

Research Article

Characterizing Pairwise Social Relationships Quantitatively: Interest-Oriented Mobility Modeling for Human Contacts in Delay Tolerant Networks

Jiaxu Chen, Yazhe Tang, Chengchen Hu, and Guijuan Wang

The Department of Computer Science & Technology, Xi'an Jiaotong University, No. 28 Xian Ning Road West, Xi'an, Shaanxi 710049, China

Correspondence should be addressed to Yazhe Tang; yztang@mail.xjtu.edu.cn

Received 19 April 2013; Accepted 20 September 2013

Academic Editor: Chih-Hao Lin

Copyright © 2013 Jiaxu Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Human mobility modeling has increasingly drawn the attention of researchers working on wireless mobile networks such as delay tolerant networks (DTNs) in the last few years. So far, a number of human mobility models have been proposed to reproduce people's social relationships, which strongly affect people's daily life movement behaviors. However, most of them are based on the granularity of community. This paper presents interest-oriented human contacts (IHC) mobility model, which can reproduce social relationships on a pairwise granularity. As well, IHC provides two methods to generate input parameters (interest vectors) based on the social interaction matrix of target scenarios. By comparing synthetic data generated by IHC with three different real traces, we validate our model as a good approximation for human mobility. Exhaustive experiments are also conducted to show that IHC can predict well the performance of routing protocols.

1. Introduction

Wireless portable devices (e.g., laptop, PDA, and cell phone) are often held by humans in a delay tolerant networking (DTN) scenario. Understanding human mobility and accordingly designing better routing protocols have drawn the attention of researchers working on DTNs. Using real trace to evaluate routing protocol performance does not allow enough flexibility to change the mobility settings in order to perform the analysis for a slightly different scenario [1]. Human mobility models, however, can depict real-life human mobility characteristics and can be used to obtain meaningful routing protocol performance results in simulations [2]. So far, a number of real-life experiments have been conducted to observe and summarize human mobility characteristics, including individual (e.g., spatio/temporal preferences), encounter-based (e.g., inter-contact time and contact duration), and social (e.g., group, community) metrics [3]. Although current human mobility models are good at reproducing individual and encounter-based metrics, their strategies of generating social metrics still need further

exploiting. For example, most existing models are able to reproduce the inherent social interactions but on a rather coarse granularity of community. Thus, the intercommunity nodes' social relationships and the social relationships of nodes that do not belong to any community are ignored. This ignorance leads to a considerable deviation of social relationships between the synthetic scenario and the real-life scenario. In addition, it will be unable to utilize the intercommunity nodes' social relationships to assist data dissemination when those mobility models are used to evaluate routing protocols.

This paper will focus on the granularity issue and propose a new mobility model that takes social relationships between each pair of nodes into account. We call this model interest-oriented human contacts (IHC) mobility model. The design principle is based on two intuitionistic and empirical observations. First, people always visit several spots periodically, and visit the more interesting spots at a higher probability. Second, the available contacts between a pair of people are always detected when the two people are relatively static. By implementing these two rules rationally, IHC is able to

reproduce social behavior among people, and characterize pairwise social relationships quantitatively. Our model is verified with three real traces on the metrics of inter-contact time, contact duration and meaningful pair-wise social relationships. Evaluation results show that the synthetic data IHC generates match the three real traces very well. In brief, we make the following contributions.

- (i) We propose IHC, which is the first human mobility model dedicated to node interest to the best of our knowledge. We not only take node interest vectors as the exclusive social relationship input but also make a similarity analysis of node interest.
- (ii) We design two methods to generate input parameters (interest vectors) based on the social interaction matrix of target scenarios. By using any of the two methods, IHC is able to reproduce social relationships on a pair-wise granularity. That is, IHC can characterize pair-wise social relationships quantitatively. As far as we know, IHC is the first such model.
- (iii) We show that IHC can generate synthetic data matching very well the three real traces on the statistical properties of inter-contact time complementary cumulative distribution function (CCDF), contact duration CCDF, and meaningful pairwise social relationships, which undoubtedly leads to a well-matched social community structure.
- (iv) We build cases to use the model by comparing the performance of three forwarding protocols working on real traces and synthetic traces generated by IHC. Simulation results indicate that IHC can predict well the performance of the forwarding protocols.

The rest of the paper is organized as follows. Section 2 introduces current research on the field of human mobility modeling. In Section 3, we describe IHC in detail, including the human mobility patterns which inspire our work, the model itself, a similarity analysis, and two methods to generate input parameters (interest vectors) based on the social interaction matrix of target scenarios. We validate our model by comparing synthetic data traces with three different real traces in Section 4. The comparison shows a good matching between IHC-generated and real traces. Section 5 shows that IHC can predict well the performance of routing protocols by comparing the performance of three DTN protocols with IHC-generated and real traces. Section 6 summarizes our conclusions and describes future work.

2. Related Work

Up to now a number of research articles on human mobility modeling are published. Some of them are dedicated to explore exhaustive human mobility itself such as [1, 4]; others are designed for assisting accurate performance evaluation of DTN forwarding protocols in simulations, such as [2, 5]. Our work belongs to the latter. To provide a good understanding of mobility model framework as well as the relationships among human mobility characteristics, mobility models, and routing protocols in a DTN, we draw Figure 1.

In Figure 1, each entity represents a human mobility characteristic: a rectangular entity represents an individual metric; an elliptical entity represents an encounter-based metric, and a circular entity represents a social metric. A bold line transversely divides Figure 1 into two portions. The entities above the bold line are taken as input parameters of mobility models, and the entities below the bold line are considered as output metrics. On the other hand, two erect dashed lines divided Figure 1 into three portions, namely individual metrics module, encounter-based metrics module and social metrics module, respectively (on order of left-to-right). These three modules are demarcated based on the attributes of the metrics, no matter where they are, above or below the bold line. Based on the extensive study of the existing work, we say that the most significant metrics that lead to human-like “inter-contact time” and “contact duration” distributions are the “pause time” and “location preference and periodic re-appearance” (two individual metrics). Therefore, individual metrics that do not affect the two encounter-based metrics are classified into individual metrics module and denoted by “...” in Figure 1. As a work for a model designed for assisting accurate performance evaluation of forwarding protocols in simulations, this paper is oblivious to such individual metrics.

Human mobility is driven by social relationships. Figure 1 indicates this intuition by exhibiting the social metrics module. The “inherent social interaction” in the input portion works together with several individual metrics, such as “pause time” and “location preference and periodic re-appearance” to generate “social metrics” in the output portion. However, early human mobility models (such as [6, 7]) only have the individual metrics module and encounter-based metrics module. As far as we know, CMM [8] is the first model that takes social relationships into consideration. Models after CMM always include the social metrics module, with different social based representations such as interaction matrix [8, 9], communities [5, 8, 9], home-points distribution [5], overlapping communities [10], and centrality [10].

Due to the lack of end-to-end paths in DTNs, routing protocols have to utilize nodes’ each chance of contact to forward packets (see Figure 1, “encounter-based protocol-dependent metrics”). The nodes’ chances of contact are unstable and hard to hold. However, in DTNs, nodes’ social relationships are steady and reliable. Hence, social-aware routing protocols utilize some social metrics, such as node community [11], node centrality [12], and node interest [13] to help with routing decisions. Among these metrics, the community is an important conception in social network theory. Therefore, most existing models reproduce the “inherent social interactions” on the granularity of (and are limited to) the community. Our model breaks this limitation and reproduces social relationships on a pair-wise granularity.

We argue that a good human mobility model should, first, be simple, second, have the ability of reproducing crucial human mobility metrics (both encounter-based and social), and third, predict well the performance of forwarding protocols on real DTNs.

