Computer Communications 33 (2010) 1056–1074

Contents lists available at ScienceDirect

Computer Communications

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

HCMM: Modelling spatial and temporal properties of human mobility driven by users' social relationships

Chiara Boldrini *, Andrea Passarella

Institute for Informatics and Telematics, National Research Council, Via G. Moruzzi 1, 56124 Pisa, Italy

ARTICLE INFO

Article history:
Received 25 November 2009
Received in revised form 18 January 2010
Accepted 19 January 2010
Available online 25 January 2010

Keywords:
Mobility models
Social network theory
Inter-contact time
Jump size
Analytical modelling

ABSTRACT

In Mobile Ad Hoc Networks (MANETs), the mobility of the network users can heavily affect the performance of networking protocols because it causes sudden connectivity changes and topological variations. This is even more important in recent promising paradigms proposed in this field, such as opportunistic and delay tolerant networks. For this reason, it is important to understand the characteristics of the user movements in order to properly handle mobility when designing networking protocols for mobile ad hoc networks. In addition, it is highly desirable to have a mobility model that accurately reproduces the user mobility, thus enabling researchers to evaluate, either analytically or by means of simulations, their protocols under realistic mobility conditions. Recently, there have been many studies aimed to uncover the nature of human movements. In this paper, based on recent literature, we identify three main properties that are fundamental to characterize human mobility. Then, we propose a mobility model (HCMM) that integrates all these three features. To the best of our knowledge, the model proposed is the first one that combines notions about the sociality of users with spatial properties observed in real users movement patterns, i.e., their preference to spend time in a limited number of popular locations and to preferentially select short distances over longer ones. We study the HCMM both through simulation and analysis. Based on this study, we highlight some of its important temporal and spatial features, and we show that they are correctly reproduced in terms of key indicators such as jump size and inter-contact time distribution.

1. Introduction

With the transition from static to mobile networks, the problem of understanding and modelling the mobility of nodes has become a key issue in the networking research area. In the last years, a very popular research topic has been that of human mobile ad hoc networks. In the conventional mobile ad hoc networking (MANET) paradigm, an Internet-like routing layer is running on the mobile nodes, and communication among two end points occurs only if there is a simultaneous multi-hop path connecting the communicating end points. Emulating a fixed-Internet routing abstraction in MANET generates too much overhead and may easily make the network collapse even in case of moderate mobility [1]. Therefore, recently the new paradigm of opportunistic networks is being explored [2]. In opportunistic networks no simultaneous multi-hop path is assumed between communicating endpoints. Disconnections and partitions are not masked by a routing protocol, but are exposed throughout the protocol stack. Thus, messages "hop" from one node to the next one only when a suitable opportunity arises to progress towards the final destination. In such an environment, knowing the main features of human mobility is clearly fundamental to design networking solutions. In this case, we come across mobility twice. First, the better the knowledge on human mobility, the more smartly its properties can be exploited in order to design efficient networking solutions. Second, once a networking protocol has been designed, a preliminary evaluation of such model must be performed before moving to the real implementation. Therefore, understanding mobility is not enough: a mobility model is needed in order to evaluate networking protocols either through analysis or simulations under realistic mobility conditions.

Recently, the problem of understanding human mobility has drawn the attention of researchers working on complex systems and mobile ad hoc networks. Different experiments have been conducted, relying on a variety of measurement technologies (surveys [3], banknote tracking [4], mobile phone traces [5], GPS equipped portable devices [6], WLAN associations [7], Bluetooth connections [8], etc.). Summarizing the most important results of all these research studies, three key properties can be identified. First, user movements are conditioned by their social relationships [3]. Specifically, the larger the social network, the higher the mobility [9], and vice versa. This implies that nodes will move more frequently and will visit more locations if they have many "friends"

* Corresponding author. Tel.: +39 050 3153504; fax: +39 050 3152593.
E-mail addresses: chiara.boldrini@iit.cnr.it (C. Boldrini), andrea.passarella@iit.cnr.it (A. Passarella).

0140-3664/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.comcom.2010.01.013

scattered all over the network. Second, users tend to visit just a few locations, where they spend the majority of their time [10]. For example, if we consider a metropolitan area, not all places show the same user densities: there will be concentration points, such as malls, universities, or highways, where a lot of people roam frequently and where users periodically come back after a period of absence. Third, users prefer shorter paths to longer ones [5], i.e., users usually travel over short distances and sometimes they move farther away. For example, short distances can be associated with commuting to and from work, while the sporadic long jumps might be due to a trip on a day off. These three properties have thus emerged as key building blocks of real user movement patterns. Surprisingly enough, as discussed hereafter, current mobility models do not fully incorporate all of them. In this paper we thus present the Home-cell Community-based Mobility Model (HCMM) in order to fill this gap.

Current proposals for human mobility modelling can be classified into two main categories. In social-based models node movements are steered using the social relationships between users [11–13]. On the opposite, location-based solutions use only the notion of preferred locations to set up the commuting schedule of nodes [10,14]. However, none of these approaches allows to satisfy all three characteristics of human mobility at the same time. The social dimension is completely absent from location-based models. Preferential locations might be present in social-based models, because nodes belonging to the same community tend to roam in the same area. However, as there is no explicit bound between social communities and physical locations, preferential locations may only appear as an "accident", and it is very difficult, often impossible, to control the associations between social communities and geographical areas. In Section 4.1 we provide an example of such issues, together with a mathematical model that shows how poorly movements in the physical space can be controlled if we only consider social relationships among users. For what concerns the preferential selection of short distances, very few models, only belonging to the location-based class, have accounted for it so far, and never including the effect of nodes' sociality.

Starting from these considerations, in this paper we propose HCMM, a mobility model that joins social and location attractions, but above all incorporates all the three driving forces of human movements that we have identified above. HCMM was born as an evolution of an existing social-based mobility model, called CMM [12], of which it inherits the social graph structure. Differently from CMM, HCMM integrates the social nature of mobility with its spatial dimension. For HCMM we provide analytical and simulation evidence of its controllability. Furthermore, we show ability to reproduce the main *temporal* and *spatial* features of real human mobility. With respect to the latter point, we consider the standard indices used to analyse the statistical properties of human mobility traces, which highlight the three key features we have discussed. Temporal properties of mobility include the *contact time*, i.e., the distribution of the duration of a contact between two nodes, and the *inter-contact time* (ICT), i.e., the distribution of the time between two consecutive contacts between nodes. Given that messages are exchanged only when nodes are in radio range of each other, the longer the contact, the more the messages that can be exchanged. In addition, given that new messages can be injected in the network at any time, a single sporadic contact cannot be enough for delivering messages. Thus, the distribution of the inter-contact times between nodes is very important: the more frequently nodes get in touch, the more the opportunities for exchanging messages. Among those two indices, inter-contact times are considered to be the most important one. For example, because they fundamentally impact on the behaviour of networking protocols [8]. Many experiments have been conducted in order to uncover the nature of inter-contact times (see Section 2) and there is now a general agreement on the fact that inter-contact times are distributed according to a power law with a final exponential cut-off. Spatial properties of human mobility include the distribution of the size of human *jumps* (or flights), defined as the path between two consecutive waypoints (i.e., points where the node stops for a while between two consecutive movements). Also in this case many experiments have been conducted (see Section 2) and, again, the jump size has been found to be reasonably approximated by a truncated power law. This has a huge impact on the way messages are spread across the network. Nodes that travel over longer distances become bridges between far away communities, and this can be decisive, e.g., for the spread of a virus [15]. In this work we focus on inter-contact time and jump size as the metrics that summarise the most relevant temporal and spatial properties of human movements, and we evaluate against them the realism of the mobility models that we consider. Specifically, for HCMM we provide analytical and simulation evidence of its ability to reproduce the distribution of inter-contact time and jump size found in real mobility traces.

The paper is structured as follows. In Section 2 we survey the existing studies on human mobility, while in Section 3 we describe existing approaches to mobility modelling. In Section 4 we introduce the CMM model, with an emphasis on one of its shortcomings that we have named "gregarious behaviour". We show analytically the emergence of this behaviour in the vast majority of settings and we explain why this is – in general – an undesired feature. Then, we introduce and analyse one at a time the additional features needed in HCMM to incorporate all the three key aspects of human mobility we have discussed. First of all, in Section 5 we show how to include bounds between social communities and physical locations in the mobility model. In Section 6 we focus on the spatial properties of the resulting mobility and we provide an analytical model that describes the jump size distribution. This model is then used to prove that additional extensions are required to reproduce realistic jump size distributions. Thus, in Section 7 we complete the definition of the HCMM model to include them. Finally, by means of simulations, we evaluate the new model and we check that the jump size and ICTs that it produces have the same features of those extracted from real mobility traces.

2. Brief history of measurement studies on human mobility

Given their impact on the delay experienced by messages, ICTs were the first property of human mobility that was extensively studied by researchers. The first available traces tracked users' associations with WLANs and were collected to study the users' usage patterns of wireless networks rather than users' mobility patterns [16,7]. For this reason, some assumptions were needed in order to extract information about inter-contact times from these traces. The strongest of these assumptions is that two users under the same AP coverage are able to communicate directly, which is in general not true. An additional problem with these traces is that they miss all contacts between nodes that do not occur within an AP radio range. Similar problems characterize the data set in [17], which tracks associations with GSM towers. The analysis of these traces yielded the first evidence for power law distributed inter-contact times. In order to fix the problems with infrastructure-based traces, targeted experiments were designed to directly collect pair-wise contacts between nodes (mostly using Bluetooth associations) in a campus [17], conference environment [18], and research lab [8]. The analysis of these traces confirmed the power law nature of ICTs [8]. Analysing these same traces, authors of [19] highlighted an exponential cut-off in the tail of the distribution of ICTs. The exponential component in pair-wise inter-contact times is also confirmed by the analysis in [20].

More recently, alongside ICTs, which relate to the temporal properties of mobility, spatial characteristics of user movements have drawn a lot of attention because of the impact that they can have on the way, e.g., viruses or messages spread across the network. Assessing the properties of human travelling patterns has not been a simple matter. At the beginning, only indirect measures were available. In [4] this indirect measure was the path travelled by a banknote, considered as an alias for the path travelled by the human that carried such banknote. The banknote experiment concluded that humans travel according to a Lévy flight. Criticism to this work was based on the consideration that if the convolutions of movements of different persons (because this was what the banknote displacements actually reproduced) are Lévy fights, it does not imply that the movements of individuals are Lévy flights. This criticism was overcome in [5], where the truncated power law distribution of jump size was confirmed using a large dataset comprising traces of mobile phone users. This same result has been obtained in [6], using GPS traces related to different settings (college campus, city, theme park, state fair). This work also highlights the fact that the power law in the movement patterns does not emerge as a result of external constraints, e.g., obstacles along the movement trajectory, but instead it seems to be generated by human intentions, as already suggested in [21].

3. State of the art on models for human mobility

Due to the importance of mobility in wireless networks, researchers have tried to come up with mobility models being both easy to handle (from a mathematical standpoint) and able to reproduce key features (from the network performance standpoint) of real traces. The first solutions proposed were based on extremely tractable mathematical models but failed to confront with realistic aspects of mobility. These models are often called random mobility models, because waypoints of consecutive movements are chosen uniformly at random. Popular examples of such models are the Random Waypoint and Random Direction mobility models [22]. Random mobility models were found to be not very realistic as soon as traces of actual human mobility became available, thus new mobility models have been proposed that try to reproduce the most important properties of real human mobility. These proposals can be categorized based on which of the three driving forces of human mobility are included in the model (social relationships as a driving force of movements, preference for particular locations as movement waypoints, preference for travelling over short than over long distances).

The organization of human society into communities [23] leads to spatial dependency among the users in the network. In fact, if two nodes A and B belong to the same community, they tend to spend more time together than nodes in different communities. This correlation is captured by mobility models that use social relationships among users as the drivers of nodes' movements. To the best of our knowledge, the first example of this class of models was presented in [11]. Users are organized in groups, then each group is associated to a physical location. The movements of users follow a predefined schedule across the locations associated with their group. This work, however, lacks a rigorous mathematical definition of the relationships among users. A more flexible social-based mobility model, called CMM, was presented in [12]. This model is the starting point of our work on mobility models and it will be described in detail in Section 4, together with its strengths and weaknesses. Here we anticipate that the CMM model does not respect the second and the third properties of human mobility. As we will show analytically in Section 4.1, users' positions under the CMM model are not predictable. Specifically, in the majority of configurations all users collapse into a single location, thus practically overthrowing the initial setting of the system. Recently, in [13] the SIMPS model has been proposed. In the SIMPS model, nodes move according to two behavioural rules: the *socialize* behaviour corresponds to a node moving towards acquaintances, the *isolate* rules implies instead a node escaping from strangers. These two behaviours alternate according to a feedback decision-making process which balance the volume of current social interactions against the volume of interactions needed by the node. While both being social-based mobility models, the social considerations on which SIMPS is built upon are quite different from those at the basis of CMM.

So called location-based models cope with the second aspect of mobility. Usually, they define a set of "preferred" locations for each user and describe the algorithm according to which users move across these locations. One of the most flexible models among the ones in this class is the Time-Variant Community model (TVC) [10]. In TVC, each user can be either in a normal movement period or in a concentration period. In each period a node is assigned to a (different) community, that represents the frequently visited location for that node. Then, each node moves within its community or across the whole network according to a two-state Markov chain, having different transition probabilities depending on the period (normal movement or concentration) in which the node is located. Basically, what authors of [10] do is to model different locations (here, communities) that realistically can have different popularity, e.g., at different time of the day (normal movements or concentration period). While being analytically tractable, the TVC model fails to explicitly include correlations across nodes' movements, typical of a social environment where friends, relatives or colleagues spend their time together.

To the best of our knowledge, there exists only two mobility models that take the third property of mobility, i.e., the preferential selection of short distances, into account. The first model, called SLAW [24], explicitly includes truncated power law jump sizes by considering waypoints separated by a power law distributed distance, and selecting the waypoint for the next movement by minimizing the distance from the current location (least action principle). The model also results in the emergence of locations that are more popular than others, thus satisfying also the second property of mobility. While being very complete, this model appears to be very complex to control and misses to consider the social properties of mobility. The second model accounting for long tailed distances is the SWIM model, proposed in [25]. Similarly to our model, SWIM relies on the concept of home location: nodes select the destination points of their movement based on their popularity among all nodes and their distance from the home point. This model has been shown to match realistic inter-contact times of real traces very accurately, but again the role of nodes' sociality has not been explored.

