ARTICLE IN PRESS

Pervasive and Mobile Computing ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Pervasive and Mobile Computing

journal homepage: www.elsevier.com/locate/pmc

Fast track article

Community Aware Heterogeneous Human Mobility (CAHM): Model and analysis

Zunnun Narmawala [a,*], Sanjay Srivastava [b]

[a] Institute of Technology, Nirma University, Ahmedabad, Gujarat, India

[b] Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar, Gujarat, India

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Mobility model
Mobile social network
Overlapping community structure
Heterogeneous popularity
Hub and gateway

ABSTRACT

Community Aware Heterogeneous Human Mobility Model (CAHM) is based on Heterogeneous Human Walk (HHW) Yang et al. (2010) mobility model. CAHM achieves heterogeneous local popularity as observed in real mobility traces which HHW fails to achieve. It also incorporates following additional properties of human mobility: preference of nearby locations, speed as a function of distance to be traveled and power-law distributed pause time. We show that these properties make significant impact on routing protocols' performance. We also propose methods based on mathematical models to identify popular nodes within community (hubs) and in entire network (gateways) from overlapping community structure itself without doing message flooding.

1. Introduction

With the rapid growth of mobile hand-held devices with Bluetooth and ad hoc WiFi connectivity, possibility of entirely new network paradigm has emerged in which encounters between these mobile devices can be exploited for opportunistic data transfer without using any fixed network infrastructure [1]. As these devices are carried by humans, their encounter patterns depend on human mobility patterns. So, knowledge of human movement behavior can be exploited to make efficient forwarding decisions [2,3]. We call this network paradigm as Pocket Switched Network (PSN) or Mobile Social Network (MSN).

Various experimental projects are undertaken to collect encounter information of devices carried by humans [4,5]. These traces can be used in simulation to evaluate and analyze performance of different protocols. While this approach generates realistic mobility patterns, its usefulness is limited as performance of a protocol can be evaluated only for limited values of network parameters for which traces are available. Nonetheless, from analysis of these traces, various statistical properties of human mobility are derived [4–7]. Well-known and widely used mobility models such as Random Way Point (RWP) [8] and Brownian Motion [9] do not exhibit these properties. So, trace-based mobility models such as Levy Walk (LW) [6], TVC [10], SWIM [11] and SLAW [12] are proposed based on these statistical properties. Although these models are able to reproduce statistical properties of real mobility traces, they assume that each node moves independent of others.

But, movement of an individual is not independent. Humans belong to various social communities like friends, family, co-workers, etc. [13,14]. These social ties significantly affect their movement. For example, individuals meet others from the same community more frequently than people of other community [15]. Mobility models such as CMM [15] and HCMM [16] incorporate this social aspect of human mobility.

* Corresponding author.
E-mail address: zunnun80@gmail.com (Z. Narmawala).

http://dx.doi.org/10.1016/j.pmcj.2014.12.008
1574-1192/

Please cite this article in press as: Z. Narmawala, S. Srivastava, Community Aware Heterogeneous Human Mobility (CAHM): Model and analysis, Pervasive and Mobile Computing (2015), http://dx.doi.org/10.1016/j.pmcj.2014.12.008

Individuals belong to multiple communities. i.e. communities in real social network overlap [14]. Further, some nodes meet more nodes in the community (locally popular nodes) or visit other communities more often than others (globally popular nodes). These properties have significant impact on the performance of forwarding strategy. But, only Heterogeneous Human Walk (HHW) model [17] attempts to incorporate these properties for generating k-clique overlapping community structure without relying on real-life social graph.

In this paper, our contributions are the following:

- We identify that HHW model does not produce heterogeneous local popularity of nodes in a community as observed in real mobility traces. We propose Community Aware Heterogeneous Human Mobility Model (CAHM) which rectifies this problem. We also incorporate following features in CAHM: (1) Levy walk nature of human mobility (2) speed of nodes as a function of distance to be traveled and (3) power-law distributed pause time.
- We propose, using mathematical models, methods to identify hub and gateway nodes from the given overlapping community structure itself without doing message flooding.
- We analyze effect of CAHM mobility model on the performance of routing protocols as compared to existing mobility models.
- We analyze properties of overlapping community structure formed by human mobility. Analysis gives important insights for designing better forwarding mechanisms.

The paper is organized as follows: In Section 2, we discuss related work. In Section 3, an overview of HHW model is given. In Section 4, we present Community Aware Heterogeneous Human Mobility (CAHM) model. In Section 5, identifying hub and gateway nodes of a community is described. Section 6 discusses simulation results and finally, we conclude in Section 7.

2. Related work

Based on whether social network dimension is incorporated in the mobility model or not, mobility models for MSN can be divided into two categories as discussed in [17]: Real-trace based models and Social-aware models.

2.1. Real-trace based models

Following are the main properties of human mobility patterns identified by analysis of various traces.

1. Aggregate inter-contact time follows power-law distribution with exponential cutoff [2,4].
2. Pause time follows power-law distribution [6].
3. Humans visit nearby locations more frequently compared to far-away locations [7].
4. Humans have location preferences and they periodically re-appear on these locations [7].
5. Speed at which humans move increases with distance to be traveled [6].

Real-trace based models try to capture features of individual's independent movement observed from real traces. Working Day Mobility (WDM) model [18] incorporates properties numbered 1 and 4 in the above list by modeling individual's mobility during a day with home sub-model, office sub-model, transport-sub model and evening sub-model. Time Variant Community (TVC) model [10] incorporates properties 1 and 4. Small World In Motion (SWIM) model [11] incorporates all of the above properties. In this model, each node is assigned a randomly and uniformly chosen point over the network area called as home. For each node, a weight is assigned to each possible destination which grows with the popularity of the place and decreases with distance from home. This weight represents the probability for the node to choose that place as the next destination. Self-similar Least Action Walk (SLAW) model [12] incorporates properties 1, 2 and 3.

2.2. Social-aware models

Following are the main properties derived from social network theory which affect human mobility.

1. Humans can be grouped into communities based on their social relationships [13].
2. Humans belong to multiple communities and so, communities overlap [14].
3. Different individuals have heterogeneous local popularity within community and heterogeneous global popularity in the social network [1].
4. Community size, number of communities in which a node is a member and overlap size approximately follow power-law distribution where overlap size is defined as number of individuals which are common in two communities [14].

