2010 IEEE/ACM International Conference on Green Computing and Communications & 2010 IEEE/ACM International Conference on Cyber, Physical and Social Computing

Human Contact Prediction Using Contact Graph Inference

Kazem Jahanbakhsh
Computer Science Department
University of Victoria
Victoria, British Columbia
Email: jahan@cs.uvic.ca

Gholamali C. Shoja
Computer Science Department
University of Victoria
Victoria, British Columbia
Email: gshoja@cs.uvic.ca

Valerie King
Computer Science Department
University of Victoria
Victoria, British Columbia
Email: val@cs.uvic.ca

***Abstract*—Predicting human mobility is considered as a challenging problem. In this paper, we formulate the problem of human contact prediction as a graph inference problem. We show the importance of using offline social information for predicting people's contacts motivated by homophily theory. We also prove that by using the small-world network properties of the contact graphs, we can reconstruct the missing part of a contact graph where only part of the graph is known. Our results are promising because they allow researchers to reconstruct the missing parts in experimentally measured human mobility traces when only partial traces are obtainable.**

I. Introduction

Predicting human mobility is complex because there are many parameters which influence the way that people move. These parameters can range from social factors such as people's occupations to the structure of the environment in which people move. Understanding the properties of human mobility has several applications in different areas. For example, it can be used to find the most efficient locations for GSM antennas. It can also be used for traffic planning in cities and public health studies of epidemic disease. Researchers have studied different aspects of human behavior such as the way they become friends, call each others by their phones, or collaborate together for publishing papers. They have found small-world network properties in most of these human-embedded networks [1].

Everybody in a society can be identified by a set of social characteristics such as occupation, affiliation, place of living, and so on. We call a person's set of social characteristics as her social profile. The homophily phenomenon in which similar people are more likely to interact with each others has been studied in social networks [2]. Similarly, we believe that people are more likely to interact with those who are socially similar to them. In our previous work, we have studied the effect of social characteristics on the people's mobility pattern in a conference setting [3].

In this paper, we focus on predicting human mobility in a conference environment. We say two people are *in contact* if they are in each other's proximity. For the first part of the paper, motivated by homophily theory we investigate the importance of social similarity on people's mobility patterns. We try to infer people's contacts by computing the similarities between their social profiles. In this part we assume that we only have information about people's social profiles while people's contacts are completely unknown. For the second part of the paper, we study a similar problem in a different setting where we try to infer the missing contacts among people when only part of the contact graph, which is the graph constructed by nodes mobility, is known. Inferring the missing part of a contact graph is important because it allows researchers to reconstruct the unobserved part of their graphs when there is a partial observation for people's contacts. To best of our knowledge, our work is the first one which addresses the problem of contact prediction in a mobile social network by using social theories. Our main contributions are:

- We show the importance of social profiles as well as the underlying structure of contact graphs in contact prediction problem.
- We also present several methods to reconstruct a contact graph when we only have information about people's social profiles or when only a partial part of the contact graph is known.

The remainder of the paper is organized as follow: Section II reviews the related work. Section III defines the problem to be tackled. Section IV describes the importance of social information on the structure of contact graphs. Section V explores the importance of using the underlying structure of contact graphs for contact prediction problem. Section VI discusses a model for contact graphs. Finally, Section VII concludes the paper.

II. Related Work

Eagle et al. [4] and Mitbaa et al. [5] have found a close relation between people's mobility patterns and their friendship network. More specifically, they have shown that people are more likely to meet their friends than strangers. Authors in [6] and [7] have also studied the properties of contact graphs. They have computed the centrality of nodes as well as their similarities by using the contact information. They have shown the effectiveness of using the small-world network properties of contact graphs for routing messages in mobile social networks. By analyzing human mobility traces, authors of [8] have analyzed the distribution of inter-contact time, that is the time gap separating two contacts between the same pair of people.

978-0-7695-4331-4/10 $26.00 © 2010 IEEE
DOI 10.1109/GreenCom-CPSCom.2010.13

Nowell and Kleinberg have studied the problem of link prediction in citation networks [9]. They have applied different methods borrowed from graph theory to find the distances between authors with respect to the underlying graph structure. Their goal was to infer the future collaborations between scientists by studying the underlying properties of the citation graph. Goldberg et al. have assessed the confidence of experimentally observed interactions among proteins by using cohesive neighborhoods and small average distance between proteins [10]. Authors of [11] have also studied the graph inference problem in metabolic networks where part of the graph is known by employing a supervised learning algorithm.

