Dwarakanath, R.; Charrier, J.; Englert, F.; Hans, R.; Stingl, D.; Steinmetz, R.: Analyzing the Influence of Instant Messaging on User Relationship Estimation. In IEEE Mobile Services, 2016, ISBN: 978-1-5090-2625-8

Analyzing the Influence of Instant Messaging on User Relationship Estimation

Rahul Dwarakanath, Jérôme Charrier, Frank Englert, Ronny Hans, Dominik Stingl, and Ralf Steinmetz
Multimedia Communications Lab
Technische Universität Darmstadt
Darmstadt, Germany
Email: dwarakan@kom.tu-darmstadt.de

***Abstract*—Recent years have seen a plethora of new mobile social networking services, given the widespread and ubiquitous availability of smartphones. However, from a user's perspective, two fundamental problems underlie these services: Undesired interruptions and privacy violations. Understanding user relationships can help smartphones to provide appropriate decision support for improved notification management and content sharing, thus mitigating these negative effects. In this work, we investigate the influence of mobile instant messaging (IM) services in estimating the type and strength of user relationships. To this end, we implemented an Android-based application to gather users' historical communication data and ran a study to collect manual assessments for each smartphone contact. Our user study shows that friends and hobby-related contacts tend to communicate more using IM services, whereas family and work-related contacts tend to use calls. Furthermore, our machine learning models estimate the *social circles* with an average accuracy of 77%, and distinguish between strong and weak relationships with an average accuracy of 76%, therein.**

***Index Terms*—User Relationships, Supervised Machine Learning, Modelling User Behaviour, Privacy**

I. Introduction

With the advent of modern mobile and ubiquitous technologies and services, the availability of the mobile users has increased considerably, facilitating enhanced productivity and improved coordination. However, two main problems threaten to underlie these services: Increased interruptability and privacy violation. Humans are known to be vulnerable to interruptions. Untimely calls, messages, as well as other notifications are some of the examples of interruptions that demand a smartphone user's attention. Undesired interruptions can result in increased mood changes, irritability, and unnecessary stress, as cited by many works so far [1]–[4]. Furthermore, sharing personal information is vital for the proper functionality of group- or community-based context-aware applications as well as online social networks (OSNs) [5], [6]. Consequently, user privacy has become a growing concern, given the perennial escalation of such services in recent times. It is paramount for one such service to facilitate a feasible and prudent sharing platform for its users.

As established in the field of human-centered computing, a user's interaction and sharing patterns with another user strongly depend on the relationship between the two [7]–[9]. Consider this application scenario to understand the core issue at hand. Bob uses an advanced context-aware communication application, ContextComm, which regulates the available communication channels based on his context. Among other services, ContextComm assists Bob while he is busy by regulating incoming calls and messages in an appropriate manner, thereby avoided undesirable interruptions. Let us assume that Bob is currently busy at the cinemas, watching a movie on a Saturday night with friends, Mark and Peter. At the same time, his co-worker Alice calls him to discuss something related to a common project. ContextComm recognizes that Alice is a co-worker and hence, decides to block the call and notify Alice that Bob is busy and will be available on Monday. A few minutes later, a close friend of Bob, Carl, calls. This time, ContextComm reacts differently – still blocks the call but informs Carl about Bob's location and status, and suggests the caller to send a text message instead. One key aspect necessary to determine the action taken by ContextComm and the level of privacy for the automatic replies – here, 'busy' or 'at the cinemas with Mark and Peter' – is the knowledge of Bob's relationship with the contacting person.

It has been proven by many studies so far that the communication history between two users has a direct correlation to the type and strength (closeness) of their relationship [10]–[14]. Our goal is to develop a smartphone service that can automatically estimate relationships based on the communication history extracted from the user's smartphone. Given the recent boom in the number of mobile instant messaging (IM) services offered, such as WhatsApp, Skype, Facebook Messenger, Threema, etc., this paper aims to understand the influence of the IM messaging channel in the estimation of user relationships.

Our main contributions are three-fold. Firstly, we developed an Android-based application to log necessary data from instant messaging services on smartphones (alongside call and SMS logs) to estimate the type and strength of user relationships. Secondly, we deployed supervised machine learning algorithms to analyze our models based on a dataset of 7191 communication events, including 6265 IM messages. In turn, we obtain 77% accuracy (κ = 0.33) in estimating the social circle, and 76% accuracy (κ = 0.42) in distinguishing the strong and weak relationships, therein. And thirdly, we individually analyze the communication channels, IM and calls, to understand their respective influence on user relationships. We observe that friends and hobby-related contacts tend to

2329-6453/16 $31.00 © 2016 IEEE
DOI 10.1109/MS.2016.17

communicate more using IM services, whereas family and work-related contacts tend to use calls.

In the following sections, after presenting some essentials in human relationships, we describe the main concept behind our approach. Based on this, we delve into the key features of our approach and the implementation of our Android-based application. We then discuss the evaluation results and our main findings, followed by some related literature, before concluding the paper with an outlook towards future work.