Community-based Mobility Model (CMM) [15] groups nodes based on social relationships among individuals. This grouping is then mapped to a topographical space. Movement of nodes is influenced by strength of social ties among individuals which may also change in time. CMM uses Caveman model [19] as artificial Social Network Model (SNM) to generate community structure. Home-cell Community-based Mobility Model (HCMM) [16] assigns home-cell to each individual which is the location where people with whom the node shares social relationships are likely to be at some point in time. After each trip, node moves to home-cell with some probability. These models incorporate only property 1 from the above list. These models also incorporate some of the properties derived from real traces. But, they do not incorporate properties 2, 3 and 4 from above list which are very important properties and have significant effect on the performance of the protocols. Social, sPatial and Temporal mobility framework (SPoT) [20] is flexible and controllable mobility framework. But, it generates only contact traces and proposal in the paper for generating movement traces is preliminary. Further, they take social graph as an input instead of generating community structure synthetically. So, it lacks flexibility of generating large number of different social graphs for simulation. A detailed review of human mobility in opportunistic networks is presented in [21]. None of the mobility models incorporate all of the above properties to generate community structure synthetically.

Heterogeneous Human Walk (HHW) model [17] incorporates all of above properties derived from social network theory to generate community structure synthetically. But, it does not incorporate important trace-based properties like human preference of nearby locations and dependence of speed at which humans move on distance to be traveled. Our model CAHM, which is the modification of HHW, incorporates all properties of above two lists which are derived from real-traces and from social network theory.

3. Overview of HHW model

There are two options to generate overlapping community structure. First is to get a social graph from some actual social network. After getting the social graph, one can apply an algorithm similar to the one proposed in [14] to identify overlapping communities but it will significantly increase implementation and computational complexity and one has to collect large number of social graphs to analyze performance of protocols. Second option is to directly construct synthetic overlapping community structure which follows all the properties found in real social network. To achieve trade-off between reality and complexity, HHW uses the second approach. In this approach, any number of different community structures can be generated using random variables.

For overlapping community structure, each individual n in the social network may belong to number of communities denoted as membership number MN_n. Further, any two communities x and y may share $S^{ov}_{x,y}$ individuals, defined as overlap size between two communities. Let us denote size of community x as S^{com}_x and probability distribution functions of membership number, overlap size and community size as $\mathrm{P(MN)}$, $\mathrm{P(S^{ov})}$ and $\mathrm{P(S^{com}} - (k-1))$ respectively where k is clique size of community as defined in [14]. Based on analysis of various real social networks, Palla et al. [14] conclude that $\mathrm{P(MN)}$, $\mathrm{P(S^{ov})}$ and $\mathrm{P(S^{com}}-(k-1))$ approximately follow power-law distribution $\mathrm{P(x)} \sim \mathrm{x}^{-\tau}$, with exponents $\tau = \mathrm{PR^{MN}}$, $\tau = \mathrm{PR^{Osize}}$ and $\tau = \mathrm{PR^{Csize}}$, respectively. Further, they report that values of $\mathrm{PR^{MN}}$ and $\mathrm{PR^{Osize}}$ are not less than 2, and the value of $\mathrm{PR^{Csize}}$ is between 1 and 1.6. HHW model uses these statistical properties to artificially construct k-clique overlapping community structure. A k-clique is a complete sub-graph of size k and k-clique community is the union of all k-cliques that can be reached from one another through series of adjacent k-cliques where two k-cliques are adjacent if they share $k-1$ nodes.

HHW model is composed of three components: (1) establishing overlapping community structure and heterogeneous local degree (2) mapping communities into geographical zones and (3) driving individual motion. These components are explained in the following three sub-sections.

3.1. Establishing k-clique overlapping community structure and heterogeneous local degree

A day (or a week or any time duration) is divided into periods, and overlapping community structures are different in each of these periods but are the same in the same period of different days. Let us define nodes having membership number larger than 2, equals 2 and equals 1 as M-3 nodes, M-2 nodes and M-1 nodes respectively. Community structure for each period is constructed as follows:

1. Generate nodes' membership numbers such that they follow $\mathrm{P(MN)}$ with exponent $\mathrm{PR^{MN}}$. Then, establish initial empty communities whose sizes S^{com} follow $\mathrm{P(S^{com}} - (k-1))$ with exponent $\mathrm{PR^{Csize}}$ such that $\sum_i MN_i = \sum_j S^{com}_j$.
2. Use all M-3 nodes to establish initial overlaps between pairs of communities.
3. Modify initial overlaps by allocating all M-2 nodes to communities such that overlaps' size follows $\mathrm{P(S^{ov})}$ with exponent $\mathrm{PR^{Osize}}$.
4. Allocate all M-1 nodes to unsaturated communities.

Please refer the original paper [17] for detailed procedure of each step. The paper claims that local heterogeneous popularity is established by assigning heterogeneous local degrees to the nodes.

3.1.1. Generating heterogeneous local degree

Let $Local^n_i$ denote local degree of node n in its community i where $Local^n_i \geq k-1$ as per the definition of k-clique community. These values are generated such that they follow power-law distribution with exponent $\mathrm{PR^{Local}}$.

3.2. Mapping communities into geographical zones

To simulate n mobile nodes in two-dimensional square plane, the model divides the plane into a grid of non-overlapping square cells. For each period, each community with size S^{com} is randomly associated with a zone composed of S^{com} connected cells. Each node n is randomly associated with $Local_i^n$ cells within the zone of its community i.

3.3. Driving individual motion

Initially, each node randomly selects one of all its associated cells and then it is located at random position inside that cell. To move, each node selects next goal inside one of its associated cells and then moves towards it by following straight path. When a node reaches its current goal, it waits for a uniformly distributed pause time and then selects and moves towards next goal. At the start of the new period, overlapping community structure, corresponding associated zone and cells change. Now, after reaching its current goal, the node selects next goal inside one of its newly associated cells of the new period.

The paper [17] demonstrates that nodes of this overlapping community structure exhibit global heterogeneous popularity and inter-contact time follows power-law distribution with exponential cutoff.

4. Community Aware Heterogeneous Human Mobility Model (CAHM)

In this section, we discuss shortcomings of the existing HHW model and propose our solution for each of them.

4.1. Incorporating Levy walk nature of human movement

In HHW, if a node is a member of more than one community, then it will have associated cells in all those communities. As per the model, the node chooses an associated cell as next goal randomly with uniform distribution irrespective of the community of the associate cell, i.e. the node chooses next goal irrespective of the distance it will have to travel. This is contrary to the finding that humans prefer short distances over long distances or in other words, human movement can be characterized as Levy walk [6]. It is also counter-intuitive. For example, a postman has to visit multiple offices in multiple buildings to deliver posts. As per existing model, the postman will move from an office in a building to another office in another building with more probability than to the office in the same building assuming that the total number of offices of other buildings is more than the number of offices in the same building.

As established in [6], distribution of distances covered in each flight by human follows power-law distribution having exponent less than 2.5 where flight can be defined as single displacement from one place to another place without a pause in between. We have incorporated this property in our model.

