On Improving Tie Strength Estimates by Aggregating Multiple Communication Channels

Narges Yousefnezhad
Aalto University
narges.yousefnezhad@aalto.fi

Marcin Nagy
Aalto University
marcin.nagy@aalto.fi

N.Asokan
Aalto University & University of Helsinki
asokan@acm.org

***Abstract*—The degree of closeness in a relationship is characterized as *tie strength*. Estimates of tie strength can be useful in many contexts, including as a parameter in access control policies or social context based services. Several papers have proposed how tie strength can be estimated by quantifying interactions in different individual communication channels such as online social networks, phone communication and face-to-face encounters. It has been conjectured by Wiese et al. [1] that considering only a single communication channel may not lead to accurate estimates of tie strengths. In this paper, we explore this conjecture by examining whether the combination of co-location events and mobile communication data can lead to better tie strength estimations than considering each channel individually. Surprisingly, our results indicate that the conjecture may not be true, but further analysis with more extensive datasets is needed to confirm the result.**

I. Introduction

Tie strength is a notion used by social scientists to represent the degree of closeness in a relationship between two people [2]. The ability to accurately estimate tie strengths among people can lead to new services or improvement of existing ones. For instance, travellers and commuters can use tie strength estimation to decide if they want to share a ride with a stranger [3]. Similarly, people can decide to share their mobile data connection with close friends, specified as the list of their contacts with tie strength values above some threshold [4]. Generally, estimation of tie strength has many important applications in user-controlled online identity authentication [5], consumer behaviour prediction systems [6], recommendation services [7] and reputation services [8].

Prior research on tie strength estimation has largely focused on using input data from a single *communication channel*. We define a communication channel as any medium that can be used for exchange of information between people. Most studies estimate tie strength based on three communication channels: online social network (OSN) interactions [9], [10], traditional telecommunication such as calls and text messages [11], [12], [13] and interactions based on physical proximity [14], [15], [16].

Intuitively, information about interactions in *different* communication channels is likely to be a more accurate predictor for tie strength values. Several previous works [2], [17], [18], [1] have touched on this topic. Want et al. [17] conducted user studies to understand how well interactions over different communication channels correlate with closeness of friendship. Hritsova et al. [18] showed that people who use multiple types of channels for communicating with each other are more likely to have higher tie strengths between them. Wiese et al. [1] showed that when only one communication channel (interactions via telecommunication networks) is considered, the resulting tie strength estimates may be incorrect. They further concluded that combining information from different communication channels can lead to more accurate estimation of tie strength values. None of above investigated concrete tie-strength computation techniques that use multiple communication channels to confirm whether the conjecture is correct.

In this paper, we explore this question by using machine learning classifiers to predict tie strengths in order to evaluate whether combination of data from different communication channels leads to a better prediction accuracy. We use an existing dataset [19]. Our results indicate that while this conjecture may be true, it cannot be claimed with statistical significance. We therefore conclude that a more extensive dataset would be needed in order to resolve this question more definitively.

II. Background

Tie strength was introduced by Granovetter in 1973 [20]. He defines strength of a tie between two people in the social network as a combination of four factors: the amount of *time* people spend with each other, *emotional intensity*, *intimacy* (mutual confiding), and *reciprocal services* that characterize the tie. Furthermore, he also divides ties into two classes: *weak ties* that link acquaintances and *strong ties* that are formed between people trusting each other. There is a lot of published prior work on tie strength estimation with a particular focus on assigning binary values (strong or weak) to ties [17], [5], [21] and labelling them [22], [8].

We now present a brief summary of recent works on tie strength estimation using a single communication channel and discuss the shortcomings of relying only on a single channel.

A. Tie strength in a single communication channel

We consider three types of communication channels: online social networks (OSN), mobile communication networks and physical proximity.

Tie strength via OSN interactions. People using OSNs often have a very large number of contacts. Although most

ISBN 978-3-901882-83-8 © 2016 IFIP

ISBN 978-3-901882-84-5, IFIP Networking 2016 © 2016 IFIP

OSNs provide the functionality of assigning social contacts to specific sets (e.g., family, acquaintances, etc.) that reflect various degrees of closeness, people usually do not bother to take advantage of such functionality. To automate this process, Gilbert et al. [23] and Spiliotopoulos et al. [24] proposed using tie strength estimation methods based on a linear combination of factors described by Granovetter as well as emotional support and social distance. Arnaboldi et al. [9] defined 19 Facebook features, found their correlation to tie strength and presented two linear models for tie strength estimation. They concluded that recency of contact between people has the highest impact on tie strength. Jones et al. [10] extracted 14 features and developed a logistic regression model to check importance of extracted features. They, however, showed that interaction frequency is the most important feature in determining tie strength.