The work presented in this paper is an extended version of our previous contributions in [26,27]. With respect to [26], in this paper we extend HCMM to account for realistic jump size distributions. With respect to [27], we complement the model for the jump size distribution of HCMM with a generic and much more detailed analytical model for computing the distribution of jump sizes. Note that this model can be used as a starting point for computing the distribution of distances under any mobility model in which node movements occur on a grid. In addition, we also present a simplified and more tractable model built on normal distributions. Being this distribution extremely convenient from a mathematical standpoint, having such approximation opens up a variety of options for further analytical investigations. In this work we also include the evaluation of the temporal properties, in terms of inter-contact times, of the traces generated by the HCMM model.

Table 1
Notation for CMM parameters.

N	Number of nodes
C	Number of communities
n	Number of nodes per community
$l_x \times l_y$	$x-y$ size of the reference area
$g_x \times g_y$	Number of cells
p_r	Rewiring probability
w_{ij}	Weight of the social link between node i and node j
SA_{c_i}	Social attraction of cell c_i
CA_{c_i}	Attractivity of cell c_i
$[v_{min}, v_{max}]$	Minimum and maximum speed

4. Pure social-based mobility model: CMM

In this section we introduce the Community-based mobility model (CMM), a social-based mobility model which is the starting point of our work. In CMM nodes (i.e., users) share social relationships with each other, and these relationships are represented through a weighted social graph. Two nodes connected by a link in the social graph are called *friends*, otherwise they are called *non-friends*. The weight of the link expresses the strength of the social relationship. From an operational standpoint, CMM starts by dividing nodes in disjoint communities. At the system start-up the social sub-graph related to each community is complete, and all nodes belonging to the same community are friends. According to the *rewiring* probability parameter (p_r), links are uniformly rewired towards randomly selected nodes in different communities. This is the method for creating external relationships, i.e., relationships between nodes belonging to different communities. This process follows the Watts-Strogatz (or Caveman) model described in [28].

The social layout of the network is then mapped onto the spatial structure of the scenario considered. In CMM nodes move in a $l_x \times l_y$ area divided into $g_x \times g_y$ cells. Each community is initially placed uniformly at random in one of the cells of the grid.[1] The friendship between nodes is used to trigger nodes' movements. The endpoint of a movement is called *goal*. The selection of the goal for the next movement consist of three phases: (i) the selection of the next goal cell, (ii) the selection of a random point inside that cell, (iii) the selection of the speed uniformly at random over $[v_{min}, v_{max}]$. Let us consider C communities, each having assigned n nodes, and let us focus on a generic node x. The probability that a cell is selected by node x as the goal cell for the next movement is proportional to the *social attraction* of the cell. The social attraction (SA_{c_i}) exerted by a generic cell c_i on node x is computed as the sum of the weights w_{xy} of the social links between node x and all its friends that either are *roaming in or moving to* cell c_i, divided by the number of nodes that are currently in c_i (denoted as $|c_i|$ in Eq. (1)).

$$SA_{c_i} = \frac{\sum_{y \text{ in or moving to } c_i} w_{xy}}{|c_i|} \quad (1)$$

A generic cell c_i is then selected as next goal cell by node x with a probability CA_{c_i} equal to

$$CA_{c_i} = SA_{c_i} \Big/ \sum_{j=1}^{C} SA_{c_j}, \quad (2)$$

where j ranges from 1 to C because we only consider cells occupied by a community.

Notation for CMM parameters is summarized in Table 1.

By definition, CMM incorporates the sociality of nodes into the design of human mobility, thus the first driving force of human movements is explicitly embedded into the model. This is not the case for the second mobility property (preferential locations). As we show below, in CMM it is very difficult to predict and control nodes' movements (e.g., in terms of the areas where nodes are expected to roam). In particular, we have detected the emergence in CMM of what we have called *gregarious behaviour* of nodes: all the nodes of a community follow the first node that has decided to exit the community. As this can happen for all communities, it is very likely that the network ends up in a state where all the nodes roam in exactly the same cell, thus nullifying every notion of community, social structure, and preferred locations. In the following we provide an analytical model for CMM that defines the probability for the gregarious behaviour to happen. In addition, we also perform a quantitative analysis that shows that the conditions under which the gregarious behaviour occurs represents the vast majority of the operating conditions of the mobility model.

4.1. Gregarious behaviour in HCMM

In order to provide a comprehensive study of the impact of the gregarious behaviour on the CMM model, in this section we design a mathematical model of the phenomenon, and then we discuss the conditions under which it occurs. The key step to model the gregarious behaviour is to find the probability that a node remains in its starting cell despite the fact that another node has moved out. To this aim, we will study how a tagged node k that goes out of its starting cell influences the movements of the other nodes of its community. Without loss of generality, we consider a simple scenario with two communities only, placed in two distinct cells. At the system start-up, n nodes are placed in node k's cell, and $f \cdot n$ nodes in the other cell (hereafter, node k's cell is denoted as starting cell, and the other cell as destination cell), f being a parameter greater than *0*. All nodes but node k have only social links within their community, while node k has also links with nodes in the other community. It is clear that this is the least favourable scenario for the gregarious behaviour to occur because all nodes but k are very tightly connected to each other. It is trivial to show that the probability of node k moving outside its starting cell at least once over n movements is $1-(CA_{start})^n$, which approaches *1* for n sufficiently high. This means that node k moves outside its starting cell at some point in time with high probability. To investigate the gregarious behaviour, we compute the *remaining probability* (P_{rem}), defined as the probability of *no other* member of node k's community to move towards the destination cell after k has moved. When P_{rem} approaches 0, at least one node in the starting cell follows node k with high probability. As will be clear from the following analysis, this may generate an avalanche effect such that all nodes in node k's community follow node k in the destination cell, thus revealing the gregarious behaviour.

We choose to study the case of a single node (k) having links outside its community, because, as we will show analytically, it represents the weaker condition for the gregarious behaviour to take place. Therefore, the P_{rem} formula computed in the following section is actually an upper bound of the remaining probability achieved in the general case.

Notation used throughout this section is stated in Table 2.

4.1.1. Analytical model

In this section we derive a closed form equation for the remaining probability (P_{rem}) and then we explore the parameter space of CMM, showing that in the vast majority of (and in the more realistic) configurations, nodes in CMM behave according to the gregarious behaviour.

Before computing P_{rem}, we first compute the probability that a single node goes out of its starting cell after node k has left. For the sake of simplicity, we hereafter assume that the degree of

[1] CMM also includes a periodical re-assignment of communities to cells, called reconfiguration, whose effect is however orthogonal to what we focus on in this paper. The interested reader is referred to [12].

Table 2
Notation used in Section 4.1.

$f \cdot n$	Fraction of nodes in the destination cell
k	Tag of the reference node
P_{rem}	Probability that no member of node k' community follows node k
C_i	Community to which node i belongs
w_k	Weight of all the social relationships between node k and every other node
$\bar{w}$	Average weight
$\bar{s}$	Expected minimum number of times that the selection algorithm is performed
l	Ratio between the length of the simulation area edge and the length of the single cell edge

friendship between node k and each other node of the same community is the same, i.e., $w_{i,k} = w_{j,k} \triangleq w_k, \forall i,j | C_i = C_j$. Remember that no node other than k has links outside its community.

Lemma 1. *The probability P_{out} that a generic node i chooses to exit its current community and to follow node k is equal to*

$$P_{out} = \frac{w_k/(fn+1)}{w_k/(fn+1) + \bar{w}}, \tag{3}$$

where $\bar{w}$ denotes the average social weight.

Proof. Applying Eq. (1), after node k starts moving outside the starting cell, each node i of its community (i.e $C_k = C_i$) is attracted by the destination cell with a social attractions equal to

$$SA^{(i)}_{(dest)} = w_{i,k}/(fn+1) \tag{4}$$

and by the starting cell with a social attraction equal to

$$SA^{(i)}_{(start)} = \sum_{j \in c_{start}} w_{i,j}/(n-1). \tag{5}$$

Note that, based on this definition in Eq. (1), a node is associated with the destination cell as soon as it *selects* that cell as its next goal.

Eqs. (4) and (5) already show two interesting features. First, the attraction of node k on the other nodes in the same community depends not only on the strength of the social relationships within the community ($w_{i,k}$), but also on the number of nodes associated with the destination cell (fn+1). The more the nodes associated with the destination cell, the less the attraction that k exerts on the other nodes. Second, it can already be shown that we are considering a worst-case scenario for the gregarious behaviour to occur. If several nodes had social relationships outside their community, there would be more nodes going outside the starting cell. The attraction exerted by the destination cell (or the joint attraction exerted by destination cells, more in general) on each remaining node will be greater than $SA^{(i)}_{(dest)}$ in Eq. (4), thus increasing the probability of at least one such node to go out of the starting cell.

Going back to our reference scenario, the probability that a generic node i in node k's community goes out of the starting cell is equal to

$$P_{out_i} = \frac{w_{i,k}/(fn+1)}{w_{i,k}/(fn+1) + \sum_{j \in c_{start}} w_{i,j}/(n-1)} \tag{6}$$

and thus $1 - P_{out_i}$ represents the probability that node i remains in the starting cell at the next step. We approximate $\bar{w} \simeq \sum_j w_{i,j}/(n-1)$, where $\bar{w}$ denotes the average of the distribution of weights between nodes of the same community. Under the assumption $w_{i,k} = w_{j,k} = w_k, \forall i,j | C_i = C_j$, Eq. (6) thus becomes Eq. (3).

Note that a special case is represented by $n = 2$. In this case, $SA^{(i)}_{(start)}$ is clearly 0 (just one node is in the starting cell after node k leaves), and therefore P_{out} is equal to 1. □

Eq. (3) shows the probability of a node getting out of the starting cell at the beginning of each user movement while node k is associated with the destination cell. To compute the remaining probability we must thus consider how many times (on average) such generic node has the opportunity of following' node k (Lemma 2). The longer node k is associated with the destination cell, the longer other nodes have the opportunity of following it. In the worst-case scenario (from the gregarious behaviour standpoint) node k comes back to the starting cell right after reaching the goal in the destination cell (this corresponds to the minimum possible time during which node k is associated with the destination cell).

Lemma 2. *The expected minimum number $\bar{s}$ of times that each node performs the goal selection algorithm while k is associated with the destination cell is equal to the ratio l between the length of the simulation area edge and the length of the single cell edge.*

$$\bar{s} = l \tag{7}$$

Proof. Considering an average case analysis, the number of times a node i (in the same community as node k) runs the next goal selection algorithm while node k is associated with the destination cell is equal to the ratio between the average time it takes for node k to reach the goal in the destination cell ($\overline{T}^{(out)}$) and the average time it takes for node i to complete a random movement inside the starting cell ($\overline{T}^{(in)}$). Under the assumption of square cells, the average length of a random movement inside the grid ($\bar{d}$) can be assumed to be proportional to the average length of a random movement inside a single cell ($\overline{d_{int}}$), with proportionality constant l (equal to the ratio between the edge of the grid and the edge of the cell). If we define $\overline{V}$ as the average speed at which nodes move, then $\frac{\overline{T}^{(out)}}{\overline{T}^{(in)}} = \frac{\bar{d}}{\overline{V}} \cdot \frac{\overline{V}}{\overline{d_{int}}} = \frac{\bar{d}}{\overline{d_{int}}} = l$ holds true. □

It is now possible to compute the closed form expression of the remaining probability.

Theorem 1. *The probability that no node follows node k once it has left its community is given by*

$$P_{rem} = \left[\left(1 - \frac{w_k/(fn+1)}{w_k/(fn+1) + \bar{w}} \right)^{l} \right]^{n-1}. \tag{8}$$

Proof. The probability that a generic node chooses a local goal for the next movement is equal to $1 - P_{out}$. On average, the probability that a local movement is chosen every time the node performs the goal selection algorithm is equal to $(1 - P_{out})^{\bar{s}}$. As there are $n - 1$ nodes in the starting cell (after node k's departure), the probability that all of them choose only local movements is equal to $(1 - P_{out})^{\bar{s}(n-1)}$. After simple substitutions, we obtain Eq. (8). □

4.1.2. Occurrence of the gregarious behaviour in CMM

In this section we explore the domain of the mobility model parameters under which the remaining probability derived in Eq. (8) triggers the gregarious behaviour. When not otherwise specified, parameters are set according to Table 3. Note that in the following we consider two alternatives for assigning relationships between nodes (weights) before rewiring. In the first one, weights are uniformly distributed between 0 (not included) and 1. In the second one, weights between friends are uniformly distributed between $1 - threshold$ and 1, with *threshold* being a configurable parameter ranging from 0 to 1. These two methods are equivalent when *threshold* is equal to 1. The second alternative is the one actually implemented in CMM because it allows to control the average social weight within a community. Without the threshold, the average social weight between friend nodes ($\bar{w}$) is stuck at 0.5,

Table 3
Default values for the model parameters.

Parameter	Value
n	10
l	5
$\bar{w}$	0.5
w_k	0.5
f	1

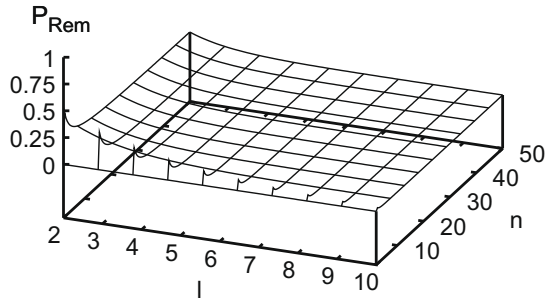

Fig. 1. P_{rem} as a function of n and l.

while, with the threshold-based approach, $\bar{w}$ is equal to $1 - threshold/2$, and it thus ranges between 0.5 and 1.

Fig. 1 illustrates the P_{rem} dependence on n (the number of the nodes of k's community), and l (the ratio between the movements duration outside and inside a cell). Figs. 2 and 3 compare selected curves from the plot in Fig. 1 (thus derived through the model) with simulation results (confidence intervals have 90% confidence level).[2] This comparison shows that the analytical model is accurate, particularly starting from medium values of n and l.

For small values of l, the grid has few cells and the duration of k's movement outside the starting cell is not so different from the duration of nodes' random movement inside a cell. Thus, a generic node i has not many opportunities of going outside the starting cell, because node k is associated with the destination cell only for a relatively small amount of time. The trend highlighted in Fig. 1 generally holds true when considering the impact of l, irrespectively of the other parameters' configurations. Therefore, we will not analyse the impact of l further on.