In CAHM, a node chooses an associated cell as next goal based on the distance it will have to travel with power-law distribution instead of choosing it randomly with uniform distribution. We calculate distances at which all associated cells of a node are located from the current cell from their location information and sort associated cells of a node based on these distances. Then, we use random variate (RV) which follows power-law distribution P(D) with exponent PR^{D} between minimum distance and maximum distance a node has to travel. We choose the associated cell as next goal whose distance from the current cell is nearest to the value generated by the random variable. As RV will generate short distance values with higher probability than long-distance values, in our model, a node prefers short distances over long distances. As each community is associated with a zone composed of connected cells, distance between any two cells within a community will most probably be less than the distance between any two cells of two different communities if simulation area is not too small. Therefore, a node will be choosing one of the associated cells of the current community as its next goal with high probability compared to the associated cells of the node in other communities which is correct behavior as we have seen in the postman example.

4.2. Treating number of cells and number of nodes in a community as separate parameters

In HHW, each community with size S^{com} is associated with a zone composed of S^{com} adjacent cells. i.e. number of nodes (N) and number of cells (C) of a community are same. So, average number of nodes associated with a cell is equal to average local degree of a community (μ). One important effect of it is that a node, on an average, meets μ number of nodes in each cell it visits. So, local popularity of nodes increases with increase in local degree at the rate proportional to μ times the rate of increase in local degree instead of increasing with same rate. As a result, there are high percentage of nodes with high local popularity than observed in real traces [1]. i.e. HHW generates too many locally popular nodes than expected. Further, μ increases with increase in community size because it follows power-law distribution with maximum value equal to community size. So, the problem is aggravated in large communities.

In CAHM, we consider C and N as separate parameters. Let C_x be the number of connected cells, μ_x be average local degree and N_x be the number of nodes in community x. Let m be the multiplier which decides denseness of all communities. Then,

$$C_x = m \times \mu_x \times N_x. \tag{1}$$

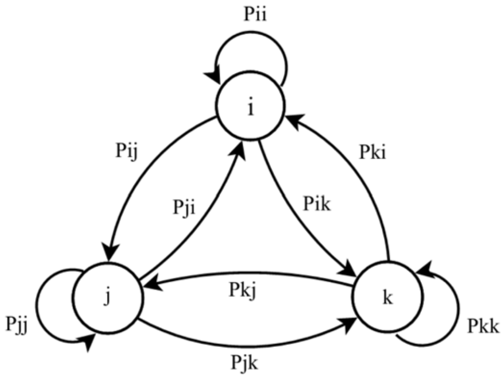

Fig. 1. Communities and movement of a node between communities represented as Markov Chain.

4.3. Calculating speed based on distance

In HHW, speed at which a node moves from one goal to the next goal is chosen from a given range uniformly regardless of the distance to be traveled. But, as found in [6], speed increases with increase in flight length because individuals use transportation to travel long distances instead of walking. They have also derived the following relation between flight time (t) and flight length (l) from various real traces.

$$t = p \times l^{1-\eta}, \quad 0 \leq \eta \leq 1. \tag{2}$$

From traces, Rhee et al. [6] have proposed $p = 30.55$ and $\eta = 0.89$ when $l < 500$ m, and $p = 0.76$ and $\eta = 0.28$ when $l \geq 500$ m. In CAHM, we also use this model to calculate speed at which a node should travel to next goal instead of choosing it uniformly from a given range.

5. Identifying hub and gateway nodes in overlapping community structure

Identifying correct hub and gateway nodes is very important for the efficiency of a protocol aiming to exploit heterogeneous popularity. Hub and gateway nodes can be identified by measuring local and global popularity of nodes through message flooding between nodes of same community and between nodes of different communities respectively as proposed in [1]. But, this approach involves considerable overhead. To reduce overhead, the paper also proposes to use cumulative average of unit-time node degree based on past encounters. The approach is called as C-Window. But, as shown in the paper, this approximation reduces protocol performance.

We identify hub and gateway nodes from overlapping community structure itself without doing message flooding. Overlapping community structure can be found in a distributed manner by using simple extension of AD-SIMPLE method [22]. AD-SIMPLE maintains only the current community of a node. When a node travels to a different community, members of the new community are added based on contact duration and previous community members are dropped through aging process. We propose to maintain these dropped community members as a separate community of the node. This simple extension can find different communities in which a node is a member instead of maintaining only the current community.

Socially-aware routing protocols such as BubbleRap [1] need to find such structure for efficient forwarding anyway. We make use of this structure for identifying hub and gateway nodes also. Each node of a community estimates its local and global popularity values for the community independently using following information: (1) community members of all communities in which the node is a member. The node finds this using simple extension of AD-SIMPLE and (2) locations the node visits in these communities which is assumed to be known to each node individually. Simulation results show that hub and gateway nodes identified through our method are closely matching with hub and gateway nodes identified through message flooding.

5.1. Hub nodes

In overlapping community structure, if a node with high local degree in a community is also member of other communities, then it is important to consider for how much fraction of time it will remain in the community. i.e. a node's local popularity in a community is dependent not only on local degree but also on the fraction of time it will spend in that community.

To find out the fraction of time for which a node will be in a particular community out of all communities in which it is a member, we model community structure as Markov chain. Communities are considered as states in Markov chain. Fig. 1 shows Markov chain of a node having membership in communities i, j and k. In the figure, P_{XY} with $X = i, j, k$ and $Y = i, j, k$

represent probability with which the node will travel from one community to another community or will remain in the same community in the next flight. If we can find these probabilities, then steady state probability vector of the node will give the fraction of time for which a node will be in each community. If locations of the communities are known, then we can calculate distances between them. It is known that human flight distances follow power-law distribution [6]. We can use this property to find out transition probabilities.

Let a node be a member of M communities and let it be currently in community $i = 0$. Let other communities be numbered from 1 to $M - 1$ in the increasing order of their distances from current community where the distance between community i and j (d_{ij}) is defined as the distance from any one of the associated cells of the node in community i to the nearest associated cell of the node in community j. Let probability of node movement from community i to community j be P_{ij}, minimum and maximum distance a node travels in a flight are $minD$ and $maxD$, d_0 is the maximum distance the node travels to reach a location within its community and D is power-law exponent. Then,

$$P_{ii} = \int_{minD}^{d_0} cx^{-D}dx$$

$$P_{ij} = \int_{d_{i(j-1)}}^{d_{ij}} cx^{-D}dx; \quad i = 0;\ j = 1, 2, \ldots, M-1 \tag{3}$$

Where $c = \dfrac{1}{\int_{minD}^{maxD} x^{-D}dx}$ is normalizing constant.

