SCIENCE CHINA
Information Sciences

• LETTER •

April 2020, Vol. 63 149101:1–149101:3
https://doi.org/10.1007/s11432-017-9701-0

Inferring explicit and implicit social ties simultaneously in mobile social networks

Guojie SONG*, Yuanhao LI, Junshan WANG & Lun DU

Key Laboratory of Machine Perception, Ministry of Education, Peking University, Beijing 100871*, China*

Received 12 February 2018/Revised 21 June 2018/Accepted 31 October 2018/Published online 6 March 2020

Citation Song G J, LI Y H, Wang J S, et al. Inferring explicit and implicit social ties simultaneously in mobile social networks. Sci China Inf Sci, 2020, 63(4): 149101, https://doi.org/10.1007/s11432-017-9701-0

Dear editor,

Social tie inferring is to determine the type of relations, which is a significant task in social network analysis. Mobile social networks based on records from mobile phones, contain both interaction and spatial data of users, which may help to infer network structure [1], mine routines [2] and recognize social ties [3]. In mobile social networks, social ties are categorized into two types: explicit and implicit. Implicit social ties exist in real life but seldom interactions can be observed, and more analysis is shown in Appendix A. Traditionally, social tie recognition researches only focus on inferring explicit ties [4,5]. But discovering implicit social ties is also meaningful, as it is a supplement to explicit tie recognition. Its applications include friend suggestion [6], unseen links detection [7] and recommendation systems [8]. Many previous researches have proved that multi-task learning can improve the performance when tasks are related. So it is reasonable to infer explicit and implicit social ties simultaneously. The first challenge of inferring explicit and implicit ties is how to reveal community features of different relation types, calculated by explicit and spatial information in mobile social networks. The second challenge is how to unify all features, including traditional and community features and construct a uniform framework.

We propose a community-based factor graph model to infer explicit and implicit ties simultaneously. We firstly give empirical analysis and observe community features in mobile social networks. Then, we propose a community factor graph (CFG), whose node layer and relation layer are for the edge level features, and community layer is for community features. Finally, we set up experiments on real data and the results show that our model has best performance.

Our contributions can be summarized as follows. (1) We propose a new type of relationship recognition problem in mobile social networks, which can infer both explicit and implicit social ties simultaneously by considering microscopic interaction behaviors and mesoscopic community features. (2) We give a deep empirical analysis on real mobile communication data and extract two kinds of community features: structural and spatial features. (3) We propose a unified CFG with the refined loopy belief propagation and simulated annealing techniques for model learning. (4) Experimental studies on real mobile data show that the performance of our model performs better than baselines on explicit and implicit social ties.

Data and observations. Data in this study is a collection of mobile phone call records and message records. After filtering, we have 28780 family groups, 1647 colleague groups and 26435835 social ties, including explicit and implicit ties, specifically introduced in Appendix A. Communities with different relation types have different features, which may help infer social ties. We propose two kinds of community features: structural and spatial features, and observe the difference of each feature between each relation type. More de-

* Corresponding author (email: gjsong@pku.edu.cn)

tails are shown in Appendix B.

• Structural features represent the inherent network topology in a community, which can be calculated by interaction information. (1) Community size. Community of families and colleagues are different in size. Families mainly have 2 to 3 members and most of colleague cliques have more than 5 members. (2) Explicit edge clustering coefficient. Families have high clustering coefficient. To the contrast, colleague communities are inclined to have low clustering coefficient.

• Spatial features represent the spatial information of a community, which are obtained by the trajectories of all individuals in this community. (1) Spatial distance. We calculate the in-group average distance of each user pair to capture the distribution of users in a community in space. Family communities have high distance at day time and low at night, while colleagues are on the contrary. (2) Spatial co-occurrence. If two users locate in the same base station in one hour, then we define this phenomenon as spatial co-occurrence. Families members usually co-occur at the same place in the morning and the evening. While colleagues have high co-occurrence at working hours, and drop to a low level in the evening. (3) Spatial semantics. We get the spatial semantics of point of information (POIs) near the base station and analyze the distribution of spatial semantics of different relation users. Users of different relationships have different spatial semantics.