TLW [7] generates movement traces using a model which is similar to Levy Walks, except that the flight lengths and

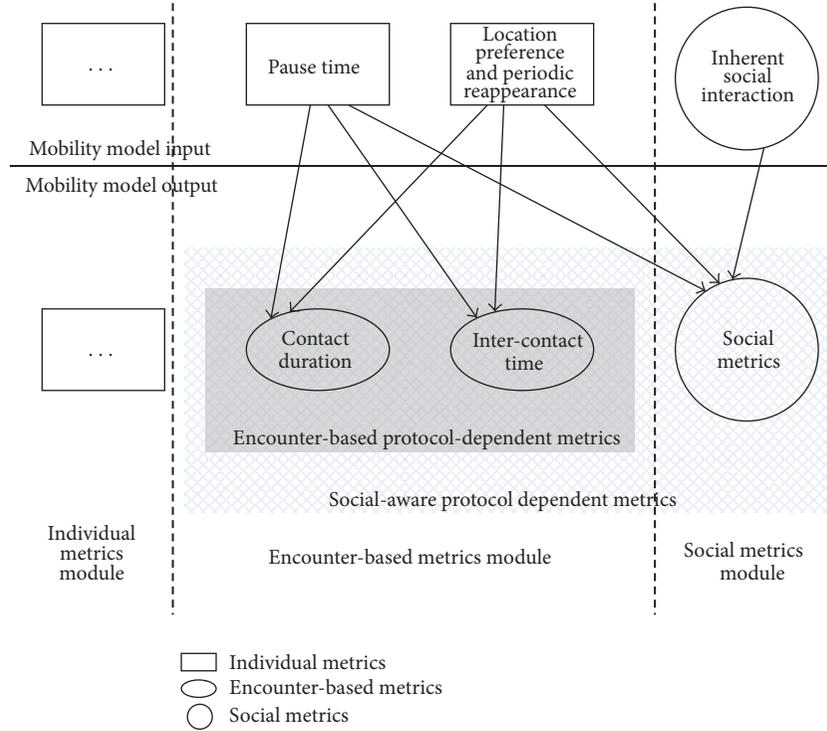


FIGURE 1: Mobility model framework.

the pause times have power-law distributions. TLW generates the inter-contact time similar to real traces. However, as each node moves independently, TLW does not capture the nodes' social behaviors. As a subsequent work to TLW, SLAW [14] correlates the walks of different nodes based on TLW, and takes heterogeneously bounded mobility areas and fractal waypoints into consideration. SLAW generates the inter-contact time and flight lengths distribution similar to real movements. However, there is no clear explanation about how to characterize nodes' social relationships. SMOOTH [15] is a mobility model with similar principle and performance to SLAW.

As the first mobility model that utilizes social network theory, CMM [8] is not concerned about individual human mobility characteristics. Besides, CMM has been proved defective: specifically, in the majority of configurations, all users collapse into a single location, this practically overthrows the initial setting of the system [9]. To erase this defect, Boldrini and Passarella [9] improve CMM by adding the following two individual human mobility characteristics: people tend to visit just a few locations, where they spend the majority of their time (the same meaning as the "location preference and periodic reappearance" in Figure 1), and people prefer shorter paths to longer ones. They validate their synthetic data with real traces and show a good matching of inter-contact time and flight lengths.

HHW [10] concerns heterogeneous human popularity. The model's input characteristics are "location preferential" (individual) and "overlapping communities" (social). Its

output metrics "inter-contact time" (encounter based) and "centrality" (social) also have a good matching with real traces.

The intuition that inspires SWIM [5] is as follows: people go more often to places not very far from their homes and where they can meet a lot of other people. This feature is actually the same as "location preference and periodic reappearance" in Figure 1, with a clear location preference weight decided by the location's distance and popularity. SWIM uses a bounded power-law-distributed "pause time." The authors validate their synthetic data with real traces and show that SWIM has a good matching of inter-contact time, contact duration, number of contacts and community structure.

Maiti et al. [1] collect numbers of human mobility patterns and explore dependencies between them. A framework is also proposed to reproduce these human mobility patterns in the model. Thakur and Helmy [3] propose a framework for the mobility model analyzing. The model generated by the framework is validated with real traces on both encounter-based metrics and social metrics. STEPS [16] is designed based on a Markov chain modeling method. The human mobility pattern it depends on is also "location preference and periodic reappearance." A good matching of both encounter-based and social metrics with real traces is shown in the literature. SAGA [17] is dedicated to the geographic diversity of the region of interest, which is different from all of the above-mentioned models. As a result, SAGA is validated with real traces on different metrics as well.

IHC also relies on “location preference and periodic re-appearance” but with a different modeling method from existing models. IHC is very simple to implement and able to reproduce statistical human mobility properties, such as inter-contact time, contact duration, and accurate meaningful social relationships on the granularity of pairwise. It also can predict well the protocol performance on real DTNs. Based on the analysis of human mobility patterns and the statistical data of real traces, IHC chooses to ignore the spatial human mobility characteristics which are non-correlative to human contacts. The inherent social interaction in the IHC input parameters is node interest, such that node interest analyzing will be more facile in an IHC simulation environment.

3. Interest-Oriented Human Contacts Mobility Model

In this section, we propose interest-oriented human contacts mobility model. We introduce the human mobility characteristics that inspire our work and describe the model in detail, followed with a similarity analysis of node interest based on IHC and two methods to generate input parameters based on the social interaction matrix of target scenarios.

3.1. Design Principle. We observe that people prefer to visit a few locations and spend plenty of time staying at such locations. In other words, a few locations bring more attractions to people. In this paper, such locations are called hotspots. In real-life scenarios, hotspots are the locations related to people’s interests, for example, the locations where we work, study, have meals or do sports. Usually, people spend a lot of time at some of the hotspots (such as work or study) and less time at others (such as having lunch or doing sports).

As people spend the majority of their time at hotspots, they spend little time on the journey. Besides, the vehicle people choose to accomplish the journey is strongly dependent on the distance of the next destination. For example, people prefer to walk to an adjacent destination, ride to a farther one and drive to a much farther one.

Further, we observe that an available contact (an available contact is considered as a contact when one person meets the other, their wireless devices can detect each other and accordingly forward messages) between a pair of people usually happens when the two people are relatively static. Suppose that, in a college, two students with active wireless devices are walking past each other. Due to the transmission range of wireless devices and relative speed of the two students, the actual contact duration is too short for wireless devices to detect each other. Only when the two students appear at the same spot and stay for a while, such as having dinner in the cafeteria or reading in the library, their wireless devices have enough time to detect each other and then forward messages.

Now that an available contact occurs only if the two people are relatively static, the geographical position where the contact happens is not significant anymore if we do not take the case of the synchronized motion (e.g., the two

people walk along with each other) of the two people into consideration (because synchronized motion involves very specific social relationships, e.g., a colleague relationship and a particular schedule, which is hard to hold for a mobility model). Since temporary passing-bys (non-available contacts) from one spot to another spot are negligible, the specific geographical position of these two spots and distance between them make no sense. Under these circumstances, the factors that impact the contact metrics of the two people (e.g., A and B) in a period of time are as follows: when A arrives at this spot; how long A will stay for (when A will leave this spot); when B arrives at this spot, how long B will stay for. If we extend this period of time to the overall runtime of the social network that A and B belong to, then the encounter based metrics of A and B (inter-contact time and contact duration) depend on the probability that A and B visit this spot and how long will A and B stay at this spot.

Consequently, theoretically, by rationally setting numbers of hotspots, the probability of visiting each hotspot, and the pause time after each arrival, it is quite promising to generate good matching statistical characteristics and social relationships with real traces.

3.2. Detailed Model. As an interest-oriented mobility model, IHC builds an environment where node interests are manifested as hotspots. In general, one interest stands for one hotspot and vice versa.

Like the most existing mobility models, in IHC, the mobility of a specific node is composed of a set of movement epochs throughout the simulation time. At the beginning of a movement epoch, the node chooses a destination and moves towards it at a speed. After arrival, the node stays at the destination spot for a time period which is known as pause time. Till the end of the pause time, the node begins to choose a new destination and start the next movement epoch. Contacts occur when one node is within the transmission range of another node. However, in IHC, only available contacts are under consideration; that is, a contact is recorded if and only if two nodes are staying at the same hotspot simultaneously. In addition, nodes spend the majority of their time at hotspots. That is, the destination of each epoch can only be chosen from hotspots. Therefore, the contact metrics of a pair of nodes in IHC are determined by their probability to visit the same hotspot and the pause time they stay there. In such a case, the network area, the node transmission range and the positions of hotspots make little sense to contact metrics in IHC, and we choose to omit them. However, in order to ensure that the node interests are mutually independent, the distance between any two hotspots (although IHC does not care about their specific geographical positions) needs to be larger than the node transmission range so that nodes visiting different hotspots will not meet each other.

Since a node has different preferences to different interests, it visits the corresponding hotspots at different probabilities. Suppose that there are n hotspots corresponding to n interests in the network area. These n node interests compose an n -dimensional interest space. Each node has an

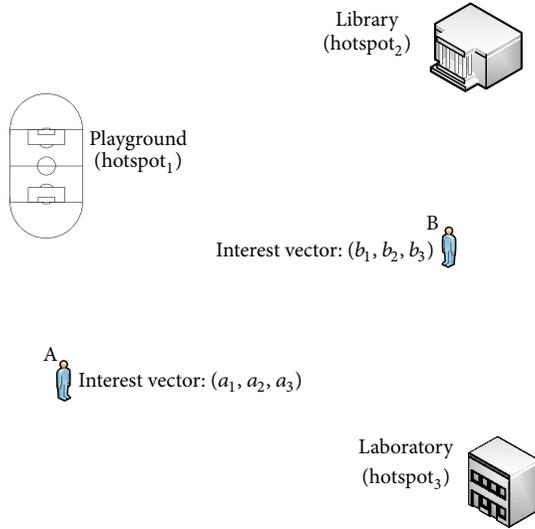


FIGURE 2: An IHC scenario containing three hotspots and two nodes.

n -dimensional vector, corresponding to a point in the interest space. The interest vector of *nodex* (x_1, x_2, \dots, x_n) means *nodex* visits hotspot i at probability x_i ($x_i \geq 0$), $i \in [1, n]$; then $\sum_{i=1}^n x_i = 1$.