Tie strength via mobile communication network interactions. Before OSNs became hugely popular, tie strength estimation research largely concentrated on interactions via (mobile) communication networks. Onnela et al. [11] examined social communication patterns based on phone calls and SMSes. They applied duration of calls for tie strength estimation to show the existence of a relationship between tie strength and local social network structure. Zhang and Dantu [12] presented an affinity model for predicting social ties relying on communication logs. Eagle et al. [25] analysed status of friendship based on mobile phone record data.

Tie strength via interactions in physical proximity. Tie strength can also be estimated based on co-location events (proximity interactions) between two people. Crandel et al. [15] found that high number of physical proximity interactions between two people corresponds to the higher probability of a strong tie between them. Bilogrevic et al. [16] used the notion of an *encounter* (defined as co-presence of two people for a sufficiently long duration) for estimating tie strength. Sekara et al. [26] presented tie strength estimation based on proximity as determined by Bluetooth encounters.

B. *Shortcomings of using a single communication channel*

Although tie strength estimation based on a single communication channel gives a fairly accurate results, applications like access control can benefit from increased accuracy. For instance, tie strength estimation based solely on physical proximity interactions is affected by the *familiar stranger* [27] phenomenon, which can causes the strength of some ties to be overestimated. Similarly, ties between people that are not usually co-located (e.g., in long-distance relationships) will be underestimated. Wiese et al. [1] showed that tie strength estimation based only on mobile communication interactions causes about 50% of strong ties to be incorrectly classified as weak ties. They concluded that there is a strong motivation for building tie strength estimation methods that connect input data from multiple communication channels.

III. MULTI COMMUNICATION CHANNEL TIE STRENGTH

We now discuss our multi communication channel tie strength estimation model. We begin with a description of the dataset we worked with, including an overview of features we use in our model. Later, we describe the three tie strength estimation models.

A. *Dataset*

We have two main requirements for the dataset to fulfill: (1) presence of at least two different communication channels and (2) ground truth about the tie strength between pairs of people. We chose the *MIT Social Evolution* dataset [19] which contains traces from everyday life of 80 students living in the dormitory on the MIT campus. The dataset includes two communication channels: physical proximity (based on Bluetooth scans) and mobile communication network interactions (logs of phone calls and SMSes). The dataset covers nine months beginning from October 2008.

Data volume. The dataset contains 372 instances (pairs of people) with mobile communication interactions and 4770 instances with physical co-presence. 367 instances have interactions in both communication channels, and can thus be used for evaluation of tie strength estimation in a multi communication channel model.

Ground truth. During the data collection campaign, participants were asked which other participants they consider to be *close friends* with. Thus, if a participant has indicated that he/she is a close friend of another participant, we recognize their tie as strong. Otherwise, we consider their tie as weak. Overall, the ground truth is skewed, as only 668 pairs out of 4770 total pairs of users in both interactions indicated strong tie.

B. *Multi-channel Tie Strength Model*

Definition. We define the multi communication channel tie strength as a tie strength between two individuals that includes communication features coming from multiple communication channels.

Figure 1 illustrates difference in tie strength estimation between single and multiple communication channel approaches. If only mobile communication channel is involved, tie strength between Bob and John cannot be estimated. Similarly, tie strength between Bob and Alice cannot be calculated if only physical proximity channel is considered. However, by aggregation of data from multiple communication channels, all possible tie strengths between them can be estimated. Furthermore, tie between Alice and John can be estimated more accurately, as it includes data coming from both communication channels.

Feature Extraction. Recall from II-A that *contact duration*, *contact frequency*, and *recency of contacts* are considered as the most important features for tie strength estimation. We use them as the basic features both in mobile communication as well as physical proximity channel. In addition, we derive several new features which are based on distribution of these basic features (e.g., percentiles of call duration).