III. Problem Definition

In this paper, our main problem is graph inference when a prior knowledge about the graph is available. In the first part of the paper, for a graph $G = (V, E)$ we assume that we have offline information (e.g. social information) about vertices $v \in V$ while the edges in $E \subset V \times V$ are totally unknown. Thus, the problem becomes inferring edges in E by using the available side information about vertices in V. In the second part of the paper, we assume that for a graph $G = (V, E)$ our vertices set is $V = V_{int} \cup V_{ext}$ where V_{int} and V_{ext} denote the internal and external vertices respectively. We also assume that all edges in $E_{known} \subset V_{int} \times (V_{int} \cup V_{ext})$ are known whereas all edges in $E_{unknown} \subset V_{ext} \times V_{ext}$ are missing ($E = E_{known} \cup E_{unknown}$). For such a partial graph, our problem becomes to infer the edges among external vertices (edges in $E_{unknown}$).

IV. Graph Inference by Using Social Information

To model the social interactions between people in a conference we can use a weighted contact graph $G = (V, E)$ where V is the set of people who attend in the social meeting and E is the set of edges between them. There is an edge $e = (u, v)$ in G between u and v if they have contacted each other at least once. In the contact graph G, we assign a weight to each edge which shows either the total number of times that u and v have seen each other or the total time period that they have spent together during the social event. This weight can represent the strength of social relation between the corresponding nodes. We also assume that contact graph G is undirected because the social interaction involves both sides. For the first part of our analysis, we assume that the set of edges E is unknown while there is social information about nodes in V.

A. Real Data Description

In this paper, we use the human mobility traces collected from two different conferences [12]. The first dataset is collected during the Infocom 2005 conference where 41 participants attended. The second dataset contains mobility traces of 79 researchers attending in Infocom 2006 conference. Both experiments lasted for three days. In these experiments, contacts between participants were recorded by using iMote sensors. These Bluetooth sensors sampled a contact between two people when they were in close proximity of each other (e.g. < 10 meters). For Infocom 2005 data, there is not any social information about participants. However, in Infocom 2006, social profiles of people who participated in the experiment were collected. These social profiles include information about participants' (1) nationalities, (2) spoken languages, (3) current affiliations, (4) city and country of residence, (5) school, and (6) research interests. Unfortunately these social profiles are not complete since some people did not give their social profiles. However, we have a brief description of social profiles for 63 people. In this section, we limit our analysis only to these subset of nodes.

As mentioned in the related work, people who are socially close are more likely to be friends. The hypothesis that we want to test can be stated as below:

> *Hypothesis: (Social proximity) Individuals who have similar social profiles are more likely to contact each other than those who do not have similar social profiles.*

To test this hypothesis, we should first define the social similarity between nodes.

B. Jacard Social Similarity

Probably the most direct way to compute the similarity between social profiles of two people is to use the Jacard index [13]. As mentioned earlier, our social profiles contain information about 6 different social dimensions. We can denote each social dimension of every node as a set of features. For example, suppose node u speaks English and Spanish while node v speaks English and French. Let us denote English, Spanish, and French languages with numbers 1, 2, and 3 respectively. We can show the spoken languages of nodes u and v with two sets $\Gamma_u^2 = \{1, 2\}$ and $\Gamma_v^2 = \{1, 3\}$ independently. By using Jacard index, we can compute the similarity between two nodes in each dimension as follow [3]:

$$\sigma_{jacard}^i(u, v) = \frac{\left|\Gamma_u^i \bigcap \Gamma_v^i\right|}{\left|\Gamma_u^i \bigcup \Gamma_v^i\right|} \tag{1}$$

, where Γ_u^i is the feature set of node u for social dimension i and $\left|\Gamma_u^i\right|$ is its cardinality. Considering the 6 different social dimensions, we can compute the total similarity between two nodes by calculating the average over all 6 dimensions as below:

$$sim_{jac}(u, v) = \sum_{i=1}^{d} \frac{\sigma_{jacard}^i(u, v)}{d} \tag{2}$$

, where $sim_{jac}(u, v)$ is the total social similarity between two nodes u and v by using the Jacard index and d is the number of social dimensions (e.g. for Infocom 2006 data: $d = 6$).