II. Estimating User Relationships

Humans exhibit different types of behaviour within the different relationships in their lives [15]–[17]. A person behaves differently with his friends, family members, co-workers, and other people involved in his daily life. Each *social circle* plays a key role in defining the nature of interaction between two people. People tend to interact with co-workers during the day on weekdays, with friends in the evenings, and with family members on the weekends, to give a few examples of varying interaction patterns. There can also be relationships that overlap multiple social circles, e.g. co-workers being good friends. Within each social circle, there exist close and distant relationships, corresponding to the strength of the relationship or *tie strength*. The tie strength is considerably indicative of the amount of information one shares with another. For example, in the scenario above, Bob may not share the same amount of information with distant friends or family members with whom he does not interact that often. A relationship can hence be perceived, in its most basic form, as a two-dimensional construct where each point describes the tie strength within a particular social circle.

Understanding the different social circles (also called *life facets* in related literature) is paramount to analyzing human relationships. In general, the social circles of work, family, and social contacts (typically, friends) comprise the most broadly accepted relationship constructs [12], [18]. The extent to which people assume different roles depends strongly on the individual, and is significantly reflected in their communication behaviour [19]. In this paper, we considered the following social circles – friend, significant other, work, family, hobby, and others, adapted from existing work in privacy-preserving sharing mechanisms [7], [9].

The term 'tie strength' can be defined as a combination of four main dimensions, comprising the amount of time spent together, the emotional intensity, the intimacy, and the reciprocal services governing the relationship [16]. Granovetter divided social relationships into three categories based on tie strength: strong, weak, and absent. Strong ties correspond to highly trusted and/or closely related contacts, whereas weak ties mostly correspond to acquaintances. Absent ties not only include unknown people, but also so-called *nodding relationships*, e.g. with one's neighbour, having minimal interactions.

A direct measurement of such theoretical dimensions is generally not practicable [14]. Instead, so-called *indicators* are used, which act as a proxy for measuring the strength of each dimension. For example, the duration of communication, frequency of interaction, diversity of the topics discussed, number of common friends or activities, and number of communication initiatives, to name a few, are some of the indicators for the strength of different dimensions of tie strength [16]. Similarly, alongside intensity and frequency of communication, temporal and locative dependencies have a particularly increasing say towards the social circle to which someone belongs.

Based on sociological references, we define the following sets of indicators to estimate user relationships: Intensity and frequency of communication (ι), temporal dependency (τ), locative dependency (λ), communication channel usage (υ), as well as relationship maintenance (ρ). A two-dimensional relationship ($\mathcal{R}$), comprising of the social circle ($\mathcal{C}$) and the tie strength ($\mathcal{S}$), can be described as a (typically, linear) function of the indicators:

$$f_{\mathcal{R}} : (\mu, \iota, \tau, \lambda, \upsilon, \rho) \rightarrow \{\mathcal{C}, \mathcal{S}\} \tag{1}$$

where, μ is a set of weighting/correlation factors for each of the indicators. In order to obtain a cogent estimate of the social circle of a given relationship as well as its tie strength, it is crucial to weigh in the different indicators appropriately. In this paper, we tackled this issue by employing the concept of supervised machine learning using labelled data, such that the extracted indicators are combined in accordance to their relevance. The following sections provide a deeper look into the devised approach and the results obtained.

III. Data Extraction and User Study

Primarily, our work in this paper deals with the understanding of how inter-user communication using IM services helps to infer the relationship between a smartphone user and his contacts. To this end and to determine the necessary weighting factors in 1, we developed an Android-based application to extract the necessary indicators from the IM-based communication data on user smartphones, as well as run a user study to obtain manual assessments of the smartphone contacts by the respective users. Given their widespread usage in related literature [11], [12], we also extracted data from the call logs, so as to validate the general assumption in the research community that a higher frequency of communication over calls can imply a stronger tie strength.

WhatsApp has been the forerunning mobile IM service over the past few years. According to a recent study, the number of messages exchanged over WhatsApp increased to 64 billion until April 2014 [20], with an exponential increase in the number of active users over the past 2 years [21]. Hence, we incorporated WhatsApp as well as the secure messaging service, Threema, in our approach. Please note that this principle can also be extended to other IM services.

One of the problems we faced was during the implementation of IM trackers for retrieving their communication histories. Most IM applications do not offer any APIs to do the same. The only work-around we could manage was to directly access the status bar notifications. This allowed us to program our data extraction application in such a way that the details of

TABLE I
THE EXTRACTED COMMUNICATION FEATURES AND THEIR CORRESPONDING DIMENSIONS

Comm. Channel	**Feature Variables**	**Indicator Set (s. (1))**
Messages	Number and avg. length	*Communication intensity and frequency, ι*
	Avg. number of emoticons per IM/SMS, % IM/SMS with emoticons, % IM read within [3, 10] mins	*Channel usage, υ*
Calls	Number, duration, days with two (or more) calls, number and max. duration of incoming calls, std. dev. for outgoing call length, number of lengthy calls	*Communication intensity and frequency, ι*
	% Calls and % call duration on [Friday, Saturday, Sunday, weekdays, before noon], % long calls on weekdays	*Temporal dependency, τ*
	Number, duration, and % at [home, work]	*Locative dependency, λ*
	% Calls to all communication [overall, weekdays, Fridays, Sundays]	*Channel usage, υ*
Combination	Number and days of all comm.	*Communication intensity and frequency, ι*
	% Calls, SMS, and IM on Saturday	*Temporal dependency, τ*
	Number of channels used	*Channel usage, υ*
All of the above in the past [14 days, overall], days since last comm.		*Relationship maintenance, ρ*

each incoming message could be read and interpreted as a log entry. Although this does not provide a complete picture of IM-based interaction, we can still draw reasonable interpretations of the relationship.