ISBN 978-3-901882-84-5, IFIP Networking 2016 © 2016 IFIP

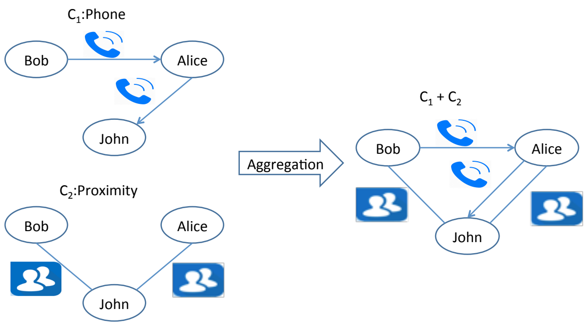

Fig. 1: Social network based on multiple communication channels.

Having extracted the features for mobile communication and physical proximity channels, we build two models, namely the **Mobile Communication-only** model and the **Proximity-only** model based on the features from respected channels. Finally, we create the new model (which we call **Aggregation**) by combining the features from the both channels.

Model description.

- ***Mobile Communication-only.*** The most important mobile communication features are identical to the top five features used by [1]. Furthermore, we define also additional 11 features describing various percentile levels of inter-communication times (i.e., time intervals between subsequent communication between two people). In total, the model includes 16 mobile communication features (see Table II for details).
  Furthermore, we assume mobile communication model follows *unidirectional* character of user interactions. This is motivated by intuitively different impact of incoming and outgoing calls and SMSes on tie strength estimation. For instance, if Alice calls Bob, it can be implied that she is interested in him, but the reverse interest cannot be proven.
- ***Proximity-only.*** Following Bilogrevic et al. [16], we use *encounter* as the primary feature. Based on it, we derive a total of 17 proximity features ranging from simple total encounter counts and mean encounter duration to more sophisticated percentile based features describing distributions of encounter and inter-encounter durations (see Table II for details).
  Unlike in the mobile communication-only model, we assume user interactions in the Proximity-only model to be *bidirectional*. This is motivated by assumption of mutual interest of two people during a co-presence event.
- ***Aggregation.*** This model includes both the Mobile Communication-only and the Proximity-only features. In total, the model has 33 features.
  Since the aggregation model is a combination of the Mobile communication-only and the Proximity model, for which notions of interactions are unidirectional and bidirectional respectively, we assume also unidirectional notion of user interactions in this model.

IV. Evaluation

This section presents accuracy of tie estimation achieved by our models. We begin with description of the dataset preparation for our evaluation. After that we describe how we checked that the dataset includes similar characteristics to the dataset used by Wiese et al. [1]. Finally, we evaluate the Aggregation model and compare accuracy of tie strength estimation with the Mobile Communication-only and the Proximity-only models.

A. Dataset preparation

Preprocessing. Recall from Section III that the Proximity-only model assumes user interactions to be bidirectional. Unfortunately results of Bluetooth scans may be asymmetrical (e.g., if Bob and Alice are co-located, only Bob's device discovers Alice presence, while her device does not discover him). This can happen for two reasons: (1) strong interference coming from neighbouring devices makes some devices unable to respond to Bluetooth discovery inquiry and (2) two parties are separated by some distance/obstacle (and in fact not co-present) and consequently Bluetooth signal between them is very weak and detectable only by one party. To compensate for this problem, we assumed the former reason and manually updated proximity data, as if the parties have been able to mutually discover themselves. On the other hand, we also checked that removal of asymmetric Bluetooth scan results (i.e., assuming the latter reason) does not change the accuracy of tie strength estimation results.

Normalization. Due to wide range of values that the features take, we apply the normalization of features with the range of 0 and 1.

B. Dataset validation

Wiese et al. [1] concluded that reliance only on the mobile communication network channel produces many errors in tie strength estimations and needs to be updated with more communication channels. Furthermore, as the dataset presented in their evaluation shows specific characteristics, we validate that our dataset exhibits similar characteristics.

They evaluated accuracy of tie strength estimation for three different input data. The first set of data (*all*) contains tie strength estimations between users and randomly chosen 70 contacts from the phone book and the Facebook contacts. The second set (*contactlist*) includes only contacts contained in user's phone book. Finally, the third set (*somecomm*) includes only contacts to which user has made at least 1 phone call or exchanged 1 SMS. Their results show a clear trend that precision of weak ties classification decreases if there are less contacts with whom user does not actively communicate (see

ISBN 978-3-901882-84-5, IFIP Networking 2016 © 2016 IFIP

Table I for details). The reason behind this performance drop is the fact that most of the ties without active communication are weak ones, thus are easier to classify.