Figure 2 maps a network area containing three hotspots (i.e., playground, library, and laboratory) and two nodes (A and B) who determine each movement epoch following IHC. Initially, each node, for example, node A, can be anywhere in the network area. Node A's first epoch soon begins: it chooses its destination from all hotspots according to its interest vector. In Figure 2, node A chooses to visit playground (hotspot₁), library (hotspot₂), and laboratory (hotspot₃) at probabilities a_1 , a_2 , and a_3 , respectively, such that $a_1 + a_2 + a_3 = 1$. Once the destination has been chosen, for example, the library (hotspot₂), node A starts moving straightly towards it with a constant time ft . A constant flight time actually indicates that the speed is proportional to the flight length (distance between the starting point and the destination) in IHC. This proportional relation is based on the observation that in real-life scenarios people spend little time on the journey by choosing different vehicles for different distances of destinations. After reaching the destination, the pause time will be determined by a variable whose probability density function (PDF) $pt()$ obeys a bounded power-law distribution as in [5]. Now node A is reading in the library and will stay for a time duration (pause time). Note that node B has also been accomplishing its movement epochs. If it is staying at the library coincidentally, both node A and node B will be able to detect this contact until one of them runs out of the pause time and begins the next epoch. Later, the other node will leave the library as well for the next epoch. Both A and B keep this kind of movements till the end of the simulation time.

It can be seen that IHC gives a clear expression on node interest and ignores specific geographical information. Such treatments lead to a much more convenient tuning of node

interest parameters. Other models do not provide a direct node interest tuning. For example, in SWIM, the probability that a node visits a spot depends on not only the spot's popularity but also the distance between the spot and the node's home. Although such settings make the preference of each node to each interest (corresponding to a spot) self-controlled, it is not easy to change the probability that a node visits a spot at will. Therefore, by ignoring specific geographical information, IHC replaces the probability of visiting a spot influenced by popularity and distance in SWIM with a single interest value. In this way, IHC gets rid of the inconvenience of altering a spot's visiting probability influenced by a home's position. In addition, IHC keeps all temporal metrics on contact and ignores the information of "the geographical position of contact," which not only needs complicated settings but also lacks corresponding information in real traces and does not affect the performance of forwarding protocols as well. In IHC, the specific geographical position of each hotspot has no influence on either contact metrics or forwarding protocols' performance as long as the distance between any two hotspots exceeds the node transmission range. As a conclusion, Table 1 summarizes all parameters and their meanings in IHC.

3.3. Similarity Analysis. Thakur et al. [18] demonstrate that people with similar behavioral principle tie together, which means that user-location coupling can be used to identify similarity patterns in mobile users. They make similarity analysis for several mobility models and show that many mobility models do not explicitly capture similarity and result in homogeneous users that are all similar to each other. Their similarity analysis is based on spatiotemporal preferences, preferential attachment to locations, and the frequency and duration of visiting these locations, which are actually the first-hand design principle of IHC. Therefore, IHC is suitable for similarity analysis inherently.

Mei et al. [13] try to utilize the cosine similarity of node interest profile to assist data forwarding in social-aware routing protocols because they believe that similar node interest profiles lead to close social interactions. However, as we have mentioned above, the mobility model they use, SWIM, cannot be used to measure node interest either accurately or conveniently. Additionally, whether cosine similarity of node interest profile can represent people's social interactions accurately is still unclear, while for IHC, node interest is taken as input parameters, thus making similarity analysis so natural that we may hopefully get meaningful conclusions.

It is generally believed that a large contact duration represents a close relationship between nodes, so social relationships are always denoted simply by contact durations [8, 12]. We also use this denotation in this paper.

Intuitively, in IHC, the social relationship between two nodes ought to be related to the interests shared by the two nodes, as only their common interests result in the two nodes' meeting at the corresponding hotspots. Based on this intuition, we conduct extensive simulations to observe what the relation between the common interests and contact durations of the two nodes is.

TABLE 1: IHC parameters.

Parameter	Meaning
num	The amount of nodes
st	Simulation time, measured in seconds
ft	Flight time, measured in seconds
$pt()$	The PDF of pause time which is measured in seconds
n	The amount of node interests
$H_i(x, y)$	The coordinate of hotspot H_i corresponding to interest $i, i \in [1, n]$
$X_j = (x_1, x_2, \dots, x_n)$	The interest vector of node $j, j \in [1, num], \sum_{i=1}^n x_i = 1$

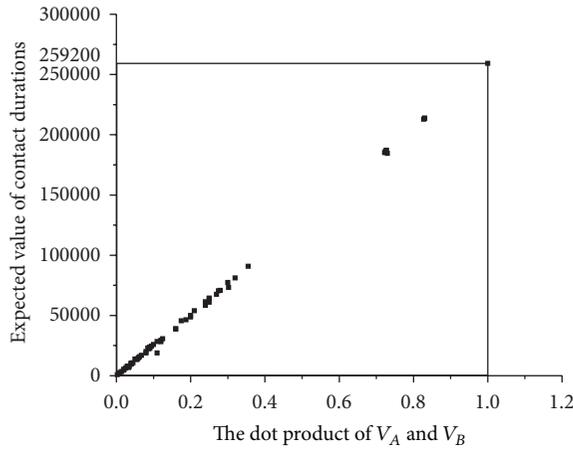


FIGURE 3: Proportional relation between a node interest metric and a social relationship metric.

In our simulations, there are only two nodes, namely, A and B, in the network area. Without specific input social interaction, pair-wise contacts are mutually independent. Thus, multiple nodes do not bring new insights. Each node has and only has 4 interests, such that node A's interest vector $V_A = (A_1, A_2, A_3, A_4)$, $A_1 + A_2 + A_3 + A_4 = 1$ and node B's interest vector $V_B = (B_1, B_2, B_3, B_4)$, $B_1 + B_2 + B_3 + B_4 = 1$. Let c_i be the amount of A and B's common interests, with value of 1, 2, 3, 4, respectively, and set different values of V_A and V_B , such as (0.25, 0.25, 0.25, 0.25), (0.3, 0.3, 0.3, 0.1), (0.1, 0.2, 0.3, 0.4), or (0.05, 0.05, 0.05, 0.85). Note that even for two specific vectors, different common interests should be assigned in each simulation. The simulation time is set as three days, that is, 259200 seconds. To get the expected value, for each scenario, we average the results of contact duration over 10000 runs using different seeds.

Excitedly, we find that there is specific relation between "the dot product of V_A and V_B " and "the expected value of A and B's contact durations," and we show the results in Figure 3. Each black point in Figure 3 represents a simulation scenario. The x -axis shows the dot product of V_A and V_B ($V_A \cdot V_B$) and the y -axis indicates the expected value of contact duration averaged over 10000 runs. The maximum of $V_A \cdot V_B$ is 1 when both nodes have only one, and the same interest. Corresponding to this specific scenario, the two nodes stay

forever at the same hotspot. Thus, their contact duration is the simulation time, 259200 s.

Figure 3 shows a proportional relation between "the dot product of V_A and V_B " (x -axis) and "the expected value of A and B's contact durations" (y -axis). The dot product of V_A and V_B is a metric derivative from A and B's interests and the expected value of A and B's contact durations is a metric, which can represent the social relationship between A and B. Figure 3 reveals a promising feature of IHC. That is, IHC may have the ability of accurately reproducing a specific contact duration matrix, which is always regarded as a social interaction map. IHC can generate a specific expected value of contact duration accurately by setting appropriate values to node interest vectors. Note that the y -axis in Figure 3 only shows the expected value of contact durations. As a complement, the distributions of the contact durations for different expected values are shown in Figure 4.

Figure 4 is graphed to assist understanding what distribution the contact durations obey for one expected value in Figure 3. Figure 4 is composed of eight subfigures. The expected value of contact duration in each subfigure is denoted by Exp. We choose Exp for eight scales to show in Figure 4, namely, 2500, 5000, 7500, 10000, 15000, 20000, 50000, and finally a very large one, 180000. As we can see in Figures 4(a)–4(d), for a small Exp (no larger than 10000), the distributions are far from the Gaussian Distribution, such that the expected value shows a considerable deviation from a randomly chosen value. In such a case, maybe IHC cannot reproduce a small contact duration accurately by tuning node interest vectors. Fortunately, a small contact duration makes nearly no sense in social network analysis. When Exp becomes larger, for example, in Figures 4(e)–4(h), the distributions look like a Gaussian Distribution, such that it will be more accurate by representing contact durations with the expected value.

3.4. Interest Vectors Generator. As shown in Table 1, we take node interest vectors as the exclusive social relationship input of IHC. The corresponding parameters, namely, n , $H_i(x, y)$, and X_j , can be derived based on the conclusion in Section 3.3. That is, the expected value of A and B's contact durations is proportional to $V_A \cdot V_B$. Consider an extreme case: when two nodes both have only one interest and their interest is the same as follows their dot production of interest vector is 1, and their contact duration is exactly the simulation time st

(see Figure 2) such that the coefficient of proportionality is $1/st$.