To better understand the behaviour with respect to n, let us rewrite Eq. (8), approximating w_k with its average value $\bar{w}$. It is easy to show that Eq. (8) becomes $P_{rem} = [(1 - (1/n+2))^l]^{n-1}$. The remaining probability of a *single* node on a single movement (1-(1/n+2)) increases with n, because a large n corresponds to a "heavy" community, that exerts a strong attraction on its members. However, as the number of nodes increases, it is more and more difficult that *all* nodes remain in starting cell. The joint effect (shown in Fig. 1) is that P_{rem} is significantly greater than 0 only for small values of n.

In Figs. 4–6 we analyse how the strength of social relationships (modelled through the parameters w_k and $\bar{w}$) impact on the remaining probability. Specifically, in order to have a complete view across the whole parameter space, in Fig. 4 we explore the whole range of possible values for the parameters, irrespective of the particular distribution used to actually assign weights. Note that, when w_k is small, i.e. when node k exerts a weak attraction on the nodes of the starting cell, the main contribution to P_{rem} is due to $\bar{w}$ (i.e., the attraction between nodes in the starting cell):

[2] Similar curves have been derived also with respect to other parameters analysed in the following (not reported here due to space reasons).

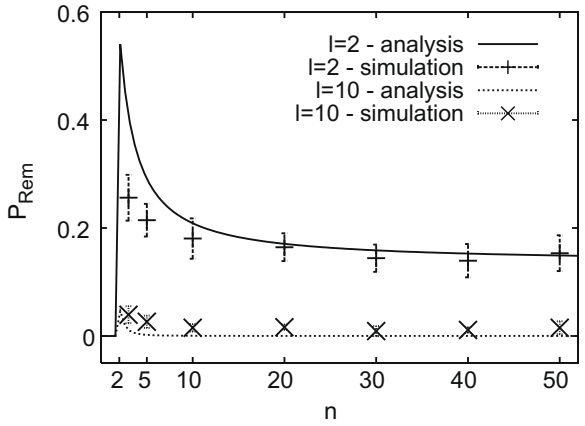

Fig. 2. Validation of the analytical model (P_{rem} as a function of n).

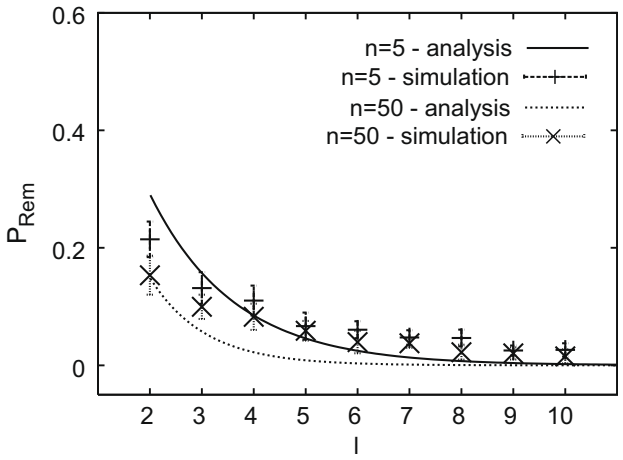

Fig. 3. Validation of the analytical model (P_{rem} as a function of l).

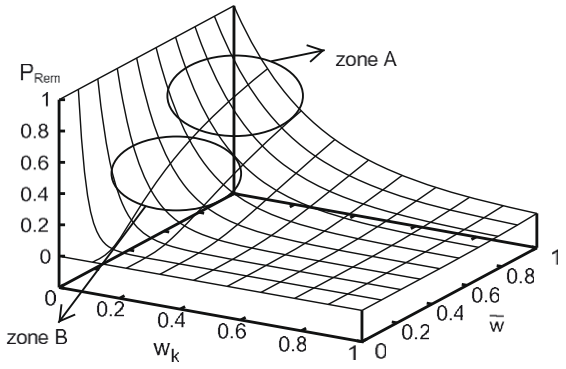

Fig. 4. P_{rem} as a function of w_k and $\bar{w}$.

if $\bar{w}$ is high, nodes are more likely to remain in their current cell. However, as soon as w_k increases beyond 0.5, P_{rem} drops approximately to 0, regardless of the weights between nodes of the same cell. This is an important results, because it highlights that the gregarious behaviour occurs in the CMM model for a large range of $(w_k, \bar{w})$ values. Remaining probabilities significantly greater than 0 are achieved only in the region (zone A in Fig. 4) with high attraction to the starting cell (high $\bar{w}$), and low attraction of node k (low w_k). However, in the following we show that these operating conditions cannot be achieved with the uniform weight assignment used by CMM, not even by controlling the average value through

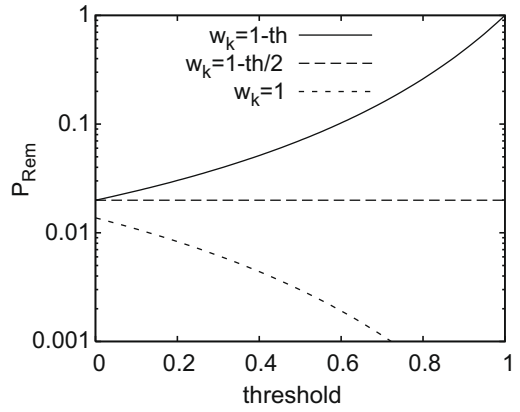

Fig. 5. P_{rem} as a function of the *threshold*.

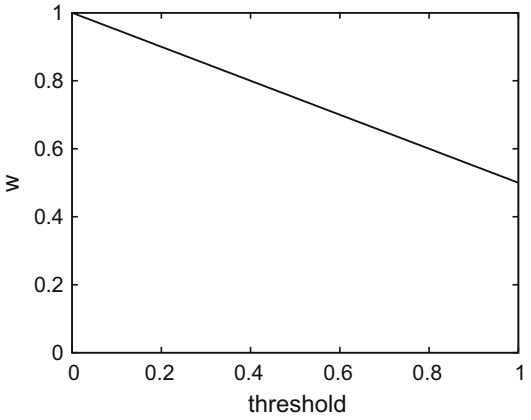

Fig. 6. $\bar{w}$ as a function of *threshold*.

the *threshold* parameters. To this aim, recalling that social weights are sampled from a uniform distribution in the interval [1 − *threshold*, 1], we plotted the remaining probability for lowest, intermediate, and highest w_k value (namely 1 − *threshold*, 1 − *threshold*/2, and 1), as the *threshold* value varies (Fig. 5). Non negligible remaining probabilities are exclusively achieved when *threshold* is high (greater than 0.8) and w_k happens to be low (=1 − *threshold*). For the operating zone A to be reached, we need also $\bar{w}$ to be high. However, as can be seen in Fig. 6, when the *threshold* value is greater than 0.8, $\bar{w}$ is below 0.6, making configurations of $(w_k, \bar{w})$ in zone A practically not feasible. Thus, the region with maximum remaining probability in the $(w_k, \bar{w})$ space where CMM can actually operate is zone B in Fig. 4, where, however, the remaining probability is still quite low.

In this first set of plots we considered f equal to 1 (i.e., the same number of nodes in the starting and destination cells). This choice is coherent with the original implementation of the CMM model, in which the number of nodes in each cell at the system start-up is almost the same. In real social networks, however, non-homogeneous communities are quite common. In the following we thus investigate the gregarious behaviour with non-homogeneously populated communities. Eq. (4) highlights how f impacts on the social attraction of the destination cell. When the destination cell is more populated than the starting cell ($f > 1$), node k's attraction is smoothed out, while, when it is less populated (f<1), the attraction is amplified.

Figs. 7 and 8 basically confirm the behaviour highlighted before. Even when the destination cell is more populated (f>1), non-negligible remaining probabilities are achieved when the attraction of node k is rather weak (low w_k values). Clearly, the greater the destination cell's population, the lower the attraction it exerts on the members of node k's community, the greater the remaining probability. Specifically, we have found that the remaining probability is always greater than 0.5 as f increases beyond 10. However, note that such high values for f are not that sensible, as f = 10 means that the destination community has 10 times more members than the starting community. Finally, Fig. 8 shows that, for more sensible ranges of f, the only chance of having high remaining probability is a very high threshold (which implies, on average, low attraction of node k, because $w_k \sim U(1 - threshold, 1)$) and a densely populated destination cell (f > 1). Note that when the threshold approaches 1, the attraction of node k becomes negligible.

The results shown in this section clearly highlight that the gregarious behaviour is a characteristic feature of the original CMM model that occurs in the vast majority of its configurations. The gregarious behaviour is both often undesired and hardly controllable. Often undesired because only for very few and very specific applications (e.g., war scenarios) this behaviour can be considered realistic. Hardly controllable because the area in which all nodes converge is not predictable but depends on the evolution of the system. In order to avoid the gregarious behaviour, and more in general to include the preferred locations property of realistic mobility, in the next section we propose the HCMM model, which, starting from the social structure defined in CMM, introduces the concept of attractions exerted by the *physical places* in which the social relationships between users usually take place (e.g., the working place). These locations become the preferred locations for the nodes of each community, thus satisfying the second property of human mobility that we have discussed in Section 1.

5. Adding spatial attraction: HCMM

In this section we extend the CMM model to include attractions exerted by the physical places in which social relations usually take place. The resulting version of the Home Cell Mobility Model, initially defined in [26], merges two fundamental aspects of human mobility: social and location attraction. The main idea behind this model is that users move towards places where the people they share social relationships with are likely to be because that is their typical "home" location (at that point in time). Social and topological structures in HCMM are the same as in CMM, but the way they are mapped onto each other is different. As in CMM, nodes are organized into social communities, and each community is initially assigned to one of the cells of the grid. Differently from CMM, HCMM, as the name suggests, uses the concept of *home cell*. Each

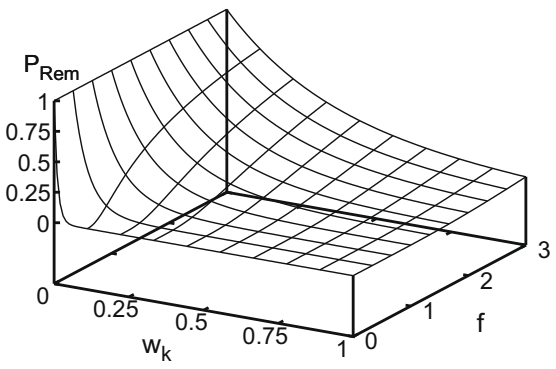

Fig. 7. P_{rem} as a function of w_k and f.

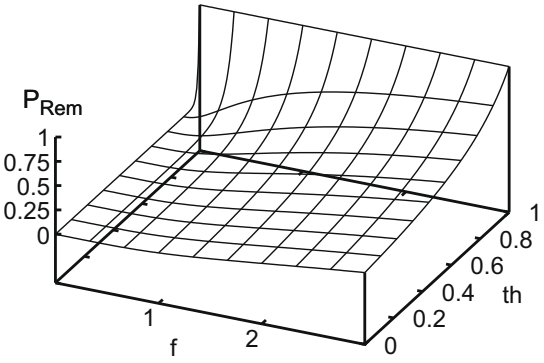

Fig. 8. P_{rem} as a function of f and *threshold*, in the worst-case scenario ($w_k = 1 - threshold$).

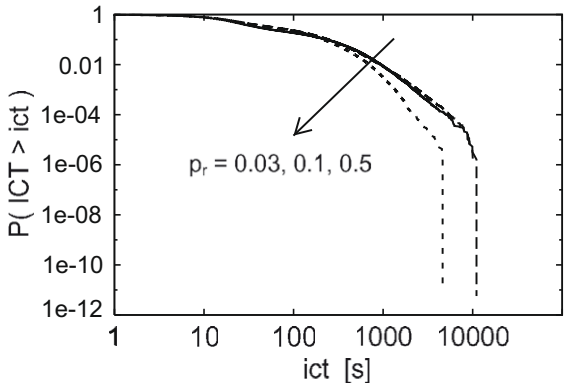

Fig. 10. Inter-contact time distribution in HCMM.

user is assumed to belong to a main social community (at any given point in time). The user's home cell is defined as the cell within which the members of the user's main social community preferentially move. As in CMM, in HCMM *when a node is located in its home cell*, it selects its cell or another cell according to the social attraction exerted on it. Unlike in CMM, in HCMM the social attraction (SA_{c_i}) exerted by a generic cell c_i is computed as the sum of the weights of the social links between node x and *all nodes that have* c_i *as their home cell*, irrespective of their current location:

$$SA_{c_i} = \frac{\sum_{y \in c_i} w_{xy}}{|c_i|} \tag{9}$$

(where the notation $y \in c_i$ means that c_i is the home cell of node y). This maps the idea of the home cells being a "proxy" for the friend nodes of the same community. A generic cell c_i is then selected as next goal cell of node x with a probability CA_{c_i} defined as in Eq. (2).

On the other hand, when x is located in a cell other than the home, the selection of the next goal in HCMM is performed as follows. With a probability p_e, node x remains in the external cell for the next movement, while it goes back to the home cell with probability $1 - p_e$. Once the next goal cell is selected, one point inside that cell is chosen uniformly at random as in CMM, and the node starts to move towards the goal. The whole process is summarised in Fig. 9 using a Markov Chain, where H_i denotes the home cell of node i.

Fig. 10 shows the distribution of the inter-contact times generated by HCMM in one representative configuration we have tsted (specifically, when 30 nodes, evenly grouped into 3 communities, move in a square of size 1250 m × 1250 m, divided in a 5 × 5 grid). The rewiring parameter is set alternately to 0.03, 0.1, and 0.5. The trend that emerges is that of a heavy-tailed behaviour up to a certain point, after which there is a sudden exponential decrease. This result matches the distribution of ICTs extracted from real movement traces, thus confirming the ability of HCMM to reproduce realistic temporal properties of human movements.

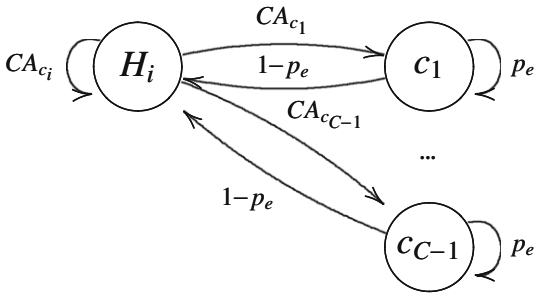

Fig. 9. Markov chain for standard HCMM.

In the next section we show how complementing sociality with spatial attraction removes the emergence of the gregarious behaviour, thus drastically improving the controllability of the model in terms of the time spent in the various physical locations.