In addition to these, we retrieved the phone location by using cell tower IDs. We decided against using GPS because most users tend to switch GPS sensing off, either out of privacy concerns or due to the higher battery drain [22]. For our purpose, we allowed the users to inform the application about logical locations (e.g. home, work, etc.), whereby the cell tower IDs for each logical location were saved for future inference. We appended the location information to each communication event in order to understand the locative dependency in a given relationship. Futhermore, the usage of emoticons in text messages as well as its impact on human emotions have had growing importance in the research community, recently [23]. We decided to analyze their impact in estimating user relationships by counting the number of emoticons used in each message.

A. Feature Extraction for Machine Learning

As mentioned earlier, we employed supervised machine learning algorithms on data extracted from a user study (details in Section III-B) and indicators – called features in the machine learning jargon – extracted from the communication logs on user smartphones. We defined a total of 75 features for the supervised machine learning algorithms: 16 from message logs (IM and SMS), 50 from call logs, and the remaining 9 as combinatorial features. The extracted features factor in the indicator sets presented in (1).

Message features: As mentioned above, we considered only incoming messages on the mobile IM applications, WhatsApp and Threema. To account for messaging intensity and frequency, we analyzed the number of messages exhanged and their average length. Temporal dependence in message-based user interaction has been proven to have a minimal influence on user relationships [17]. We decided to evaluate its usage by measuring the time taken by a user to respond to IM messages. Given the API-related problem mentioned earlier, we decided to measure the time taken for a user to react to the received messages (by opening the corresponding IM application) after turning the screen on. Thereby, we measured the percentage of IM messages "read" within 3 and 10 minutes. In addition, we also analyzed the number and proportion of emoticons used in IM messages. For the SMS messages, we restricted our analysis to the usage of emoticons, given the steady decrease in SMS popularity in recent years [11].

Call features: We primarily used established features from related work [12], [18] to analyze the communication based on the call logs. These basically included the number and duration of calls overall, as well as data based on temporal conditions. For example, we analyzed the call intensity during the week and on weekends, also distinguishing between Fridays, Saturdays, and Sundays, so as to cover varying behaviours in different social circles. Furthermore, we examined call activity at 'home' and at 'work', to analyze the locative dependence.

Combinatorial features: Given the inherent dependence among the various available communication channels, we also considered features based on a combination of these channels. For example, certain text messages may warrant an immediate call back to sender. We measured the number of communication activities on each channel, number of channels used, days since last communication over any channel, and also temporal dependence on channel usage.

We analyzed all of the above features for both short-term and long-term views, in order to account for relationship maintenance [12]. Table I shows the various features used in our work, along with their significance towards the dimensions of user relationships.

B. User Study

Our user study was performed over a time span of one month (July 2015). We obtained participants by advertising our application among fellow work colleagues as well as students at the university. The participants were asked to install our Android-based application and provide the application the rights to access their contact information, call and SMS logs, as well as the status bar notifications. There was no additional incentive provided for participation.

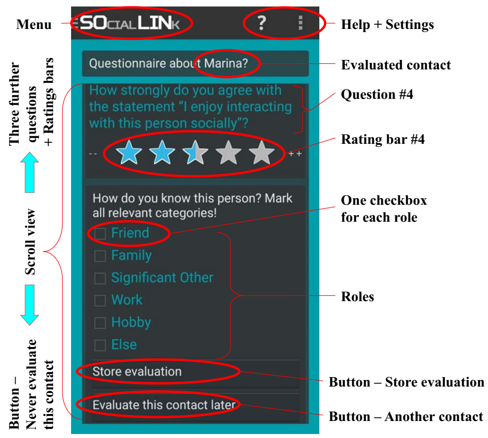

Fig. 1. Design and description of the GUI for contact assessment

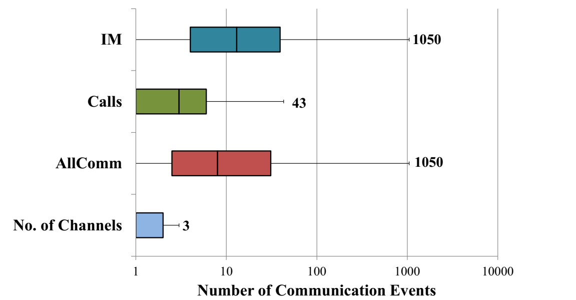

Fig. 2. Distribution of the extracted communication events per participant over the different channels

For the ground truth data, the participants were asked to assess a selection of their contact persons. For a reliable model, we chose a balanced set of strong and weak ties from the contact list. Given the relevance of call duration with the tie strength [11], we chose the initial selection of contacts based on the median of their aggregated call duration over the past 90 days from the call log. The other suggested contacts were chosen randomly from the contact list based on the user's *runtime* interaction on the observed communication channels.

The participants were asked to classify the selected contacts into pre-defined labels corresponding to the social circle(s) they fulfill (s. Sec. II). We allowed the participants to choose more than one label or even skip certain contacts. The latter was done to account for uncertain or irrelevant contacts like ex-partners, old classmates, or even entries for one's "home phone" or "parents", indicating one contact entry for a group of people.