C. Social Foci Distance

We can also think about each dimension of social profiles as a social focus. Sharing any social focus between two nodes *u* and *v* can increase the likelihood that they contact each other in the conference. Adapting the focus distance proposed by Kleinberg to our work, we can define another social distance

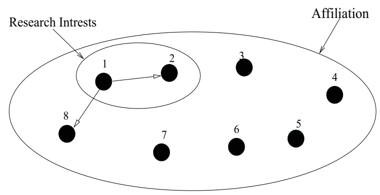


Fig. 1. Nodes Belonging to Multiple Foci

TABLE I
CORRELATIONS

$\rho_{nc,jac}$	$\rho_{cd,jac}$	$\rho_{nc,foc}$	$\rho_{cd,foc}$
0.10	0.30	0.17	0.32

between two nodes *u* and *v* as the size of the smallest social focus that includes both of them [14]. We call this new distance as the social foci distance which is denoted by $d_f(.,.)$:

$$d_{foc}(u,v) = min\,|\{F|u,v \in F\}| \quad (3)$$

, where F is the focus set to which both *u* and *v* belong. Note that there is a super group which contains all nodes. Using the foci distance, we can formulate the foci similarity between two nodes u and v by Equation 4:

$$sim_{foc}(u,v) = \frac{1}{d_{foc}(u,v)} \quad (4)$$

According to Equation 4, the social similarity between two nodes has an inverse relation with their foci distance. To highlight the main difference between social foci and Jacard similarities, we have shown a set of 8 nodes in Figure 1. Suppose all of these 8 nodes have the same affiliation, nationality, school, language, and city. Furthermore, suppose all nodes share the same research interests except 1 and 2 which have a similar set of interests that is different from the rest of nodes (nodes 3 to 8). By using the Jacard index, we can show that $sim_{jac}(1,2) = 1.0$ and $sim_{jac}(1,3) = 0.83$. Thus, based on the Jacard index node 1 is almost at the same distance from both nodes 2 and 3. However, the foci distance gives us $sim_{foc}(1,2) = 0.5$ and $sim_{foc} = 0.125$. As we can see the foci distance shows a closer distance between $(1,2)$ than $(1,3)$ because both nodes 1 and 2 share the same interests. Therefore, the foci distance can separate those nodes which are socially close from other nodes more significantly.

D. Max Social Similarity

Watts et al. have used a set of social characteristics to identify nodes in a social network [15]. They have defined the social distance between two nodes as the minimum distance over all dimensions. Combining their social distance with Equation 1, we can introduce a new social similarity as below:

$$sim_{max}(u,v) = \max_i \sigma^i_{jacard}(u,v) \quad (5)$$

Here, we assume all dimensions have the same weights, and if two nodes are similar in any dimension, they are assumed to be socially close to each other. As the first step, we calculate the Pearson correlation coefficient between social similarity of all pairs of nodes and their total number of contacts, and total contact durations for Infocom 2006. Let nc, cd, jac, and foc variables denote the number of contacts, contact duration, jacard similarity, and foci similarity respectively. We can compute the correlation coefficients between each possible pair of these variables as shown in Table I. The obtained correlation coefficients show a positive dependency between the contact pattern for a pair of nodes and their social similarity which in turn supports our hypothesis.

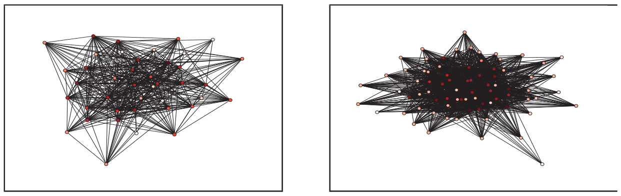
E. Prediction Based on Social Similarity

As we have seen in the previous section, there is a dependency between the pattern of interactions among nodes and their social similarities. We can interpret the total number of contacts or the total contact duration between two nodes as their level of interaction. To construct the contact graph G, we add an edge between two nodes if they have seen each other at least once. We assign to each edge a weight which shows the total contact duration in the conference. People can randomly contact each other during the conference; therefore, if we include all contacts in the contact graph, we will have almost a complete graph as shown in Figure 2.