Community-based recognition model. A mobile social network with communities can be represented as a graph $G = (V, E, R, C)$, in which V represents vertex set, E represents the edge set, R represents the relation set and C represents the community set. For all relations, $\boldsymbol{X} = \{\boldsymbol{x}_i\}$ denotes their features. Community set C consists of several communities $c_i = (V_i, E_i)$, $V_i \subset V$, $E_i \subset E$, and $\boldsymbol{W} = \{\boldsymbol{w}_i\}$ denotes their features. Let $L = \{l_i\}$ represent the label set, $Y = \{y_i\}$ be the relation type of relations and $Z = \{z_i\}$ be the relation type of communities, $y_i, z_i \in L$. Our goal is to infer the type of all relations, which is to learn the function:

$$f : G = (V, E, R, C) \to Y. \tag{1}$$

Inspired by the traditional factor graph, we propose our CFG model, shown in Appendix C.1. There are three different layers in CFG, indicating user nodes, user relations and communities of users respectively. The first layer, node layer consists of all the user nodes. In the second layer, relation layer, each node corresponds to a relation in node layer. What makes our model special is the third layer, community layer. Each node in community layer represents a specific group of users in node layer. As a community of users means that a group of people which have the same relation type, hence all relations in this group are labeled the same.

We describe CFG in formalism under the context of a mobile social network:

$$p(Y|G) = \prod_{i,j} F_R(y_i, \boldsymbol{x}_i) F_C(z_j, \boldsymbol{w}_i), \tag{2}$$

where relation factor $F_R(y_i, \boldsymbol{x}_i)$ represents the posterior probability of relation y_i given $\boldsymbol{x}_i$, and community factor $F_C(z_j, \boldsymbol{w}_i)$ represents the posterior probability of community z_j given $\boldsymbol{w}_j$.

Furthermore, each factor contains two parts: attribute and correlation. We choose exponential-linear functions and have the relation factor and the community factor as

$$F_R(y_i, \boldsymbol{x}_i) = f(y_i, \boldsymbol{x}_i) g(y_i, G(y_i)), \tag{3}$$

$$F_C(z_j, \boldsymbol{w}_j) = h(z_j, \boldsymbol{w}_j) k(z_j, G(z_j)). \tag{4}$$

In (3), feature function $f(y_i, \boldsymbol{x}_i)$ represents the probability of y_i given the features $\boldsymbol{x}_i$, and $g(y_i, G(y_i))$ denotes the likelihood value of the correlation between relations, where $G(y_i)$ is the set of correlated relations to y_i. We use traditional relation features $\boldsymbol{x}_i$ of relation r_i. In (4), feature function $h(z_j, \boldsymbol{w}_j)$ represents the probability of z_i given the community features $\boldsymbol{w}_i$ and $k(z_j, G(z_j))$ denotes the likelihood value of the correlation between communities, where $G(y_i)$ is the set of correlated communities to z_j. Community features are $\boldsymbol{w}_i$ divided into structural and spatial features. More descriptions are shown in Appendix C.2.

Learning CFG is to give an estimate of parameter $\theta = (\lambda, \alpha, \beta, \gamma)$, so that the likelihood function reaches the maximum value on labeled input. The log-likelihood objective is defined as

$$O(\theta) = \ln p(Y|G), \tag{5}$$

where Z is the normalization factor. We apply Newton-Raphson method to solve the objective. Though the CFG have multiple layers, the parameters in different layers do not affect each other, so that each layer can be calculated individually. Each layer has a complex structure and we use loopy belief propagation (LBP). More descriptions and the algorithm are shown in Appendix C.3.