Since our dataset does not contain the notion of the Facebook contacts and the phone book, we must generate a *simulated phone book* which includes all users contained in the dataset. We have following inputs:

- $fullbook$**:** includes all possible pairs of dataset users. Our dataset contains 80 users, so there are $80*79 = 6320$ pairs in total. If there is a recorded communication between a pair of users, we assign mobile communication-only features for them, otherwise we set values for all features to zero.
- $someEn1$**:** includes all possible pairs of users in the dataset with at least one recorded co-location event. It has 4403 pairs in total.
- $someEn10$**:** includes all possible pairs of users in the dataset with at least 10 recorded co-location events. It has 3893 pairs in total.

We validated that our dataset shows similar trends to Wiese et al.'s by evaluating our inputs using the Weka Toolkit [28]. We balanced ground truth using Synthetic Minority Oversampling Technique (SMOTE) [29] and used implementation of the Sequential Minimal Optimization (SMO) [30] with 10-fold cross-validation as the classifier. As the strong tie in our dataset is indicated by "close friend" label, for comparison we chose $2 - mediumstrong$ class condition from Wiese et al. which classifies tie strength into two classes (strong-medium ties and weak ties). Our dataset exhibits similar performance drop trend as reported by Wiese et al., thus it can be used for evaluation of multi communication channel tie strength estimation (see Table I for details).

TABLE I: Comparison of performance drop in precision of weak ties classification for Wiese et al. [1] and our dataset.

		Strong ties		Weak ties	
		Precision	**Recall**	**Precision**	**Recall**
Wiese et al.	all	0.693	0.420	0.764	0.920
	$contactlist$	0.683	0.460	0.680	0.843
	$somecomm$	0.707	0.724	0.488	0.467
Our dataset	$fullbook$	0.928	0.338	0.615	0.976
	$someEn1$	0.936	0.398	0.547	0.964
	$someEn10$	0.943	0.425	0.51	0.959

C. Aggregation analysis

Evaluation settings. Now we present the accuracy values for the Mobile Communication-only, the Proximity-only and the Aggregation models. Recall from III-A that only 367 pairs of users appear both in the mobile communication network channel and the physical proximity channel. Thus, to ensure equal input data in comparison of models, we use only data belonging to pairs of users appearing in both communication channels. We balanced ground truth using SMOTE and used SMO with 5-fold cross-validation as the classifier (10-fold cross-validation was not possible due to low number of input pairs).

TABLE II: Attributes and their weights in proximity and communication models.

Model	Attribute	Weight
Communication	Total duration of call	-0.0616
	Count of days with at least 1 call	-0.3544
	Count of calls	-0.0179
	Count of calls and SMSes	-0.0741
	Call duration mean time	-1.5441
	Mean time between two calls	-0.4804
	90th percentile of time between two calls	-1.534
	75th percentile of time between two calls	0.4843
	50th percentile of time between two calls	0.0186
	25th percentile of time between two calls	-0.9565
	90th percentile of call duration	-0.0237
	75th percentile of call duration	-0.0192
	50th percentile of call duration	-0.4215
	25th percentile of call duration	-0.311
	Count of days since last call	0.0298
Proximity	Count of co-location events	-0.1276
	Count of all encounters	0.4214
	Mean encounter time	-0.082
	Count of encounter days	0.1409
	95th percentile of encounters	-0.4808
	90th percentile of encounters	-0.3974
	80th percentile of encounters	-0.5826
	75th percentile of encounters	-0.1368
	50th percentile of encounters	0.3816
	25th percentile of encounters	0.6159
	90th percentile of time between two encounters	0.3849
	75th percentile of time between two encounters	0.1433
	50th percentile of time between two calls	0.092
	Mean time between two encounters	0.1584
	Count of non-encounter co-location events	0.53
	Count of days since last co-location	0.1467
	Sum of all times between two encounters	-0.2853

Results. The accuracy for the Mobile Communication-only, the Proximity-only and the Aggregation model equal 72.49%, 62.23% and 72.71% respectively. The Aggregation model obtains a 10% accuracy improvement over the Proximity-only model. However, results achieved by the Mobile Communication-only and the Aggregation models are almost equal with a slight edge for the latter (see Table III for details). Thus, although the Aggregation model achieves the best accuracy, the significant accuracy gain anticipated by Wiese et al. [1] is not observed with this dataset.

