Delphine Reinhardt, Franziska Engelmann, Andrey Moerov, Matthias Hollick: *Show Me Your Phone, I Will Tell You Who Your Friends Are: Analyzing Smartphone Data To Identify Social Relationships.* In: Proceedings of the 14th ACM International Conference on Mobile and Ubiquitous Multimedia (MUM), pp. 75-83, 2015.

Show Me Your Phone, I Will Tell You Who Your Friends Are: Analyzing Smartphone Data To Identify Social Relationships

Delphine Reinhardt
University of Bonn and
Fraunhofer FKIE
Bonn, Germany
delphine.reinhardt@cs.uni-bonn.de

Franziska Engelmann
TU Darmstadt
Darmstadt, Germany
fengelmann@seemoo.tu-darmstadt.de

Andrey Moerov
University of Bonn
Bonn, Germany
moerov@informatik.uni-bonn.de

Matthias Hollick
TU Darmstadt
Darmstadt, Germany
mhollick@seemoo.tu-darmstadt.de

ABSTRACT
Access control is a key principle to protect user privacy online. The combination of both the wealth of user-generated data in online social networks and overly complex user interfaces lead to a high user burden for privacy control, hence making the observance of the above principles difficult. We investigate how communication metadata on smartphones can facilitate providing tailored suggestions for restricted audience groups, thus limiting the sharing of data to the intended users only. To this end, we have performed a user study collecting a dataset including contact names, calls, SMS, MMS, and e-mail on personal smartphones in everyday use. In this paper, we examine which are the key features determining the social relationship category of a contact using machine learning. We obtain promising results for an automated classification of contacts into work-related, family-related and other social-interaction-related, thus enabling the possibility of user assistance for privacy control. Obtaining a more fine-grained categorization of the latter category into acquaintances, friends, and university-mates is shown to be difficult, since these categories blur in our study group.

CCS Concepts
•Human-centered computing → Ubiquitous and mobile devices; Empirical studies in ubiquitous and mobile computing; *Field studies;*

Author Keywords
Interpersonal relationships mining

INTRODUCTION
Facebook counts an increasing number of mobile users [19, 7], who create and upload a tremendous amount of user-generated content. In the current state-of-the-art, users can control who is authorized to access the uploaded content by either manually defining a group of individuals or using predefined groups. Predefined groups include categories, such as "everyone", "friends", and "friends of friends".

However, the current access control proposed to the users is inappropriate, as demonstrated in [12, 9, 23, 13, 24]. In particular, the existing solution remains complex and burdensome [12]. As a result, users may either misapply it or even not apply it at all, so that content may be shared with an unintended audience [9, 23]. This potentially puts the users' privacy at risk and might have negative consequences, such as dismissals [2] or loss of health-insurance benefits [1] in the most severe cases. A further drawback of the current method is that the provided access rules do not reflect the dynamics of relationships between users [13]. In addition to create groups, users hence need to manually maintain them, increasing the related overhead. As an expected result, users rarely do it [24].

To assist users in identifying appropriate contacts to share content with, an approach is to leverage mobile communication data already available on the users' phone to classify their contacts according to different categories as detailed in [21]. For example, call and SMS metadata are considered in [15] to categorize users' contacts into family, coworkers, and social contacts (e.g., schoolmates or neighbors). Within the scope of this paper, we follow the same direction and make the following contributions:

1. We have implemented an Android mobile app in order to log contact names, calls, SMS, MMS, and e-mails on the users' mobile phone. We collected 56,680 communication items related to the 1,895 contacts of our 19 participants. To allow further processing, we asked the participants to classify each contact in one of the following categories: acquaintances, coworkers, family, friends, and uni/schoolmates.
2. We analyze the collected dataset according to different dimensions, such as distribution between different communication channels and temporal distribution depending on the considered relationship categories. Next, we examine the information gain obtained by the extracted features and observed that call-based features provide the key information about the relationship category to which contacts belong.
3. We finally examine the performance of well-established classifiers according to different settings. We first compare our results to [15] by adopting their classification and hence, merging acquaintances, friends, and uni/schoolmates into a unique category. Our results show that the classifiers manage to classify the given contacts with an accuracy of up to 86.8% in the best case. Moreover, we observe that e-mail-based features improve the performance of all considered classifiers when considering five social categories.

The rest of this paper is structured as follows. We first present the dataset collection, before detailing the different classification steps. We then discuss the results and related work, before making concluding remarks.

DATASET COLLECTION

After review and approval by our ethics committee, 19 participants installed our app for three to four weeks, starting from October, the 28th, 2014. No incentives were provided to the participants. In what follows, we detail the participants' demographics as well as the data collection settings.

Demographics

14 of our 19 participants are male. Their age ranges between 23 and 29 ($m = 25, SD = 1.5$). Most of them are students (74%) or employees (16%). One participant is unemployed and another is following dual studies. 74% are studying or working in the fields of natural sciences, computer sciences, or engineering. Other working areas are production (5%), humanities (5%), and social and healthcare (5%). All participants own their Android phone for at least one year and indicated to use it around 2.89 hours daily. On a scale from 1 (beginner) to 5 (expert), the participants indicated to have an experience level of 4.21 ($SD = 0.79$).