In this section, we assume that for the graph $G = (V,E)$ the edge set E is unknown; however, we have the information about social profiles of nodes in V. By having social profiles, we can calculate the social similarities between all possible pairs in V by applying the three described social similarities. We can store the similarity results in separate lists called L_{sim}. Next, we sort each similarity list in a decreasing order. These sorted lists are our predictor results which is going to be used to infer the edges in E.

For inferring missing links, we extract from the original L_{sim} those pairs which are similar at least as much as a prespecified threshold thr_{sim}. We store these pairs in a separate list called L^{temp}_{sim}. Let us assume that there are l_{sim} pairs in the L^{temp}_{sim}. Meanwhile, we store all observed edges in G according to our real data in another list called L_{obs}. We similarly sort L_{obs} list in a decreasing order based on the total contact duration for each pair. We extract the first l_{sim} pairs from L_{obs} and store them in a separate list called L^{temp}_{obs}. Finally, we count the number of matched pairs between L^{temp}_{sim} and L^{temp}_{obs} lists. This number shows the percentage of correct predictions for thr_{sim}. Our intuition is that larger values of thr_{sim} should predict the strong social interactions more significantly than a random predictor.

To evaluate the performance of our predictor, we use different threshold values for similarities. For each thr_{sim}, we compute the percentage of correct predictions versus the fraction of all node pairs such as (u,v) for which we have $sim(u,v) >= thr_{sim}$. Figure 3 shows the results of our three social similarities. As we can see in Figure 3, the performances of three social similarities are statistically more significant than random predictor. Note that a random predictor guesses the missing links completely at random. Figure 3 also shows that



(a) Infocom 2005 (b) Infocom 2006

Fig. 2. Contact graphs with different threshold values

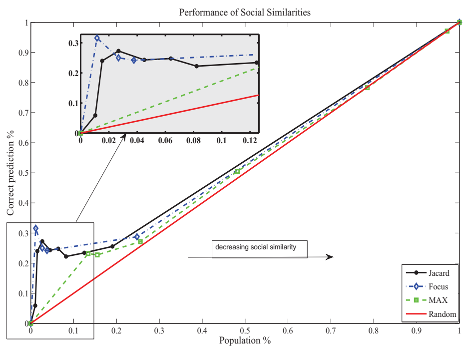


Fig. 3. Performance of social profiles (Infocom06)

for large values of thr_{sim}, the percentage of match between our predictors and the observed edge set is higher than a random predictor. These results clearly support our social proximity hypothesis. Looking more closely at Figure 3, we can see that the focus distance performs better than the Jacard and MAX especially for large values of thr_{sim}. Note that in Figure 3, as we decrease the similarity threshold (using a larger percentage of the population), the effect of social profiles diminishes and our results become more similar to random predictor. This is because there are only a few number of pairs which are very close to each other. For instance, the total number of pairs which are at least 10% similar is less than 20% of all possible pairs.

For our three proposed social similarities, we have assumed that all social dimensions have similar weights. However, this assumption may not hold in the real world. In a conference setting, for a given pair of nodes having the same affiliation or sharing similar interests may have higher weight on the social similarity than sharing the same country of residence. Having human mobility data for a set of people with a detailed version of their social profiles will allow us to study the effect of different social dimensions on human mobility patterns. We have plan to pursue this task as the future work.

V. Graph Inference by Using Contact Graph Properties

In the previous section we have shown that nodes do not meet each others uniformly at random whereas there is tendency for nodes to contact those ones which are socially similar to them. In this section we want to answer our second question. For a contact graph $G = (V, E)$, we randomly assign labels to vertices either as internal or external nodes. We also assume that we only have information about edges in $E_{known} \subset V_{int} \times (V_{int} \cup V_{ext})$. This means that all edges among external nodes are unknown. Our task is to infer the edges among external nodes by using the known part of the graph G. Moreover, we assume that there is not any information about social profiles of nodes.

We can formulate this problem as a graph inference problem where only part of the contact graph is known. Our main intuition is to use the underlying properties of the contact graph to infer the missing edges among external nodes. We assume that for each edge in E_{known} we know the total time period that the end nodes have spent together as well as the total number of times that they have contacted each other during the conference.