Additionally, the participants were asked to answer a set of four questions on a 10-step Likert scale from 0 to 100, pertaining to their tie strength towards each of the selected contacts. These four questions were chosen based on previous work in the field of sociology [24], related to the matters such as the amount of social interaction, willingness to borrow money, and general closeness. The linear sum of the answers to each question corresponds to the tie strength between the participants and their respective contacts. Fig. 1 shows an excerpt from the application user interface used for the assessment of the selected contacts.

IV. Evaluation and Results

In total, 12 participants took part in our user study. However, two of the participants failed to use the application for more than 10 days. We decided to eliminate their instances due to insufficient data, lest there is a falsification of the machine learning classifiers. We discarded the instances of two more participants, who returned default values (50) for each of the questions in their contact assessments. We presume that these values do not indicate their true assessment. In the end, we obtained 249 instances from the datasets of the remaining 8 participants, each having assessed an average of 31 contacts (min = 4, max = 58, σ = 18.96).

Table II (List *AllComm*) describes the obtained dataset from our user study, indicating the distribution of the assessed contacts over the pre-defined social circles. We can clearly see that most of the assessed contacts were classified as friends and very few of the contacts were classified as the significant others. The number of work and family contacts is also paltry. In the ensuing discussion, we do not consider the analysis of the contacts classified as 'significant other' and 'others'.

In total, our extracted dataset contains 7191 communication events, including 680 calls, 246 SMS messages, and 6265 IM messages. Fig. 2 presents the average distribution of the extracted communication events per participant over the two main channels - calls and IM. To better understand the usage of the two communication channels in the different social circles, we divided the original dataset into two sets: instances with at least one call, and instances with at least one IM message, resp (s. Table III). The row values '% Rel. change' indicate the deviation in the number of instances in each social circle from a linear decrease (with respect to the overall set) in the presence of only calls and IM, resp. We observe that 'family' and 'work' instances show a clear tendency towards call-based interaction (+49.28% and +26.45%, resp.), whereas 'friend' and 'hobby' instances show a significant tendency towards IM-based interaction (+10.91% and +22.95%, resp.). Furthermore, we also observe a significant negative deviation in 'family' and 'work' instances towards IM-based interaction (-24.47% and -49.49%, resp.), and in 'hobby' instances towards call-based interaction (-59.63%), indicating the lack of preference towards these communication channels in the corresponding social circles.

For the rest of our analysis, we categorized the user-assessed contacts into different lists based on the communication channel used (we merged SMS and IM interactions into one). In Table II, lists *OnlyC* and *OnlyM* include data obtained exclusively from call logs and messaging services, respectively. In both cases, we excluded data obtained from the combinatorial features, so as to obtain a fair understanding of each channel's influence.

TABLE II
THE DIFFERENT DATASETS USED FOR ANALYSIS – DATASET *AllComm* INCLUDES ALL COMMUNICATION FEATURES; DATASET *OnlyC*: EXCLUDES MESSAGING AND COMBINATORIAL FEATURES; DATASET *OnlyM*: EXCLUDES CALL LOG AND COMBINATORIAL FEATURES.

Dataset Properties			**Social Circles**					
List	**#Features**	**All Instances**	**Friend**	**Work**	**Family**	**S.O.**	**Hobby**	**Others**
AllComm	75	249	120	17	36	2	71	34
OnlyC	50							
OnlyM	16							

TABLE III
CHANNEL USAGE BASED ON SOCIAL CIRCLE. THE RELATIVE CHANGE SHOWS HOW MUCH THE NUMBER OF INDIVIDUAL INSTANCES DEVIATES FROM A LINEAR DECREASE IN THE PRESENCE OF ONLY THE CORRESPONDING COMMUNICATION CHANNEL.

		All	**Friend**	**Work**	**Family**	**S.O.**	**Hobby**	**Others**
Original Set		249	120	17	36	2	71	38
Those using calls	⇒	139	69	12	30	1	16	17
	Linear decrease	139	66.99	9.49	20.10	1.12	39.63	21.21
	% Rel. change	–	+3.00	+26.45	+49.28	-10.43	-59.63	-19.86
Those using IM	⇒	174	93	6	19	2	61	23
	Linear decrease	174	83.86	11.88	25.16	1.40	49.61	26.55
	% Rel. change	–	+10.91	-49.49	-24.47	+43.10	+22.95	-13.38

A. Analysis of Social Circles

In order to achieve reliable models, for each classification test, we set aside 30% of the original (randomized) dataset as the test set, resulting in 70% for training. As can be observed in Table II (List *Allcomm*), there is a vivid imbalance in the number of instances for the different social circles, showing a clear proclivity for 'friend' instances. The social circle classifier would tend to classify a contact as 'friend', given the low number of representative instances for other social circles. Thus, we resampled each training set randomly before each test so as to balance out the classes of each social circle and have roughly equal number of representative instances in the training set. To this end, we used the resampling technique with replacement available in the WEKA toolkit [25], since each new instance is independent of the other entries.