Settings

Before installing the app on their phones, we distributed an agreement form to the participants in order to inform them about both data collection and processing modalities. Note that participants were able to opt out at anytime. Once the participants agreed to conduct the study, we assisted them in installing our Android app on their own phones. Participants were first asked to create an account including user name and password. They then entered their login data for each e-mail provider they agreed to include in the collection process. By default, the data summarized in Tab. 1 are automatically collected in the background and periodically uploaded to a server maintained by our university when Wi-Fi was available. The participants could however decide to stop (and later resume) the data collection or change the uploading settings. In order to cater for transparency, the participants were able to access statistics about the collected data, such as the number of logged contacts, SMS, or e-mails. Communication and storage were secured to protect the participants' personal information and respect both their privacy and the privacy of their contacts.

Table 1: Summary of the collected data. Receiver type includes to, cc, bcc)

	Caller/ sender	Callee/ receiver	Receiver type	Duration/ length	Type	Date
Calls	x	x		x		x
SMS	x	x		x		x
MMS	x	x		x	x	x
E-mails	x	x	x	x	x	x

Table 2: Distribution of the collected data by communication channels and relationship categories

	Calls	SMS	MMS	E-mails	$\sum$
ACQ	424	2,356	0	3,494	6,274
WORK	277	324	0	6,906	7,507
FAM	2,436	2,234	58	2,003	6,731
FRIENDS	49,04	15,177	42	12,446	32,569
UNI	166	2,475	2	956	3,599
$\sum$	8,207	22,566	102	25,805	56,680

Note that additional data could be collected using an implementation tailored to rooted devices. However, our goal was to provide an app that can be used by a large user base without any restrictions. Moreover, we originally aimed at including Facebook messages in our collection process, as they often represent a substantial share in the participants' communication. To be able to do it, our participants would have needed to create a personal developer account in Facebook—the only solution to get around the denied permissions imposed by Facebook to applications not building on a Facebook-branded client [8]. While we have implemented such a solution, the user efforts would go beyond the limits of what is acceptable.

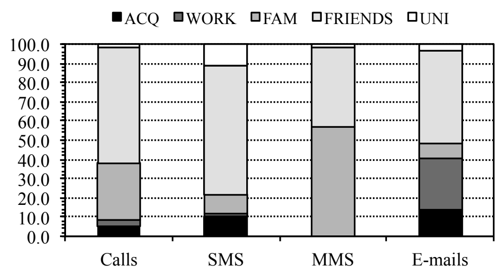
After three to four collection weeks, the participants were invited to sort the logged contacts by both eliminating irrelevant contacts and merging duplicates, i.e., phone numbers or email addresses belonging to the same user. They finally classified them into only one of the proposed relationship categories: *acquaintances* (ACQ), *coworkers* (WORK), *family* (FAM), *friends* (FRIENDS), and *school/university* (UNI).

CLASSIFICATION

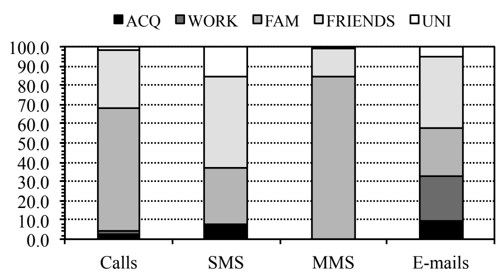
In this section, we first present the dataset contributed by the participants of our study. We next detail the feature extraction, before commenting on the classifier design.

Table 3: 169 features extracted from the dataset (#: number, DUR: duration, AVG: average, STD: standard deviation, lengthly-calls: duration greater than 2 times the average duration)

Logs	Factors	Variables
Calls	Intensity	Total {#, DUR}, total #lengthy-calls
	Regularity	{AVG, STD, MIN, MAX} # calls per week {last month, last half year, whole interval}, # days called/days logged, {AVER, MAX} DUR {all, outgoing, incoming}
	Temporal tendency	# and DUR weekend, weekday/total # or total DUR, {#, DUR} for each of the week/total # or DUR, {# and DUR} {early morning, morning, afternoon, evening, early night, late night}/total # or DUR
	Maintenance cost	{#, DUR} calls for past {2 weeks, 3 months} / total calls
SMS, MMS, E-mails	Intensity	Total {length, #} of messages
	Regularity	{AVG, STD, MIN, MAX} # {last month, last half year, whole period}, # days communicated/days logged, {AVG, MAX} length {all, outgoing, incoming}
	Temporal tendency	# and length weekend, workday/total # or total length, {#, length} for each of the week/total #, length, {# and length} {early morning, morning, afternoon, evening, early night, late night}/total #
	Maintenance cost	{#, length} for past {2 weeks, 3 months}/total {#, length}



(a) Original distribution



(b) Normalized distribution

Figure 1: Distribution of the collected data by communication channels and relationship categories (in percent)

Dataset

Tab. 2 and Fig. 1(a) show the original distribution of the collected data depending on the communication channels and the considered relationship categories. In total, 1,895 phone contacts, 8,207 calls, 25,805 e-mails, 22,566 SMS, and 102 MMS messages were collected. As expected, the overall number of contacts for each relationship category is different: 30% are classified as acquaintances, 27% as friends, 23% as coworkers, 14% as uni/school, and 6% as family members. As a result, we normalize the dataset and show the resulting distribution in Fig. 1(b). In more details, the participants had at least 181 calls and exchanged at least 56 SMS, 1 MMS, and 92 e-mails (see Tab. 4). At the exception of MMS, our participants communicated more with their friends than the remaining relationship categories. In contrast, more MMS are exchanged with family members. Since only 102 MMS have been logged in total, this trend may however not be as representative as compared to the other communication channels.