Let $\mathbf{w}$ be the steady state probability vector of this Markov chain then it is known that [23]

$$w_0 + w_1 + \cdots + w_{M-1} = 1$$

$$\mathbf{wP} = \mathbf{w}. \tag{4}$$

From these equations, steady state vector $\mathbf{w}$ can be found which represents the fraction of time for which the node will be in each community. Let for the community i, local popularity of the node in community i be LPt_i and the number of places the node visits in community i be $Local_i$. Then,

$$LPt_i = w_i \times Local_i. \tag{5}$$

Each node in a community estimates its local popularity using the above method. Then, community nodes exchange their local popularity values with each other. From this, each node identifies given percentage of nodes having higher local popularity than other nodes as hub nodes of the community independently.

5.2. Gateway nodes

Intuitively, for a node to qualify as gateway node of a community, it should move from one community to other community frequently. It should also spend more time in the community to carry packets of the community to other communities. Moreover, a node having membership in larger communities should be preferred over a node having membership in smaller communities.

Mathematically, a node having less average self-transition probability will move between communities more frequently. Let average self-transition probability of a node be P_{avg}, transition matrix of the node be $\mathbf{P}$ and M be the number of communities in which the node is a member. Then,

$$P_{avg} = \frac{1}{M} \sum_{i=0}^{i=M-1} P_{ii}. \tag{6}$$

Similarly, a node having high steady state probability w_i for community i will spend more time in community i. Let S_i represent summation of sizes of communities in which a node is member except community i.

Let global popularity of a node in community i be GPt_i. Then,

$$GPt_i = (1 - P_{avg}) \times w_i \times S_i. \tag{7}$$

Each node in a community estimates its global popularity using above method. Then, community nodes exchange their global popularity values with each other. From this, each node identifies given percentage of nodes having higher global popularity than other nodes as gateway nodes of the community independently.

Please cite this article in press as: Z. Narmawala, S. Srivastava, Community Aware Heterogeneous Human Mobility (CAHM): Model and analysis, Pervasive and Mobile Computing (2015), http://dx.doi.org/10.1016/j.pmcj.2014.12.008

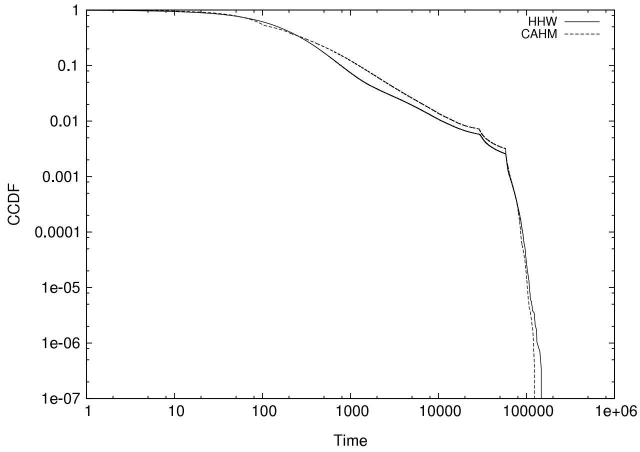

Fig. 2. CCDF of inter-contact time.

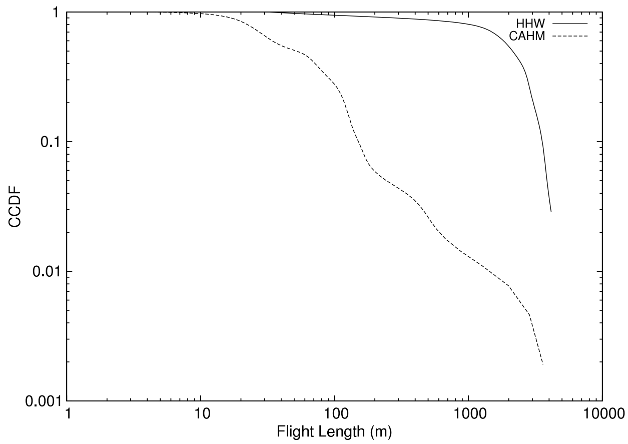

Fig. 3. CCDF of flight lengths.

6. Simulation results

We have implemented HHW and CAHM model in ONE simulator [24]. We simulated HHW and CAHM model with following scenario. There are 200 nodes in a simulation plane of 5000 m $\times$ 5000 m, divided into a grid of 62,500 cells of 20 m $\times$ 20 m each. The transmission range of node is 20 m. For HHW, the speed of a node is uniformly distributed between 1 and 6 m/s. For CAHM, speed follows Eq. (2). Pause time is generated using power-law distribution with exponent 2 between range 0 and 1000 s. We generated 4-clique communities. i.e. with $k = 4$. We set $PR^{MN} = 3$, $PR^{Osize} = 2$, $PR^{Csize} = 1.2$, $PR^{Local} = 2.5$ and flight length exponent $PR^{D} = 2$. All these values are in the range recommended for these exponents in the literature from various real traces [6,13,14]. For comparison, we used same community structures for both HHW and CAHM model.

To verify that in CAHM also, inter-contact time distribution is power-law with exponential cutoff, we simulated HHW and CAHM model for two days and each day has been divided into three identical periods of 8 h each. We generated three different overlapping community structures for each period using random variables which follow power-law distribution with exponents for different quantities as specified earlier. As shown in Fig. 2, Complementary Cumulative Distribution Function (CCDF) of inter-contact time of CAHM matches with HHW which is shown to be matching with CCDF of inter-contact time of real traces [17] following power-law distribution with exponential cutoff.

To show how a node with membership in multiple communities in HHW prefers long distances over short distances, we selected a node which is member of 11 different communities and recorded flight lengths of all its flights both in HHW and CAHM. As shown in Fig. 3, flight lengths in CAHM follow power-law distribution while in HHW, flight lengths do not follow power-law distribution.

Please cite this article in press as: Z. Narmawala, S. Srivastava, Community Aware Heterogeneous Human Mobility (CAHM): Model and analysis, Pervasive and Mobile Computing (2015), http://dx.doi.org/10.1016/j.pmcj.2014.12.008

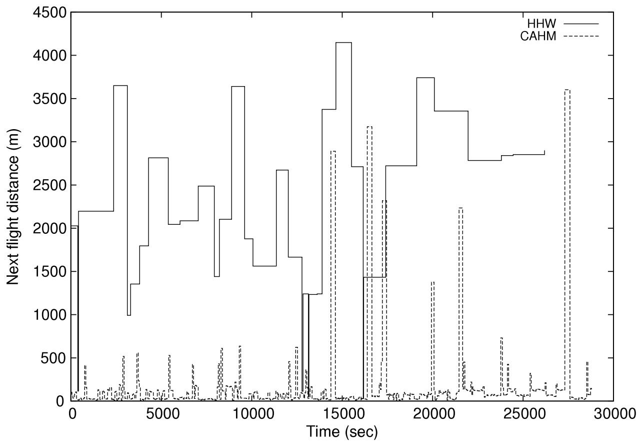

Fig. 4. Distance covered in each flight vs. time (large distance implies inter-community transition).