TABLE III: Classification performance with different models

Model	F-measure		Accuracy
	Strong ties	**Weak ties**	
Mobile Communication-only	0.688	0.754	72.49%
Proximity-only	0.656	0.581	62.23%
Aggregation	0.69	0.756	72.71%

Statistical Analysis. To verify that there is enough evidence to accept our claims about results, we run the statistical test. Table IV lists the accuracies of five folds for each of the three models.

We verify our claims by testing two hypotheses:

Null Hypothesis 1 (H1): There isn't any significant difference between accuracy results achieved by the Mobile Communication-only model and the Aggregate model.

Null Hypothesis 2 (H2): There isn't any significant difference between accuracy results achieved by the Proximity-only model and the Aggregate model.

ISBN 978-3-901882-84-5, IFIP Networking 2016 © 2016 IFIP

TABLE IV: Accuracy for Each Fold

	Accuracy		
	Mobile Communication-only	**Proximity-only**	**Aggregation**
fold 1	74.74%	56.84%	74.74%
fold 2	76.6%	57.45%	75.53%
fold 3	76.59%	60.64%	76.59%
fold 4	65.96%	63.83%	65.96%
fold 5	70.21%	57.45%	71.28%

We found that difference in accuracy between the Aggregate model and the Proximity-only method is statistically significant in the 95% confidence interval ($t^*_{H2} = -5.792$). However, there is not any significant difference between the accuracy of the Mobile Communication-only model and the Aggregate model in the 95% confidence interval ($t^*_{H1} = 0$).

V. Related Work

Motivated by constant increase in the use of OSNs, the research community has worked on several solutions for the social relationships based access control. Fogues et al. [31] reviewed some *Relationship-based Access Control (ReBAC)* models and specified their features. One feature which can be used for differentiating relationships in these models is tie strength. Carminati et al. [32] defined several *access control rules* and leverage relationship types for determining numerical values for strength of friendship.

Another group of research activities concern mappings of tie strength estimations between social networks. *WeMeddle*, the Twitter application showed that a predictive model for tie strength can be generalized to other social media [21]. Tang et al. [22] described a transfer-based factor graph (TranFG) model that can be used to learn and infer tie strength across heterogeneous networks.

Another set of activities is related to prediction of online social network evolution. Wang et al. [33] discovered that online and offline movement patterns have strong correlation with each other and measured that both patterns can be used for *link prediction*. They also observed that tie strength has more correlation with offline proximity than online measures. On the other hand, Kahanda et al. [34] investigated *link strength prediction* in online social networks. They derived four categories for social features and showed that network transactional features (e.g. wall posts) are the most important one.

Another field of research studies mechanisms for*trust inference* based on tie strength estimation. In [35], Seyedi et al. introduced a *proximity-based* method for bootstrapping trust values, and showed by experiment that trust values are relevant to tie strength using the MIT Reality Mining dataset. TidalTrust [36], SUNNY [37], H-OSTP [38], SocialTrust [39], FuzzyTrust [40] algorithms are examples for inferring trust in social networks.

Onnela et al. [11] examined social communication patterns based on phone calls and SMSes. They applied duration of calls for tie strength estimation to prove existence of a relationship between tie strength and local social network structure.

VI. Conclusion

In this paper, we evaluated the three new tie strength estimation models. Two of them are based on a single communication channel (the Mobile Communication-only and the Proximity-only models), while the third one (the Aggregate model) is constructed by merging all the features provided by the first two models. We evaluated performance of these models using the MIT Social Evolution dataset. Our results show a significant accuracy improvement of the Aggregate model in comparison to the Proximity-only model. However, the gain between the Aggregate model and the Mobile Communication-only model is negligible.

Based on obtained results, we cannot confirm (with this dataset) the hypothesis stated by Wiese et al. [1] that usage of multiple communication channels improves accuracy of tie strength estimation. However, their hypothesis cannot be dismissed either, as the dataset used by us contains communication data only between people that have participated in the collection campaign, thus it may not be fully representative. In addition, there are no other publicly available datasets that fulfil requirements of having multiple communication channels and verified ground truth. Finally, construction of the new dataset is also not a trivial task. In order to have a more meaningful dataset than the MIT Social Evolution dataset, it must be able to correlate identities of users (both actively participating in the dataset construction process as well as accidentally encountered) over multiple communication channels.

Acknowledgments

The authors would like to thank Mika Juuti for helping with using Weka Toolkit functions and machine learning techniques.

References

[1] Jason Wiese, Jun-Ki Min, Jason I Hong, and John Zimmerman, "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*. ACM, 2015, pp. 765–774.