Table 4: Minimum, quartiles, and maximum for each communication channel and per participant

	Calls	SMS	MMS	E-mails
Min	181	56	1	92
Q_1	387	227	2	363
Q_2	474	486	3	779
Q_3	577	1,838	7	1,541
Max	747	5,635	54	3,809

Table 5: Mapping time periods and corresponding hours

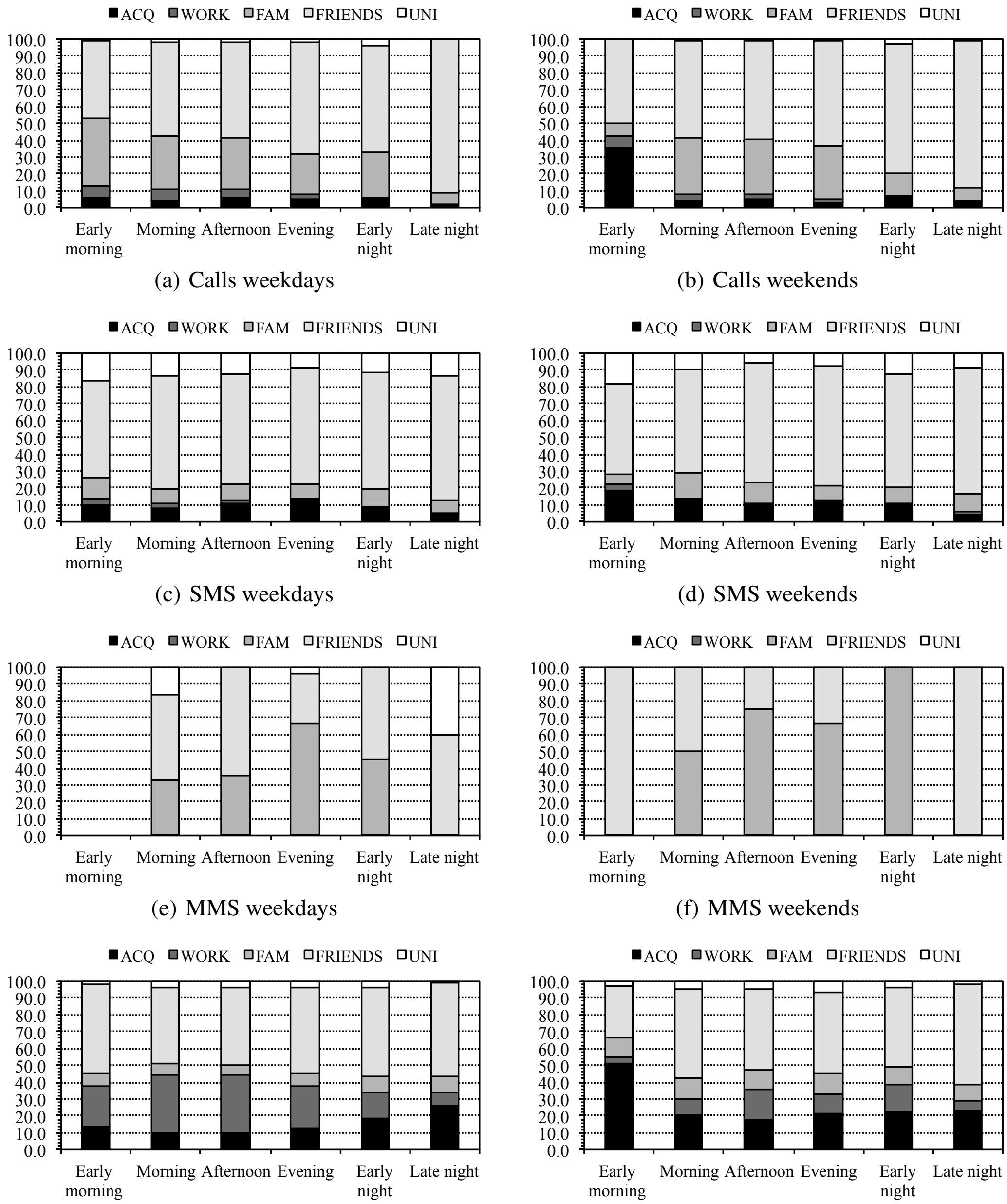
Time periods	Corresponding hours
Early morning	5.00 am - 8.59 am
Morning	9.00 am - 12.59 am
Afternoon	1.00 pm - 4.59 pm
Evening	5.00 pm - 8.59 pm
Early night	9.00 pm - 12.59 pm
Late night	1.00 am - 4.59 am

Moreover, most interactions are based on a unique channel (73% of the contacts). Only 1% of the contacts communicate using the four considered channels.

We next examine the influence of the temporal factor on the communication patterns. To this end, we use the time periods adopted in [15] and compiled in Tab. 5. Fig. 2 presents the corresponding results. During the weekends, we observe that the distribution between the relationship categories is different in the early morning as compared to the other time periods. This difference is mainly due to the low number of exchanges logged during this time slot. At the exception of e-mails, all communication channels show a similar distribution between weekdays and weekends. As expected, more e-mails are exchanged with coworkers during weekdays. As mentioned earlier, the temporal distribution of the MMS does not show the same diversity as the other channels due to the low number of exchanged MMS.

