table 2. Descriptive summary of social capital scores

	Minimum	Maximum	Median	Mean
Bridging Social Capital	27	50	40	40.45
Bonding Social Capital	21	45	35	35.85
Overall Social Capital	55	93	76.30	75

3.2 Demographic Descriptors

The participants were also asked questions about their demography (age, gender, marital status, ethnicity, level of education, and family income level).

3.3 Characterizing Phone-based Social Behavior

At a conceptual level, social capital has been connected with an individual’s relative position in the network [30]. On a more granular level, the influence of strong and weak ties in a network has also been connected to social capital [52]. Such features have been operationalized over online social networks [25] and recently over phone networks in different contexts [13, 53]. At an empirical level, multiple recent efforts have quantified different aspects of individual behavior based on phone metadata with a goal to identify personal traits and well-being [13, 36, 42, 45, 24, 27].

Based on a survey of existing literature we characterize phone use features into five categories to understand different aspects of an individual’s social behavior. In defining these features (total 26), we integrate some of the more obvious usage patterns (total number of calls/sms) with more nuanced/exploratory aspects of phone usage while still keeping the number of features manageable.

A summary of the features considered in this study is presented in Table 3.

Table 3. Summary of features considered in this study.

Types of features	Prior literature support	Features
Communication Frequency	Coleman[8]; Ellison et al. [12]; Min et al. [30]; deMontjoye et al. [9]	Call_Count, Sms_Count, Distinct_users_call, Distinct_users_sms, Distinct_users_called_outgoing Total_call_duration, Number_of_calls_per_day
Reciprocity	Putnam [34]; Williams [45]; Ghosh et al. [18]; Kovanen et al. [23]	InOut_ratio Call_response_rate Call_response_latency
Tie Strength	Lin [24]; Granovetter [17]; Gilbert et al. [19]	Call_strong_ties_ratio Sms_strong_ties_ratio Call_strong_ties_value Sms_strong_ties_value
Diversity, Loyalty, and Novelty in interactions	Putnam [34]; Granovetter [17]; Singh et al. [37]; Eagle et al. [11]	Call_Entropy Sms_Entropy Call_Loyalty Sms_Loyalty Number_of_new_contacts_called Number_new_contacts_outgoing Time_spend_with_new_contacts
Temporal Rhythms	Abdullah et al. [1]; Saeb et al. [35]; deMontjoye et al. [9]	DayNight12am_call DayNight12am_sms Auto_regularity_1 Auto_regularity_7 Auto_regularity_31

3.3.1 *Communication Frequency.* Prior research links the frequency of interactions with an individual’s network with their social capital [12, 16]. From the perspective of phone usage, interaction frequency can be measured as a function of the total calls/sms exchanged by the individual [36, 13]. This includes the calls/sms made to and received from distinct individuals, the total duration of calls, and the number of calls made per day. A high count on these measures would imply that the person is more socially active.

3.3.2 *Reciprocity.* Along with the frequency of communication, the ease with which communication is conducted can also be a significant predictor of an individual’s social capital [41, 54]. People who are easy to contact can be said to engage with a wider social network than those who remain unavailable. Here, ease of access is operationalized using features related to accepting or rejecting calls [24, 29]. This includes the ratio of incoming to outgoing calls, call response rate (percentage of missed calls responded back), and the time taken to respond to missed calls. These features taken together provide an understanding of how keen is the individual to engage with calls initiated by others in the network.

3.3.3 *Tie Strength.* Significant prior literature connects strong and weak ties with social capital [30, 23]. Since frequency of interaction is an important predictor of tie strength [25], to quantify the relative role of different types of ties in one’s social network we approximate “strong ties” as the top-third most frequent of their contacts [45, 56]. For both calls and SMS, we quantify the number of interactions that take place with these “strong ties” as well as the relative percentage of all interactions that take place with these contacts.