We performed 10 such runs for each class (original dataset randomized before each run). In each run, we analyzed the performance of our model on the corresponding test set, and averaged the obtained classification results. Based on empirical observations on our datasets, we chose the Random Forest machine learning algorithm for our analysis. This can be accredited to our considerably diverse feature set, corresponding to the varied implications of the relationship indicators stated in (1). We analyzed our results based on the evaluation metrics – accuracy, Cohen's kappa (κ; the measure of agreement between the predicted model and the observed data, without considering the random agreements), and the F1 positive and negative scores (the ability of a prediction model to tell apart the same class from the non-class instances in a dataset, and vice versa). We also included an information gain analysis (based on Ranker [25]), which is indicative of the correlation factors assigned by the models to each feature as per (1).

Table IV presents the obtained results from the models created for the classification of each social circle. The considerably low average accuracy in the 'friend' circle across all datasets can be accredited to the fact that the perception of friendship is quite diversified across communities, hence depicting varied interaction patterns. However, the F1 positive score reveals a drop of 6.5% for the *OnlyC* dataset (from the *AllComm* dataset), while remaining roughly constant for the *OnlyM* dataset. This indicates that 'friend' instances exhibit a more distinct messaging pattern over the others, as seen in the information gain analysis (s. Table IV).

The models for the circles, 'family' and 'work', perform much better when the *OnlyC* dataset is used, producing high accuracy values of 82.97% and 90.95% (κ = 0.44 and 0.33), respectively. However, their performance goes down considerably when the *OnlyM* dataset is used, resulting in an average accuracy decrease of 17% and 21%, resp. Given that both the F1 positive and negative scores drop considerably in each case, we see that there exists a lack of distinct messaging patterns in both these social circles. This is also reflected in the feature information gain (s. Table IV), where the top three features for both circles are based on calls.

In case of the 'hobby' category, we obtain an average accuracy of 76.08% (κ = 0.44) using all communication features (*AllComm*). We observe that the average accuracy drops by 5.4% for *OnlyC*, whereby the F1 positive score remains relatively similar to that of *AllComm*. This indicates that 'hobby' contacts exhibit distinct calling patterns, as indicated by the lower call intensity and frequency in the information gain analysis. Furthermore, while we observe relatively similar average accuracy values for both *OnlyC* and *OnlyM*, the F1 positive score dips significantly for the *OnlyM* dataset – by 12.97% in comparison to *AllComm* – indicating the lack of distinct messaging patterns. This can probably be caused by similar messaging patterns between 'friend' and 'hobby' contacts.

TABLE IV

EVALUATION RESULTS FOR THE DIFFERENT SOCIAL CIRCLES. *Accuracy analysis*: AVERAGED OVER 10 RUNS. FOR EACH RUN, THE ORIGINAL DATASET WAS RANDOMIZED AND SPLIT INTO 70%-30% SETS FOR TRAINING AND TESTING, RESPECTIVELY. *Information gain analysis (based on Ranker [25])*: TOP 4 COMMUNICATION FEATURES FOR EACH SOCIAL CIRCLE. THE SIGN IN THE BRACKETS INDICATES CORRELATION VALENCY.

Social Circle	**Accuracy Analysis**				**Information Gain Analysis**	
	Metric	**AllComm**	**OnlyC**	**OnlyM**	**Communication Feature**	**Info Gain (Corr. Valency)**
Friend	*Avg. acc.*	61.22%	61.08%	62.84%	No. of IM messages	0.095 (+)
	κ	0.2307	0.2138	0.2489	No. of all comm. events	0.08 (+)
	F1 pos.	59.01%	52.53%	57.90%	No. of all comm. events [14 days]	0.077 (+)
	F1 neg.	62.85%	66.83%	66.40%	No. of IMs [14 days]	0.077 (+)
Family	*Avg. acc.*	81.49%	82.97%	65.68%	% Calls on Saturday	0.209 (+)
	κ	0.3780	0.4442	0.1216	% Call duration on Saturday	0.179 (+)
	F1 pos.	48.33%	54.26%	30.43%	Total duration of calls	0.172 (+)
	F1 neg.	88.69%	89.39%	77.03%	% Calls to overall comm.	0.152 (+)
Work	*Avg. acc.*	90.68%	90.95%	69.32%	% Call duration on weekdays	0.319 (+)
	κ	0.2806	0.3319	0.0632	% Calls to overall weekday comm.	0.311 (+)
	F1 pos.	32.67%	37.75%	15.08%	% Calls on weekdays	0.311 (+)
	F1 neg.	94.94%	95.11%	81.23%	No. of IM messages	0.3 (-)
Hobby	*Avg. acc.*	76.08%	70.68%	68.92%	% Calls to overall comm.	0.184 (-)
	κ	0.4413	0.4205	0.2674	No. of calls	0.184 (-)
	F1 pos.	61.45%	63.45%	48.48%	Total duration of calls	0.18 (-)
	F1 neg.	82.45%	75.17%	77.59%	% IMs read within 3 mins [14 days]	0.15 (+)

TABLE V

EVALUATION RESULTS FOR THE TWO TIE STRENGTH DATASETS. *Accuracy analysis*: AVERAGED OVER 10 RUNS. FOR EACH RUN, THE ORIGINAL DATASET WAS RANDOMIZED AND SPLIT INTO 70%-30% SETS FOR TRAINING AND TESTING, RESPECTIVELY. *Information gain analysis (based on Ranker [25])*: TOP 4 COMMUNICATION FEATURES FOR EACH DATASET. THE SIGN IN THE BRACKETS INDICATES CORRELATION VALENCY.