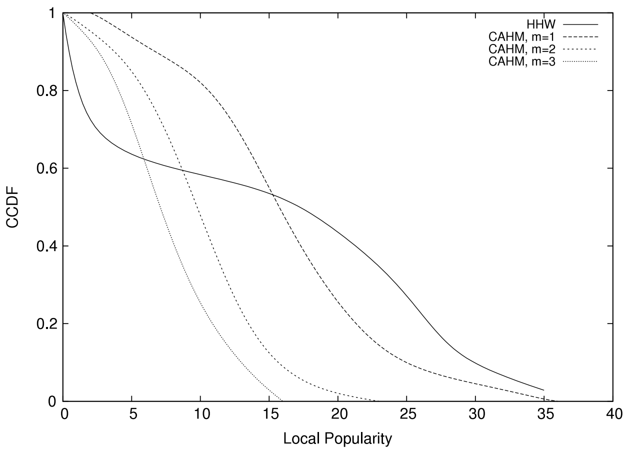

Fig. 5. CCDF of local popularity.

As shown in Fig. 4, in HHW, a node takes high number of consecutive long flights. i.e. it jumps from one community to another community while in CAHM, the node moves within one community most of the time before jumping to another community.

In simulation, in order to measure local popularity, we count the number of encounters between all pairs of nodes in a community. We define node's popularity as the number of nodes which it has encountered more number of times than average number of encounters in the community.

Fig. 5 shows CCDF of local popularity of nodes in a community of size 70 in HHW and in CAHM with multiplier $m = 1, 2$ and 3 (Eq. (1)). As conjectured, it is evident from the figure that in HHW, there are too many nodes with high local popularity than expected while CAHM generates local popularity as expected. Further, in CAHM, with increase in the value of m (i.e. with decrease in denseness) local popularity of all nodes decreases. So, by changing the value of m, one can control local popularity of nodes in a community.

6.1. *Effect of mobility models on the performance of routing protocols*

In order to analyze effect of CAHM mobility model on the performance of routing protocols, we compared delivery ratio and delay of Epidemic routing [25] and BubbleRap [1] protocols with CAHM, HHW, CMM and RWP mobility models. HHW and CMM incorporate some of the properties of human mobility listed in Section 2 while RWP does not incorporate any of these properties. Epidemic routing does not exploit any of the properties of human mobility while BubbleRap exploits community structure and heterogeneous popularity of nodes to achieve better performance.

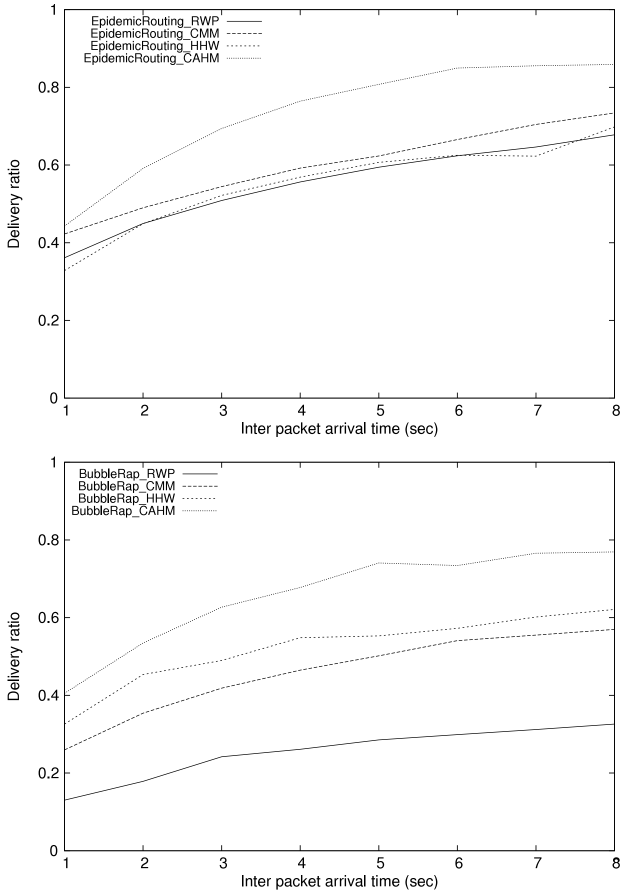

Fig. 6. Delivery ratio vs. inter packet arrival time.

Implementation of Epidemic routing is available in ONE simulator. For CMM, Musolesi et al. [15] have provided an implementation which generates mobility scenario for Network Simulator (NS-2) simulator. We have implemented conversion class in ONE simulator which moves nodes in ONE simulator as per mobility scenario generated for NS-2.

We move 200 nodes in the area of 13 000 m × 13 000 m. In CAHM, speed of individual flight is calculated based on Eq. (2). The equation and values of terms in the equation are derived in [6] from mobility traces of four different places. For our simulation scenario, with CAHM, average speed turned out to be 8.23 m/s. So, we set average flight speed as 8.23 m/s with uniform distribution with range [0.35, 16.11] for other mobility models. In CAHM, pause time is power-law distributed with exponent 2.39, minimum pause time 25 s and maximum pause time 9500 s. These values are derived in [6] from mobility traces of a university campus referred as 'Campus I' in the paper. For our simulation scenario, with CAHM, average pause time turned out to be 82.32 s. So, we set average pause time as 82.5 s with uniform distribution with range [25, 140] for other mobility models. In CAHM, flight length exponent is 1.29. Transmission speed of nodes is 3 MBps and transmission range is 40 m. Nodes are not buffer constrained. Packets are of 8 kB size. We vary inter packet arrival time from 1 to 8 s.

For HHW, CAHM and CMM, the number of communities is 13. For HHW and CAHM, cell size is 250 m × 250 m and remaining parameters are as specified at the start of the section. For CMM, the number of rows and columns are 13. It is chosen such that node density within a community remains same in HHW, CAHM and CMM. Re-wiring probability for CMM is 0.1.

It is evident from Fig. 6 that delivery ratio of both Epidemic routing and BubbleRap routing is significantly greater with CAHM mobility model as compared to other three mobility models. As delivery ratio of BubbleRap with CAHM is much higher than delivery ratio with HHW, it can be said that delivery ratio increases significantly with CAHM because of heterogeneous local popularity of nodes as observed in real mobility traces and also because of following additional properties of human mobility which are not incorporated in HHW and CMM but are part of CAHM: Levy walk nature of human mobility, speed of nodes as a function of distance to be traveled. Fig. 7 shows that with CAHM, average delivery delay of both Epidemic routing and BubbleRap routing is minimum as compared to other mobility models.