Feature Extraction

After having analyzed the dataset, we next aim at selecting relevant machine learning features. To this end, we first consider the following factors initially introduced in [15]: (1) intensity, (2) regularity, (3) temporal tendency, and (4) maintenance costs, as indicators for the strength of the social relationships. For example, the call duration and frequency can



(a) Calls weekdays (b) Calls weekends (c) SMS weekdays (d) SMS weekends (e) MMS weekdays (f) MMS weekends (g) E-mails weekdays (h) E-mails weekends

Figure 2: Temporal distribution of the participants' communication for each channel and relationship category

provide insights about the intensity and regularity of the interactions between contacts, respectively. Similarly, the times and days at which contacts interact (i.e., the temporal tendency) might leak information about the nature of their relationships. Finally, we also look at efforts, such as time elapsed between two interactions, provided by the participants to stay in contact with others as a measure of their relationship strength. Inspired from [15], we hence choose the features compiled in Tab. 3, which includes 26 features for calls, 45 for MMS and 49 for both SMS and e-mails based on the aforementioned dataset.

We then analyze the extracted features by firstly measuring their respective information gain for the five considered categories. The greater the information gain, the more useful a feature is to classify the contacts into the different relationship categories. The results presented in Tab. 8 show that eight out of the nine top features are based on calls that thus represent a key classification information. Secondly, we apply the correlation-based feature subset selection approach [11] and obtain a list of 29 recommended features. Among them, a majority are based on calls and e-mails (12 and 13, respectively). No features based on MMS are included in the final selection.

Classifier Design

Our dataset is unbalanced, i.e., the number of contacts is different in each relationship category. In our case, the ratio between cardinalities of the major class (ACQ) and the minor one (FAM) is 4.8. Since most classifiers are designed for balanced datasets, we apply three techniques to either under-sample or over-sample the original dataset *DATA*. The generated *UNDER* dataset corresponds to under-sampling the major class of *DATA* by randomly resampling the instances without replacement. Note that we resample the instances three times with different random seeds. The *SMOTE* dataset is constructed by using the synthetic minority oversampling approach introduced in [3]. For its construction, we choose five neighboring instances to create a new sample. Finally, *SMUNDER* is the combination of the two aforementioned approaches, i.e., synthetic instances are first generated, before conducting three runs to resample them without replacement.

We further select three classification algorithms, namely SVM, decision tree C4.5, and naïve Bayes based on [15]. We finally run a 10-fold cross validation for each pair of dataset and classification model using WEKA [10]. Note that we choose RBF kernel for the SVM model and perform a grid search on C and γ to optimize their values.

Evaluation Results

We conduct our evaluation in two steps. We first compare our results with those obtained in [15]. To this end, we hence merge acquaintances, friends, and uni/schoolmates into the same category called *socials* (SOC) as in [15]. In a second step, we consider the five original categories labelled by the participants, i.e., ACQ, WORK, FAM, FRIENDS, and UNI.

Three Categories: WORK, FAM, and SOC

As detailed above, we first generate three additional datasets based on the original dataset noted *DATA-3*: *UNDER-3*, *SMOTE-3*, and *SMUNDER-3*. Tab. 9 shows the resulting distribution of the number of contacts for each dataset. Based on these datasets, we apply the chosen classification algorithms and show their performance in Tab. 6. We can observe that the SVM shows the best performance independently of the datasets. Overall, the best results are obtained with SMOTE-3. In this case, precision, recall, and F-measure are equal to 86.8% with a kappa of 0.80. Tab. 7 provides a detailed view of the results including the corresponding confusion matrix. The SVM algorithm performs slightly better in identifying family members followed by coworkers and socials. However, the difference between the considered relationship categories remains minimal.

In Tab. 10, we compare the performance of the SVM algorithm with the ZeroR algorithm. This algorithm serves as baseline, since it always predicts that all contacts belong to the largest category. For all datasets, the SVM algorithm performs better than the ZeroR algorithm. The difference in accuracy between both algorithms is the largest for SMOTE-3. Overall, the highest accuracy (86.8%) is reached with the combination of the SMOTE-3 dataset and the SVM algorithm. In comparison, [15] achieves 87.1% accuracy in similar settings, i.e., when only considering those contacts participants have communicated with and leveraging only communication-based features. Our results support the findings of [15] based on a completely independent dataset.

Five Categories: ACQ, WORK, FAM, FRIENDS, and UNI

Next, we investigate the performance of the algorithms when disassembling the SOC category into ACQ, FRIENDS, and UNI. By doing so, we expect a lower accuracy, as we are increasing the number of categories from three to five. Simultaneously, offering more categories would refine the contact categorization and hence, enable the expression of additional sharing preferences. Finding the right number of categories to achieve sufficient accuracy remains challenging and requires additional efforts that are considered as future work.