3.3.4 *Diversity, Loyalty, and Novelty in Interactions.* The relationship between a diverse network and an increased social capital is well documented in literature [41, 23]. Prior literature has also connected phone-

based diversity and loyalty with an individual's propensity to cooperate [45]. From the perspective of phone usage, we quantify diversity as:

$$D_i = -\sum_j p_{ij} \log_b p_{ij} \quad (1)$$

where p_{ij} is the percentage of communication (call/sms) between individual 'i' and 'j', and 'b' is the total number of contacts. This diversity score is based on Shannon Entropy and is a measure of how evenly the individual communicates across different contacts [15]. An individual with low diversity distributes her communication unevenly across contacts, whereas an individual with high diversity spends time evenly across many contacts. We similarly, define "loyalty" as the percentage of communication (call/sms) occurring only with the top three contacts of an individual [47]. A high loyalty score (close to 1) would indicate a tendency to prefer repeated interactions with a small number of favourite connections.

Further, an important aspect of social capital is the growth of networks. Hence, we also consider "new contacts" i.e. those who are not present in the first four weeks of the data collection period. We characterize the number of new contacts, number of such new contacts for outgoing calls, and the time spent in talking to these new contacts.

3.3.5 Temporal Rhythms. While mobile phones allow us to be always connected and available, prior research has documented a difference in the interactions made during phases of the day (e.g. morning, evening, night). Further, interactions during late night (past midnight) have been shown to have predictive power on mental health and well-being [1, 27, 42]. Hence, to understand the rhythms in user activity i.e. consistency across different phases and to capture the late night communication activity, we quantify temporal phone usage pattern as the ratio of calls/sms received during AM Phase (12 AM – 1:59 AM) to calls/sms received during the PM phase (12 PM – 11:59 PM). In the considered dataset individuals tended to have more communication during the PM phase. Hence a higher score for this metric would indicate a more even spread of communication across 24 hours and also indicate more late night and before noon activity.

Further, multiple recent studies have connected regularity in an individual's behavior with aspects of their personality, financial outcomes, and academic performance [13, 46, 10]. We model the regularity for an individual based on ARIMA (autoregressive integrated moving average) time-series coefficients [48, 9]. Specifically, we define regularity scores based on coefficients quantifying the degree to which knowing a person's number of calls made on an "earlier" day (yesterday, previous week, and previous month) can predict their number of phone calls for today.

4 RESULTS – INFERRING SOCIAL CAPITAL

4.1 Predictive Model

Predicting social capital level is a regression problem; that is, predicting an outcome variable (i.e., social capital level) from a set of input predictors (i.e., phone-based features). We utilize Lasso (Least Absolute Shrinkage and Selection Operator) [48] regularized linear regression model as our predictive model. Lasso minimizes the sum of squared errors, with a bound on the sum of the absolute values of the coefficients. Thus, Lasso automatically selects more relevant features and discards redundant features. We have $D = 26$ input features and $N = 55$ training cases. Lasso has been specifically designed to help identify relevant features (i.e., predictors) in such settings (N not much greater than D), while avoiding overfitting [5].

We use Pearson's correlation coefficient (r) between the model's predicted value and the "ground truth" survey value to quantify the prediction performance. This is also closely related to the coefficient of determination (R^2) of the model [35]. Both, r and R^2 values range from 0 to 1, where 1 indicates that the model perfectly fits the data. We also report the mean absolute error (MAE). A smaller MAE is preferred because it indicates that the predictions are closer to the ground truth. The MAE is in the same unit as the predicted variable and is hence relatively easy to interpret.

4.2 Prediction Results

We evaluate nine different prediction models in this work. These correspond to three different target variables (overall social capital, bonding social capital, and bridging social capital) and three different approaches (phone feature-based, demography-based, and phone + demography based). Demography based features have been connected with social capital in numerous studies (e.g., [55]) and hence we use this approach as a baseline. Specifically, the difference between the demography-based model and the phone-features based model helps us interpret the predictive power of phone-based features. Further, the combined phone + demography model quantifies the performance in scenarios where demography information is already available (e.g., to the phone company, automatically derived via phone features [57], or surveyed once to allow for repeated social capital inferences over time.)