5.1. Predictability of users' movements

In this section we analytically investigate the ability of CMM and HCMM to produce a predictable outcome in terms of the expected time spent by nodes inside ($E[T_{in}]$) and outside ($E[T_{out}]$) their home cell. These characteristic times are particularly relevant because nodes, when moving outside the starting community, meet new nodes and therefore increase their transmission opportunities. On the contrary, when nodes roam for a long time (or for good, in the extreme case) within their starting community, they have to leverage the mobility of other more social nodes (if any), to deliver messages to nodes of the other communities. Being able to control $E[T_{in}]$ and $E[T_{out}]$ is therefore an important property of mobility models for opportunistic networks.

For ease of presentation, we still assume to have just two cells, the starting cells and the destination cell. The destination cell can jointly represent all cells other than the starting cell. We assume that *all* links can be rewired at the system start-up (with probability p_r). Therefore, we do not assume any difference between a tagged node (node k) and the other nodes any more. We also do not focus any more on the event of a particular node exiting the starting cell. Notation used in this section is stated in Table 4.

With respect to internal and external movements, CMM and HCMM can be modelled with a two-state discrete Markov chain (Fig. 11), where the "IN" state corresponds to a node roaming within its starting cell and the "OUT" state corresponds to a node being outside the starting cell. For both CMM and HCMM the following proposition holds, which follows directly from the definition of the two models (Eq. (2)).

Proposition 1. *The transition probabilities p_{in} and p_{out} for both CMM and HCMM are given by*

$$p_{in} = \frac{SA_{start}^{(out)}}{SA_{dest}^{(out)} + SA_{start}^{(out)}}$$
$$p_{out} = \frac{SA_{dest}^{(in)}}{SA_{dest}^{(in)} + SA_{start}^{(in)}}, \tag{10}$$

where $SA_{start}^{(in)}$ and $SA_{dest}^{(in)}$ ($SA_{start}^{(out)}$ and $SA_{dest}^{(out)}$) represent the social attractions, namely, of the starting and destination cells when the node is in the IN (OUT) state.

CMM and HCMM differ in how the attractions of the starting cell and of the external cell are computed. Let us start with the case of CMM. Note that, with respect to the analysis in Section 4, here

Table 4
Notation used in Section 5.1.

p_e	Probability of remaining in an external cell for the next movement
$E[T_{in}]$	Expected time spent inside the home cell
$E[T_{out}]$	Expected time spent outside the home cell
q	Number of nodes belonging to the starting cell that are currently outside
q'	Number of nodes belonging to the destination cell that are currently outside
$SA_{start}^{(in)}$, $SA_{dest}^{(in)}$	Social attraction of the starting and destination cell in the IN state
$SA_{start}^{(out)}$, $SA_{dest}^{(out)}$	Social attraction of the starting and destination cell in the OUT state

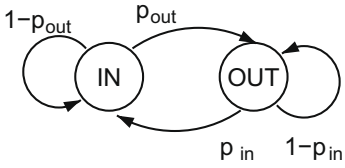

Fig. 11. Node's status in HCMM and CMM.

we consider the more general case of an arbitrary number of nodes roaming in the starting and destination communities, as well as the case of an arbitrary number of nodes having social relationships outside their home cell. The following theorem holds.

Theorem 2. (Social attractions for CMM). *The attractivity of the starting cell and of the destination cell under the CMM model is as follows:*

$$\begin{cases} SA_{start}^{(C,in)} = \frac{(n-1-q)(1-p_r)+\frac{(n-1)p_r}{fn}q'}{n-q+q'} \cdot \bar{w} \\ SA_{dest}^{(C,in)} = \frac{q(1-p_r)+\frac{(n-1)p_r}{fn}(fn-q')}{fn-q'+q} \cdot \bar{w} \\ SA_{start}^{(C,out)} = \frac{(n-q)(1-p_r)+\frac{(n-1)p_r}{fn}q'}{n-q+q'} \cdot \bar{w} \\ SA_{dest}^{(C,out)} = \frac{(q-1)(1-p_r)+\frac{(n-1)p_r}{fn}(fn-q')}{fn-q'+q} \cdot \bar{w} \end{cases}, \tag{11}$$

where q (q') are the number of nodes of the starting (destination) cell that are currently roaming in the destination (starting) cell.

Proof. In the case of CMM, the attraction to a cell dynamically depends on the number of nodes actually being in that cell. For the sake of simplicity, we carry on the analysis under the hypothesis that q nodes of the starting cell are roaming in the destination cell, and q' nodes of the destination cell are roaming in the starting cell. In addition, we assume that the initial number of nodes in each cell is, respectively, n and $f\,n$. Therefore, in the starting cell there are $n-q$ nodes initially assigned to the community in the starting cell that are still in that cell, and q' nodes belonging to the community in the destination cell and that have moved to the starting cell. Analogously, there are $fn-q'$ nodes initially assigned to the destination community that still are in the destination cell, while q nodes in the destination cell come from the community in the starting cell. In order to compute the social attractions, we need to compute the expected fraction of nodes that are friends of a tagged node k belonging to the community assigned to the starting cell. Let us consider the case of this node being in the IN state. Being the rewiring process uniform, the expected number of social links still pointing towards nodes initially assigned to the same community is $(n-1-q)(1-p_r)$ for those nodes that are still in the starting cell, and $(1-p_r)q$ for those that are currently in the destination cell. Of the initial $n-1$ "internal" links of node k (remember that initially communities are fully connected), $p_r(n-1)$ have been rewired towards nodes belonging to the destination community. Of these $p_r(n-1)$ links, on average the number of links rewired towards a node of the destination cell that is currently still there is $(n-1)p_r\frac{fn-q'}{fn}$, while the average number of links rewired to nodes of the destination cell that have moved to the starting cell is $(n-1)p_r\frac{q'}{fn}$. When we consider the case of node k having moved to the destination cell (thus being currently in the OUT state), the computation is very similar to what explained above, except for the presence of node k that now must be accounted for in the counter for nodes in the destination cell. As an example, let us consider the social attraction exerted by the destination cell on node k when it is in the IN state. This attraction is determined by the $q(1-p_r)$ friends belonging to the starting community that have moved to the destination cell and by the $(n-1)p_r\frac{fn-q'}{fn}$ friends belonging to the destination community and that are still in the destination cell. According to Eq. (1), we can write $SA_{dest}^{(C,in)}$ as $\frac{\sum_{j=1}^{q(1-p_r)}w_{kj}+\sum_{j=1}^{(n-1)p_r\frac{fn-q'}{fn}}w_{kj}}{fn-q'+q}$. For the sake of simplicity, we approximate $\sum_{j=1}^{z}w_{kj}$ with $z\bar{w}$, i.e., we substitute the sum of z continuous random variables (w_{ij}) with z times their expected value. After applying this simplification, we get the expression for $SA_{dest}^{(C,in)}$ in Eq. (11). Based on the above line of reasoning, it is possible to derive also $SA_{start}^{(C,in)}$, $SA_{start}^{(C,out)}$, and $SA_{dest}^{(C,out)}$. □

Computing the attractions under the HCMM model is much simpler.

Theorem 3. (Social attractions for HCMM). *The attractivity of the starting cell and of the destination cell under the HCMM model is given by the following:*

$$\begin{cases} SA_{start}^{(IN)} = \bar{w} \\ SA_{dest}^{(IN)} = \frac{p_r(n-1)\bar{w}}{fn} \\ SA_{start}^{(OUT)} = 1-p_e \\ SA_{dest}^{(OUT)} = p_e \end{cases}. \tag{12}$$

Proof. For HCMM we are interested only in the IN state, because here is when social attractions play a role. When the system is in the IN state, the attraction to the starting cell depends only on the nodes having the starting cell as home cell (Eq. (9)). In addition to node k, the number of nodes that share the starting cell as home cell is $n-1$ (the reference node itself must not be considered), therefore $SA_{start}^{(IN)} = \frac{\sum_{j=1}^{(n-1)(1-p_r)}w_{kj}}{n}$. The social attraction of the destination cell can be computed according to Eq. (9). The expected number of social links rewired to the destination community is $p_r(n-1)$. Therefore the social attraction of the destination cell is given by $SA_{dest}^{(IN)} = \frac{\sum_{j=1}^{p_r(n-1)}w_{kj}}{fn}$. By applying the approximation $\sum_{j=1}^{z}w_{kj} \simeq z\bar{w}$, we obtain $SA_{start}^{(IN)}$ and $SA_{dest}^{(IN)}$ as in Eq. (12). The formula for $SA_{start}^{(OUT)}$ and $SA_{dest}^{(OUT)}$ follow directly from the definition of the HCMM model. □

By substituting Eqs. (11) and (12) in Eq. (10), we obtain the following corollary.

Corollary 1. *The transition probabilities for CMM and HCMM are given by:*

$$\begin{cases} p_{in}^{cmm} = \frac{(fn+q-q')(fn(n-q)(-1+p_r)-(-1+n)p_rq')}{p_r((-1+f)n+2q-2q')(fn(n-q)-(-1+n)q')+fn(-(n-q)(-1+fn+2q)+(1+n-2q)q')} \\ p_{out}^{cmm} = -\frac{(n-q+q')(fnq+p_r(fn(-1+n-q)-(-1+n)q'))}{p_r((-1+f)n+2q-2q')(fn(-1+n-q)-(-1+n)q')+fn(q+2q(-n+q)+fn(1-n+q)+(-1+n-2q)q')} \end{cases}. \tag{13}$$

$$\begin{cases} p_{in}^{hcmm} = 1-p_e \\ p_{out}^{hcmm} = \frac{(-1+n)p_r}{fn+(-1+n)p_r} \end{cases}. \tag{14}$$

Finally, the expected time spent in each state of the Markov chain in Fig. 11 can be computed as follows.

Theorem 4. (Sojourn time). *The expected time spent in the IN and OUT state by a node moving according to the two-state Markov chain in Fig. 11 is given by the following:*

$$\begin{cases} E[T_{in}] = \frac{p_{in}(1-p_{out})}{p_{in}+p_{out}} \cdot \overline{T}^{(local)} \\ E[T_{out}] = \frac{p_{out}(1-p_{in})}{p_{in}+p_{out}} \cdot \overline{T}^{(local)} + \frac{p_{in}p_{out}}{p_{in}+p_{out}} \cdot 2\overline{T}^{(trans)} \end{cases} \quad (15)$$

Proof. The limiting probabilities $\pi_{in} = p_{in}/(p_{in}+p_{out})$, and $\pi_{out} = p_{out}/(p_{in}+p_{out})$ for the reference Markov chain follows from the application of standard Markov theory [29]. The expected time spent in the IN (or OUT) state can be computed by conditioning on the fact that the node has just entered the IN (or OUT) state according to the following:

$$\begin{cases} E[T_{in}] = P[E_{IN}] \cdot E[T_{in}|E_{IN}] = \pi_{out}p_{in} \cdot E[T_{in}|E_{IN}] \\ E[T_{out}] = P[E_{OUT}] \cdot E[T_{out}|E_{OUT}] = \pi_{in}p_{out} \cdot E[T_{out}|E_{OUT}] \end{cases} \quad (16)$$

where E_{IN} and E_{OUT} denote the events "the node enters the IN state" and "the node enters the OUT state", respectively, while $\pi_{out}p_{in}$ and $\pi_{in}p_{out}$ are the probabilities of these events. In fact, e.g., the probability of entering state IN is equal to the probability of not being in state IN joined with the probability of selecting state IN for the next step. $E[T_{in}|E_{IN}]$ and $E[T_{out}|E_{OUT}]$ corresponds to the expected time spent in the IN and OUT states given that the node has just entered that state. The number of consecutive roaming steps in the same state are distributed according to a geometric law with success probability equal to p_{out} for the IN case and p_{in} for the OUT case. Therefore their expected value is equal to $\frac{1-p_{out}}{p_{out}}$ and $\frac{1-p_{in}}{p_{in}}$, respectively. The duration of each step both in the IN and OUT state can be approximated with $\overline{T}^{(local)}$, while the duration of the transitions between the states can be approximated with $\overline{T}^{(trans)}$. If we assume that the duration of the transition from the IN to the OUT state (and vice versa) contributes to the expected time spent outside the IN state, then the following equalities hold:

$$\begin{cases} E[T_{in}|E_{IN}] = \frac{1-p_{out}}{p_{out}} \cdot \overline{T}^{(local)} \\ E[T_{out}|E_{OUT}] = \frac{1-p_{in}}{p_{in}} \cdot \overline{T}^{(local)} + 2\overline{T}^{(trans)} \end{cases} \quad (17)$$

In both CMM and HCMM local movements within the same cell are equivalent to Random Waypoint epochs. Therefore, the expected T_{local} can be computed as in [30]. T_{trans} is equal to L_{trans}/V, i.e, T_{trans} is equal to the ratio of two independent random variables, the transition length (with probability density f_L) and the velocity of nodes (with probability density f_V). The formula for the probability density function of the ratio Z of two independent continuous random variables X and Y is well-known [31]. Unfortunately, there is no closed form for the probability density of L_{trans} (see Section 6.1, where $L_{trans} = P_{c_i c_j}(d)$). Therefore, for the sake of convenience, we use the approximation proposed in the proof of Lemma 2. Finally Eq. (15) follows after substituting Eq. (17) into Eq. (16). □

Theorem 4 is very important because it highlights the impact of Corollary 1 on the controllability of CMM and HCMM. In CMM the closed form expression of $E[T_{in}]$ and $E[T_{out}]$ depend on the dynamic evolution of the users' movements, because p_{in}^{cmm} and p_{out}^{cmm} depend, according to Corollary 1, on q and q', which are not model parameter, but change based on the actual nodes' positions. Therefore, in CMM it is very hard to set model parameters to achieve a desired nodes' behaviour as far as nodes' physical positions are concerned. On the other hand, in HCMM $E[T_{in}]$ and $E[T_{out}]$ do *not* depend on the dynamic evolution of the system, but depend only on f, n, p_r, and p_e, i.e., on the configuration parameters of the model. This means that HCMM, while retaining the social theoretical approach of CMM, also provides simple knobs to control the time spent by nodes in the preferred physical locations. These remarks are highlighted in Fig. 12, which plots $E[T_{in}]$ and $E[T_{out}]$ for CMM and HCMM as functions of q (time is normalised with respect to $\overline{T}^{(local)}$ and f, n, p_r, and p_e are the same for both CMM and HCMM).