Dataset	**Accuracy Analysis**				**Information Gain Analysis**	
	Metric	**AllComm**	**OnlyC**	**OnlyM**	**Communication Feature**	**Info Gain (Corr. Valency)**
StrongWeak	*Avg. acc.*	75.81%	75.95%	74.59%	No. of days of comm.	0.146 (+)
	κ	0.4188	0.3715	0.2338	Days since last comm.	0.129 (-)
	F1 pos.	58.61%	53.05%	38.23%	Total duration of calls	0.117 (+)
	F1 neg.	82.85%	83.75%	83.89%	No. of calls	0.116 (+)
VerystrongWeak	*Avg. acc.*	80.13%	81.60%	71.47%	No. of channels used	0.189 (+)
	κ	0.3680	0.3338	0.2323	Avg. no. of emoticons per IM [14 days]	0.184 (+)
	F1 pos.	48.71%	44.15%	39.51%	No. of days of comm.	0.169 (+)
	F1 neg.	87.63%	88.89%	80.91%	% Calls, SMS, and IMs on Saturday	0.169 (+)

B. Analysis of Tie Strength

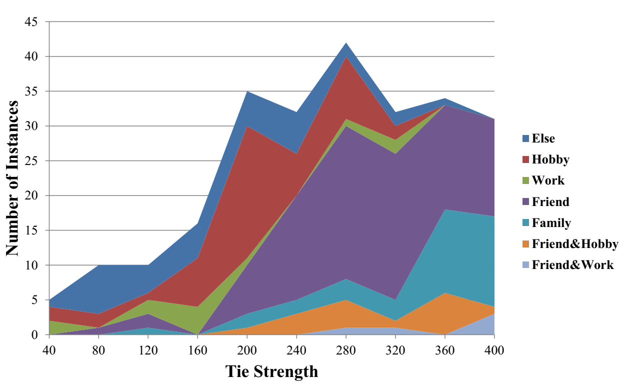

Fig. 3. Distribution of tie strength over different relationship types

Fig. 3 shows the sum of the responses to the tie strength questions, showing a general tendency towards a higher tie strength (mean = 250.6, σ = 97.9). This suggests that people generally tend to maintain phonebook entries of contacts who have an above-average tie strength. We also clearly see that the strong relationships are predominantly either among friends or family members, whereas purely work and hobby-related contacts generally have a weaker/moderate tie strength. However, we also observe that work and hobby-related contacts who are also friends tend to exhibit a high tie strength ($mean_{workfriend}$ = 360, $mean_{hobbyfriend}$ = 307.5).

Similar to existing work on tie strength analysis, we employed a nominal approach, where we grouped the assessed contacts into 'strong' and 'weak' ties. We used a threshold of 80% (320) of the summed tie strength to distinguish between the two, obtaining 72 strong ties and 177 weak ties (*StrongWeak*). We also grouped the assessed contacts into 'very strong' and 'weak' ties, with the threshold at 90% (360), as done by related work [11], [26]. In doing so, we obtained 38 very strong ties and 211 weak ties (*VerystrongWeak*).

Table V presents the evaluation results for the tie strength analysis based on the Random Forest algorithm. We observe that the performance of our models is best when we make use of all the communication features, as seen in the average F1 positive scores of 58.61% and 48.71% for the two datasets, resp., showing at least a 5% improvement over *OnlyC* and *OnlyM*. The information gain analysis reflects the same, where the main contributing features are combinatorial. Interestingly, the F1 negative score for all three feature sets is quite similar in case of *StrongWeak*, while the F1 positive score drops considerably for *OnlyM*. This can be accredited to the lack of representative instances for the 'strong' class with respect to messages. This also led to a lower average recall rate of 32.69%, as compared to 64.13% in the *AllComm* case (recall results not shown in the table).

Upon observing the information gain analysis in Table V, we see that the number of channels used per contact as well as the number of emoticons used in the IM messages have a considerable say in distinguishing very strong ties. For the *StrongWeak* dataset, the number of days of communication as well as the days since last communication tend to weigh in most in distinguishing strong ties from the weak ones.

V. Discussion

Upon observing our results, we see that the two communication channels – calls and messages – have a significant influence on the estimation of user relationships. In particular, we observe that the inclusion of IM-based communication data in the estimation of user relationships helps in distinguishing 'friend' instances from 'family' and 'work' social circles, given their distinct IM interaction patterns. Using our models, we obtained an average accuracy of 77.36% (κ = 0.33) using all communication features (*AllComm*) for the estimation of the social circles. The estimation of the 'strong' and 'weak' ties resulted in an average accuracy of 75.81% (κ = 0.42) using the *AllComm* dataset.

Overall, given the relatively low number of participants in our user study and the low number of instances in our original dataset, the results obtained are only representative of a small section of smartphones users in the world. The low κ and F1 positive values are indicative of the lack of sufficient characteristic instances. Also, while our user relationship estimation algorithm analyzes a sparse view of the relationships themselves, a more extensive user study will help in improving the developed models, significantly.

Given the collection and analysis of sensitive user data, two settings are relevant when discussing user privacy in our approach – firstly, data extraction during our field study, and secondly, the execution of our application in day-to-day usage. For the field study, a centralized data collection is required to obtain manually labelled contacts together with indicators extracted from the communication data. In order to preserve privacy, we used a pseudonymization scheme based on hashing, together with feature extraction directly on the phone. Furthermore, we informed the users about the amount and information content of the collected data, and provided means for automatically leaving the study at any point of time.