A. Number of Common Neighbors (NCN)

We can assume that our contact graph G supports the triadic closure property which is common in social networks [1]. For instance, if node u meets node v and w often, then v and w are more likely to contact each other too. Based on this fact, we count the number of common neighbors (NCN) between external nodes to compute their similarity [9]. Our intuition is that if two nodes have a large NCN, they will be more likely to interact together. Note that since we do not know the edges among external nodes, they can only have neighbors in V_{int}. We generate a list for all possible pairs of external nodes with the number of common neighbors for each pair. Finally, we sort the NCN list called L_{NCN} in a decreasing order. This sorted list is the output of NCN predictor for inferring missing edges.

B. Shortest Path (SP)

It is known that social networks have low diameters [1]. By using the inverse of total time that two nodes have spent together, we can assign a weight to each known edge in E_{known}. Thus, the smaller value for an edge weight means that the corresponding end nodes met each other frequently during the conference. For finding the similarity among external nodes, we compute the total weight of the edges in the shortest paths among them. This gives us a list of external nodes pairs with the total weight of the shortest path between them. Let us denote this list with L_{SP}. We sort this list in an increasing order. The L_{SP} list can be used for inferring the missing edges.

C. Random Walk (RW)

Motivated by Page Rank algorithm used for finding the importance of web pages [16], we can propose another method which explores a larger subset of paths between two nodes. Let us assume that we want to find the similarity between two external nodes u and v. We can consider a random walk which starts from u and moves to a neighbor of the current node with probability of α at each step. The random walk returns to the starting node u with probability of $1 - \alpha$. This guarantees that we do not explore those parts of the graph which are far from u and v. At each step, we choose a random neighbor

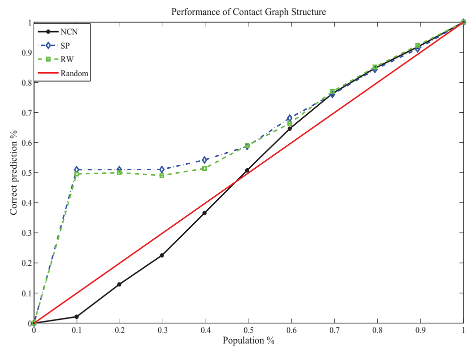


Fig. 4. Performance of contact graph structure (Infocom05)

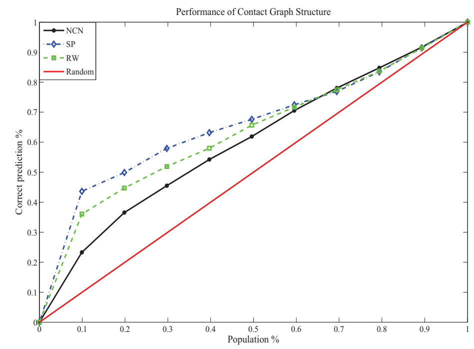


Fig. 5. Performance of contact graph structure (Infocom06)

with a probability which is proportional to the weight of the corresponding edge. For this method, we use the total contact duration which two node have spent together as the weight of the edge between them. Thus, at each step, the random walk is more likely to move to a node which is similar to the current node. To find the similarity between u and v, we compute the stationary probability of v for the described random walk. Finally, we can generate a sorted list of external nodes by using the computed stationary probabilities (in a decreasing order). Let us denote this sorted list with L_{RW}.

We simulate a partial graph by randomly choosing a subset of nodes as external ones and remove the edges among them. In our simulation, we randomly pick 50% of nodes as external nodes and remove all edges among them. We use both Infocom 2005 and 2006 data for our simulations. To evaluate the performance of our predictors, we follow the same steps as in Section IV. By using different percentage of the whole population, we compute the percentage of matches between the lists generated by the three described methods with the sorted version of observed edges among external nodes. Note that we sort the list of observed edges among external nodes based on their total contact duration time.

Figures 4 and 5 compare the performances of NCN, SP, and RW with a random predictor. Our results show that both SP and RW predict the missing edges statistically better than a random predictor. As we can see, the SP predictor outperforms others and the RW predictor has the second rank. These results show the importance of using the contact graph structure for inferring the missing links. This also proves that there is an underlying mechanism governing the formation of links in a contact graph that clearly cannot be explained by a purely random process. Figure 4 also shows that NCN does not work efficiently for Infocom 2005 data. To explain this, we have to consider the fact that our dataset for Infocom 2005 contains a small subset of nodes where most of nodes have almost similar degrees even before we remove edges among external nodes. Thus, using number of common neighbors cannot provide useful information for inferring missing edges.