6. Analytical model for jump size distribution

So far, we have analysed the HCMM model in terms of its temporal properties, of which we have highlighted the controllability and predictability. In the remaining of the paper we study the spatial properties of HCMM. We prove that the model defined in Section 5 is not able to reproduce realistic spatial properties of mobility and we improve it accordingly. To this aim, we first introduce an analytical model that describes the distribution of jump size under a generic grid-based mobility model, and then we specialize this model to the HCMM case.

The intuition of what we formally prove in the rest of the section is as follows. Analysing HCMM from the point of view of jump sizes, we identify two main kinds of movements: movements within the same cell (or *internal*) and movements from one cell to another (or *external*). External movements are generally longer than internal movements, because they are not confined to a single cell. External movements are determined by the number of external links: the more the social links towards nodes belonging to different communities, the higher the number of external movements, because nodes tend to be more attracted outside the home cell. The number of external links is determined by the rewiring process. This process is uniform across different communities, because the node towards which a link is rewired is selected uniformly at random. In addition, when communities have the same number of members, all the nodes start with the same number of links, and these links have the same probability of being rewired. These two considerations imply that, on average, all the communities, and therefore the cells associated with them, have the same probability of being selected as goals for external movements. Therefore, in HCMM, there is no preferential selection of short paths (closer cells) over long paths (distant cells), and this is in contrast with the spatial characteristics of real human mobility. In the following, this argument is formalized using a mathematical model, in which we denote the jump size distribution as $P(d)$. Specifically, in Section 6.1 we provide a framework for computing $P(d)$ that is common to all grid-based mobility models, then (Section 6.2) we specialize this framework for HCMM. Notation used throughout this section is stated in Table 5.

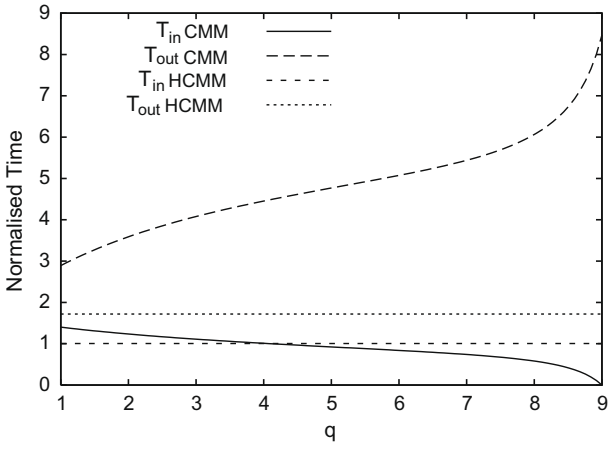

Fig. 12. Average time in the IN and OUT states as functions of q.

Table 5
Notation used in Section 6.

$P(d)$	Jump size distribution
$P_{c_ic_j}(d)$	Jump size distribution for paths between cell c_i and cell c_j
$P_{int}(d)$	Jump size distribution for paths inside the same cell
a	Cell width and height
P_{x_1}, P_{x_2}	Points in a one-dimensional space
P_1, P_2	Points in a bidimensional space
p_{ij}	Probability of having a movement between cell c_i and cell c_j
p_{int}	Probability of having a movement inside a cell
n_e	Average number of links rewired towards nodes belonging to a different home cell for each node
$n_e^{c_j}$	Average number of links rewired towards nodes having as home cell a generic cell c_j
$P(c_i)$	Steady state probability of being in cell c_i
$P(c_j\|c_i)$	Probability of a transition from c_i to c_j

6.1. Distances between random points in different cells

In this section we focus on the general framework for $P(d)$. As described before, under the assumption that nodes can move either between two random points within the same cell or between two random points in different cells, the set of the pairs of cells available for movements is the same for all nodes. Then, according to the rules of each mobility model, only a subset of these pairs can be selected by each node that is taking mobility decisions. Thus, we first study the distribution of jump size for movements between any pair of cells, and later we restrict the possible movements to the subset of cells that can actually be selected under the HCMM model. Without loss of generality, we focus on two tagged cells on the grid, c_i and c_j, for which we compute the jump size distribution $P_{c_ic_j}(d)$. In order to completely characterize this framework, we also need to compute the jump size distribution for internal movements, denoted as $P_{int}(d)$. As all cells have the same size, $P_{int}(d)$ is the same for all cells. In this work we use the closed form formula for P_{int} provided in [30] for the distribution of the distance between two random points in a rectangle, and we focus on the analytical description of $P_{c_ic_j}(d)$, using the same approach as in [32].

For the sake of convenience, in this section we drop the subscript c_ic_j of $P_{c_ic_j}(d)$ and we assume that cells are $a \times a$ wide (square cells). We refer our coordinate system to either of the two cells, and we express the relative position of the two cells using k_1 and k_2 as in Fig. 13. In our analysis, first we consider the unidimensional case (random points associated with different segments) and then we use the obtained results to solve the bidimensional case.

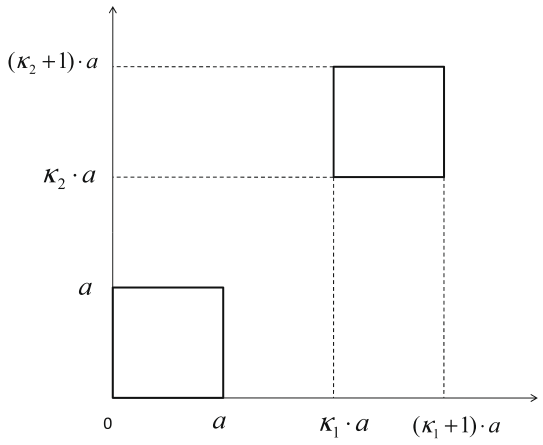

Fig. 13. Squares over a grid.

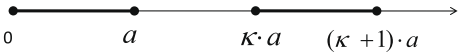

Fig. 14. Points on different segments.

6.1.1. Random points in different segments

We consider the case of two points P_{x_1} and P_{x_2}, uniformly distributed on two unidimensional segments: $[0,a]$ for P_{x_1}, $[ka, (k+1)a]$ for P_{x_2} (Fig. 14).

The probability density functions defining the positions of the two points are:

$$f_{P_{x_1}}(x_1) = \begin{cases} \frac{1}{a} & \text{if } 0 \leqslant x_1 \leqslant a \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

$$f_{P_{x_2}}(x_2) = \begin{cases} \frac{1}{a} & \text{if } ka \leqslant x_2 \leqslant (k+1)a \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

Theorem 5. (Distance distribution in one dimension). *The probability density function $P(d)$ of the distance d between P_{x_1} and P_{x_2} is given by*

$$P(d) = \begin{cases} \frac{a-|d-ak|}{a^2} & \text{if } (k-1)a \leqslant d \leqslant (k+1)a \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

Proof. First we compute the joint probability distribution of P_{x_1} and P_{x_2} which is given by:

$$f_{P_{x_1}P_{x_2}}(x_1,x_2) = \begin{cases} \frac{1}{a^2} & \text{if } 0 \leqslant x_1 \leqslant a \\ & \text{and } ka \leqslant x_2 \leqslant (k+1)a \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

We are interested in the distribution of the distance between P_{x_1} and P_{x_2}, defined as $D = |x_2 - x_1|$. Its cumulative distribution function $P(D \leqslant d)$ can be obtained by integrating the joint probability in Eq. (21) over the domain $\mathscr{D} = |x_2 - x_1| \leqslant d$. This implies solving Eq. (22).

$$P(D \leqslant d) = \begin{cases} 0 & \text{if } d < 0 \\ \frac{1}{a^2}\int_{ka-d}^{a}\int_{ka}^{x_1+d} dx_2dx_1 & (k-1)a \leqslant d \leqslant ka \\ \frac{1}{a^2}\left(\int_0^{(k+1)a-d}\int_{ka}^{x_1+d} dx_2dx_1 + \right. & \\ \quad \left. + \int_{(k+1)a-d}^{a}\int_{ka}^{(k+1)a} dx_2dx_1\right) & ka < d \leqslant (k+1)a \\ 1 & d > (k+1)a \end{cases} \tag{22}$$

The result is given in Eq. (23):

$$P(D \leqslant d) = \begin{cases} 0 & \text{if } d < 0 \\ \frac{(a-ak+d)^2}{2a^2} & (k-1)a \leqslant d \leqslant ka \\ -\frac{a^2(-1+k(2+k))-2a(1+k)d+d^2}{2a^2} & ka < d \leqslant (k+1)a \\ 1 & d > (k+1)a \end{cases} \tag{23}$$

Differentiating Eq. (23) with respect to d, we get the probability density function for D in the unidimensional case □

6.1.2. Random points in different cells

In this section we consider the bidimensional case shown in Fig. 13. We are interested in the distance between two points $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$, that is given by:

$$D = \|P_2 - P_1\| = \sqrt{|y_2 - y_1| + |x_2 - x_1|} = \sqrt{D_x^2 + D_y^2} \tag{24}$$

From Section 6.1.1 we know that the distribution of D_x and D_y is given by Eq. (20). Thus, we can compute the joint distribution

$$f_{D_xD_y}(d_x,d_y) = f_{D_x}(d_x)f_{D_y}(d_y) = \\ = \begin{cases} \frac{(a-|d_x-ak_1|)(a-|d_y-ak_2|)}{a^4} & a(-1+k_1) \leqslant d_x \leqslant a(1+k_1) \\ & a(-1+k_1) \leqslant d_y \leqslant a(1+k_2) \\ 0 & \text{else} \end{cases} \quad (25)$$

In order to compute $P(D \leqslant d)$ we integrate Eq. (25) over the domain $\sqrt{D_x^2+D_y^2} \leqslant d$, which is equivalent to $D_x^2+D_y^2 \leqslant d^2$. The domain corresponds to the area within a circumference centered in the origin of the axes and having radius equal to d. Unfortunately we could not obtain a general, closed form expression for $P(D \leqslant d)$. The only exact solution available is numerical, and it is the one that we use in Section 6.2. In the next section we propose an approximation for $P(D \leqslant d)$ that could possibly be used as a substitute closed form approximated expression of $P(D \leqslant d)$.

6.1.3. Model approximation

Again considering square cells, we call X_d the random variable associated with the jump size when a node is moving between two cells having the distance between their centers equal to d. We have numerically checked that, for any value of d except for $d = 0$,[3] X_d has a bell-shaped curve. As an example, the dotted line in Fig. 15 shows the PDF of X_d for $d = 100\sqrt{2}$.

The bell shape suggests the possible fitting with a normal distribution. In order to find an estimate of the parameters μ and σ of the normal distribution, we apply the Maximum Likelihood Estimation (MLE) method. The PDF of the random variable N_d, distributed according to $N_d(\mu,\sigma)$ with $d = 100\sqrt{2}$, is shown in Fig. 15.

By using the normal distribution to approximate the exact distribution of the jump size, we introduce two types of errors. On the tail, some values that were not present in the exact distribution can now appear, due to the infinite tails of the normal distribution against the finiteness of those of the exact distribution. On the body, the likelihood of certain values to appear can be slightly different from the case of the exact distribution. These errors can be quantified in order to estimate how good our approximation is. The error on the tail of the distribution corresponds to the probability of drawing values outside the domain of X_d from N_d, and this can be easily quantified as

$$e_{tail} = P(N_d < b_{left}) + P(N_d > b_{right}), \quad (26)$$

where b_{left} and b_{right} are the left and right boundaries on the values of X_d. In the case shown in Fig. 15, $e_{tail} = 0.0005$. The error introduced in $[b_{left}, b_{right}]$ can be estimated by computing the root mean squared error (RMSE) over the reference interval. In this case, using a sampling interval equal to 0.001, we obtain a RMS value of 0.00025.

6.2. Jump size distribution for HCMM

In this section we propose a way for computing the jump size distribution under the HCMM model. We assume that the initial placement of communities on the grid is given, and that the same home cell can be associated with only one community. We consider a $g_x \times g_y$ grid scenario with C communities, each having assigned n nodes. Again, the rewiring probability is denoted with p_r and the probability of having two consecutive trips within an external cell with p_e.

In HCMM, for all external trips the home cell is either the source or the destination of the movement. Therefore, the possible distances travelled by each node depend on the distance between its home cell and the other cells of the scenario. In addition to that, all nodes belonging to the same home cell are statistically equivalent, as they all have the same probability of being rewired to any other cell. From these considerations follows the proposition below.

[3] The case $d = 0$ corresponds to local movements, i.e., consecutive waypoints chosen within the same cell. As already discussed, in the case of local movements we already have a closed form expression for the distance between random points.

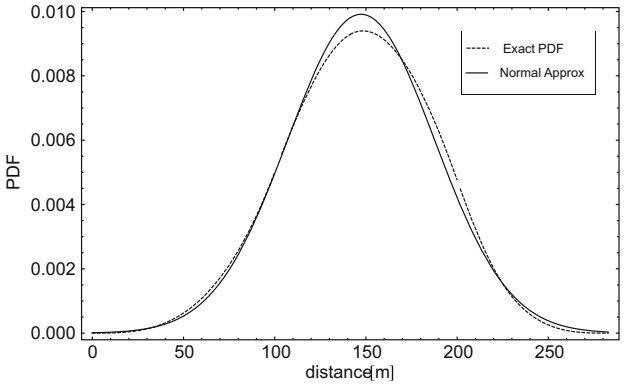

Fig. 15. Exact PDF of jump size versus normal approximation.

Proposition 2. (Jump size distribution). *The distribution of jump sizes in HCMM is the result of the separate contributions $P_i(d)$ of the different home cells. With C communities, each having a different home cell, $P_i(d)$ can be written as follows:*

$$P(d) = \frac{1}{C}\sum_{i=1}^{C} P_i(d) \quad (27)$$

Let us denote with c_i the home cell of community i. All the formulas that we give hereafter refer to a tagged home cell c_i and are valid for all its nodes, because all the nodes of the same home cell are statistically equivalent. For the sake of convenience, we implicitly discard empty cells from the model, because, in HCMM, nodes cannot move towards empty cells and therefore their contribution to $P(d)$ is zero.

In HCMM nodes travel from one random point in a cell to another random point in another cell. The distance between a random point associated with c_i and a random point within another cell c_j follows a distribution $P_{c_ic_j}(d)$ that can either be computed exactly but not in a closed form (Section 6.1), or be approximated using, e.g., a normal distribution (Section 6.1.3). In the following we use the exact numerical solutions for $P_{c_ic_j}(d)$. For what concerns internal movements, we use for $P_{int}(d)$ the formula provided in [30]. The distributions of external and internal movements taken together can describe exhaustively all the possible paths that a node can travel.