We apply leave-one-subject-out cross validation [5] to determine the parameters for Lasso and the weights for each feature in each evaluation. The results for the evaluation in terms of correlation coefficients between predicted and actual scores for the target variable are summarized in Table 4. (The results for R^2 follow a similar pattern with squared values.)

Table 4. Correlation between the predicted values based on different models and the real values of overall social capital, bonding social capital, and bridging social capital.

	Phone Features based Model	Demography based Model	Phone + Demography based Model
Overall Social Capital	0.70	0.20	0.79
Bonding Social Capital	0.71	0.16	0.80
Bridging Social Capital	0.63	0.30	0.78

As can be seen in Table 4, the predicted social capital strongly correlates with the ground truth with $r = 0.70$ ($p < 0.01$). The results for bridging and bonding social capital also follow a similar trend and the corresponding r values are 0.71 and 0.63 respectively. As a baseline for comparison we also consider models that focus only on the available demographic features (age, gender, marital status, ethnicity, level of education, and family income level). While demographics explain some variance in the social capital levels, it can be clearly seen that phone-based features provide much higher predictive performance than that obtained by using only demography-based features. This underscores the value of phone-based features in automatically predicting social capital scores for individuals.

Lastly, the models that use both phone and demography based data yield predictions with correlations in the range of 0.78 to 0.80. This suggests that phone features are not merely replacements for demography-based features but rather, add complementary information. The corresponding R^2 values for these models are in the range 0.62, 0.64, and 0.61 respectively indicating that the models capture more than 60% of the variance in the social capital levels.

Table 5. Mean Absolute Error (MAE) for the predicted values based on different models and the real values of overall social capital, bonding social capital, and bridging social capital.

	Phone Features based Model	Demography based Model	Phone + Demography based Model
Overall Social Capital	4.96	6.70	4.39
Bonding Social Capital	2.47	3.45	2.11
Bridging Social Capital	3.39	4.21	2.70

We also report the Mean Absolute Errors between the actual and predicted values in Table 5. For instance, the MAE for the predicted social capital score using phone and demographic data was found to be 4.39, indicating that the predictions are within ± 4.39 of the “ground truth” or survey based values of social capital scores. We also note that the phone based models consistently yield lower errors than demography based models, and combining the phone and demography data yields the models with the lowest error rates. Note also that the values for social capital score can theoretically range between 20 and 100, hence a range of ± 4.39 as obtained by the predictive models here can be considered a reasonable approximation of the survey based score.

4.3 Correlation Analysis

To understand the relative effect of different phone-based features on bridging, bonding, and overall social capital, we undertook a *post-hoc* Pearson’s correlation analysis between the social capital scores obtained from the survey and phone-based features. In the interest of space we focus on the associations found to be significant for at least one of the three variables of interest (overall social capital, bridging social capital, and bonding social capital) and the results are summarized in Table 6.

Table 6. Bivariate correlation between phone-based features and overall social capital, bonding social capital, and bridging social capital.

	Bridging Social Capital	Bonding Social Capital	Overall Social Capital
InOut_ratio	-0.35**	-0.08	-0.27
Distinct_users_call	0.30*	0.07	0.23
Distinct_users_sms	0.28*	0.09	0.23
Distinct_users_called_outgoing	0.34*	0.10	0.27*
Call_response_latency	-0.29*	-0.09	-0.23
Number_new_contacts_outgoing	0.27*	0.08	0.21
DayNight12am_sms	0.27	0.28*	0.31*

For bridging social capital, the highest association was found with the InOut ratio of calls ($r = -0.35$). The negative association underscores the importance of outgoing calls in social capital. From the perspective of social capital (especially, bridging social capital), it is just as important to initiate relationships as it is to be available to other members of the network. Similarly, we note that the number of individuals contacted in outgoing calls is positively associated with bridging social capital ($r = 0.34$). Further, the number of distinct users in one’s network (call or SMS) and new contacts are positively associated with bridging social capital. Lastly, giving prompt response to missed calls is also associated with higher bridging social capital.