It is evident from Figs. 6 and 7 that difference in delivery ratio and delivery delay of Epidemic routing and BubbleRap with CAHM, HHW and CMM model is less pronounced as compared to with RWP even though Epidemic routing does around 50%

Please cite this article in press as: Z. Narmawala, S. Srivastava, Community Aware Heterogeneous Human Mobility (CAHM): Model and analysis, Pervasive and Mobile Computing (2015), http://dx.doi.org/10.1016/j.pmcj.2014.12.008

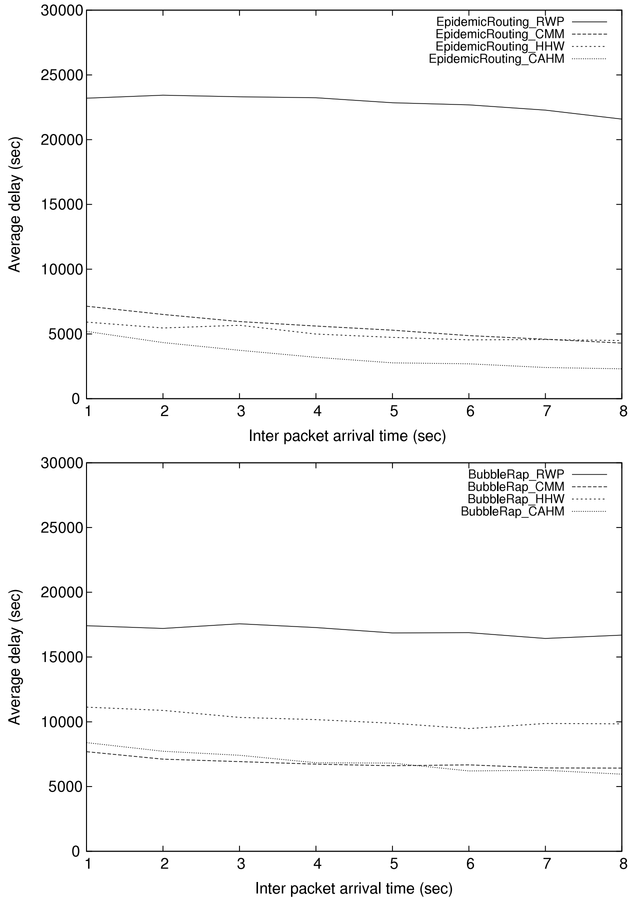

Fig. 7. Average delay vs. inter packet arrival time.

more transmissions per packet than BubbleRap routing in our simulation. The results confirm that exploiting community structure and heterogeneous node popularity significantly improves performance. The results also show that due to less realistic mobility models, simulation significantly under estimates performance of protocols.

6.2. Validation of hub identification model

To validate steady state probabilities found using Eq. (4), in simulation, we kept log of amount of time a node has spent in each community in which it is member. Dividing these times by total simulation time gives the fraction of time a node has spent in each community. It turns out that, these values are very closely matching with those found using Eq. (4).

To validate our hub identification method, we generated overlapping community structure with parameters specified at the start of this section. The implementation generated 13 communities with some random seed. Then, using Eq. (5) we calculated local popularity of all member nodes for each community and ordered them in a separate list for each community. We compared the ordered lists with lists generated using following two methods: (1) we measured local popularity of nodes in each community in simulation as explained earlier in this section and ordered them. (2) For each community, we sent packets from different sources of a community to different destinations of the same community through Epidemic routing in simulation. Then, we counted the number of times a node of the community is on shortest paths followed by packets to reach to destinations and ordered nodes based on these counts.

Then, we used Spearman's rank correlation coefficient (ρ) [26] to compare two ordered lists. Average of ρ of all communities comes out to be 0.9890 for our method and the first method. Similarly, for our method and the second method, average of ρ of all communities is 0.9980. As, $\rho = 1$ means ordered list of our estimate is perfect monotone function of ordered list of measured local popularity values, the result confirms that our method identifies correct hub nodes.

6.3. Validation of gateway identification method

Using Eq. (7), we calculated global popularity values of all member nodes for each community and ordered them in a separate list for each community. We compared the ordered lists with lists generated using following two methods:

Please cite this article in press as: Z. Narmawala, S. Srivastava, Community Aware Heterogeneous Human Mobility (CAHM): Model and analysis, Pervasive and Mobile Computing (2015), http://dx.doi.org/10.1016/j.pmcj.2014.12.008

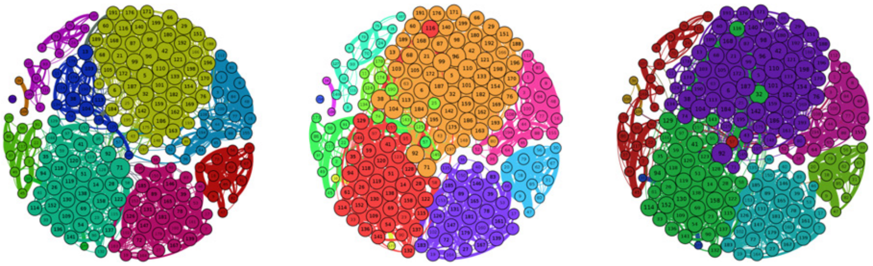

Fig. 8. Community structure detected by modularity algorithm in consecutive time intervals using Gephi.

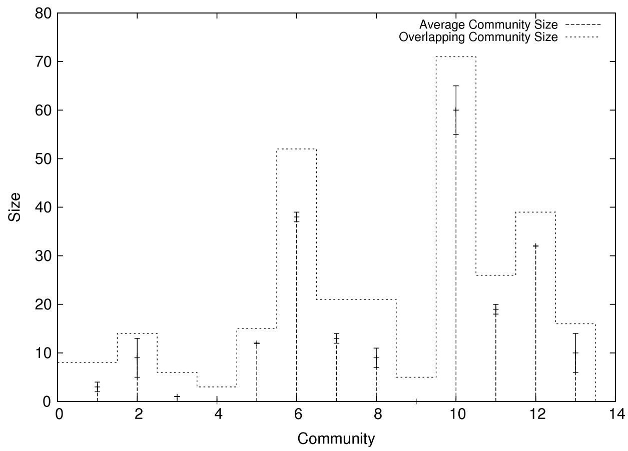

Fig. 9. Size vs. community.

(1) we accumulated encounter information of all nodes through simulation. From the resultant weighted graph, for each community, we calculated betweenness centrality values of nodes having membership in other communities and ordered these nodes as per their betweenness centrality values. While calculating betweenness centrality for a node, we considered only those shortest paths for which source is in the community and destination is in another community. (2) For each community, we sent packets from different sources of a community to different destinations of other communities through Epidemic routing in simulation. Then, we counted the number of times a node of the community is on shortest paths followed by packets to reach to destinations and next hop is from other community. We ordered nodes based on these counts.