Proposition 3. (Contribution of internal and external movements).

$$P_i(d) = \sum_{\substack{j=1 \\ j\neq i}}^{C} p_{ij} * P_{c_ic_j}(d) + p_{int} * P_{int}(d), \quad (28)$$

where p_{ij} is the probability of having a movement from cell i to cell j, $P_{c_ic_j}(d)$ is the probability density function of the length of such movement, p_{int} is the probability of having a movement within the same cell, and $P_{int}(d)$ is the probability density function of the length of an internal movement.

Proof. Proposition 3 directly follows from the application of the law of total probability to $P_i(d)$. In fact, we can express the distribution of movements of nodes belonging to community c_i as the composition of the distributions $P_{c_ic_j}(d)$ of external movements (for any $j \neq i$) and the distribution $P_{int}(d)$ of the movement within the same cell. In HCMM, however, not all paths are allowed: only

cells friend nodes belong to can be selected as the goal of a movement. Therefore we have to weight $P_{c_ic_j}(d)$ and $P_{int}(d)$ with the probabilities p_{ij} and p_{int} of actually having that component according to HCMM rules. Eq. (28) shows the final formula for $P_i(d)$. □

p_{ij} and p_{int} express the probability of moving outside and inside the home cell, and in HCMM they depend on the social attractivity of the home cell and of the other cells. We compute this social attractivity in Lemma 3.

Lemma 3. (Cell attraction). *The attraction of cell c_j on a generic node having c_i as home cell is given by*

$$CA_{c_j} = \frac{p_r(n-1)}{n(C-1)}, \tag{29}$$

and the attraction of the home cell c_i can be expressed as

$$CA_{c_i} = \frac{n - p_r(n-1)}{n}. \tag{30}$$

In Eqs. (29) and (30), *n_e corresponds to the average number of external links for each node, while n_e^c gives the average fraction of those external links that are rewired to each community.*

Proof. Let us assume an unweighted social graph: if a link between two nodes exists, then its weight is 1, otherwise it is 0. Recall that the social attraction SA_{c_j} exerted by a cell c_j on a generic node x is given by the sum of the weights of the social links between node x and the nodes having c_j as their home cell. These social links are the result of the rewiring process. The rewiring process follows a Binomial distribution with probability of success equal to p_r over a sequence of $n-1$ experiments, corresponding to the $n-1$ links that a node initially has with the nodes in the same community. The number of rewired links for a node x is thus a realization of the Binomial distribution $B(n-1, p_r)$. Using directly this distribution would further complicate the analysis. Therefore, for the sake of tractability, in this paper we focus on the average case. This choice introduces some approximations into the obtained results. In particular, the smaller n, the greater the error. The average number of external links of each node in community c_i is therefore

$$n_e = (n-1) * p_r \tag{31}$$

Conversely, on average each node will have $n - n_e$ links with other nodes of the same community. As the rewiring process is uniformly distributed among all communities, the number of links between each node in c_i and a generic community c_j is given by

$$n_e^{c_j} = n_e^c = \frac{(n-1) * p_r}{C-1}, \tag{32}$$

where we dropped the index j because the formula does not depend on the community j chosen. If all weights are either 0 or 1, using the average number of external links n_e and n_e^c, we can compute the social attraction SA_{c_j} for every generic cell c_j of the grid, according to Eq. (9). In order to complete the proof, we need to compute the cell attraction CA_{c_j} of cell c_j, which is defined as $SA_{c_j}/\sum_{z=1}^{C} SA_{c_z}$ (Eq. (2)). After substituting SA_{c_j} in this formula, we obtain Eqs. (29) and (30). □

Now that we know the formulas for CA_{c_i} and CA_{c_j}, we proceed to computing p_{ij} and p_{int} of Eq. (28). For p_{ij} the following lemma holds.

Lemma 4. (p_{ij}). *The probability p_{ij} of a movement from the home cell c_i to another cell c_j or vice versa is equal to*

$$p_{ij} = \frac{2(n-1)(1-p_e)p_r}{C-1(n(1-p_e)+(n-1)p_r)} \tag{33}$$

Proof. The probability p_{ij} can be expressed in terms of the transition probabilities of the Markov chain corresponding to HCMM. In particular, we are interested in the probability $P(c_j|c_i)$ of selecting c_j as the next goal cell given that the node is currently in its home cell c_i, and to the probability $P(c_i|c_j)$ of returning to the home cell c_i given that the node is currently in c_j. These probabilities must be multiplied by the probability $P(c_i)$ that a node is currently in its home cell and the probability $P(c_j)$ that a node is currently in cell c_j, respectively:

$$p_{ij} = P(c_j|c_i) * P(c_i) + P(c_i|c_j) * P(c_j) \tag{34}$$

In HCMM $P(c_i|c_j) = 1 - p_e$, while $P(c_i)$ and $P(c_j)$ correspond to the steady state probabilities of the Markov chain representing the HCMM evolution, i.e.,

$$P(c_i) = \frac{1-p_e}{(1-p_e)+(1-P(c_i|c_i))} \tag{35}$$

$$P(c_j) = \frac{P(c_j|c_i)}{(1-p_e)+(1-P(c_i|c_i))} \tag{36}$$

In HCMM $P(c_i|c_i) = CA_{c_i}$. If we substitute Eq. (30) into this formula, we can rewrite Eqs. (35) and (36) as

$$P(c_i) = \frac{1-p_e}{(1-p_e)+\frac{(n-1)p_r}{n}} \tag{37}$$

$$P(c_j) = \frac{P(c_j|c_i)}{(1-p_e)+\frac{(n-1)p_r}{n}} \tag{38}$$

Now the only component missing is $P(c_j|c_i)$. In HCMM $P(c_j|c_i)$ is equal to the cell attraction of c_j. By simply plugging Eq. (29), we obtain

$$P(c_j|c_i) = \frac{p_r(n-1)}{n(C-1)} \tag{39}$$

By simple substitutions into Eq. (34) we get the probability of having a movement between c_i and c_j: □

Corollary 2. *p_{ij} does not depend on the distance d travelled by nodes.*

Corollary 2 points out that, in the average case, movements towards near cells are as likely as movements towards distant cells in HCMM. This is in contrast with what has been shown about human mobility in [33] regarding preferential selection of short distances, and it calls for an extension of HCMM in this direction.

Going back to Eq. (28), now that we have computed p_{ij}, only p_{int} is left.

Lemma 5. (p_{int}). *The probability p_{int} of having a movement inside a cell is equal to*

$$p_{int} = -1 + 2p_e + \frac{2n(1-p_e)^2}{n(1-p_e)+(n-1)p_r}. \tag{40}$$

Proof. The procedure is similar to what we have described in the previous proof. The probability p_{int} of having a movement inside a cell is equal to the probability that the next movement will be within the home cell c_i given that the node is currently in c_i, plus the probability that the next movement will be within c_j given that the node is currently in c_j (Eq. (41)).

$$p_{int} = P(c_i|c_i) * P(c_i) + \sum_{i=1}^{C-1} P(c_j|c_j) * P(c_j) \tag{41}$$

$P(c_i)$ and $P(c_j)$ are given by Eqs. (37) and (38). $P(c_i|c_i)$ and $P(c_j|c_j)$ are equal to $P(c_i|c_i) = CA_i \sim \frac{n-p_r(n-1)}{n}$ and $P(c_j|c_j) = p_e$. After some substitutions we get

$$p_{int} = \frac{n-p_r(n-1)}{n} * \frac{1-p_e}{(1-p_e)+\frac{(n-1)p_r}{n}} + \sum_{i=1}^{C-1} p_e * \frac{P(c_j|c_i)}{(1-p_e)+\frac{(n-1)p_r}{n}} \tag{42}$$

By substituting the expression for $P(c_j|c_i)$ (Eq. (39)), we can rewrite p_{int} as in Eq. (40). □

Finally, using Lemmas 3–5 and Eqs. (27) and (28), we can compute the expression for $P(d)$.

Theorem 6. (Jump size distribution). *The distribution of jump sizes in HCMM is given by*

$$P(d) = \frac{1}{C}\sum_{i=1}^{C}\sum_{\substack{j=1\\ j\neq i}}^{C}\frac{2(n-1)(1-p_e)p_r}{(C-1)(n(1-p_e)+(n-1)p_r)} * P_{c_ic_j}(d) + \left(-1+2p_e+\frac{2n(1-p_e)^2}{n}(1-p_e)+(n-1)p_r\right) * P_{int}(d) \tag{43}$$

Proof. Eq. (27) follows from simple substitutions. □

$P_{int}(d)$ corresponds to the distribution of the jump size when a node is performing a movement within the same cell. These movements are analogous to Random Waypoint [34] movements, and therefore we can use for $P_{int}(d)$ the equation derived in [30] for the RWP model. This equation, valid for a rectangle of size $a \times b$, is given below for the convenience of the reader.

$$P_{int}(d) = \frac{4l}{a^2b^2} * \begin{cases} \frac{\pi}{2}ab - ad - bd + \frac{1}{2}d^2 & \text{for } 0 \leqslant d \leqslant b \\ ab\arcsin\frac{b}{d} + a * \sqrt{d^2-b^2} + & \\ -\frac{1}{2}b^2 - al & \text{for } b \leqslant d \leqslant a \\ ab\arcsin\frac{b}{d} + a * \sqrt{d^2-b^2} + & \\ -\frac{1}{2}b^2 - ab\arccos\frac{a}{d} + & \\ +b\sqrt{d^2-a^2} - \frac{1}{2}a^2 - \frac{1}{2}d^2 & \text{for } a \leqslant d \leqslant \sqrt{a^2+b^2} \\ 0 & \text{otherwise.} \end{cases} \tag{44}$$

With regard to $P_{c_ic_j}(d)$, there is no closed form for generic c_i and c_j, as discussed in Section 6.1.2. This implies that also for $P_i(d)$ only numerical solutions are available. As an alternative, one could resort to approximate solutions of $P_i(d)$ using the approximation proposed in Section 6.1.3.

6.2.1. Model validation

In this Section we compare the distribution given in Theorem 6 with the empirical distributions obtained through simulations. We consider a 10×10 grid on which 4 communities are placed, each having 50 nodes assigned. We set the rewiring probability to 0.1, v_{min} and v_{max} to 9 and 10, respectively, and we let the simulation run for 500,000 s to ensure stationarity. As the model requires as input the initial positions of communities, let us consider the scenario in which communities are placed, respectively, in positions (10, 1), (7, 4), (3, 4) and (1, 10) on a 10×10 grid (Fig. 16).

Analytical and simulative results are shown in Fig. 17, where the empirical probability distribution is obtained using the Kernel density estimation method [31]. The two densities, analytical and simulative, are very close to each other, thus showing that the analytical model is accurate. Note also that the first bell-shaped curve is associated with the shortest path that a node can travel, i.e., to local movements within a cell. In this case, local movements are statistically predominant across all movements of a node. The other bell curves are due to external movements. As only cells where friend communities are placed can be visited by a given node, not all possible distances on a $g_x \times g_y$ grid are represented in the plot, but just those at which a friend community is placed. Note also that all bell-shaped curves, apart from the bigger one, are at about the same height. Again, this confirms the fact that the standard HCMM has no mechanism for prioritizing the choice of the distances.

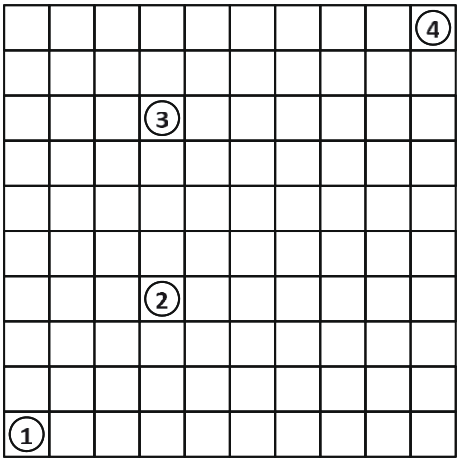

Fig. 16. Reference scenario.

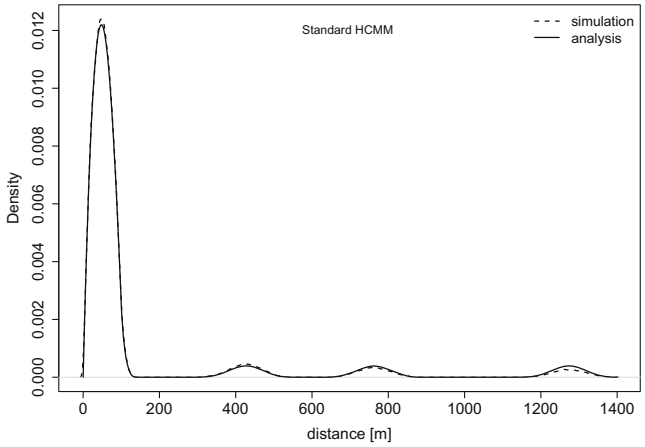

Fig. 17. Simulation vs. analysis – standard HCMM.

7. Accounting for spatial properties in HCMM

Corollary 2 in Section 6.2 showed, from a mathematical standpoint, that in HCMM there is no preferential selection of short distances, and the reason is that the rewiring process is uniformly distributed between communities. Simulations also confirmed this behaviour (Section 6.2.1). Given the important role of the jump size distribution on network performance [15], it would be highly desirable for HCMM to be able to include this third property of human mobility into the model. Therefore, in this section we complete the definition of HCMM in order to include such a feature.

The idea is the following. Being the uniform selection of the next goal cell the source of the non-preferential selection, we decided to remove it by explicitly including the preference for short distances over longer ones. In more details, when a node with home cell c_i has to select the target cell for an external trip, the attraction exerted on that node by a generic cell c_j is considered to be inversely proportional to the power of the average distance $d_{c_ic_j}$ between the home cell of the node and the cell c_j itself. The longer the distance, the weaker the attraction. Note that, as in standard HCMM, only cells for which the social attractivity is greater than zero are selected. Recalling that CA_{c_i} denotes the probability of selecting the home cell for the next movement (starting from

within the home cell itself), the probability of selecting cell c_j for the next movement is now:

$$CA_{c_j}^{dist} = \begin{cases} (1 - CA_{c_i}) \frac{d_{c_i c_j}^{-\alpha}}{\sum_{z=1}^{C-1} d_{c_i c_z}^{-\alpha}} & \text{if } SA_j > 0 \\ 0 & \text{otherwise} \end{cases}, \quad j \neq i \tag{45}$$

The rest of the transition probabilities remains as explained in Section 5, and thus the complete HCMM model can be described according to the Markov chain in Fig. 18. Note that social attraction here has the role of driving movements within or outside the home cell, while the distance among cells is the key driver for selecting the particular cell towards which an "external" movement is directed. Still, "external" cells can be selected only if some social relationships exist with nodes whose home cell is that particular cell.