Suppose the num nodes are $node1, node2, \dots, nodenum$, and their interest vectors are X_1, X_2, \dots, X_{num} , respectively. Denote the sum of elements of vector X by $sum(X)$. Denote the contact duration matrix of the scenario which we want to reproduce by D , such that D is a $num \times num$ matrix. Then the contact duration between $nodei$ and $nodej$ is D_{ij} . The following equation set holds:

$$\begin{aligned} sum(X_i) &= 1, \quad i \in [1, num]; \\ X_i \cdot X_j &= \frac{D_{ij}}{st}, \quad i, j \in [1, num], \quad i \neq j. \end{aligned} \quad (1)$$

The equation set has $n \times num$ variables, $(num \times (num - 1)/2 + num)$ equations. Obviously a properly selected n can make this equation set have solutions. Approximate solutions can be derived with the Levenberg Marquardt algorithm. The solutions include the parameter settings of n and X_j . $H_i(x, y)$ can be anywhere in the network area as long as the distance between any two hotspots is larger than r .

However, the above method of choosing amount of interests and nodes' interest vectors, named as method-1, may not generate accurately small contact durations when the simulation time is not long enough. The reason can be deduced in Figure 3. For a small contact duration expected value, the smaller the sample size (amount of contacts between a pair of nodes) is, the harder the control of contact duration value (because the distribution is far from the Gaussian Distribution) is. To tackle this problem, we provide another method of choosing amount of interests and nodes' interest vectors, namely, method-2.

The detailed method-2 is as follows.

- (1) Let $n = num$. To simulate the small contact durations in scenarios with a short simulation time, we assume that the amount of node interests equals the number of nodes. In other words, it can be regarded that each node has a home spot which the node visits at a high probability. Under this circumstance, a node's interest vector (x_1, x_2, \dots, x_n) means that this node visit $node1$'s home at probability x_1 , $node2$'s home at probability x_2 , and $noden$'s home at probability x_n . Therefore, if two nodes have a large contact duration value, one node will certainly visit the other node's home at high probability, and the value of the probability is determined by the contact duration of these two nodes in the corresponding scenario.
- (2) Designate a public spot. We use a spot that is visited by all nodes at a specific probability to generate all small contact durations in the contact duration matrix of the scenario. Now, the value of n is actually $num + 1$. A node's interest vector becomes $(p, x_1, x_2, \dots, x_n)$, where p presents the probability at which the node visits the public spot. Generally, p holds the same value for all nodes.
- (3) Assign an initial interest vector value to each node. Initially, we assign $nodei$'s interest value as $(p, 0, \dots,$

TABLE 2: The three experimental data sets.

Dataset name	Infocom 06 Trace-1	Infocom 05 Trace-2	Cambridge Trace-3
Device	iMote	iMote	iMote
Network type	Bluetooth	Bluetooth	Bluetooth
Duration (days)	3	3	11
Granularity (sec)	120	120	600
Devices number	98 (78 mobile)	41	54 (36 mobile)

$x_i = 1 - p, \dots, 0$, $1 \leq i \leq n$. That is, initially, each node only visits two hotspots: the public spot and its own home.

- (4) Set a threshold to the target scenario's contact duration matrix. For the contact duration matrix, we set a threshold T_{th} and select all the values that are no less than T_{th} to reproduce in IHC. The value of T_{th} is chosen intuitively and empirically, assuring that contact duration larger than T_{th} is considerable and meaningful to represent a close social relationship.
- (5) Tune up all nodes' interest vector value based on the values exceeding T_{th} in the target scenario's contact duration matrix. Since all contact duration values no larger than T_{th} are generated by the visiting of the public spot, the remaining contact duration values (exceeding T_{th}) can be generated by tuning up the probability of home spots in nodes' interest vector values. Generally, we deal with the nodes one by one on the order of node ID from 1 to num . That is, for $node1$, the interest value is $(p, x_1 = 1 - p, 0, \dots, 0)$. Then, we search the $node1$'s list in contact duration matrix; if $nodea$ and $node1$'s contact duration CD_{1a} exceeds T_{th} , $nodea$'s interest vector will be updated as $(p, x_1 = CD_{1a}/(st(1 - p)), 0, \dots, x_a = 1 - p - x_1, \dots, 0)$ and so on. Different treatments can also be conducted as long as the contact duration values exceeding T_{th} are all held and for each node's interest vector $(p, x_1, x_2, \dots, x_n)$, $p + \sum_{i=1}^n x_i = 1$ is assured.

We can derive the interest vectors by using method-1 or method-2 if we want to reproduce a real-life scenario using IHC. However, in general, there are totally two cases when we need to determine the interest vectors of IHC. In the other case, if we want to set up just a simulation scenario, the interest vectors can be set as we need, for example, random values.

4. Model Verification

In order to show the accuracy of IHC in simulating real-life scenarios, we compare IHC with three real traces whose data is gathered from experiments done with wireless devices carried by people. These three traces are known as Infocom 06 trace (trace-1) [19], Infocom 05 trace (trace-2) [20], and Cambridge trace (trace-3) [21]. More details of the three real traces are shown in Table 2.

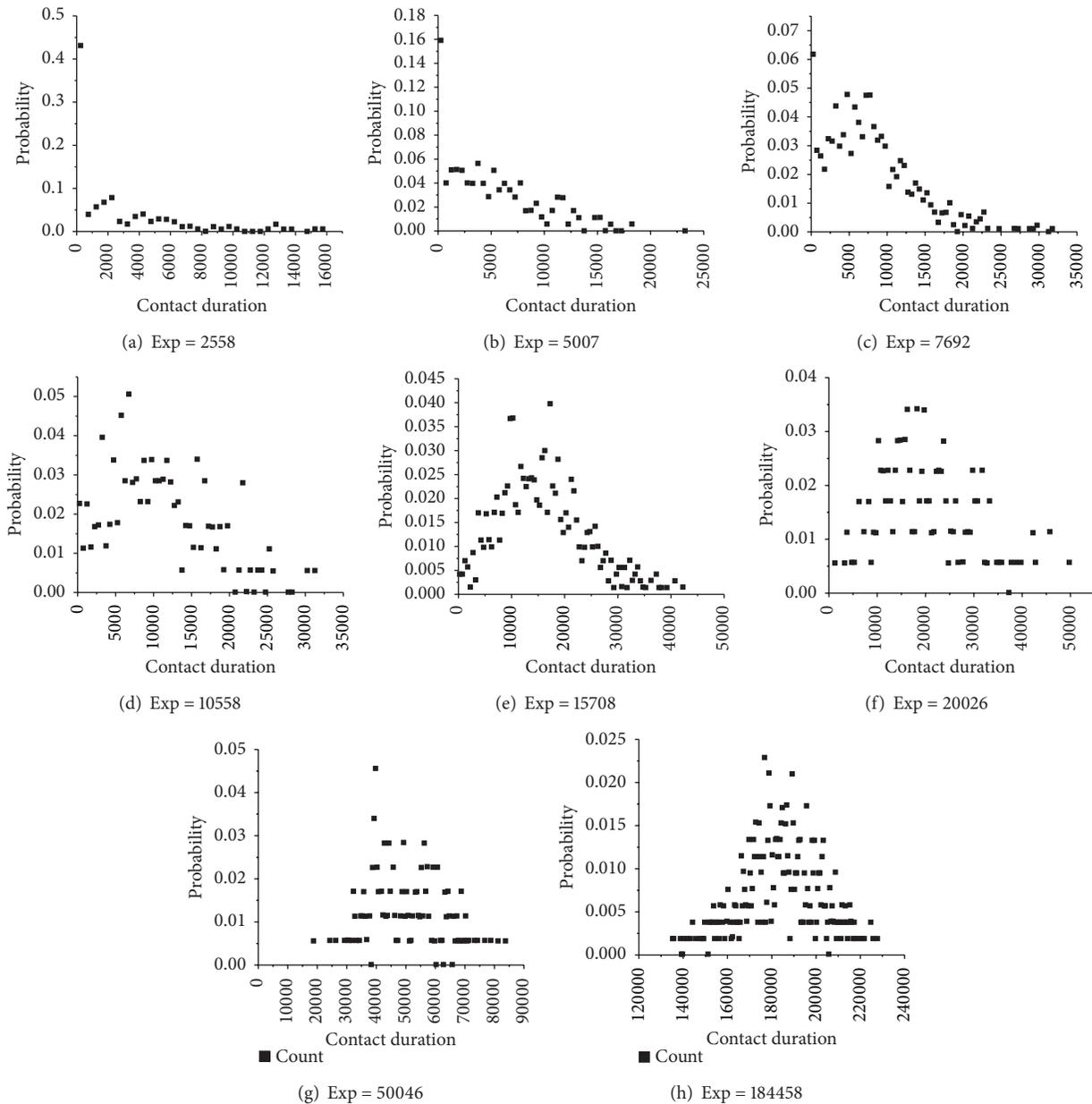


FIGURE 4: Distributions of contact durations under different expected values.