Average of ρ of all communities comes out to be 0.9770 for our method and the first method. Similarly for our method and the second method, average of ρ of all communities is 0.9880. The result confirms that our method identifies correct gateway nodes.

6.4. *Analysis of overlapping community structure*

Analysis of properties of overlapping community structure formed by human mobility can give important in-sights for designing better forwarding mechanisms. We analyze properties such as actual community sizes, probability of community existence and fraction of hub and gateway nodes present in the community on an average.

We created 172 separate contact graphs based on the number of encounters between all pairs of nodes for 172 intervals of 1000 s. An edge between two nodes is added if the number of encounters is greater than some threshold. We imported this graph in Gephi [27]. Then, we ran modularity algorithm which is an implementation of the algorithm presented in [28] with resolution $= 1$ [29] to find out communities. Fig. 8 shows communities found in first three consecutive time intervals. Nodes with same color belong to the same community. Size of a node is proportional to node degree. A node with different color in a cluster of differently colored nodes means that the node's community has changed. Further, communities thus found closely match with the community structure of mobility model. It shows that CAHM is successful in creating intended community structure of human beings.

Fig. 9 shows average of sizes of each community in 172 different intervals. As, in CAHM, a node can be part of multiple communities but modularity algorithm will put it with only one of these communities, community sizes found through modularity algorithm are less than community sizes generated in CAHM.

Fig. 10 shows that smaller communities are transient and they get merged with bigger communities most of the time. For this, we found community structures using modularity algorithm for 172 different intervals and counted how many number of times a community is detected by the algorithm. From the figure it is evident that for 13 communities with power-law

Please cite this article in press as: Z. Narmawala, S. Srivastava, Community Aware Heterogeneous Human Mobility (CAHM): Model and analysis, Pervasive and Mobile Computing (2015), http://dx.doi.org/10.1016/j.pmcj.2014.12.008

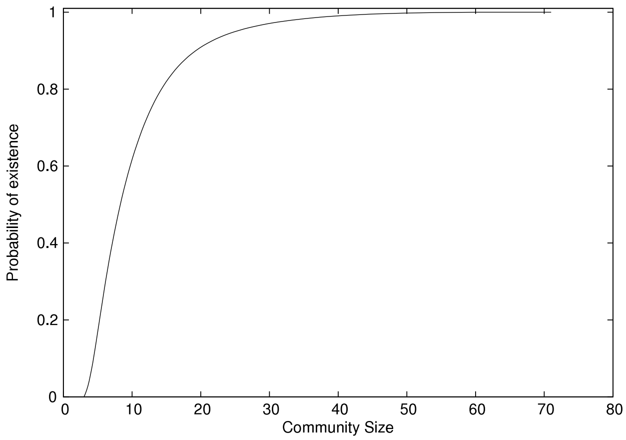

Fig. 10. Probability of existence vs. community size.

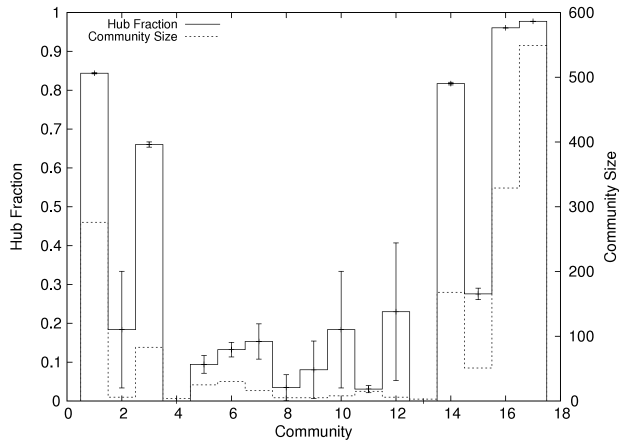

Fig. 11. Hub fraction vs. community number.

community sizes having total 200 nodes, knee point is around community size of 20. The result is very useful for community based forwarding as it serves as a guideline for deciding whether a group of nodes should be considered as part of separate community or not.

Fig. 11 shows average of the fraction of identified hub nodes present in communities in different intervals. For this result, community structure from 1000 nodes is created. We identified first 30% of nodes from the sorted list of locally popular nodes as hub nodes. Community size is shown on the second y axis. It is evident that higher fraction of hub nodes remain present in larger communities with less standard deviation and lower fraction of hub nodes remain present in smaller communities with high standard deviation. This information can be used while deciding percentage of nodes to be considered as hub nodes and also for other forwarding decisions.

Fig. 12 shows average of fraction of identified gateway nodes present in communities in different intervals. The result is similar to the result for hub nodes. Further, it is evident from two results that fraction of gateway nodes present in a community is much lower than fraction of hub nodes present.

7. Conclusion

HHW mobility model is based on heterogeneous popularity and overlapping community structure in social network. In this paper, we identify shortcomings of HHW model and propose Community Aware Heterogeneous Human Mobility Model (CAHM) with four modifications: incorporation of Levy walk nature of human mobility, treatment of the number of cells and the number of nodes in a community as separate parameters, calculation of speed based on distance to be traveled and power-law pause time. Simulation results demonstrate that CAHM successfully generates flight lengths with power-law distribution while in HHW flight lengths are uniformly distributed. Further, movement of individuals in CAHM is as per rational human behavior of preference of nearby locations over far-away locations while in HHW it is not. Results also establish that CAHM generates desired heterogeneous local popularity of nodes while HHW generates too many highly popular nodes.

Please cite this article in press as: Z. Narmawala, S. Srivastava, Community Aware Heterogeneous Human Mobility (CAHM): Model and analysis, Pervasive and Mobile Computing (2015), http://dx.doi.org/10.1016/j.pmcj.2014.12.008

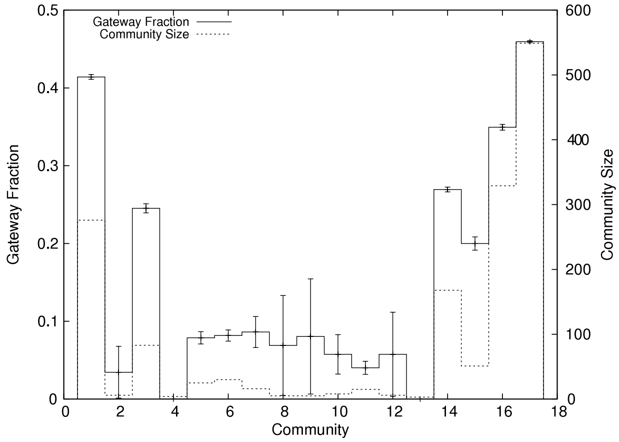

Fig. 12. Gateway fraction vs. community number.