Prior research has highlighted that bridging and bonding social capital might be associated with very different aspects of an individual’s life [41] and this is also reflected in the associations observed here. For bonding social capital, we find a significant correlation between SMS messages received after 12 am and bonding scores ($r = 0.28$). A possible interpretation is that messages received late at night are likely to be from close family or friends, therefore, a higher number of such messages would imply that the individual had access to a larger number of strong ties in their network.

We notice that the associations between overall social capital and phone features are interpretable based on those found for bridging and bonding social capital. The results underscore the value of initiating calls, talking to multiple individuals in such calls, and access to contacts with whom one can interact even at less common hours (i.e. during AM phase).

5 DISCUSSION

The first research question (RQ1) for this work aimed at understanding the associations (if any) between long-term phone-use patterns and an individual's social capital. Based on the reasonably high explanatory power at a collective level (regression analysis) and multiple significant associations at an individual level (correlation analysis), we report that many of the phone-based features may indeed be intricately associated with individual social capital.

The results of correlation analysis are largely in line with prior literature [41, 30]. At the same time, they expand the understanding of these ties and the associations to how they manifest themselves over smartphones. The smartphone is a convenient, highly accessible, and capable device that is well suited to share and disclose information. It is a two-way device, creating and consuming information, is highly personal, and is almost always available. Based on Beale's (2005) work, this study treats people connected via the smartphone as part of a social network and interactions performed within this network communicative of existing bridging and bonding ties [6].

This is in line with Altman's Information Exchange Theory [4], which suggests that individuals need to exchange information in order to build and maintain healthy ties. This has also been tested and found in social network studies in the past [16]. We have found evidence of a similar phenomenon occurring over phone based social networks [6]. The Social Exchange Theory [18] also stresses on the importance of rules or norms existing within a network that assist the evolution of relationships in a network. The theory discusses the importance of norms as the "guidelines of interactional processes" [18]. The relevance of this theory has been observed in user behavior online shopping networks [44], we find smartphone interactions are guided by a similar set of rules or norms.

Different networks (or different ties in the network) may have different social capital implications. Multiple associations point to the value of size of the network as well as initiating communication. We also find the delay in responding to calls to be associated with social capital. Reciprocity is an important aspect of maintaining network relationships and has been linked with an increase in social capital [41, 45]. By responding to missed calls immediately an individual can show willingness to be accessed by others within the network. Conversely, a delay in response may be demonstrative of a lack of interest in others in the network. The Social Exchange Theory presents reciprocity as a guideline for interactional behavior, [18] for example, a person who receives a favor from another in a group is likely to give a favor in return. Therefore, characteristics of relationships—such as reciprocity can be used as a predictor of social capital.

A feature of potential interest is the ratio of phone usage during the AM phase (midnight – 11:59 am) to the PM phase. In the considered dataset, the AM phase corresponds to the less socially active time period, and a higher score on this feature corresponds to more than usual activity during this period. Mobile phones allow an individual the possibility of being constantly connected to one's family and friends [11], however, etiquette dictates that most communications are made in more common or "regular" hours. Therefore, the analysis of phone-use patterns at irregular times could provide useful insights into an individual's social network. In past studies, late night communication has been connected with stress and potentially the need for social support [1]. This suggests a future research direction where the temporality of interactions between individuals should be studied for its associations with strength of ties, as well as diversity in network structure.