As in the previous evaluation, we first generate three additional datasets based on the original dataset *DATA-5*: *UNDER-5*, *SMOTE-5*, and *SMUNDER-5*. Tab. 12 displays the resulting distribution for each relationship category. Additionally, Tab. 11 presents the classification results when (1) ignoring e-mail-based features and (2) considering them. They show that the classifiers better perform when e-mails are taken into consideration. Again, we observe that SVM performs better for all tested datasets than the other chosen algorithms. Moreover, the best results are obtained with the SMOTE-5 dataset with $C = 1.7$ and $\gamma = 1.0$. In the best case, precision, recall, and F-measure are around 58% with Kappa equals to 0.47.

As expected, the obtained accuracy decreases when the number of categories increases. This observation is valid for both the ZeroR and SVM algorithm as illustrated in Tab. 13. The SVM algorithm however clearly outperforms the ZeroR algorithm, which systematically classifies contacts into the most populated category. Tab. 14 shows that most contacts are classified in the right social category with a few exceptions. For the DATA-5 dataset, a majority of family members are categorized as friends, while the same number of uni/schoolmates are classified as such and as friends. With UNDER-5, almost the same number of acquaintances is recognized as coworkers and as acquaintances, while family members are also identified as friends. Based on our results, we observe that the differences between family members and friends are insufficient to reliably distinguish both categories. The same observation can also be made for uni/schoolmates and friends as well as acquaintances and coworkers. These results may however be influenced by the demographics of our participants' sample as discussed in detail in the next section.

Note that we have tested a two-level classification, in which we (1) classify the contacts into WORK, FAM, and SOC, and (2) classify the SOC category into ACQ, FRIENDS, and UNI. The two-step process however does not improve the results, which have therefore not been included in this manuscript.

In summary, we have shown that including e-mails as a new communication channel improves the performance of the

Table 6: Classification results for the three relationship categories: WORK, FAM, and SOC. For UNDER-3 and SMUNDER-3, we show the mean and standard deviation of the three runs

Evaluated dataset	Algorithm	Precision (%)	Recall (%)	F-measure (%)	Kappa
DATA-3	Naïve Bayes	70.6	61.1	62.2	0.28
	Decision tree C4.5	73.8	77.3	74.1	0.40
	SVM	74.9	77.9	75.2	0.42
UNDER-3	Naïve Bayes	58.1 (3.5)	58.0 (1.9)	55.1 (2.6)	0.37 (0.03)
	Decision tree C4.5	57.0 (1.3)	57.6 (1.1)	56.8 (1.6)	0.36 (0.02)
	SVM	62.1 (4.4)	62.0 (4.3)	60.7 (4.9)	0.43 (0.07)
SMOTE-3	Naïve Bayes	73.4	70.9	70.2	0.56
	Decision tree C4.5	83.5	83.6	83.5	0.75
	SVM	86.8	86.8	86.8	0.80
SMUNDER-3	Naïve Bayes	69.7 (0.3)	68.0 (0.4)	67.2 (0.4)	0.52 (0.01)
	Decision tree C4.5	79.6 (0.6)	79.7 (0.6)	79.6 (0.6)	0.69 (0.01)
	SVM	81.0 (0.2)	81.0 (0.2)	80.9 (0.2)	0.71 (0.01)

Table 7: Detailed results and confusion matrix for the SVM algorithm applied to the SMOTE dataset when considering WORK, FAM, and SOC

SMOTE-3				Predicted as		
Class	Precision	Recall	F-measure	WORK	FAM	SOC
WORK	0.857	0.852	0.854	1104	29	163
FAM	0.917	0.939	0.928	17	1127	56
SOC	0.834	0.821	0.828	167	73	1103
Mean	0.868	0.868	0.868	Accuracy = 86.8%		

Table 8: Top nine features based on the information gain computed using the Ranker search method

Feature	Information gain
Total number of calls	0.20981
Number of days called divided by days logged calls	0.20951
Total duration of calls	0.19114
Average number of calls per week	0.17921
Maximum length of calls	0.1731
Standard deviation of the number of calls per week	0.17253
Average length of calls	0.17251
Maximum number of calls per week	0.16947
Average length of e-mails	0.16505

Table 9: Number of contacts for each relationship category in the different datasets. Note that each contact belongs to one category. For UNDER-3 and SMUNDER-3, we show the total number of contacts for the three runs

Method	WORK	FAM	SOC
DATA-3	432	120	1343
UNDER-3	327	346	350
SMOTE-3	1296	1200	1343
SMUNDER-3	2573	2514	2680

Table 10: Accuracy comparison between ZeroR and SVM classifiers for WORK, FAM, and SOC

Method	ZeroR (%)	SVM (%)
DATA-3	70.9	77.9
UNDER-3	34.2	68.0
SMOTE-3	35.9	86.8
SMUNDER-3	34.5	81.3

classifiers. However, decomposing the social category into friends, acquaintances, and uni/schoolmates leads to a reduction in the performance of the classifiers. This result may be due to the choice of the categories or the communication patterns of the participants. It would therefore be interesting to further investigate if different categories or groups of participants would lead to improved results. As already mentioned, proposing different categories would allow the users to refine their sharing preferences and hence hopefully allow to optimize the audience of their posts. Simultaneously, the more categories, the higher the probability that the contacts may be misidentified. Finding a balance between both aspects is considered as future work.