The general lack of locative dependence in the communication patterns can be attributed to privacy concerns, as shown in a study performed by Consolvo et al. [8], despite our measures to mitigate these concerns by using cell-tower information. Furthermore, given the impact of emoticons on user relationship estimation, a more thorough analysis of text messages will allow for better results.

In general, human relationships are complex and dynamic in nature. The concept of friendship is quite diverse, ranging from best friends from childhood to schoolmates to after-work colleagues to acquaintances at a party. Communication between two users can take place over different communication channels, as we proved in this paper. While an analysis of IM-based communication data had not been researched earlier in this regard, the complexity of relationship estimation increases when multiple communication channels are used, or when the interaction takes place in person than on the phone, as in the case with flat-sharing roommates or elderly relatives [11].

And finally, while the social circle and tie strength of a relationship do have a direct correlation with the interaction and sharing patterns within the relationship, there are additional factors that can affect the same, owing to the surrounding conditions, the context of the user, as well as the nature of the users themselves. For instance, people communicate differently in the presence of other people, as well as based on their current activity. The interaction patterns of extroverts and introverts differ considerably, as well [27]. People of different ages, occupations, and cultural backgrounds can exhibit different interaction patterns, too. Adapting such relationship estimation services to dynamic user behaviour is an open research question.

VI. Related Work

Most efforts in related literature have tried to estimate user relationships in terms of tie strength or social circle either based on their smartphone communication history or their interactions on online social networks.

Min et al. [12] created a model that assesses call and SMS logs on user smartphones to determine the social circle to which their contacts belong: family, work, or social. They employed supervised machine learning mechanisms to build a model that resulted in a maximum accuracy of around 87%. Reinhardt et al. [18] implemented a similar approach with call, SMS, MMS, and E-mail logs to obtain an accuracy of around 86%. Yu et al. [28] analyzed interpersonal relationships based on cellphone network data, obtaining a maximum accuracy of 73%. Their main focus was on using location information, such as co-location and temporal communication information, to classify users into different relationship groups. Backstrom and Kleinberg [13] analyzed relationship status information on Facebook and attempted to predict if a given relationship is a romantic partnership or not, achieving a performance accuracy of around 79% in distinguishing single and married users.

Coming to tie strength, Gilbert and Karahalios [10] as well as Spiliotopoulos et al. [26] built models based on Facebook data to determine if a 'friend' is a weak or a (very) strong tie, obtaining an accuracy of 85% and 66%, resp. Wiese et al. [29] tackled this issue by analyzing the communication history in call and SMS logs, achieving an accuracy of around 91%. By doing so, they could reconfirm the assumption that frequent and intense communication indicates strong ties. However, in their follow-up paper [11], they further analyzed their results to establish that "a lack of communication does not necessarily indicate a weak tie".

Our work distinguishes itself from the rest, in that we restricted ourselves exclusively to data obtained on a user smartphone, and particularly focused on extracting IM communication data in a privacy-preserving manner. We provide an individual analysis of the two main communication channels – calls and instant messaging – thus highlighting their impact on estimating the social circle and tie strength of user relationships. Furthermore, unlike some of the above-mentioned efforts, our evaluation setup is more robust: we (randomly) split the obtained dataset into training and test sets, and only resampled the training set, leaving the test set untouched. This ensures better quality of the generated models, avoiding the case of overfitting.

VII. Conclusion

The perennial progress in the fields of mobile communications and ubiquitous computing has inadvertently led to increased interruptions and privacy concerns for smartphone users. It is paramount for the modern applications to understand the different user relationships and adapt the services provided, accordingly. In this paper, we analyzed the influence of instant messaging interaction patterns in estimating the social circle and tie strength for a given relationship. Based on data collected from a user study with 12 participants and 249 assessed contacts, we observe that friends and hobby contacts predominantly communicate via IM, whereas family and work contacts prefer to call. We also observe that the number of communication channels used for a contact and the number of emoticons used in IM messages have a considerable say in distinguishing very strong ties. Overall, we obtained an average accuracy of 77% in estimating the social circle(s), and 76% in distinguishing the strong and weak ties therein. As part of future work, we plan to run a larger user study with varied demographic backgrounds to obtain more representative and extensive results. In addition, we also plan to extend our analysis of messaging content beyond emoticons, by undertaking a sentiment and image analysis in order to extract the mood behind the messages exchanged.

Acknowledgment

This work has been [co-]funded by the Social Link Project within the Loewe Program of Excellence in Research, Hessen, Germany.

References

[1] G. Mark, D. Gudith, and U. Klocke, "The Cost of Interrupted Work: More Speed and Stress," in *ACM HCI*, 2008, pp. 107–110.

[2] A. Mehrotra, M. Musolesi, R. Hendley, and V. Pejovic, "Designing Content-driven Intelligent Notification Mechanisms for Mobile Applications," in *ACM UbiComp*, 2015, pp. 813–824.

[3] B. P. Bailey and J. A. Konstan, "On the Need for Attention-aware Systems: Measuring Effects of Interruption on Task Performance, Error Rate, and Affective State," *Computers in Human Behavior*, vol. 22, no. 4, 2006.