The second research question for this work (RQ2) focused on the feasibility of a data analytics algorithm to automatically infer individual social capital based on phone metadata. Based on the results of the Lasso regression models, we report that phone-based features can perform reasonably well at estimating the social capitals as obtained via survey methods. The correlation between the predicted and the actual overall social capital scores was 0.7 and this went up to 0.79 in case demography data were also available. As many of the demography variables are often available to phone service and app providers, can be inferred using phone metadata [19], or require one-time input, they can often be used in conjunction with phone based features. The resulting model can explain 61% of the variance observed in social capital scores and with a MAE of ± 4.39 indicates that the predictions are within ± 4.39 of the "ground truth" or survey based values of social capital scores

This work is intended to start a conversation on the topic of automated inference of social capital, rather than be a final word on it. It considers a relatively high number of features (26) given the modest sample size (55). This limits the statistical power available for some of the analysis. For instance, the data is not suitable for stricter Bonferroni correction with correlation analysis. Hence, we will be cautious in generalizing the results to larger populations until they are verified at scale. We try to ameliorate these concerns by using Lasso regression that is specifically designed to avoid overfitting in such scenarios and consider the correlation analysis only as a post-hoc interpretation mechanism.

We note the moral and ethical considerations in building a social capital score based solely on passive data collection [49]. However, this score is intended to help individuals gain an awareness of their social network and the resources they potentially have access to. We also recognize the privacy implications of using an individual's smartphone metadata, however, we suggest using explicit opt-in measures that allow individuals to retain control of their information (e.g. via OpenPDS [14]). While a more nuanced policy debate is required, rather than abstaining from presenting such outcomes, we choose to raise awareness about these new prospects and informing the policy debate around them.

The phone based predictions obtained here are not an exact replication of the scores obtained by the survey. However, the associations found paint a vision forward and the explanatory power achieved is comparable to those found in similar social computing studies [16, 27]. Given the limitation of sample size, we believe the correlation scores (0.70-0.80) obtained here show a promising direction for the automated prediction of an individual's social capital.

Despite certain limitations, this paper makes an important contribution towards the literature on social computing. While previous efforts have studied self-reported interconnections between social behavior and social capital, there is no existing effort that has studied the potential of using automated phone-based data to infer and predict the social capital scores for individuals.

6 IMPLICATIONS AND FUTURE DIRECTIONS

Social capital has a significant influence on a wide-ranging set of socioeconomic phenomena [2, 40, 41]. Just like an ability to check one's credit score or fitness scores periodically allows one to plan aspects of one's life better, an ability to measure evolving social capital score could yield multiple benefits. The ability to track one's social capital via the phone, without having the need for complicated surveys or generator methods makes it a convenient way to get an ongoing assessment of how their daily interactions impact their social capital and potentially even suggest how it can be increased. Given the associations between social capital and multiple other facets of human life (e.g., mental health, well-being) an ability to keep track of one's score could help individuals plan their lifestyles and activities better. For instance, a user may choose to build up towards specific goals in terms of social capital in preparation for important life events (e.g., job search). Similarly, a user who has faced issues with social support or loneliness in the past, may keep an eye out for the trends in their social capital and take pre-emptive corrective actions, where appropriate. An automated social capital score can also act as a certification of an individual's social credentials. It reflects the individual's accessibility to resources through social networks and relationships. Thus, in certain conditions, an individual may choose to divulge this information to potential partners and collaborators.

Further, given the intricate associations between social capital and multiple societal outcomes (e.g., civic engagement, innovation, and economic mobility) an ability to continuously infer the social capital levels for billions of phone-users can have significant epistemological advantages as well as societal impact. It could allow for creating more nuanced understanding of the impact of social behaviors (e.g., rejecting phone calls) on social capital, expose phenomena that can simply not be studied in lab-settings (e.g., contagion in social capital across networks of millions of users). Similarly, the proposed approach could be used to replace the costly implementation of manual social capital surveys across countries, like the one initiated recently by the OECD [38], with phone based measurements, and eventually create real-time census of social capital levels in different cities and nations.