DISCUSSIONS

We have seen in the previous section that existing classifiers do not succeed in classifying all contacts in the right relationship categories. This result may be explained by different factors. Firstly, most participants having contributed to the dataset are students. While some of them may have a part-time job, they have fewer and less intense interactions in a professional context with coworkers than for instance full-time employees have. Moreover, students can more easily communicate with, e.g., their friends, during traditional working hours than employees can do. As a result, the main differences between the proposed relationship categories are melting and make the classification more difficult. In the future, we therefore plan to extend the data collection by further diversifying the participants' demographic backgrounds.

Moreover, the participants decided which e-mail accounts they were willing to include in the data collection process. As excepted, most of them excluded their professional e-mail account(s). By doing so, contact names and communication

Table 11: Differences in the classification performance when (1) ignoring e-mail-based features and (2) including them for the five considered relationship categories. For UNDER-5 and SMUNDER-5, we present both the mean and the standard deviation for the three runs

Evaluated dataset		Algorithm	Precision (%)	Recall (%)	F-measure (%)	Kappa
DATA-5	No emails	Naïve Bayes	39.6	35.0	30.3	0.15
		Decision tree C4.5	38.2	39.6	37.2	0.18
		SVM	41.0	41.4	35.4	0.23
	Emails	Naïve Bayes	44.2	41.2	39.2	0.25
		Decision tree C4.5	51.1	52.3	50.6	0.36
		SVM	55.1	55.1	53.2	0.40
UNDER-5	No emails	Naïve Bayes	36.4 (2.3)	34.9 (2.4)	29.9 (1.8)	0.18 (0.02)
		Decision tree C4.5	34.0 (2.5)	36.0 (2.6)	39.5 (2.4)	0.19 (0.03)
		SVM	34.7 (1.0)	37.7 (1.9)	32.4 (0.8)	0.20 (0.02)
	Emails	Naïve Bayes	40.2 (1.6)	40.8 (1.7)	38.1 (1.9)	0.26 (0.02)
		Decision tree C4.5	40.0 (0.9)	40.2 (0.6)	39.5 (0.4)	0.25 (0.01)
		SVM	43.4 (1.4)	44.4 (0.7)	43.2 (1.7)	0.30 (0.01)
SMOTE-5	No emails	Naïve Bayes	43.1	32.3	28.5	0.16
		Decision tree C4.5	41.3	38.6	36.3	0.21
		SVM	48.8	43.3	40.2	0.30
	Emails	Naïve Bayes	47.9	45.6	44.1	0.32
		Decision tree C4.5	54.6	54.8	54.5	0.43
		SVM	58.4	58.1	58.2	0.47
SMUNDER-5	No emails	Naïve Bayes	40.5 (2.7)	40.3 (1.4)	34.9 (1.6)	0.26 (0.02)
		Decision tree C4.5	42.5 (2.0)	41.7 (1.4)	38.2 (1.7)	0.27 (0.01)
		SVM	46.3 (0.9)	44.8 (1.2)	40.6 (1.3)	0.31 (0.01)
	Emails	Naïve Bayes	46.9 (0.7)	45.9 (1.1)	43.6 (1.4)	0.33 (0.01)
		Decision tree C4.5	52.3 (0.4)	52.1 (0.4)	52.1 (0.4)	0.40 (0.01)
		SVM	55.9 (1.5)	54.5 (1.2)	54.8 (1.2)	0.43 (0.02)

Table 12: Number of contacts per relationship category for the considered datasets. For UNDER-5 and SMUNDER-5, we show the total number of contacts for the three runs

Method	ACQ	WORK	FAM	FRIENDS	UNI
DATA-5	576	432	120	509	258
UNDER-5	284	306	293	305	288
SMOTE-5	576	432	360	509	516
SMUNDER-5	629	579	593	608	606

Table 13: SVM and ZeroR accuracy comparison for five relationship categories

Method	ZeroR (%)	SVM (%)
DATA-5	31.0	55.1
UNDER-5	20.7	43.9
SMOTE -5	24.1	58.1
SMUNDER-5	20.9	55.7

patterns with coworkers were not logged. By adding them, we would also expect a better classification of this category, especially based on the time of the day at which e-mails would be exchanged.

We have also observed that few MMS were exchanged between our participants and their contacts. Indeed, this communication channel has been widely replaced by free alternatives, such as WhatsApp or Facebook messages. However, those are more difficult to access. Concerning WhatsApp, the scheme used to encrypt the local database was changed after an exploit has been made public [26] shortly before we implemented our app. As a result, it was not possible to access WhatsApp messages using an unrooted phone. As mentioned before, logging Facebook messages would also only have been possible by asking the participants to create a developer account in Facebook. Because of that, we further missed a relevant part of the participants' communication that may have help to better classify their contacts in the correct relationship categories.

As already mentioned in [15], it is also difficult to judge without conducting additional studies whether the obtained accuracy is enough for the users to accept and use the proposed solution. Similarly, balancing the tradeoff between loss in accuracy and increased number of categories requires also further efforts. Finally, our dataset and analysis share the same limitations as mentioned in [25]. We may fail in observing and interpreting the evolution of relationships. For example, after an intense communication phase, exchanges between contacts may become rarer. In this case, either the social tie may be broken or it may be even stronger as the contacts may spend more time together and thus do not need to communicate via their mobile phones. Additionally, we do not tailor the communication models based on the participants' demographics.