Delivery ratio of Epidemic routing and BubbleRap routing increases by 23%–26% with CAHM as compared to HHW in our simulation setup. Average delivery delay of these protocols decreases by 40%–50% with CAHM as compared to HHW. These results show that additional properties of human mobility incorporated in our mobility model such as heterogeneous popularity of nodes as observed in real life traces, Levy walk nature of human mobility and speed as a function of distance have significant impact on performance of routing protocols.

Our analysis of overlapping community structure establishes that small communities are transient. As per simulation results, threshold for the same is around 10% of total number of nodes in the network. Further, higher fraction of hub and gateway nodes remain present in larger communities with less standard deviation as compared to smaller communities. Also, fraction of gateway nodes present in a community is much lower than fraction of hub nodes present. These result give important in-sights for designing better forwarding mechanisms for mobile social networks.

We also propose methods based on mathematical models to identify hub and gateway nodes of communities from overlapping community structure itself without doing message flooding. Simulation results show that our models correctly identify hub and gateway nodes.

References

[1] P. Hui, J. Crowcroft, E. Yoneki, Bubble Rap: social-based forwarding in delay-tolerant networks, IEEE Trans. Mob. Comput. 10 (11) (2011) 1576–1589.
[2] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, J. Scott, Impact of human mobility on opportunistic forwarding algorithms, IEEE Trans. Mob. Comput. 6 (6) (2007) 606–620.
[3] C. Boldrini, M. Conti, A. Passarella, Impact of social mobility on routing protocols for opportunistic networks, in: World of Wireless, Mobile and Multimedia Networks, 2007, WoWMoM 2007, in: IEEE International Symposium on, IEEE, 2007, pp. 1–6.
[4] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, C. Diot, Pocket switched networks and human mobility in conference environments, in: Proceedings of the 2005 ACM SIGCOMM Workshop on Delay-Tolerant Networking, ACM, 2005, pp. 244–251.
[5] N. Eagle, A.S. Pentland, Reality mining: sensing complex social systems, Pers. Ubiquitous Comput. 10 (4) (2006) 255–268.
[6] I. Rhee, M. Shin, S. Hong, K. Lee, S.J. Kim, S. Chong, On the Levy-walk nature of human mobility, IEEE/ACM Trans. Netw. (TON) 19 (3) (2011) 630–643.
[7] M.C. Gonzalez, C.A. Hidalgo, A.-L. Barabasi, Understanding individual human mobility patterns, Nature 453 (7196) (2008) 779–782.
[8] E. Hyytiä, H. Koskinen, P. Lassila, A. Penttinen, J. Roszik, J. Virtamo, Random waypoint model in wireless networks, Networks and Algorithms: complexity in Physics and Computer Science, Helsinki.
[9] A. Nain, R. Groenevelt, E. Altman, Relaying in mobile ad hoc networks: the Brownian motion mobility model, Wirel. Netw. 12 (5) (2006) 561–571.
[10] W.J. Hsu, T. Spyropoulos, K. Psounis, A. Helmy, Modeling spatial and temporal dependencies of user mobility in wireless mobile networks, IEEE/ACM Trans. Netw. 17 (5) (2009) 1564–1577.
[11] A. Mei, J. Stefa, SWIM: a simple model to generate small mobile worlds, in: INFOCOM 2009, IEEE, 2009, pp. 2106–2113.
[12] K. Lee, S. Hong, S.J. Kim, I. Rhee, S. Chong, SLAW: a new mobility model for human walks, in: INFOCOM 2009, IEEE, 2009, pp. 855–863.
[13] M.E. Newman, The structure and function of complex networks, SIAM Rev. 45 (2) (2003) 167–256.
[14] G. Palla, I. Derényi, I. Farkas, T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society, Nature 435 (7043) (2005) 814–818.
[15] M. Musolesi, C. Mascolo, Designing mobility models based on social network theory, ACM SIGMOBILE Mob. Comput. Commun. Rev. 11 (3) (2007) 59–70.
[16] C. Boldrini, A. Passarella, HCMM: modelling spatial and temporal properties of human mobility driven by users' social relationships, Comput. Commun. 33 (9) (2010) 1056–1074.
[17] S. Yang, X. Yang, C. Zhang, E. Spyrou, Using social network theory for modeling human mobility, IEEE Netw. 24 (5) (2010) 6–13.
[18] F. Ekman, A. Keränen, J. Karvo, J. Ott, Working day movement model, in: Proceedings of the 1st ACM SIGMOBILE Workshop on Mobility Models, ACM, 2008, pp. 33–40.
[19] D.J. Watts, Small Worlds: the Dynamics of Networks Between Order and Randomness, Princeton University Press, 1999.
[20] D. Karamshuk, C. Boldrini, M. Conti, A. Passarella, SPoT: representing the social, spatial, and temporal dimensions of human mobility with a unifying framework, Pervasive Mob. Comput. 11 (2014) 19–40.
[21] P. Pirozmand, G. Wu, B. Jedari, F. Xia, Human mobility in opportunistic networks: characteristics, models and prediction methods, J. Netw. Comput. Appl. 42 (2014) 45–58.
[22] E. Borgia, M. Conti, A. Passarella, Autonomic detection of dynamic social communities in opportunistic networks, in: 2011 The 10th IFIP Annual Mediterranean Ad Hoc Networking Workshop (Med-Hoc-Net), IEEE, 2011, pp. 142–149.
[23] C.C.M. Grinstead, J.L. Snell, Introduction to Probability, American Mathematical Soc., 1997.

Please cite this article in press as: Z. Narmawala, S. Srivastava, Community Aware Heterogeneous Human Mobility (CAHM): Model and analysis, Pervasive and Mobile Computing (2015), http://dx.doi.org/10.1016/j.pmcj.2014.12.008

[24] A. Keränen, J. Ott, T. Kärkkäinen, The ONE simulator for DTN protocol evaluation, in: Proceedings of the 2nd International Conference on Simulation Tools and Techniques, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009, p. 55.
[25] A. Vahdat, D. Becker, et al., Epidemic routing for partially connected ad hoc networks, Tech. Rep., Technical Report CS-200006, Duke University, 2000.
[26] W. Pirie, Spearman rank correlation coefficient, Encyclopedia of statistical sciences.
[27] M. Bastian, S. Heymann, M. Jacomy, Gephi: an open source software for exploring and manipulating networks, in: ICWSM, 2009, pp. 361–362.
[28] V.D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, J. Stat. Mech. Theory Exp. 2008 (10) (2008) P10008.
[29] R. Lambiotte, J.-C. Delvenne, M. Barahona, Laplacian dynamics and multiscale modular structure in networks. ArXiv Preprint arXiv:0812.1770.

Please cite this article in press as: Z. Narmawala, S. Srivastava, Community Aware Heterogeneous Human Mobility (CAHM): Model and analysis, Pervasive and Mobile Computing (2015), http://dx.doi.org/10.1016/j.pmcj.2014.12.008