[4] R. Dwarakanath, D. Stingl, and R. Steinmetz, "Improving Inter-user Communication: A Technical Survey on Context-aware Communication," *PIK-Praxis der Informationsverarbeitung und Kommunikation*, vol. 38, no. 1-2, 2015.

[5] P. Alves and P. Ferreira, "Radiator: Context Propagation Based on Delayed Aggregation," in *ACM CSCW*, 2013, pp. 249–260.

[6] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell, "A Survey of Mobile Phone Sensing," *Communications Magazine, IEEE*, vol. 48, no. 9, 2010.

[7] A. Khalil and K. Connelly, "Context-aware Telephony: Privacy Preferences and Sharing Patterns," in *ACM CSCW*, 2006, pp. 469–478.

[8] S. Consolvo, I. E. Smith, T. Matthews, A. LaMarca, J. Tabert, and P. Powledge, "Location Disclosure to Social Relations: Why, When, & What People Want to Share," in *ACM HCI*, 2005, pp. 81–90.

[9] J. S. Olson, J. Grudin, and E. Horvitz, "A Study of Preferences for Sharing and Privacy," in *ACM HCI*, 2005, pp. 1985–1988.

[10] E. Gilbert and K. Karahalios, "Predicting Tie Strength with Social Media," in *ACM CHI*, 2009, pp. 211–220.

[11] J. Wiese, J.-K. Min, J. I. Hong, and J. Zimmerman, "You Never Call, You Never Write: Call and SMS Logs Do Not Always Indicate Tie Strength," in *ACM CSCW*, 2015, pp. 765–774.

[12] J.-K. Min, J. Wiese, J. I. Hong, and J. Zimmerman, "Mining Smartphone Data to Classify Life-Facets of Social Relationships," in *ACM CSCW*, 2013, pp. 285–294.

[13] L. Backstrom and J. Kleinberg, "Romantic Partnerships and the Dispersion of Social Ties: A Network Analysis of Relationship Status on Facebook," in *ACM CSCW*, 2014, pp. 831–841.

[14] P. V. Marsden and K. E. Campbell, "Measuring Tie Strength," *Social Forces*, vol. 63, no. 2, 1984.

[15] S. G. Roberts and R. I. Dunbar, "Communication in Social Networks: Effects of Kinship, Network Size, and Emotional Closeness," *Personal Relationships*, vol. 18, no. 3, 2011.

[16] M. Granovetter, "The Strength of Weak Ties: A Network Theory Revisited," *Sociological Theory*, vol. 1, no. 1, 1983.

[17] S. Bell, A. McDiarmid, and J. Irvine, "Nodobo: Mobile Phone as a Software Sensor for Social Network Research," in *IEEE VTC*, 2011, pp. 1–5.

[18] D. Reinhardt, F. Engelmann, A. Moerov, and M. Hollick, "Show Me Your Phone, I Will Tell You Who Your Friends are: Analyzing Smartphone Data to Identify Social Relationships," in *ACM MUM*, 2015, pp. 75–83.

[19] S. D. Farnham and E. F. Churchill, "Faceted Identity, Faceted Lives: Social and Technical Issues with Being Yourself Online," in *ACM CSCW*, 2011, pp. 359–368.

[20] "Cumulative daily mobile message volume of whatsapp messenger as of april 2014 (in billions)," http://www.statista.com/statistics/258743/daily-mobile-message-volume-of-whatsapp-messenger/, accessed: 2016-03-31.

[21] "Number of monthly active whatsapp users worldwide from april 2013 to september 2015 (in millions)," http://www.statista.com/statistics/260819/number-of-monthly-active-whatsapp-users/, accessed: 2016-03-31.

[22] "Why Does GPS Use More Battery Than Any Other Antenna Or Sensor In A Smartphone?" http://www.forbes.com/sites/quora/2013/08/06/why-does-gps-use-more-battery-than-any-other-antenna-or-sensor-in-a-smartphone/, accessed: 2016-03-31.

[23] S. L. Rojas, U. Kirschenmann, and M. Wolpers, "We Have no Feelings, We Have Emoticons ;-)," in *IEEE ICALT*, 2012, pp. 642–646.

[24] A. Marin and K. N. Hampton, "Simplifying the Personal Network Name Generator Alternatives to Traditional Multiple and Single Name Generators," *Field Methods*, vol. 19, no. 2, 2007.

[25] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA Data Mining Software: An Update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, 2009.

[26] T. Spiliotopoulos, D. Pereira, and I. Oakley, "Predicting Tie Strength with the Facebook API," in *ACM PCI*, 2014, pp. 1–5.

[27] Y.-A. de Montjoye, J. Quoidbach, F. Robic, and A. S. Pentland, "Predicting Personality using Novel Mobile Phone-based Metrics," in *Social computing, Behavioral-cultural Modeling and Prediction*, 2013, pp. 48–55.

[28] M. Yu, W. Si, G. Song, Z. Li, and J. Yen, "Who Were You Talking to – Mining Interpersonal Relationships from Cellphone Network Data," in *IEEE/ACM ASONAM*, 2014, pp. 485–490.

[29] J. Wiese, J.-K. Min, J. I. Hong, and J. Zimmerman, "Assessing Call and SMS Logs as an Indication of Tie Strength," 2014, Technical Report.