4.1. Trace Data. We illuminate how we use the trace data as follows.

First, we only care for the data generated by mobile and homogeneous nodes. For example, in trace-1, there are totally 98 iMotes in the experiment, but 20 of them are long range (around 100 meters) and static (deployed throughout the area or placed in lift of the hotel). The remaining 78 iMotes are carried by participants of the Infocom student workshop, with transmission range around 30 meters. Hence, these 78 iMotes are our research objects in this experiment. In order to find neighbor iMotes, each iMote performs periodic desynchronized scanning. The scanning takes approximately 5 to 10 seconds with time granularity between two consecutive scanning 120 seconds. An iMote cannot respond to any

request when it is active such that the synchronization needs to be avoided. In this experiment, a contact is defined as a period of time where all successive scanning by one iMote receive a positive answer by another. That is, a contact can only be confirmed after at least two scanning. Given that the scanning granularity is 120 s, the speed of a pedestrian is around 1 m/s, and the transmission range is around 30 m, it indicates that a moving iMote can hardly detect a contact such that the rationality of the feature “ignoring the contacts of moving nodes” in IHC is supported.

Second, symmetrize the contact duration matrix. In the three experiments, due to the interference and other limitations, non-mutual sightings are always created. As a result, the inter-contact time and contact duration are not

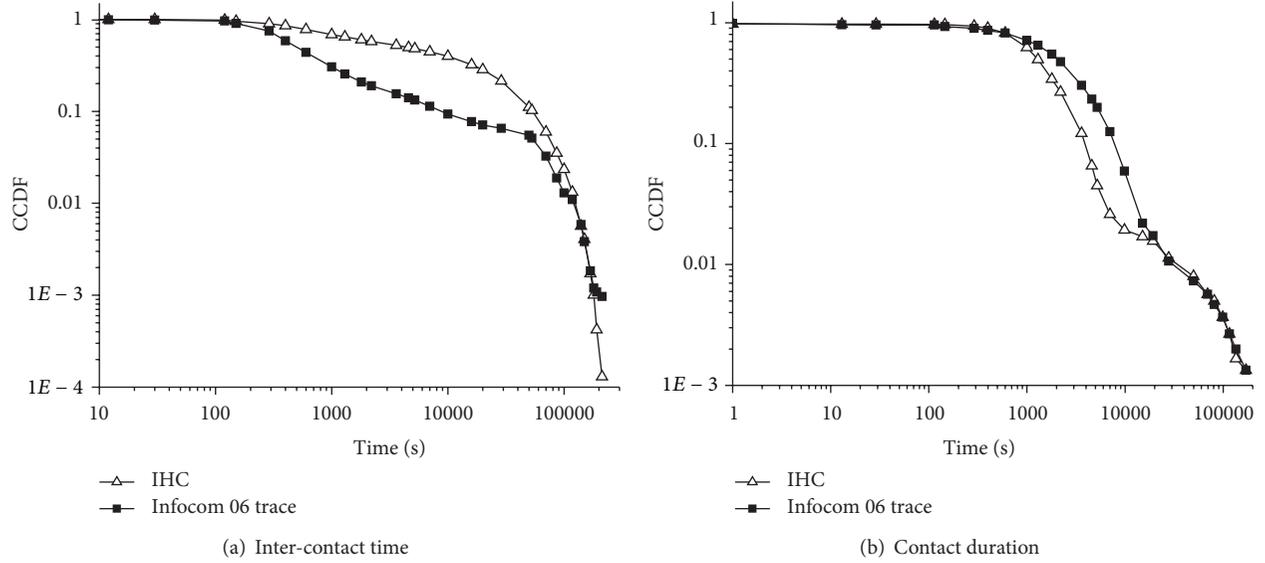


FIGURE 5: Reproduce Infocom 06 trace using IHC.

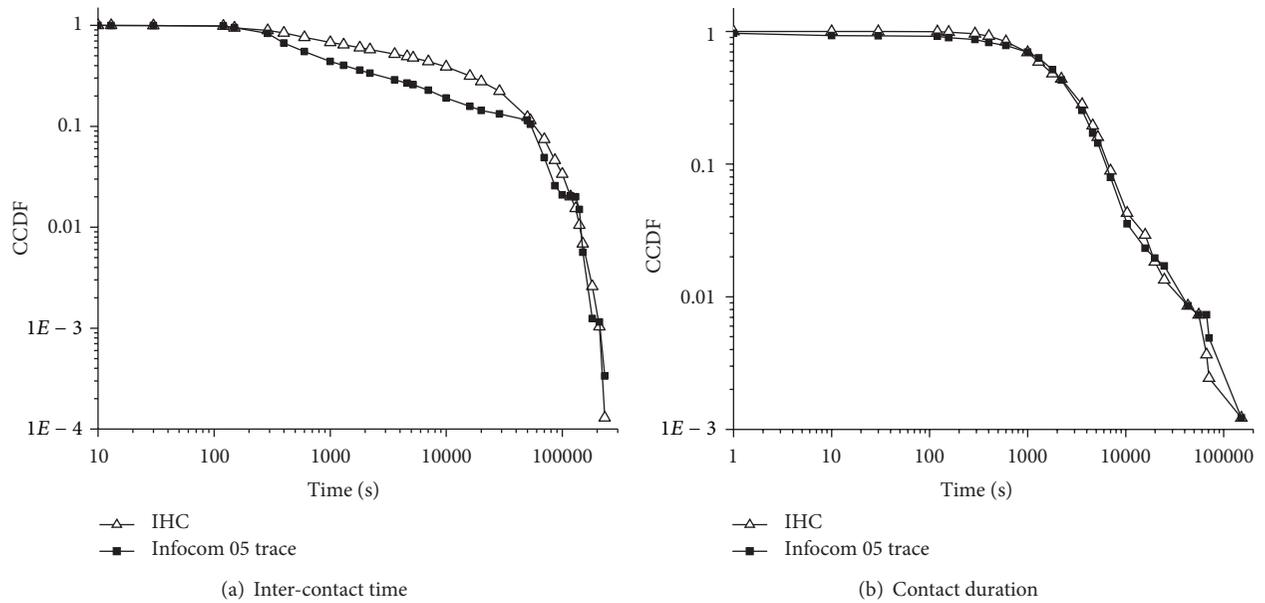


FIGURE 6: Reproduce Infocom 05 trace using IHC.

symmetrical. We keep all inter-contact times detected by the mobile iMotes. However, for the contact duration between a pair of iMotes, we take the maximum of their detected results as the value. That is, if iMote A has detected that its contact duration with B is CD_{A-B} , while iMote B detected the value as CD_{B-A} , we will take $\max(CD_{A-B}, CD_{B-A})$ as the contact duration value between iMotes A and B.

Note that these two treatments are conducted for all the three real traces.

4.2. Simulation Environment and Parameter Settings. IHC takes parameters listed in Table 1 as input. To compare IHC with real traces, we make the output text files containing

records on contact metrics and social relationships, including:

- (i) `inter-contact time.txt`: recording all inter-contact times between any two nodes;
- (ii) `contact duration.txt`: recording all contact durations between any two nodes;
- (iii) `interaction matrix.txt`: recording all contact durations between any two nodes in a matrix.

As we build a discrete event simulator of IHC with VC++6.0, we are able to change the output of the simulator to observe each event, such as a node starting moving or finish

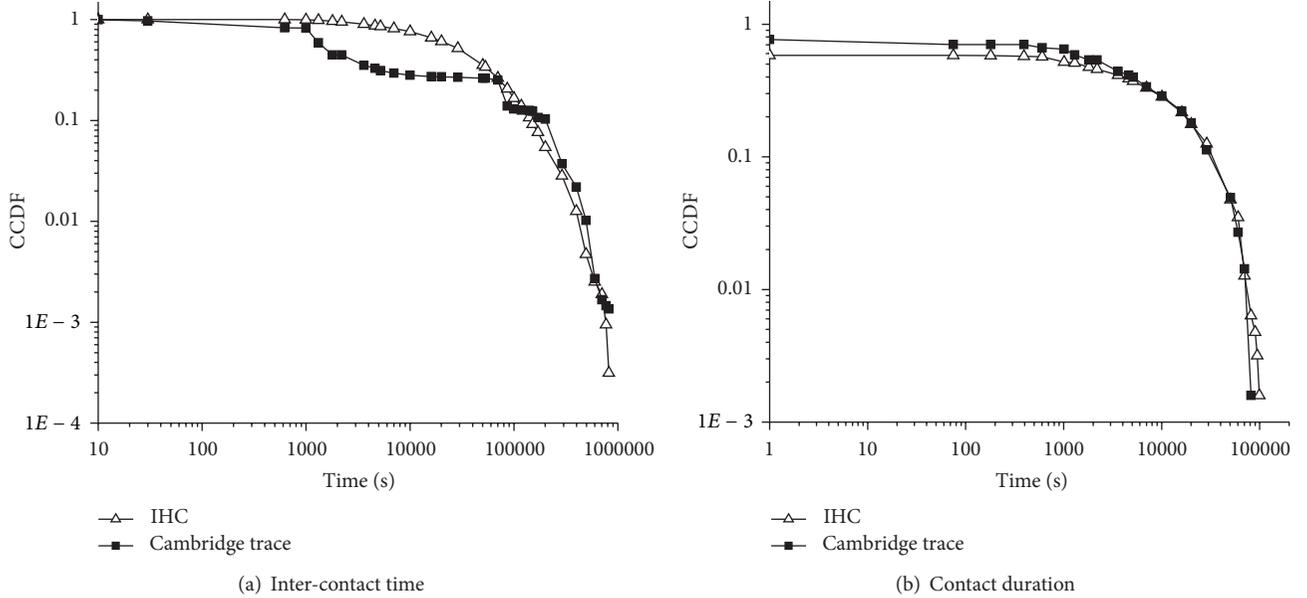


FIGURE 7: Reproduce Cambridge trace using IHC.