7.1. Jump size distribution for the complete HCMM

In the following we complement the analysis in Section 6 by computing the jump size distribution under the complete HCMM model. Specifically, Propositions 2 and 3 still hold true also for the complete HCMM. Lemma 4 must be replaced with Lemma 6.

Lemma 6. (p_{ij} for complete HCMM). *The probability p_{ij} of a movement from the home cell c_i to another cell c_j or vice versa is equal to*

$$p_{ij} = \frac{2(n-1)n(1-p_e)p_r w_{c_i c_j}^{\alpha}}{n(1-p_e) + (n-1)p_r}, \tag{46}$$

where, for the sake of readability, $w_{c_i c_j}^{\alpha}$ substitutes $\frac{d_{c_i c_j}^{-\alpha}}{\sum_{z=1}^{C-1} d_{c_i c_z}^{-\alpha}}$

Proof. The proof is analogous to the proof for Lemma 4, but for $P(c_j|c_i)$. In fact, in the case of the complete HCMM, $P(c_j|c_i)$ depends on the average distance between c_j and the home cell c_i according to the following equation:

$$P(c_j|c_i) = (1 - CA_i) \frac{d_{c_i c_j}^{-\alpha}}{\sum_{z=1}^{C-1} d_{c_i c_z}^{-\alpha}} = n_e \frac{d_{c_i c_j}^{-\alpha}}{\sum_{z=1}^{C-1} d_{c_i c_z}^{-\alpha}}. \tag{47}$$

For the sake of readability, let us refer to $\frac{d_{c_i c_j}^{-\alpha}}{\sum_{z=1}^{C-1} d_{c_i c_z}^{-\alpha}}$ as $w_{c_i c_j}^{\alpha}$. CA_i is computed according to Lemma 3. Then, by simple substitutions, we obtain Eq. (46). Note here that p_{ij} depends on the distance through $w_{c_i c_j}^{\alpha}$. □

Lemma 5 is substituted by the following lemma.

Lemma 7. (p_{int} for complete HCMM). *The probability p_{int} of having a movement inside a cell is equal to*

$$p_{int} = \frac{(1-p_e)(-n + (n-1)p_r)}{n(1+p_e) + (-1+n)p_r} + \sum_{i=1}^{C-1} \frac{(n-1)p_e p_r w_{c_i c_j}^{\alpha}}{1 - p_e + \frac{(n-1)p_r}{n}}. \tag{48}$$

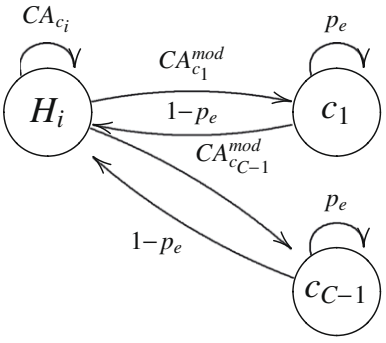

Fig. 18. Markov chain for modified HCMM.

Proof. The proof follows the same lines of the proof for Lemma 5. Here we only have to use the expression for $P(c_j|c_i)$ under the complete HCMM model (Eq. (47)). Then, after simple substitutions, we obtain Eq. (48). □

Finally, we provide the expression for the jump size distribution under the complete HCMM mobility model.

Theorem 7. (Jump size distribution for complete HCMM). *The distribution of jump sizes in the complete HCMM is given by*

$$P(d) = \frac{1}{C} \sum_{i=1}^{C} \sum_{\substack{j=1 \\ j \neq i}}^{C} \frac{2(n-1)n(1-p_e)p_r w_{c_i c_j}^{\alpha}}{n(1-p_e) + (n-1)p_r} * P_{c_i c_j}(d) + \left(\frac{(1-p_e)(-n + (n-1)p_r)}{n(1+p_e) + (-1+n)p_r} + \sum_{i=1}^{C-1} \frac{(n-1)p_e p_r w_{c_i c_j}^{\alpha}}{1 - p_e + \frac{(n-1)p_r}{n}} \right) * P_{int}(d). \tag{49}$$

Proof. Eq. (49) follows from Lemmas 6 and 7 after simple substitutions in Eq. (3). □

We have validated the proposed model for the HCMM using the same settings as in Section 6.2.1. Fig. 19 shows that, by adding the spatial dimension to HCMM, the two curves overlap almost perfectly. When comparing this plot with Fig. 17, the preference for short distances can be clearly noticed.

8. Statistical properties of HCMM traces

In this Section we evaluate the complete HCMM model by means of simulations in order to verify the temporal and spatial characteristics of the traces that are generated . In particular, we are interested in evaluating if prioritizing the distances chosen by the nodes is enough for obtaining a truncated power law distribution of the jump size. Let us first highlight a general property of HCMM (as well as any other models over finite physical spaces). In HCMM external movements are bounded by the size of the scenario, while local movements are bounded by the size of the cells. In addition, in HCMM nodes are bounded to move where their friend nodes are (or are suppose to be). This implies that not all cells, but only the ones where friend nodes are, can be selected. This choice is actually a sort of subsampling: not all the distances can be chosen, on the contrary, their set heavily depends on the social configuration of the mobility model. In the remaining of the Section we show how a power law behaviour emerges, or not, depending on the configuration of the mobility model. In summary, we show that there is a power law behaviour when

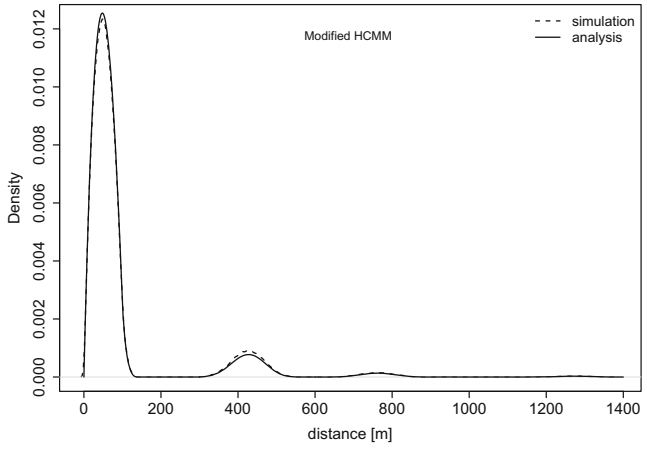

Fig. 19. Simulation vs. analysis – complete HCMM.

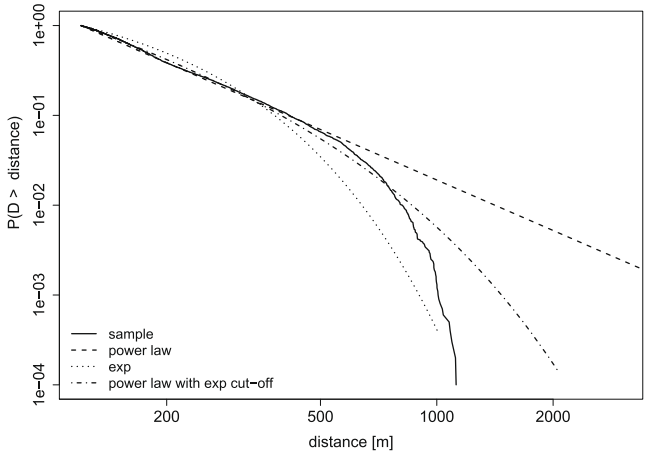

Fig. 20. MLE fitting – 100 communities.

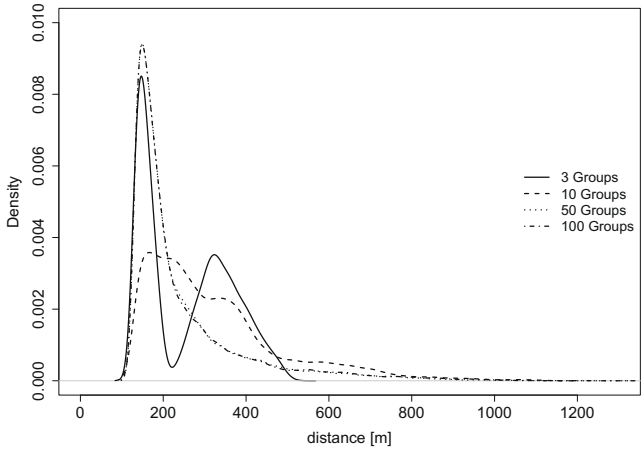

Fig. 21. Kernel density estimate with varying number of communities.

the density of the scenario, both in terms of communities and external social links, is sufficient to explore "enough" distances.

Our reference scenario is a 1000 m × 1000 m, divided into a 10 × 10 grid. The number of groups, the number of nodes, and the rewiring probability are varied in each set of simulations. The exponent α that we use for Eq. (45) is 3. The same results hold true for different values of α. We analyse the probability density function (PDF) and the complementary cumulative distribution function (CCDF) for the distances measured when the mobility model is in its steady state, and aggregated over all nodes.

8.1. Jump size distribution

We start with a dense scenario, where all cells have a community assigned to them. These 100 communities have 10 nodes each. In addition each node is friend of all communities (we force the rewiring process to guarantee such condition). In this case, there is no subsampling effects on the characteristics distances: (i) all cell combinations are possible (i.e., every characteristic distance is represented); (ii) each node has a non-zero probability of visiting all the cells of the grid. In Fig. 20, the continuous black line shows the empirical CCDF of the jump size obtained, while the other curves are the Maximum Likelihood Estimation fitting for the exponential, power law and power law with exponential cut-off case. The first part of the CCDF is very accurately fitted with the power law (with or without cut-off) distribution. On the tail, the CDF decreases very rapidly and it is delimited by the tail of the power law with cut-off and the exponential distribution. We can conjecture that, in its final part, the jump size distribution has an exponential decay. Analogously to what found in [35], this behaviour can be due to the bounded domain over which the simulation is performed. No distance can be present that goes beyond the boundaries of the scenario (here $1000\sqrt{2}$) and this is a good explanation for the rapid decay of the tail. This results are also consistent with [33], where a power law with an exponential decay was suggested as a very good approximation of human travelling patterns.

Now we evaluate the effect of removing some groups from the grid. Specifically, we consider the case in which 3, 10, 50, and 100 groups are present. In this case, some distances related to unoccupied cells are not selected. Therefore, the power law selection of the complete HCMM model will consider only a subset of all the possible distances, thus creating what we have called subsampling effect. This effect is evident in the probability density function in Fig. 21, where, for ease of reading, the contribution of the local movements has been removed. While with many groups, there is a constantly decreasing curve that is the result of the mixture in Eq. (27), with fewer groups we can clearly distinguish the single bell-shaped curves corresponding to the available distances. The existence of such bell-shaped components typically results in less clear power law distributions.

In the previous set of simulations, each node had as many friends as the number of available groups. By reducing the number of occupied cells, we have reduced the set of possible distances a node can travel. In the next set of simulations we reduce the average number of external friends that a node can have by setting the rewiring probability to 0.1. In this case, there is a further subsampling on the possible set of distances travelled by the users: not all the cells in which there is a group can be selected, but only those in which the node has at least one friend. Having a look at the PDF of jump sizes in Fig. 22, we can see that, with rewiring 0.1, even in the case of 100 communities the single bells are clearly visible. Specifically, each bell corresponds to the curve associated with a characteristic distance for which there is at least one friend. We can conclude that, when the rewiring is low, i.e., when the number of external links is low and there are only few external cells towards which a node can move, the bounded domain predominates over the power law behaviour.

A reduction in the set of characteristic distances can be obtained also by fixing the number of groups and varying the rewiring parameter. When the number of groups is equal to the number of cells, variations in the value of the rewiring parameter directly

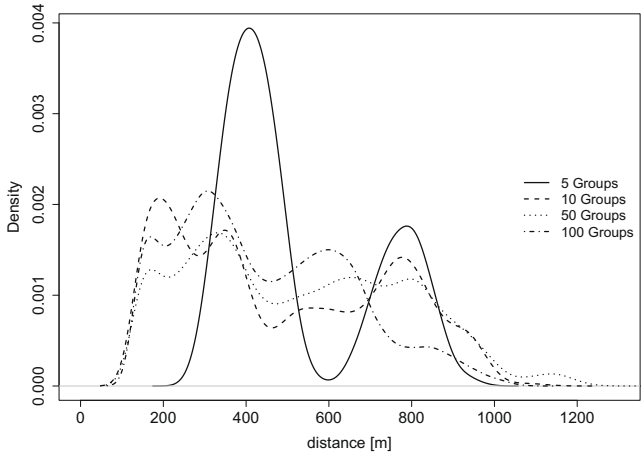

Fig. 22. Kernerl density estimate of jump size with rewiring 0.1.

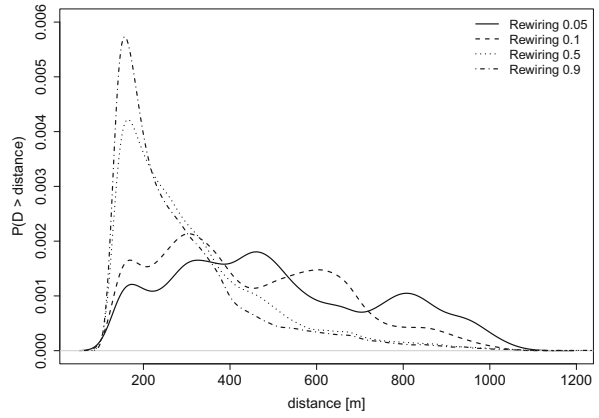

Fig. 23. Kernel density estimate of jump size with varying rewiring.

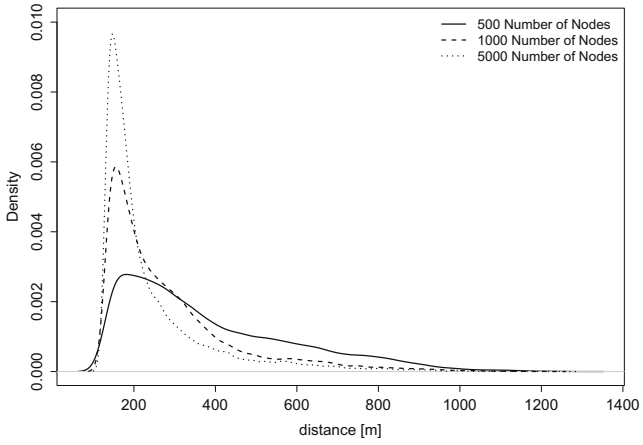

Fig. 24. Kernel density estimate of jump size with varying number of nodes.