RELATED WORK

Different approaches have already been followed to infer and classify social links. Early works, such as [22], have primarily focused on categorizing social ties based on offline communication and behaviors. With the emergence of new technologies, the analysis of social relationships has progressively moved to the online domain. For example, communication patterns based (only) on e-mails have been observed in [6, 14], while mobile phone communication at network level have been considered in [17, 18].

Table 14: Detailed performance for the SVM algorithm (e-mail-based features included) and the five relationship categories. For UNDER-5 and SMUNDER-5, we show the results of one run only

(a) DATA-5				Predicted as				
Class	Precision	Recall	F-measure	ACQ	WORK	FAM	FRIENDS	UNI
ACQ	0.582	0.576	0.579	332	80	3	118	43
WORK	0.624	0.650	0.637	69	281	0	58	24
FAM	0.556	0.042	0.078	19	8	5	87	1
FRIENDS	0.502	0.680	0.578	86	47	1	346	29
UNI	0.452	0.310	0.368	64	34	0	80	80
Mean	0.551	0.551	0.532	Accuracy = 55.1%				

(b) UNDER-5				Predicted as				
Class	Precision	Recall	F-measure	ACQ	WORK	FAM	FRIENDS	UNI
ACQ	0.441	0.303	0.359	30	31	4	11	23
WORK	0.478	0.735	0.579	9	75	4	5	9
FAM	0.407	0.258	0.316	5	13	24	38	13
FRIENDS	0.407	0.447	0.426	10	15	23	46	9
UNI	0.432	0.432	0.432	14	23	4	13	41
Mean	0.433	0.439	0.425	Accuracy = 43.9%				

(c) SMOTE-5				Predicted as				
Class	Precision	Recall	F-measure	ACQ	WORK	FAM	FRIENDS	UNI
ACQ	0.550	0.523	0.536	301	80	24	85	86
WORK	0.581	0.637	0.608	65	275	7	50	35
FAM	0.750	0.683	0.715	24	15	246	63	12
FRIENDS	0.489	0.513	0.500	78	55	44	261	71
UNI	0.601	0.595	0.598	79	48	7	75	307
Mean	0.584	0.581	0.582	Accuracy = 58.1%				

(d) SMUNDER-5				Predicted as				
Class	Precision	Recall	F-measure	ACQ	WORK	FAM	FRIENDS	UNI
ACQ	0.511	0.431	0.468	91	30	13	49	28
WORK	0.624	0.630	0.627	27	121	5	17	22
FAM	0.724	0.688	0.706	5	5	139	48	5
FRIENDS	0.398	0.544	0.460	30	18	22	105	18
UNI	0.588	0.502	0.542	25	20	13	45	104
Mean	0.569	0.557	0.560	Accuracy = 55.7%				

Instead of large scale analysis, further works have concentrated on user-based analysis and leveraged the users' personal device as main collecting platform. One of the first approaches in this area is the Reality Mining project [4], which logged calls, current cell tower IDs, application usage, phone status, as well as nearby Bluetooth devices using participating devices. Based on the collected dataset, the authors aim at inferring the nature of the friendship relationships (i.e., non-friendship, asymmetric friendship, or symmetric friendship). [16] builds on the same categories, but exclusively utilizes data related to explicit communication between users, such as short messages and calls. In contrast, FriendSensing [20] focuses on Bluetooth-based proximity information. The study conducted in [5] further confirms the validity of mobile phone data as information source to identify existing friendships.

Our work primarily draws on [15]. Compared to the aforementioned studies, our approach and that of [15] do not only focus on friendships, but consider more relationship categories, respectively. In [15], the authors concentrate on family, coworkers, and social contacts in general (including, e.g., school, hobby, neighborhood). In comparison, we further divide the social contacts into friends, acquaintances, and uni/school mates to attempt to refine the classification. Our work also differs from [15] when considering the logged data. In addition to contacts, calls, and SMS, already logged in [15], we extend the collection to e-mails and MMS.

CONCLUSIONS

Using online social networks in a privacy-concious manner is difficult, and using mobile devices—the predominant way to interact with online social networks—to do so further complicates matters. One possible solution is to share data only in well-defined contexts, i.e., solely with the intended audience. Setting this audience in platforms such as Facebook is cumbersome, and hence many users resort to privacy-unconscious sharing behavior. We propose to utilize communication metadata available on smartphones to help with the categorization of users into social relationship categories, thus facilitating the sharing of content with restricted audience groups. To this end, we performed a user study collecting a dataset including contact names, calls, SMS, MMS, and e-mail on personal smartphones in everyday use of 19 users. Based on this sample and using machine learning, we identify that metadata about calls as well as e-mails constitute the key features determining the social relationship category of a contact. We obtain a classification accuracy of 86.8% for an automated contact classification into three main classes: work-related, family-related and other social-interaction-related, thus enabling the possibility of user assistance for privacy control. Our results also show that obtaining a more fine-grained categorization of latter category into acquaintances, friends, and university-mates is difficult. This is partially due to the fact that these categories blur in our study group. Further metadata, e.g. collected from WhatsApp, Twitter, or Facebook appears to be a promising avenue to increase the quality of the categorization, however, as of today, these services do not allow this type of data mining.

ACKNOWLEDGMENTS

Our thanks go to the participants of our study and A. García Bouso for her feedback.