moving and two nodes meeting each other or departing from each other.

Parameters of IHC are chosen and tuned up based on the scenarios which we want to simulate, for example, Infocom 06 trace (trace-1). For the parameters of num and st , we assign them exactly the same values as trace-1, that is, 78 and 3 days (259200 seconds). ft is set to be 10 seconds based on the intuition that people spend few time on the journey. The pause time, which makes the best output (e.g., inter-contact time) matching with the real traces, is a bounded power law over the range of [120s, 4800s] with *slope* 6, denoted by (*slope*, *lower_bound*, *upper_bound*) in Table 3. Among them, the *lower_bound* affects the head of the inter-contact time CCDF, the *upper_bound* affects the tail of the inter-contact time CCDF, and the *slope* weakly affects the *slope* of inter-contact time CCDF in a very small range. The *lower_bound* is determined by scanning granularity because it is the scanning granularity that strongly affects the head of inter-contact time CCDF of the real traces. The values of *slope* and the *upper_bound* are determined by matching between real trace inter-contact time CCDF and simulation results.

Table 3 summarizes all parameter settings of the three scenarios. $H_i(x, y)$ and X_j are too expatiatory to show in Table 3 and thus omitted, since we have indicated the specific method to get them in detail in Section 3.4.

4.3. Simulation Results. We show the simulation results of inter-contact time and contact duration of Infocom 06 trace, Infocom 05 trace, and Cambridge trace in Figure 5, Figure 6, and Figure 7, respectively. Figures 5–7 validate that IHC can generate statistical metrics that approximate real traces. For a quantitative comparison, we calculate the Jensen-Shannon divergence between the distributions of the real traces and IHC traces in Table 4, as well as the corresponding results of SWIM traces whose data can be found in [5], since SWIM is

a very outstanding work on human mobility modeling. The results shown in Table 4 indicate that our model outperforms SWIM in the accuracy of reproducing inter-contact time and contact duration.

We draw meaningful social relationships in real and IHC traces in Figures 8, 9, and 10, which are weighted undirected graphs. A vertex in the graph (Figures 8, 9, and 10) represents the node with the same ID in the networks. The edge between two vertices indicates that the social relationship (contact duration) between these two nodes exceeds a certain threshold. For Infocom 06 trace, T th is assigned as 20000 seconds because only the top 1.665% (50 out of 3003) largest contact durations are larger than 20000. For Infocom 05 trace, T th is set to be 10000 seconds because only the top 3.9% (32 out of 820) largest contact durations are larger than 10000. The threshold for Cambridge trace is chosen with similar principle with the value of 50000. The weights of the edges are calculated as the ratio of contact duration between the two nodes to the network simulation time, retaining two decimal places.

Figures 8, 9, and 10 visually show the social relationship similarity between real traces and corresponding IHC-generated ones. For a quantitative view, we conduct Mantel Test on the real and IHC-generated social interaction matrices where the raw data Figures 8, 9, and 10 comes from. Mantel Test measures the correlativity between two matrices. Since Figures 5(b)–7(b) have shown a very similar scale of social interaction matrices between real traces and corresponding IHC-generated ones, a high correlativity can complementarily prove that the IHC-generated social interaction matrices are very similar to the real ones. The Mantel Test results are shown in Table 5.

Figures 5(b)–7(b) and Table 5 prove that IHC can accurately reproduce the overall social relationships in real-life scenarios. Further, Figures 8, 9, and 10 indicate that

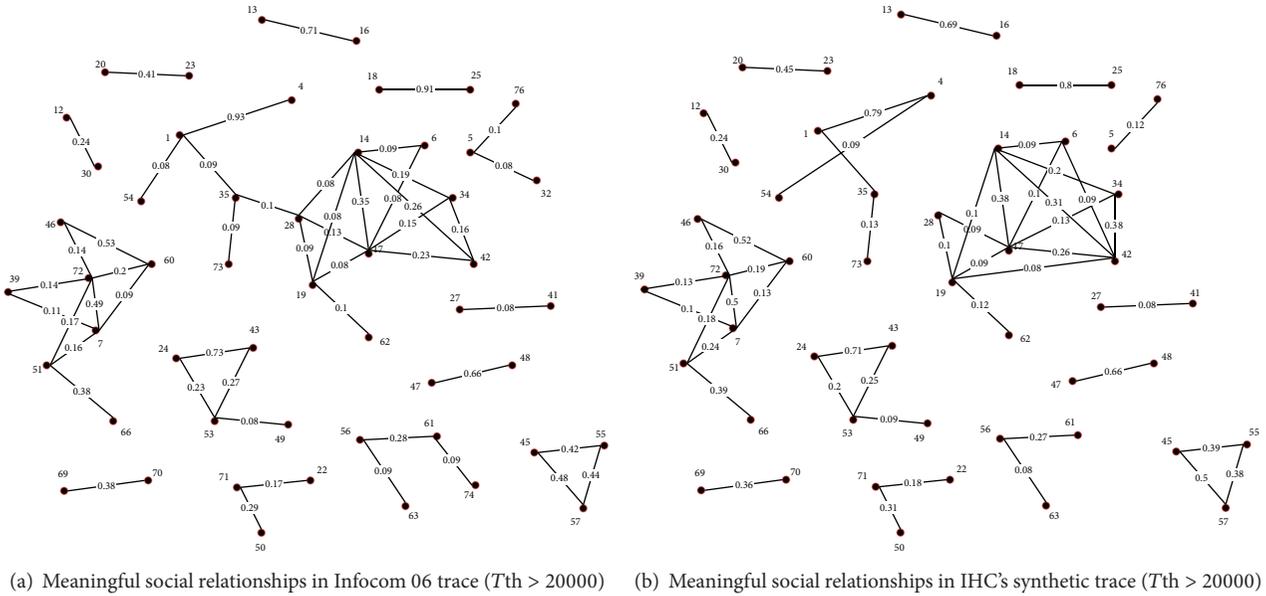


FIGURE 8: Comparisons between Infocom 06 trace and IHC: meaningful social relationships.

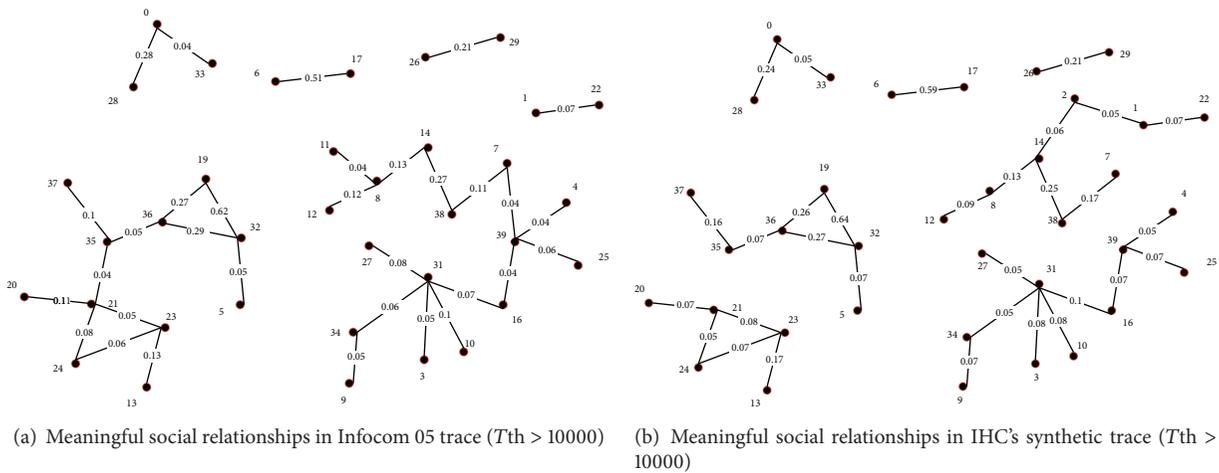


FIGURE 9: Comparisons between Infocom 05 trace and IHC: meaningful social relationships.