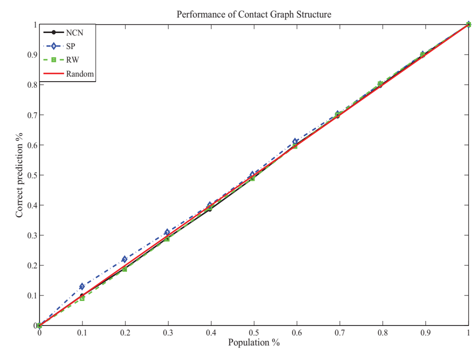


Fig. 6. Performance of contact graph structure (synthetic mobility with $q = 1.0$)

VI. Contact Graph Model

In the previous sections, we have seen that the contact graphs have a structure which is clearly different from a random contact graph. Looking more closely at Figures 3, 4, and 5, we can recognize a local maximum in the beginning of the graphs. In this region where chosen nodes are very similar, we can see that the maximum difference between our predictors and the random predictor happens. In [1], Watts and Strogatz have proposed a graph structure to model small-world networks. Their graph models have properties of both random graphs and structured graphs (e.g. a lattice). We can use the same idea to model the structure of a contact graph. Let us assume that our network has N nodes. We can generate a social graph for these N nodes by following the same model as Watts where we place N nodes on a one-dimensional lattice. We connect every node to its k nearest neighbors with respect to the lattice structure. Let us call these links short range links. Then, we rewire each short range link with a probability p to add random graph properties to our social graph. We call these links long range links. Both short and long range links can be considered as those nodes which are similar to each other. The resulting social graph can be used as the underlying structure by which we generate our contacts.

To generate contacts, we randomly pick a node u from all N

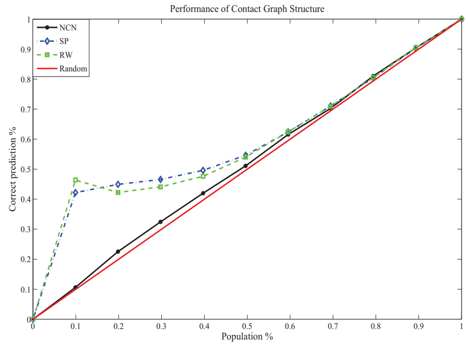


Fig. 7. Performance of contact graph structure (synthetic mobility with $q = 0.2$, $p = 0.2$, and $k = 5$)

possible nodes in our generated social graph. Next, we choose the node v which is going to meet u with a probability q uniformly at random from all N nodes. This models the fact that nodes may contact each other at random. With probability of $1 - q$, we choose the peer node for u from its neighbor set in the social graph structure. This models the fact that similar nodes are more likely to see each other. Following these steps, we can generate a contact graph by having k, p, and q parameters. Intuitively speaking, larger values for q makes the resulting contact graph to be similar to a random graph whereas small values for q appreciate the homophily process in the generated contact graph.

For simulating a contact graph, we assume that our network has $N = 100$ nodes. For the first run, we simulate a contact graph with $q = 1.0$ in which nodes contact each other uniformly at random. We pick 50% of nodes as the external nodes. Figure 6 shows the results of NCN, SP, RW predictors. As we can see, none of predictors performs better than a random predictor because there is not any structure in the simulated contact graph. For our second test, we choose $k = 5$, $p = 0.2$, and $q = 0.2$ to simulate another contact graph. The results of our three predictors are shown in Figure 7. Interestingly, in this case we can see similar patterns for SP and RW predictors as the ones which we have observed in the real data. Thus, if nodes are more likely to see the ones which are similar to them (e.g. their neighbors on the underlying social graph), we can explain the observed local maximums which appears in the beginning of our predictors graphs.