Further work in this direction includes a more nuanced policy conversation as well as validation at scale with more, diverse, participants and additional phone features (e.g. communication app usage and bluetooth

based face to face interactions). This research needs to be followed by studies that explore the relationships between smartphone metadata and social capital with nationally representative samples. A larger sample size would also support statistical analysis of a larger number of features some of which are motivated by the current study. For example, we would like to study the temporal activity variations at a finer resolution and understand its interactions with individual social capital.

With refinements, the proposed approach could allow billions of users to keep track of their evolving social capital scores and receive helpful suggestions. At a societal level, researchers could study social capital related phenomena at unprecedented scales and ultimately guide data-driven policy making.

ACKNOWLEDGMENTS

We would like to thank Cecilia Gal, Padampriya Subramnian, Ariana Blake, Suril Dalal, Sneha Dasari, and Christin Jose, for help with conducting the study and processing the data.

REFERENCES

- [1] Saeed Abdullah, Mark Matthews, Elizabeth L. Murnane, Geri Gay, and Tanzeem Choudhury. 2014. Towards circadian computing: early to bed and early to rise makes some of us unhealthy and sleep deprived. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing* 673-684.
- [2] Paul S. Adler, and Seok-Woo Kwon. 2002. Social capital: Prospects for a new concept. *Academy of Management Review*, 27(1), 17–40.
- [3] Nadav Aharoni, Wei Pan, Cory Ip, Inas Khayal, and Alex Pentland. 2011. Social fMRI: Investigating and shaping social mechanisms in the real world. *Pervasive and Mobile Computing*.
- [4] Irwin Altman, and William W. Haythorn. 1965. "Interpersonal exchange in isolation." *Sociometry* 411-426.
- [5] Michael A. Babyak. 2004. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic medicine*. 66(3), 411-421.
- [6] Russell Beale. 2005. Supporting social interaction with smart phones. *IEEE Pervasive computing* 4(2) 35-41.
- [7] Joshua Blumenstock, Gabriel Cadamuro, and Robert On. 2015. Predicting poverty and wealth from mobile phone metadata. *Science*, 350(6264), 1073-1076
- [8] Pierre Bourdieu. 1986. The forms of capital. In J. G. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241–258). New York: Greenwood
- [9] Moira Burke, Robert Kraut, and Cameron Marlow. 2011. Social capital on Facebook: Differentiating uses and users" In *Proceedings of the SIGCHI conference on human factors in computing systems*. 571-580.
- [10] Yi Cao, Defu Lian, Zhihai Rong, Jiayu Shi, Qing Wang, Yifan Wu, Huaxiu Yao, and Tao Zhou. 2017. Orderliness Predicts Academic Performance: Behavioral Analysis on Campus Lifestyle. arXiv preprint arXiv:1704.04103
- [11] Mary Chayko. 2016. *Superconnected: The Internet, Digital Media, and Techno-Social Life*. SAGE Publications
- [12] James S. Coleman. 1988. Social capital in the creation of human capital. *The American Journal of Sociology*, 94, 95–120