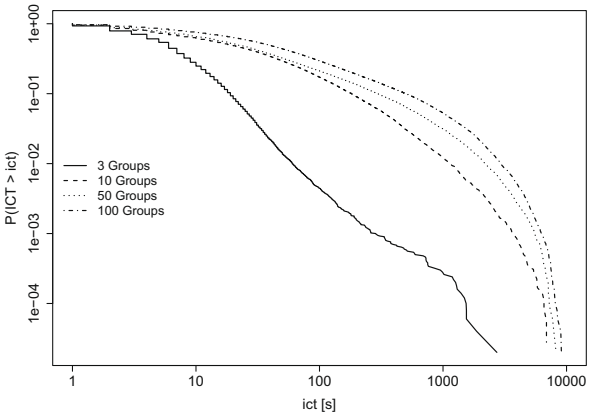

Fig. 25. ICTs when all communities are friend.

control the subsampling that is performed. This is shown in Fig. 23, where we go from a non-power-law behaviour with low rewiring (high subsampling) to clear preference for short distances with rewiring 0.9.

We have seen that the distribution of jump sizes heavily depends on the number of external friends of nodes. While the rewiring parameter constitutes an fundamental knob to control such a number, it is not the only factor. In fact, the rewiring probability says that x percent of the existing links will be rewired. In the presence of low density groups (few members per group), the number of available links between the nodes of the same community are low, and so will be the number of links rewired . The more the nodes, the less the subsampling, the clearer the power law behaviour: this effect is clearly shown in Fig. 24.

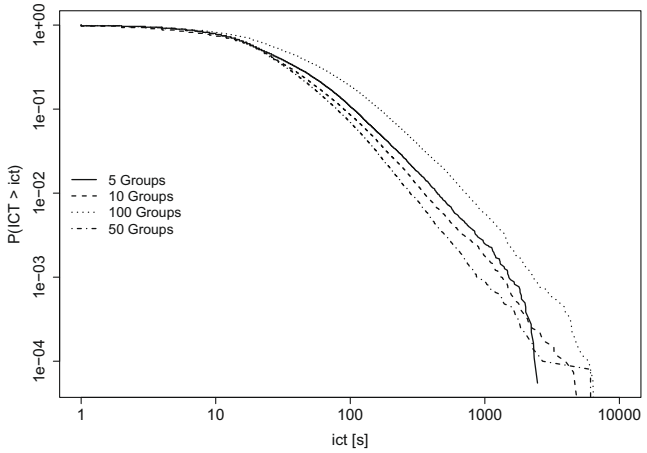

Fig. 26. ICTs with rewiring 0.1 and varying number of groups.

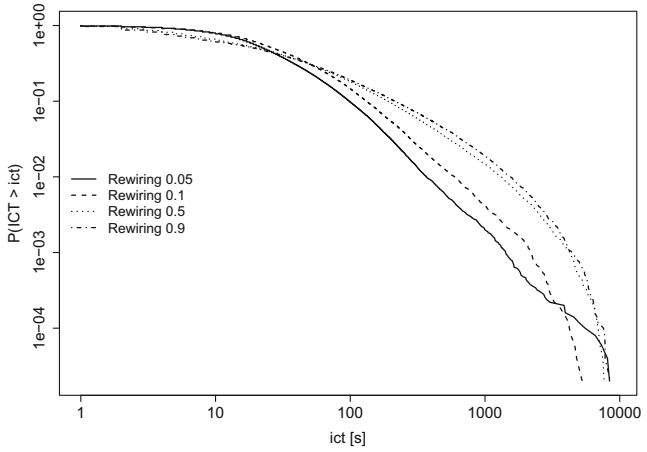

Fig. 27. ICTs with varying rewiring.

Concluding, in this Section we have shown that plugging a preferential selection of distances in the mobility models is a good method for obtaining jump sizes that follows a truncated power law. However, this is true only when the scenario is dense enough, in terms of both external links and number of communities. Otherwise, the subsampling effect due to the bounded domain distorts the power law shape giving way to bumpy curves that are no more power laws.

8.2. Inter-contact times distribution

In this paper we have previously discussed the importance of inter-contact times in mobility models, and in Section 2 we have surveyed the effort spent so far in order to uncover the nature of inter-contact times in real human mobility. In this section we give a plot for each scenarios studied in Section 8.1 and we describe the effect of such configuration on ICTs.

Let's start with the case where all nodes have friends in all communities and we vary the number of groups. Fig. 25 shows that a clear power law behaviour is present with few groups on the scenario and that a power law with exponential cut-off better describes the scenarios with more groups.

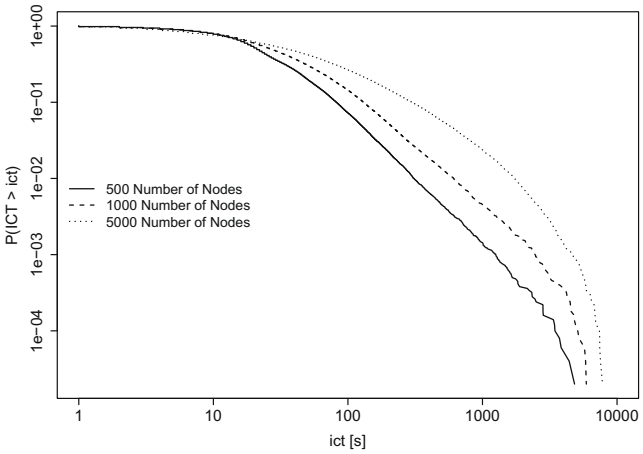

Fig. 28. ICTs with varying number of nodes and rewiring 0.1.

If we reduce the rewiring probability, the power law behaviour is still present (Fig. 26). In fact, the lower the rewiring probability, the fewer the external friends of nodes: some nodes will not have external friends at all, some other will have external friends, but very few (on average ~ 1 external friend per node). Thus the attraction of the home cell is predominant and the inter-contact times with external friends are long. Even worse, the closer (in a topological sense) friends will be preferred over the more distant ones, thus increasing the duration of inter-contact times with this latter class of friends.

When we fix the number of groups to 100 and we vary the rewiring probability (Fig. 27) or the number of groups (Fig. 28), the results are the same. We have a clear power law behaviour when the number of external friends is low (low rewiring or small number of nodes) that changes to a power law with exponential cut-off as the number of external friends increases.

In all the configurations, the inter-contact times generated by HCMM model follows either a power law or a power law with exponential cut-off distribution. This result is consistent with the feature of realistic ICTs highlighted in [8,19]. Thus, the HCMM model is able to reproduce realistic temporal properties of human mobility.

9. Conclusions

In this paper we have identified three main properties of human mobility, namely, social attraction, location attraction, and preference for short distances. The mobility model that we have proposed (HCMM) is the first model, to the best of our knowledge, that accounts for all of them at the same time. HCMM has been described starting from a pure social-based model through incremental steps, where, at each stage, we have mathematically justified the need for extending the model features. Specifically, we have provided a mathematical model for the gregarious behaviour of CMM, for the controllability of HCMM, and for the distribution of distances under both the standard and complete HCMM. Finally we have checked that HCMM is able to reproduce the main features of human mobility in terms on ICTs and jump size. For what concerns the jump size, we have also highlighted the impact of the subsampling effect on grid-based models. As future work, we plan to exploit the analytical framework that we have provided for the HCMM model in order to study from a mathematical standpoint some general properties of grid-based mobility model and the performance of networking protocols for mobile ad hoc networks.

Acknowledgments

This work was partially funded by the European Commission under the HAGGLE (027918) and SOCIALNETS (217141) FET Projects.

References

[1] M. Conti, S. Giordano, Multihop ad hoc networking: the reality, IEEE Communications Magazine 45 (4) (2007) 88–95.
[2] L. Pelusi, A. Passarella, M. Conti, Opportunistic networking: data forwarding in disconnected mobile ad hoc networks, IEEE Communications Magazine 44 (11) (2006) 134–141.
[3] J. Silvis, D. Niemeier, R. D'Souza, Social networks and travel behavior: report from an integrated travel diary, in: 11th International Conference on Travel Behaviour Research, Kyoto, 2006.
[4] D. Brockmann, L. Hufnagel, T. Geisel, The scaling laws of human travel, Nature 439 (7075) (2006) 462–465.
[5] M. Gonzalez, C. Hidalgo, A. Barabasi, Understanding individual human mobility patterns, Nature 453 (7196) (2008) 779–782.
[6] I. Rhee, M. Shin, S. Hong, K. Lee, S. Chong, On the levy-walk nature of human mobility, in: Proceedings of the INFOCOM 2008, The 27th Conference on Computer Communications. IEEE, 2008, pp. 924–932.
[7] T. Henderson, D. Kotz, I. Abyzov, The changing usage of a mature campus-wide wireless network, Computer Networks 52 (14) (2008) 2690–2712.
[8] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, J. Scott, Impact of human mobility on opportunistic forwarding algorithms, IEEE Transactions on Mobile Computing (2007) 606–620.
[9] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, A. Tomkins, Geographic routing in social networks, Proceedings of the National Academy of Sciences 102 (33) (2005) 11623–11628.
[10] W. Hsu, T. Spyropoulos, K. Psounis, A. Helmy, Modeling time-variant user mobility in wireless mobile networks, in: Proceedings of the IEEE INFOCOM, Citeseer, 2007, pp. 758–766.
[11] K. Herrmann, Modeling the sociological aspects of mobility in ad hoc networks, in: Proceedings of the 6th ACM International Workshop on Modeling Analysis and Simulation of Wireless and Mobile Systems, ACM, New York, NY, USA, 2003, pp. 128–129.
[12] M. Musolesi, C. Mascolo, Designing mobility models based on social network theory, SIGMOBILE Mobile Computing and Communications Reviews 11 (3) (2007) 59–70.
[13] V. Borrel, F. Legendre, M. De Amorim, S. Fdida, Simps: using sociology for personal mobility, IEEE/ACM Transactions on Networking (TON) 17 (3) (2009) 831–842.
[14] J. Ghosh, M.J. Beal, H.Q. Ngo, C. Qiao, On profiling mobility and predicting locations of wireless users, in: REALMAN '06: Proceedings of the Second International Workshop on Multi-hop ad hoc Networks: From Theory to Reality, ACM, New York, NY, USA, 2006, pp. 55–62.
[15] P. Wang, M. Gonzalez, C. Hidalgo, A. Barabasi, Understanding the spreading patterns of mobile phone viruses, Science 324 (5930) (2009) 1071–1076.
[16] M. Balazinska, P. Castro, Characterizing mobility and network usage in a corporate wireless local-area network, in: MobiSys '03: Proceedings of the First International Conference on Mobile Systems, Applications and Services, ACM, New York, NY, USA, 2003, pp. 303–316.
[17] N. Eagle, A. Pentland, Reality mining: sensing complex social systems, Personal and Ubiquitous Computing 10 (4) (2006) 255–268.
[18] 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.
[19] T. Karagiannis, J.-Y. Le Boudec, M. Vojnović, Power law and exponential decay of inter contact times between mobile devices, in: MobiCom '07: Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Networking, ACM, New York, NY, USA, 2007, pp. 183–194.
[20] W. Gao, Q. Li, B. Zhao, G. Cao, Multicasting in delay tolerant networks: a social network perspective, in: MobiHoc '09: Proceedings of the 10th ACM International Symposium on Mobile ad hoc Networking and Computing, ACM, New York, NY, USA, 2009, pp. 299–308.
[21] A. Barabasi, The origin of bursts and heavy tails in human dynamics, Nature 435 (7039) (2005) 207–211.
[22] T. Camp, J. Boleng, V. Davies, A survey of mobility models for ad hoc network research, Wireless Communications and Mobile Computing 2 (5) (2002) 483–502.
[23] P. Hui, E. Yoneki, S.Y. Chan, J. Crowcroft, Distributed community detection in delay tolerant networks, in: MobiArch '07: Proceedings of the Second ACM/IEEE International Workshop on Mobility in the Evolving Internet Architecture, ACM, New York, NY, USA, 2007, pp. 1–8.
[24] K. Lee, S. Hong, S.J. Kim, I. Rhee, S. Chong, Slaw: A new mobility model for human walks, in: Proceedings of the INFOCOM 2009, The 28th Conference on Computer Communications. IEEE, 2009, pp. 855–863.
[25] A. Mei, J. Stefa, Swim: a simple model to generate small mobile worlds, in: Proceedings of the INFOCOM 2009, The 28th Conference on Computer Communications. IEEE, 2009, pp. 2106–2113.

[26] C. Boldrini, M. Conti, A. Passarella, Users mobility models for opportunistic networks: the role of physical locations, in: IEEE WRECOM 2007, 2007, pp. 1–6.
[27] C. Boldrini, M. Conti, A. Passarella, The sociable traveller: human travelling patterns in social-based mobility, in: 7th ACM International Symposium on 7th ACM International Symposium on Mobility Management and Wireless Access (MOBIWAC'09), ACM, 2009, pp. 1–8.
[28] D. Watts, Small Worlds: The Dynamics of Networks Between Order and Randomness, Princeton University Press, 1999.
[29] S. Ross, Introduction to Probability Models, Academic Press, 2006.
[30] C. Bettstetter, H. Hartenstein, X. Pérez-Costa, Stochastic properties of the random waypoint mobility model, Wireless Networks 10 (5) (2004) 555–567.
[31] F. Dekking, C. Kraaikamp, H. Lopuha, L. Meester, A Modern Introduction to Probability and Statistics – Understanding Why and How, Springer Texts in Statistics, Springer, 2005.
[32] A. Mathai, P. Moschopoulos, G. Pederzoli, Random points associated with rectangles, Rendiconti del Circolo Matematico di Palermo 48 (1) (1999) 163–190.
[33] M. González, C. Hidalgo, A. Barabási, Understanding individual human mobility patterns, Nature 453 (7196) (2008) 779–782.
[34] D. Johnson, D. Maltz, Dynamic source routing in ad hoc wireless networks, Kluwer International Series in Engineering and Computer science (1996) 153–179.
[35] Cai, Han, Eun, Do Young, Crossing over the bounded domain: from exponential to power-law inter-meeting time in MANET, in: ACM MOBICOM 2007, 2007, pp. 159–170.