REFERENCES

1. CBC News. 2009. Depressed Woman Loses Benefits over Facebook Photos. Online: http://www.cbc.ca (accessed in 06.15). (2009).
2. CBS News. 2011. Did the Internet Kill Privacy? Facebook Photos Lead to a Teacher Losing her Job; What Expectations of Privacy Exist in the Digital Era? Online: http://www.cbsnews.com (accessed in 06.15). (2011).
3. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. 2002. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* 16 (2002).
4. N. Eagle and A. S. Pentland. 2006. Reality Mining: Sensing Complex Social Systems. *Personal Ubiquitous Computing* (2006).
5. N. Eagle, A. S. Pentland, and D. Lazer. 2009. Inferring Friendship Network Structure by Using Mobile Phone Data. *Proceedings of the National Academy of Sciences* (2009).
6. H. Ebel, L.-I. Mielsch, and S. Bornholdt. 2002. Scale-free Topology of E-mail Networks. *Physical review E* 66, 3 (2002).
7. Facebook, Inc. 2015a. Facebook Statistics. Online: http://newsroom.fb.com/company-info/ (accessed 06.15). (2015).
8. Facebook, Inc. 2015b. Permissions with Facebook Login. Online: https://developers.facebook.com/docs/facebook-login/permissions/v2.2 (accessed in 06.15). (2015).
9. L. Fang and K. LeFevre. 2010. Privacy Wizards for Social Networking Sites. In *Proceedings of the 19th ACM International Conference on World Wide Web (WWW)*.
10. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. 2009. The WEKA Data Mining Software: An Update. *ACM SIGKDD Exploration Newsletter* 11, 1 (Nov. 2009).
11. M. A. Hall. 1998. *Correlation-based Feature Subset Selection for Machine Learning*. Ph.D. Dissertation. University of Waikato, Hamilton, New Zealand.
12. M. Johnson, S. Egelman, and S. M. Bellovin. 2012. Facebook and Privacy: It's Complicated. In *Proceedings of the 8th ACM Symposium on Usable Privacy and Security (SOUPS)*.
13. P. G. Kelley, R. Brewer, Y. Mayer, L. F. Cranor, and N. Sadeh. 2011. An Investigation into Facebook Friend Grouping. In *Proceedings of the 13th IFIP TC 13 International Conference on Human-computer Interaction (INTERACT)*.
14. G. Kossinets and D. J. Watts. 2006. Empirical Analysis of an Evolving Social Network. *Science* (2006).
15. J.-K. Min, J. Wiese, J. I. Hong, and J. Zimmerman. 2013. Mining Smartphone Data to Classify Life-facets of Social Relationships. In *Proceedings of the 16th ACM Conference on Computer Supported Cooperative Work (CSCW)*.
16. S. H. Mirisaee, S. Noorzadeh, A. Sami, and R. Sameni. 2010. Mining Friendship from Cell-Phone Switch Data. In *Proceedings of the IEEE 3rd International Conference on Human-Centric Computing (HumanCom)*.
17. S. Motahari, O. J Mengshoel, P. Reuther, S. Appala, L. Zoia, and J. Shah. 2012. The Impact of Social Affinity on Phone Calling Patterns: Categorizing Social Ties from Call Data Records. In *Proceedings of the 6th Workshop on Social Network Mining and Analysis (SNA-KDD)*.
18. J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási. 2007. Structure and Tie Strengths in Mobile Communication Networks. In *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*.
19. E. Protalinski. 2012. Facebook has over 425 Million Mobile Users. Online: http://www.zdnet.com/article/facebook-has-over-425-million-mobile-users (accessed 06.15). (2012).
20. D. Quercia and L. Capra. 2009. FriendSensing: Recommending Friends Using Mobile Phones. In *Proceedings of the 3rd ACM Conference on Recommender Systems (RecSys)*.
21. D. Reinhardt, F. Engelmann, and M. Hollick. 2015. Can I Help You Setting Your Privacy? A Survey-based Exploration of Users' Attitudes towards Privacy Suggestions. In *Proceedings of the 13th ACM International Conference on Advances in Mobile Computing and Multimedia (MoMM)*.

22. Henri Ed Tajfel. 1978. *Differentiation between Social Groups: Studies in the Social Psychology of Intergroup Relations*. Academic Press.

23. Y. Wang, G. Norcie, S. Komanduri, A. Acquisti, P. G. Leon, and L. F. Cranor. 2011. "I Regretted the Minute I Pressed Share": A Qualitative Study of Regrets on Facebook. In *Proceedings of the 7th Symposium on Usable Privacy and Security (SOUPS)*.

24. J. Wiese, P. G. Kelley, L. F. Cranor, L. Dabbish, J. I. Hong, and J. Zimmerman. 2011. Are You Close with Me? Are You Nearby?: Investigating Social Groups, Closeness, and Willingness to Share. In *Proceedings of the 13th ACM International Conference on Ubiquitous Computing (UbiComp)*.

25. J. Wiese, J.-K. Min, J. I. Hong, and J. Zimmerman. 2015. "You Never Call, You Never Write": Call and SMS Logs Do Not Always Indicate Tie Strength. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work (CSCW)*.

26. S. Yegulalp. 2014. Android Exploit Eats up WhatsApp Chat Databases. Online: http://www.infoworld.com (accessed in 06.15). (2014).