- [13] Yves-Alexandre de Montjoye, Jordi Quoidbach, Florent Robic, and Alex Sandy Pentland. 2013. Predicting personality using novel mobile phone-based metrics. In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*. 48-55.
- [14] Yves-Alexandre de Montjoye, Erez Shmueli, Samuel S. Wang, and Alex Sandy Pentland. 2014. openpds: Protecting the privacy of metadata through safeanswers. *PLoS one*, 9(7).
- [15] Nathan Eagle, and Alex Sandy Pentland. 2006. Reality mining: sensing complex social systems. *Personal and ubiquitous computing*. 10(4), 255-268.
- [16] Nicole B. Ellison, Charles Steinfield, and Cliff Lampe. 2007. The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*. 12(4), 1143-1168.
- [17] Nicole B. Ellison, Jessica Vitak, Rebecca Gray, & Cliff Lampe. 2014. Cultivating social resources: The relationship between bridging social capital and Facebook use among adults. *Journal of Computer-Mediated Communication*. 19(4), 855-870.
- [18] Richard M. Emerson. 1976. "Social exchange theory." *Annual review of sociology*. 335-362.
- [19] Bjarke Felbo, Pål Sundsøy, Alex Sandy Pentland, Sune Lehmann, and Yves-Alexandre de Montjoye. 2015 Using deep learning to predict demographics from mobile phone metadata. arXiv preprint arXiv:1511.06660
- [20] Claude S. Fischer. 2005. Bowling alone: What's the score? *Social Networks*. 27(2), 155–167
- [21] Henk Flap, Tom Snijders, Beate Völker, and Martin Van Der Gaag. 1999. Measurement instruments for social capital of individuals. *Questionnaire items*.
- [22] Jim Giles. 2012. Making the links. *Nature*, 488(7412).
- [23] Mark S. Granovetter. 1973. The strength of weak ties. *American Journal of Sociology*. 78(6), 1360-1380
- [24] Isha Ghosh, and Vivek K. Singh. 2016. Predicting Privacy Attitudes Using Phone Metadata. In *Proceedings of 2015 International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction (SBP15)*. 51-60.
- [25] Eric Gilbert, and Karrie Karahalios. 2009. Predicting tie strength with social media. In *Proceedings of the SIGCHI ACM Conference on Computer Supported Cooperative Work (CSCW)*. 211-220.
- [26] Trudy Harpham, Emma Grant, and Elizabeth Thomas. 2002. Measuring social capital within health surveys: key issues. *Health policy and planning*. 17(1) 106-111.
- [27] Chi Jumin, Jo Hyungeun, and Ryu Jung-hee. 2010. Predicting Interpersonal Relationships based on Mobile Communication Patterns. In *2010 ACM Conference on Computer Supported Cooperative Work (CSCW)*, 487-488.
- [28] Jaemin Jung, Sylvia Chan-Olmsted, and Youngju Kim. 2013. From access to utilization: Factors affecting smartphone application use and its impacts on social and human capital acquisition in South Korea. *Journalism & Mass Communication Quarterly*. 90(4), 715-735.
- [29] Lauri Kovanen, Jari Saramaki, and Kimmo Kaski. 2010. Reciprocity of mobile phone calls. arXiv preprint arXiv:1002.0763
- [30] Nan Lin. 1999. Building a network theory of social capital. *Connections*. 22(1), 28-51.
- [31] Nan Lin, Karen S. Cook, and Ronald S. Burt. (Eds.). 2001. *Social capital: Theory and research*. Transaction Publishers
- [32] Nan Lin, and Mary Dumin. 1986. Access to occupations through social ties. *Social networks*. 8(4), 365-385.