[2] Nils Jeners, Petru Nicolaescu, and Wolfgang Prinz, "Analyzing tie-strength across different media," in *On the Move to Meaningful Internet Systems: OTM 2012 Workshops*. Springer, 2012, pp. 554–563.

[3] Blerim Cici, Athina Markopoulou, Enrique Frias-Martinez, and Nikolaos Laoutaris, "Assessing the potential of ride-sharing using mobile and social data: A tale of four cities," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, New York, NY, USA, 2014, UbiComp '14, pp. 201–211, ACM.

[4] Marcin Nagy, Thanh Bui, Emiliano De Cristofaro, N Asokan, Jörg Ott, and Ahmad-Reza Sadeghi, "How far removed are you? scalable privacy-preserving estimation of social path length with social pal," *arXiv preprint arXiv:1412.2433*, 2014.

[5] Tiffany Hyun-Jin Kim, Akira Yamada, Virgil Gligor, Jason Hong, and Adrian Perrig, "Relationgram: Tie-strength visualization for user-controlled online identity authentication," in *Financial Cryptography and Data Security*, pp. 69–77. Springer, 2013.

[6] Vikas Mittal, John W Huppertz, and Adwait Khare, "Customer complaining: the role of tie strength and information control," *Journal of Retailing*, vol. 84, no. 2, pp. 195–204, 2008.