IHC is also able to characterize pairwise social relationships quantitatively. For example, there exist some close social relationships that cannot be detected by a community detection algorithm (e.g., k -clique [22], $k > 2$), such as the relationship between 13 and 16 and 18 and 25 in Figure 8(a). This kind of relationship is defined as “friendship” in [12]. IHC has the ability of reproducing the “friendship” in the target scenario. Further, the inter/intra-community social relationships and pairwise social relationships belonging to no communities that IHC generates all match real traces very well. As far as we know, no model has such a feature. Note here that we only compare and show pair-wise social relationships of real traces and IHC traces, since other mobility models cannot reproduce the social relationships on the basis of pair-wise.

5. Building Cases to Use IHC

In this section, we build cases to use our model. We compare the performances of forwarding protocols running with real traces and our simulated scenarios to validate that IHC can be used to predict protocols’ performance. We use the three real traces (Infocom 06 trace, Infocom 05 traces and Cambridge trace) and the three corresponding synthetic traces generated by IHC as the network environments. Our goal is to validate that IHC is able to predict the performance of forwarding protocols rather than evaluating which forwarding protocol performs better. Therefore, the protocols we choose, that is, Epidemic Forwarding [23] and Spray and Wait [24], which are very mature and get extensively utilized in DTNs and

TABLE 3: Parameter settings.

Scenario	Infocom 06	Infocom 05	Cambridge
<i>num</i>	78	41	36
<i>st</i>	259200	259200	950400
<i>ft</i>	10	10	10
<i>pt: (slope, lower_bound, upper_bound)</i>	(6, 120, 4800)	(6, 120, 7200)	(2.45, 600, 14400)
Interest vectors generator	Method-2	Method-2	Method-1
<i>n</i>	79	42	36

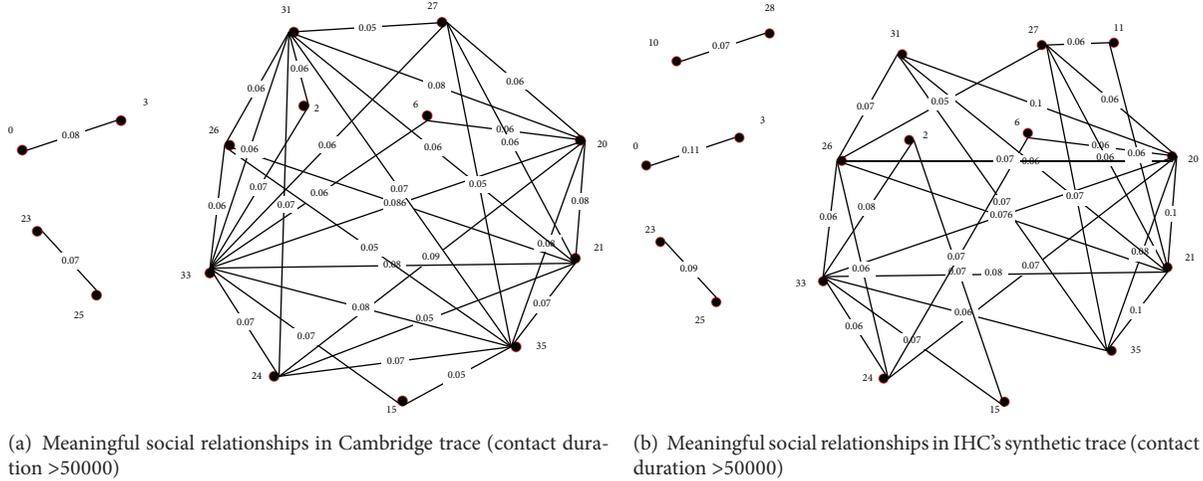


FIGURE 10: Comparisons between Cambridge trace and IHC: meaningful social relationships.

TABLE 4: Jensen-Shannon divergence between distributions of the real and IHC traces, comparing with the corresponding results of SWIM traces.

Trace	Infocom 06	Infocom 05	Cambridge
Intercontact time (IHC)	0.048	0.037	0.043
Intercontact time (SWIM)	0.049	0.062	0.058
Contact duration (IHC)	0.022	0.004	0.014
Contact duration (SWIM)	0.18	0.21	0.15

TABLE 5: Mantel Test results on the real and IHC-generated social interaction matrices.

Scenario	Infocom 06	Infocom 05	Cambridge
Correlativity	0.9496	0.9407	0.9256

BUBBLE [12], which is a sophisticated social-aware protocol are appropriate for our goal.

As in [5, 25], we choose two metrics to evaluate the performance of forwarding protocols. They are delivery cost (*cost*) and packet delivery ratio (*pdr*). The former indicates the price of forwarding a data packet successfully and accounts for the efficiency of the protocol. The delivery cost is calculated by the ratio of “the amount of received control packets plus the amount of data packets’ replicas” to “the amount of received data packets”. The packet delivery ratio, instead, is actually the successful rate of forwarding data packets and accounts for the effectiveness of the protocol. Packet delivery ratio is

calculated as the ratio of “amount of received data packets” to “amount of generated data packets”.

The following settings are validated for each scenario: a set of messages is generated with sources and destinations chosen uniformly at random with interval of 20 minutes, as we simulated the overall periods, that is, 3 days or 11 days, which is significantly different from that of [5] where each simulation runs only for 3 hours (choosing 3 hours out of 3 days or 11 days incurs too many uncertainties). However, the interest vector setting in Section 4.2 is based on the overall experiments duration (i.e., 3 days and 11 days). In IHC, all movement epochs are consecutive, but the actual movement epochs differ greatly in daytime and nighttime. Therefore, for a more meaningful simulation and for making statistical analysis in separate scenario, we divide the experiment duration into fragments equally. Concretely, the scale of the fragment is chosen as 12 hours to capture daytime and nighttime motions, respectively. That is, Infocom 06 and Infocom 05 scenarios are divided into 6 fragments and the Cambridge scenario is divided into 22 fragments. More importantly, interest vectors are dynamic and the values are derived using “Interest Vectors Generator” based on the current experiment fragment. To avoid end-effects, no messages are generated in the last hour; the time-to-live of messages is set as 1 hour. The accumulated forwarding protocols’ results are shown in Figure 11. That is, in each simulation in Figure 11, the statistical metrics (*cost* and *pdr*) are continuously calculated except that interest vectors change with the alternate fragments. Table 6 shows