In normal probability graph, the inferring task is to maximize the joint probability based on θ: $y_* = \text{argmax}(P(y|x, \theta))$. When inferring explicit ties, we assist simulated annealing (SA) to exchange users in communities to compute the approximate solution with community features. When coming to implicit ties, it is unaffordable to calculate each

Table 1 Results of explicit and implicit tie recognition

	Model	Family			Colleague		
		Precision	Recall	F1-score	Precision	Recall	F1-score
Explicit	Logistic regression	0.8494	0.4619	0.5984	0.9032	0.2500	0.3916
	Classification and regression trees	0.8992	0.6783	0.7733	0.9285	0.4062	0.5652
	PLP-FGM	0.6682	0.8033	0.7296	0.9527	0.8129	0.8773
	Factor graph	0.7064	0.8651	0.7778	0.9553	0.8117	0.8777
	Community factor graph	0.8418	0.9269	**0.8823**	0.9693	0.9278	**0.9481**
Implicit	Logistic regression	0.8571	0.0517	0.0976	0.8727	0.0250	0.0486
	Classification and regression trees	1	0.1121	0.2016	0.9955	0.0244	0.0477
	Multilayer perception	0.4327	0.1940	0.2679	0.5020	0.0010	0.0021
	Community factor graph	0.5192	**0.5198**	**0.5195**	0.5491	**0.5912**	**0.5694**

node pair to determine whether there is an implicit ties or not. We use an approximate solution that when there is a new community, all the uninteracted edges are in consideration as implicit ties. Algorithms are shown in Appendix C.4.

The complexity of learning our model is $O(k_1(|E||R| + |E_C||R|))$ and the complexity of inferring is $O(k_2|E||R|)$, where k_1, k_2 are the number of iterations and $|E_C|$ is the number of edges between communities.

Experiments. We conduct the experiments on real world data containing more than 10000 users. As for a mobile social network, its nodes are users, its edges are interaction information and its relation types between nodes include no-relation, family or colleague. Table 1 and Appendix D show the result of explicit and implicit tie recognition for each kind of relation. For explicit tie recognition, our CFG model reaches the highest F1-score, of 10% promotion, confirming the superiority of our model. For implicit tie recognition, our model have the highest recalls and F1-scores among all the methods, about 2 and 10 times promotion on family and colleague implicit relationship respectively, thanks to community features. Then we conduct the experiment to prove that each kind of features is useful and make contribution to the improvement of F1-score as shown in Appendix E. we also infer the colleague ties without and with implicit ties. With the help of implicit ties, the recall and F1-score both improve more than 0.025. It demonstrates that when inferring explicit and implicit ties together, those enlarged ties help to infer more explicit ones.

Conclusion. We explore the explicit and implicit tie recognition problem in mobile social networks and observe two categories of community features. Based on the analysis that different relation types have different community features, we propose a CFG to infer both explicit and implicit social ties. In the model, a community layer is built with community features. We evaluate our model and confirm that using our community features based model can reach a better recognition result.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61876006, 61572041).

Supporting information Appendixes A–E. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Eagle N, Pentland A, Lazer D. From the cover: inferring friendship network structure by using mobile phone data. Proc Natl Acad Sci USA, 2009, 106: 15274–15278
2 Wang D S, Pedreschi D, Song C M, et al. Human mobility, social ties, and link prediction. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2011. 1100–1108
3 Qin T, Shangguan W F, Song G, et al. Spatio-temporal routine mining on mobile phone data. ACM Trans Knowl Discov Data, 2018, 12: 1–24
4 Tang J, Lou T C, Kleinberg J. Inferring social ties across heterogenous networks. In: Proceedings of the 5th ACM International Conference on Web Search and Data Mining, 2012. 743–752
5 Tang W B, Zhuang H L, Tang J. Learning to infer social ties in large networks. In: Proceedings of Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Berlin, 2011. 381–397
6 Qin Y, Yu Z T, Wang Y B, et al. Detecting micro-blog user interest communities through the integration of explicit user relationship and implicit topic relations. Sci China Inf Sci, 2017, 60: 092105
7 Krebs V E. Mapping networks of terrorist cells. Connections, 2002, 24: 43–52
8 Taheri S M, Mahyar H, Firouzi M, et al. Extracting implicit social relation for social recommendation techniques in user rating prediction. In: Proceedings of the 26th International Conference on World Wide Web Companion, 2017. 1343–1351