ISBN 978-3-901882-84-5, IFIP Networking 2016 © 2016 IFIP

[7] Katrina Panovich, Rob Miller, and David Karger, "Tie strength in question & answer on social network sites," in *Proceedings of the ACM 2012 conference on computer supported cooperative work*. ACM, 2012, pp. 1057–1066.
[8] Wenbin Tang, Honglei Zhuang, and Jie Tang, "Learning to infer social ties in large networks," in *Machine Learning and Knowledge Discovery in Databases*, pp. 381–397. Springer, 2011.
[9] Valerio Arnaboldi, Andrea Guazzini, and Andrea Passarella, "Egocentric online social networks: Analysis of key features and prediction of tie strength in facebook," *Computer Communications*, vol. 36, no. 10, pp. 1130–1144, 2013.
[10] Jason J Jones, Jaime E Settle, Robert M Bond, Christopher J Fariss, Cameron Marlow, and James H Fowler, "Inferring tie strength from online directed behavior," *PloS one*, vol. 8, no. 1, pp. e52168, 2013.
[11] J-P Onnela, Jari Saramäki, Jorkki Hyvönen, György Szabó, David Lazer, Kimmo Kaski, János Kertész, and A-L Barabási, "Structure and tie strengths in mobile communication networks," *Proceedings of the National Academy of Sciences*, vol. 104, no. 18, pp. 7332–7336, 2007.
[12] Huiqi Zhang and Ram Dantu, "Predicting social ties in mobile phone networks," in *Intelligence and Security Informatics (ISI), 2010 IEEE International Conference on*. IEEE, 2010, pp. 25–30.
[13] Frank Bentley and Ying-Yu Chen, "The composition and use of modern mobile phonebooks," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2015, pp. 2749–2758.
[14] Mangesh Gupte and Tina Eliassi-Rad, "Measuring tie strength in implicit social networks," in *Proceedings of the 4th Annual ACM Web Science Conference*. ACM, 2012, pp. 109–118.
[15] David J Crandall, Lars Backstrom, Dan Cosley, Siddharth Suri, Daniel Huttenlocher, and Jon Kleinberg, "Inferring social ties from geographic coincidences," *Proceedings of the National Academy of Sciences*, vol. 107, no. 52, pp. 22436–22441, 2010.
[16] Igor Bilogrevic, Kévin Huguenin, Murtuza Jadliwala, Florent Lopez, Jean-Pierre Hubaux, Philip Ginzboorg, and Valtteri Niemi, "Inferring social ties in academic networks using short-range wireless communications," in *Proceedings of the 12th ACM workshop on Workshop on privacy in the electronic society*. ACM, 2013, pp. 179–188.
[17] Hua Wang, Vincent Chua, and Michael A Stefanone, "Social ties, communication channels, and personal well-being a study of the networked lives of college students in singapore," *American Behavioral Scientist*, p. 0002764215580590, 2015.
[18] Desislava Hristova, Mirco Musolesi, and Cecilia Mascolo, "Keep your friends close and your facebook friends closer: a multiplex network approach to the analysis of offline and online social ties," *arXiv preprint arXiv:1403.8034*, 2014.
[19] Anmol Madan, Manuel Cebrian, Sai Moturu, Katayoun Farrahi, et al., "Sensing the" health state" of a community," *IEEE Pervasive Computing*, , no. 4, pp. 36–45, 2012.
[20] Mark S Granovetter, "The strength of weak ties," *American journal of sociology*, pp. 1360–1380, 1973.
[21] Eric Gilbert, "Predicting tie strength in a new medium," in *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, 2012, pp. 1047–1056.
[22] Jie Tang, Tiancheng Lou, and Jon Kleinberg, "Inferring social ties across heterogenous networks," in *Proceedings of the fifth ACM international conference on Web search and data mining*. ACM, 2012, pp. 743–752.
[23] Eric Gilbert and Karrie Karahalios, "Predicting tie strength with social media," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2009, pp. 211–220.
[24] Tasos Spiliotopoulos, Diogo Pereira, and Ian Oakley, "Predicting tie strength with the facebook api," in *Proceedings of the 18th Panhellenic Conference on Informatics*. ACM, 2014, pp. 1–5.
[25] Nathan Eagle, Alex Sandy Pentland, and David Lazer, "Inferring friendship network structure by using mobile phone data," *Proceedings of the National Academy of Sciences*, vol. 106, no. 36, pp. 15274–15278, 2009.
[26] Vedran Sekara and Sune Lehmann Jørgensen, "The strength of friendship ties in proximity sensor data," *PL o S One*, vol. 9, no. 7, 2014.
[27] Stanley Milgram, "The Familiar Stranger: An Aspect of Urban Anonymity," in *The Individual in a Social World*, pp. 51–53. Addison-Wesley, Reading, MA, USA, Aug. 1977.
[28] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten, "The weka data mining software: an update," *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, 2009.
[29] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *J. Artif. Int. Res.*, vol. 16, no. 1, pp. 321–357, June 2002.
[30] Chih-Chung Chang and Chih-Jen Lin, "Libsvm: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, May 2011.
[31] Ricard Fogues, Jose M Such, Agustin Espinosa, and Ana Garcia-Fornes, "Open challenges in relationship-based privacy mechanisms for social network services," *International Journal of Human-Computer Interaction*, vol. 31, no. 5, pp. 350–370, 2015.
[32] Barbara Carminati, Elena Ferrari, Raymond Heatherly, Murat Kantarcioglu, and Bhavani Thuraisingham, "Semantic web-based social network access control," *computers & security*, vol. 30, no. 2, pp. 108–115, 2011.
[33] Dashun Wang, Dino Pedreschi, Chaoming Song, Fosca Giannotti, and Albert-Laszlo Barabasi, "Human mobility, social ties, and link prediction," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2011, pp. 1100–1108.
[34] Indika Kahanda and Jennifer Neville, "Using transactional information to predict link strength in online social networks," *ICWSM*, vol. 9, pp. 74–81, 2009.
[35] Amir Seyedi, Rachid Saadi, and Valérie Issarny, "Proximity-based trust inference for mobile social networking," in *Trust Management V*, pp. 253–264. Springer, 2011.
[36] Jennifer Ann Golbeck, "Computing and applying trust in web-based social networks," 2005.
[37] Ugur Kuter and Jennifer Golbeck, "Sunny: A new algorithm for trust inference in social networks using probabilistic confidence models," in *AAAI*, 2007, vol. 7, pp. 1377–1382.
[38] Guanfeng Liu, Yan Wang, Mehmet A Orgun, and Ee-Peng Lim, "A heuristic algorithm for trust-oriented service provider selection in complex social networks," in *Services Computing (SCC), 2010 IEEE International Conference on*. IEEE, 2010, pp. 130–137.
[39] James Caverlee, Ling Liu, and Steve Webb, "Towards robust trust establishment in web-based social networks with socialtrust," in *Proceedings of the 17th international conference on World Wide Web*. ACM, 2008, pp. 1163–1164.
[40] Mohsen Lesani and Saeed Bagheri, "Fuzzy trust inference in trust graphs and its application in semantic web social networks," in *Automation Congress, 2006. WAC'06. World*. IEEE, 2006, pp. 1–6.

ISBN 978-3-901882-84-5, IFIP Networking 2016 © 2016 IFIP