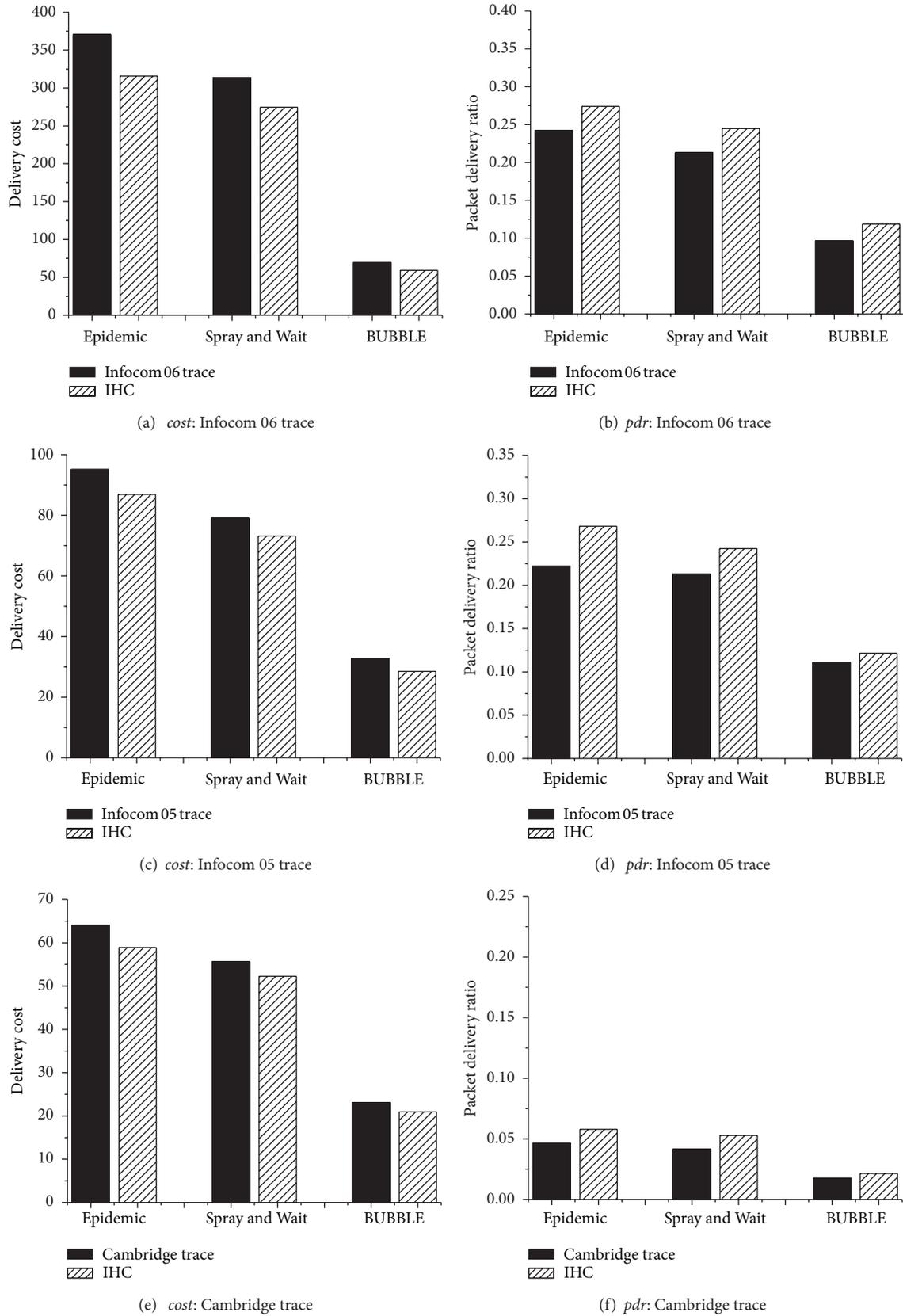


FIGURE 11: Performance of forwarding protocols (interest vectors change with the alternate fragments).

TABLE 6: The average error percentage of all fragments in each separate scenario.

Scenario	Infocom 06	Infocom 05	Cambridge
<i>cost</i> (Epidemic)	-0.14	-0.08	-0.08
<i>cost</i> (Spray & Wait)	-0.11	-0.07	-0.06
<i>cost</i> (BUBBLE)	-0.14	-0.12	-0.13
<i>pdr</i> (Epidemic)	0.10	0.21	0.24
<i>pdr</i> (Spray & Wait)	0.15	0.17	0.25
<i>pdr</i> (BUBBLE)	0.23	0.09	0.22

the average error percentage of all fragments in each separate scenario. In our simulations, IHC-generated traces always lead to lower *cost* and higher *pdr* than the corresponding real ones. Thus in Table 6 the average error percentage of *cost* are all negative and those of *pdr* are all positive. However, the error percentage results in Table 6 have small absolute value, showing that each of the three forwarding protocols has similar performance in both real and synthetic traces generated by IHC.

Figure 11 is composed of six subfigures, namely Figures 11(a)–11(f). Figures 11(a)–11(b), 11(c)–11(d), and 11(e)–11(f) depict the performance of forwarding protocols in Infocom 06 trace, Infocom 05 trace, and Cambridge trace, respectively. In each subfigure, we draw six pillars representing consecutively the performance (corresponding to the subfigure, such as *cost* or *pdr*) of Epidemic Routing in the real trace and synthetic trace, Spray and Wait in the real trace and synthetic trace and BUBBLE. Figure 11 shows that the trend of the protocols in the real traces is the same as that of the corresponding synthetic ones. That is, the ones that perform better in the real world do the same things in the IHC-generated one. Figure 11 and Table 6 both indicate that IHC can predict well the performance of all the three protocols. As a result, IHC is a good model for protocol validation; the performance of protocols in the real life scenarios can be accurately predicted by running the protocols on the synthetic traces generated by IHC.

6. Conclusions

In this paper, we propose a mobility model, IHC. IHC merges a few human mobility characteristics and is very simple to implement. IHC takes node interest as input to reproduce nodes' social relationships. Correspondingly, we explore 2 methods to generate node interest vectors based on a contact duration matrix. Through the comparisons with real-life human mobility metrics of inter-contact time and contact duration, we validate that IHC can generate synthetic traces that approximate real traces. Being different from any existing mobility models, IHC has the ability of characterizing pairwise social relationships quantitatively. Further simulations have been conducted to show that IHC can predict the performance of forwarding protocols well.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (NNSFC) under Grant no. 61170245.

References

- [1] R. R. Maiti, A. Mallya, and N. Ganguly, "Characterizing Mobility Models for Human Movement".
- [2] C. Zhao and M. L. Sichitiu, "N-body: social based mobility model for wireless ad hoc network research," in *Proceedings of the 7th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON '10)*, June 2010.
- [3] G. S. Thakur and A. Helmy, "COBRA: A Framework for the Analysis of Realistic Mobility Models," 2012.
- [4] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'11)*, pp. 1082–1090, August 2011.
- [5] S. Kosta, A. Mei, and J. Stefa, "Large-scale synthetic social mobile networks with SWIM," *IEEE Transactions on Mobile Computing*, 2012.
- [6] W.-J. Hsu, T. Spyropoulos, K. Psounis, and A. Helmy, "Modeling time-variant user mobility in wireless mobile networks," in *Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM '07)*, pp. 758–766, May 2007.
- [7] I. Rhee, M. Shin, S. Hong, K. Lee, S. J. Kim, and S. Chong, "On the levy-walk nature of human mobility," *IEEE/ACM Transactions on Networking*, vol. 19, no. 3, pp. 630–643, 2011.
- [8] M. Musolesi and C. Mascolo, "Designing mobility models based on social network theory," *ACM SIGMOBILE Mobile Computing and Communication Review*, vol. 11, no. 3, pp. 59–70, 2007.
- [9] C. Boldrini and A. Passarella, "HCM: modelling spatial and temporal properties of human mobility driven by users' social relationships," *Computer Communications*, vol. 33, no. 9, pp. 1056–1074, 2010.
- [10] S. Yang, X. Yang, C. Zhang, and E. Spyrou, "Using social network theory for modeling human mobility," *IEEE Network*, vol. 24, no. 5, pp. 6–13, 2010.
- [11] T. Abdelkader, K. Naik, A. Nayak, N. Goel, and V. Srivastava, "SGBR: a routing protocol for delay tolerant networks using social grouping," *IEEE Transactions on Parallel and Distributed Systems*, 2012.
- [12] P. Hui, J. Crowcroft, and E. Yoneki, "BUBBLE Rap: social-based forwarding in delay-tolerant networks," *IEEE Transactions on Mobile Computing*, vol. 10, no. 11, pp. 1576–1589, 2011.
- [13] A. Mei, G. Morabito, P. Santi, and J. Stefa, "Social-aware stateless forwarding in pocket switched networks," in *Proceedings of the 30th IEEE International Conference on Computer Communications (INFOCOM '11)*, pp. 251–255, April 2011.
- [14] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong, "SLAW: a mobility model for human walks," in *Proceedings of the 28th Conference on Computer Communications (INFOCOM '09)*, pp. 855–863, April 2009.
- [15] A. Munjal, T. Camp, and W. C. . Navidi, "SMOOTH: a simple way to model human walks," in *ACM SIGMOBILE Mobile Computing and Communications Review*, pp. 34–36, 2010.

- [16] A. D. Nguyen, P. Sénac, V. Ramiro, and M. Diaz, “STEPS-an approach for human mobility modeling,” in *Proceedings of the 10th International IFIP TC 6 Networking Conference*, vol. 6640, pp. 254–265, Springer, Valencia, Spain, 2011.
- [17] B. Astuto, A. Nunes, K. Obraczka, and A. Rodrigues, “SAGA: socially-and geography-aware mobility modeling framework,” in *Proceedings of the 15th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pp. 367–376, ACM, 2012.
- [18] G. S. Thakur, A. Helmy, and W.-J. Hsu, “Similarity analysis and modeling in mobile societies: The missing link,” in *Proceedings of the 5th ACM Workshop on Challenged Networks (CHANTS '10)*, pp. 13–20, September 2010.
- [19] <http://crawdad.cs.dartmouth.edu/cambridge/haggle/imote/infocom2006>.
- [20] <http://crawdad.cs.dartmouth.edu/cambridge/haggle/imote/infocom>.
- [21] <http://crawdad.cs.dartmouth.edu/upmc/content/imote/cambridge>.
- [22] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, “Uncovering the overlapping community structure of complex networks in nature and society,” *Nature*, vol. 435, no. 7043, pp. 814–818, 2005.
- [23] A. Vahdat and D. Becker, “Epidemic routing for partially connected ad hoc networks,” Tech. Rep. CS-200006, Duke University, 2000.
- [24] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, “Spray and wait: An efficient routing scheme for intermittently connected mobile networks,” in *Proceedings of the ACM SIGCOMM Conference on Computer Communications*, pp. 252–259, August 2005.
- [25] P. Costa, C. Mascolo, M. Musolesi, and G. P. Picco, “Socially-aware routing for publish-subscribe in delay-tolerant mobile ad hoc networks,” *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 5, pp. 748–760, 2008.