- [33] Nan Lin, Yang-chih Fu, and Ray-May Hsung. 2001. Measurement techniques for investigations of social capital. *Social capital: theory and research*. New York
- [34] Lynne McCallister, and Claude S. Fischer. 1978. A procedure for surveying personal networks. *Sociological Methods & Research*. 7(2), 131-148.
- [35] Lukas Meier, Sara Van De Geer, and Peter Bühlmann. 2008. The group lasso for logistic regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 70(1), 53-71.
- [36] Jun-Ki Min, Jason Wiese, Jason I. Hong, and John Zimmerman. 2013. Mining smartphone data to classify life-facets of social relationships. In *Proceedings of the 2013 ACM Conference on Computer Supported Cooperative Work (CSCW)* 285-294.
- [37] David C. Mohr, Jennifer Duffecy, Ling Jin, Evette J. Ludman, Adam Lewis, Mark Begale, and Martin McCarthy Jr. 2010. Multimodal e-mental health treatment for depression: a feasibility trial. *Journal of medical Internet research*. 12(5), e48
- [38] OECD (2013). The OECD measurement of social capital project and question databank [Data file]. Retrieved from <http://www.oecd.org/std/social-capital-project-and-question-databank.htm>
- [39] Kyung-Gook Park, Sehee Han, and Lynda Lee Kaid. 2013. Does social networking service usage mediate the association between smartphone usage and social capital? *New Media & Society*. 15(7), 1077-1093
- [40] Robert D. Putnam. 1995. Bowling alone: America's declining social capital. *Journal of Democracy*. 6(1), 65-78.
- [41] Robert D. Putnam. 2000. *Bowling Alone*. New York: Simon & Schuster
- [42] Sohrab Saeb, Mi Zhang, Christopher J. Karr, Stephen M. Schueller, Marya E. Corden., Konrad P. Kording, and David C. Mohr. 2015. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. *Journal of medical Internet research*. 17(7), e175
- [43] Akane Sano, Andrew J. Phillips, Z. Yu Amy, Andrew W. McHill, Sara Taylor, Natasha Jaques, Charles A. Czeisler, Elizabeth B. Klerman, and Rosalind W. Picard. 2015. Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones. In *Wearable and Implantable Body Sensor Networks (BSN), 2015 IEEE 12th International Conference*. 1-6.
- [44] Wen-Lung Shiau, and Margaret Meiling Luo. 2012. Factors affecting online group buying intention and satisfaction: A social exchange theory perspective. *Computers in Human Behavior* 28(6), 2431-2444.
- [45] Vivek K. Singh, & Rishav R. Agarwal. 2012. Cooperative phoneotypes: exploring phone-based behavioral markers of cooperation. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 646-657.
- [46] Vivek K. Singh, Burcin Bozkaya, and Alex Sandy Pentland. 2015. Money walks: implicit mobility behavior and financial well-being. *PLoS one*. 10(8), e0136628.
- [47] Vivek K. Singh, Laura Freeman, Bruno Lepri, and Alex Sandy Pentland. 2013. Predicting spending behavior using socio-mobile features. In *2013 International Conference on Social Computing*. 174-179.
- [48] Robert Tibshirani. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 267-288.
- [49] Zeynep Tufekci. 2015. Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. *Journal on Telecomm. & High Tech. L.*, 13, 203.
- [50] Martin Van Der Gaag and Tom AB Snijders. 2005. The Resource Generator: social capital quantification with concrete items. *Social networks*. 27(1), 1-29.
- [51] William W.S. Wei. 1994. *Time series analysis*. Reading: Addison-Wesley publishers
- [52] Barry Wellman., and Scot Wortley. 1990. Different strokes from different folks: Community ties and social support. *American journal of Sociology*. 96(3), 558-588.
- [53] Jason Wiese, Jun-Ki Min, Jason I. Hong, and John Zimmerman. 2015. You never call, you never write: Call and SMS logs do not always indicate tie strength. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, 765-774.
- [54] Dimitri Williams. 2006. On and off the 'net: scales for social capital in an online era. *Journal of Computer Mediated Communication*, 11(2), 593-628
- [55] Connie Yuan, and Geri Gay. 2006. Homophily of network ties and bonding and bridging social capital in computer-mediated distributed teams. *Journal of Computer-Mediated Communication*. 11(4), 1062-1084.
- [56] Huiqi Zhang, and Ram Dantu. 2010. Predicting social ties in mobile phone networks. In *Intelligence and Security Informatics (ISI), 2010 IEEE International Conference*. 25-30.
- [57] Erheng Zhong, Ben Tan, Kaixiang Mo, and Qiang Yang. 2013. User demographics prediction based on mobile data. *Pervasive and mobile computing*. 9(6), 823-837.

Received April 2017; revised July 2017; accepted November 2017.