VII. Conclusion

In this paper, we have formulated the problem of social contact prediction as a graph inference problem. First, we have shown that we can predict the edges of a contact graph when we only have social profiles of people, but we do not have any information about the contact graph. Next, we studied the effectiveness of using the underlying properties of the contact graphs for contact prediction problem where only a part of the contact graph is known. In both settings, our prediction results are statistically more accurate than a random predictor. All of the proposed contact predictors are based on well known properties of social graphs. Finally, motivated by small-world networks model, we have shown that the performance of our predictors can be justified by considering the role that homophily process plays on the structures of contact graphs.

One interesting direction for future work is to study the performance of contact prediction by using a weighted version of social similarity which assigns different weights to distinct social dimensions. We also have plan to revalidate the results of our predictors by using other available human mobility datasets for other settings.

References

[1] D. J. Watts and S. H. Strogatz, “Collective dynamics of ’small-world’ networks.” *Nature*, vol. 393, no. 6684, pp. 440–442, June 1998.

[2] M. McPherson, L. S. Lovin, and J. M. Cook, “Birds of a feather: Homophily in social networks,” *Annual Review of Sociology*, vol. 27, no. 1, pp. 415–444, 2001. [Online]. Available: http://dx.doi.org/10.1146/annurev.soc.27.1.415

[3] K. Jahanbakhsh, G. C. Shoja, and V. King, “Social-greedy: a socially-based greedy routing algorithm for delay tolerant networks,” in *MobiOpp ’10: Proceedings of the Second International Workshop on Mobile Opportunistic Networking*. New York, NY, USA: ACM, 2010, pp. 159–162.

[4] N. Eagle, A. Pentland, and D. Lazer, “Inferring social network structure using mobile phone data,” *Proceedings of the National Academy of Sciences (PNAS)*, vol. 106(36), pp. 15 274–15 278, July 2009.

[5] A. Mtibaa, A. Chaintreau, J. LeBrun, E. Oliver, A.-K. Pietilainen, and C. Diot, “Are you moved by your social network application?” in *WOSP ’08: Proceedings of the first workshop on Online social networks*. New York, NY, USA: ACM, 2008, pp. 67–72.

[6] E. M. Daly and M. Haahr, “Social network analysis for routing in disconnected delay-tolerant manets,” in *MobiHoc ’07: Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing*. New York, NY, USA: ACM, 2007, pp. 32–40.

[7] P. Hui, J. Crowcroft, and E. Yoneki, “Bubble rap: social-based forwarding in delay tolerant networks,” in *MobiHoc ’08: Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing*. New York, NY, USA: ACM, 2008, pp. 241–250.

[8] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, “Impact of human mobility on the design of opportunistic forwarding algorithms,” in *Proceedings of the 25th IEEE International Conference on Computer Communications (INFOCOM)*, Barcelona, Spain, April 2006.

[9] D. Liben-Nowell and J. Kleinberg, “The link prediction problem for social networks,” in *CIKM ’03: Proceedings of the twelfth international conference on Information and knowledge management*. New York, NY, USA: ACM, 2003, pp. 556–559.

[10] D. S. Goldberg and F. P. Roth, “Assessing experimentally derived interactions in a small world,” *Proceedings of the National Academy of Sciences (PNAS)*, vol. 100, no. 8, pp. 4372–4376, February 2003.

[11] J.-P. Vert and Y. Yamanishi, “Supervised graph inference,” in *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 2005, pp. 1433–1440. [Online]. Available: http://eprints.pascal-network.org/archive/00001405/

[12] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau, “CRAWDAD trace for infocom2006,” http://crawdad.cs.dartmouth.edu, May 2009.

[13] P. Jaccard, “Étude comparative de la distribution florale dans une portion des alpes et des jura,” *Bulletin del la Société Vaudoise des Sciences Naturelles*, vol. 37, pp. 547–579, 1901.

[14] J. Kleinberg, “Small-world phenomena and the dynamics of information,” in *In Advances in Neural Information Processing Systems (NIPS) 14*. MIT Press, 2001, p. 2001.

[15] D. J. Watts, P. S. Dodds, and M. E. J. Newman, “Identity and search in social networks,” May 2002. [Online]. Available: http://arxiv.org/abs/cond-mat/0205383

[16] S. Brin and L. Page, “The anatomy of a large-scale hypertextual web search engine,” *Computer Networks and ISDN Systems*, vol. 30, no. 1–7, pp. 107–117, 1998. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.42.3243