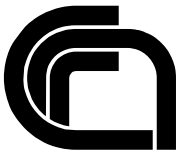

Consiglio Nazionale delle Ricerche

The sociable traveller: human travelling patterns in social-based mobility

C. Boldrini, M. Conti, A. Passarella

IIT TR-07/2009

Technical report

Luglio 2009

Istituto di Informatica e Telematica

The sociable traveller: human travelling patterns in social-based mobility

Chiara Boldrini
IIT-CNR
Via G.Moruzzi 1
Pisa, Italy
chiara.boldrini@iit.cnr.it

Marco Conti
IIT-CNR
Via G.Moruzzi 1
Pisa, Italy
marco.conti@iit.cnr.it

Andrea Passarella
IIT-CNR
Via G.Moruzzi 1
Pisa, Italy
andrea.passarella@iit.cnr.it

ABSTRACT

Understanding how humans move is a key factor for the design and evaluation of networking protocols and mobility management solutions in mobile networks. This is particularly true for mobile scenarios in which conventional single-hop access to the infrastructure is not always possible, and multi-hop wireless forwarding is a must. We specifically focus on one of the most recent mobile networking paradigms, i.e., opportunistic networks. In this paradigm the communication takes place directly between the personal devices (e.g., smartphones and PDAs) that the users carry with them during their daily activities, without any assumption about pre-existing infrastructures. Among all mobility characteristics that may affect the performance of opportunistic networks, the users' travelling patterns have recently gained a lot of attention due to their impact on the spreading of both viruses and messages in such a network. In this paper we consider a social-based mobility model (HCMM) and we extend this model to account for the typical travelling behaviour of users. To the best of our knowledge, the resulting mobility model is the first model in which movements driven by social relations also match statistical features of travelling patterns as measured in reality. Finally, we evaluate our proposal through simulations over a wide range of scenarios, emphasizing the effect of finite sampling on the obtained results.

Categories and Subject Descriptors

C.4 [**Performance of Systems**]: Network Architecture and Design; C.2.1 [**Network Architecture and Design**]: Wireless Communication

General Terms

Algorithms, Design, Performance

Keywords

mobility models, power law distribution, human movements

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
MOBIWAC'09 Tenerife, Canary Islands, Spain
Copyright 200X ACM X-XXXXX-XX-X/XX/XX ...$10.00.

1. INTRODUCTION

One of the most intriguing direction of current networking research is that of *pervasive* (or *ubiquitous*) computing. A future is imagined where user devices will be small, lightweight, powerful and self-configuring. The conventional Internet will be a part of a larger network, in which mobility will be commonplace. Access to the Internet from mobile devices will not be limited to conventional single hop solutions (such as WiFi Access Points or cellular networks), but will also be enabled by "spontaneous" networking paradigms, in which ad hoc networks will be formed by mobile users' devices. This idea of transparent ubiquitous technology relies on a communication paradigm that is necessarily ubiquitous as well. This extends conventional strategies involving cable deployment or set up of an infrastructure only. The new keywords are *wireless* and *infrastructureless*. In such a network, the concept of a user docked to its desk while using his personal computer is not the only paradigm anymore. In a pervasive environment, devices like smartphones or PDA are indeed carry-on personal computers that are always with their owners during their daily activities and move with them.

The work on infrastructureless wireless networks started with Mobile Ad hoc Networks (MANETs). Traditional MANETs dealt with mobility, but never leveraged it: mobility was an accident rather than a feature. In MANETs the communication takes place directly between the nodes participating to the networks, without the need for an infrastructure, but they required the nodes to be connected with a multi-hop path at the time the communication takes place. In some cases, this is simply not possible. In a sparse environment, for example, nodes can be connected to different subnetworks which get in touch only sporadically. The messages having the source and destination nodes in different subnetworks, and generated when the two nodes are not in touch, will never be delivered. Opportunistic and Delay Tolerant networks (DTN) try to address this problem. Specifically, they study networking solutions for real ubiquitous environments. In opportunistic networks multi hop paths are built opportunistically, and each node carrying a message exploits other peers it gets directly in touch to forward the message for the intended destination.

As opportunistic networks exploit mobility of users as a key feature, a solid understanding of users mobility patterns is a requirement in ubiquitous networking environments, both for mobility management solutions and for the design and evaluation of networking protocols. This holds true for at least two main reasons. First, having an accurate knowl-

edge on the characteristics of human mobility is a prerequisite to design protocols for environments where the communication takes place between moving entities. Second, when these protocols have been developed, they must be tested on the scenario for which they were intended. This implies that realistic mobility models must be available for plugging them into simulators and, therefore, using them for evaluating the performance of networking protocols. Many characteristics of mobility have caught the interest of researcher on networking, above all the contact and inter-contact times, that directly impact on the performance of, e.g., routing protocols. However, recently also the distances travelled by users have gained a lot of attention, due to the seminal work of [10]. In this paper we focus on this metric.

For what concerns the first point, i.e., understanding how humans move, many papers in the last few years have tried to statistically describe the features of mobility from different points of view. In particular, there has been a long discussion on the nature of contact and inter-contact times [8] [15] [9] [21]. The contact time (CT) corresponds to the duration of a contact between two users: depending on the scenario considered, it can be the duration either of a bluetooth association or of the staying under the coverage of the same Access Point or of the calls between the users of the same mobile phone company. The inter-contact time (ICT) is the time elapsed between two consecutive contacts (again the exact semantic depends on the scenario under study). There is no final agreement on the distribution that better approximates the shape of CTs and ICTs. However, modeling them with a power law with exponential cut-off seems to be the most popular approach. While CTs and ICTs are important, e.g., for the delay of messages, the travelling patterns of the users of a network can seriously determine the way in which, e.g., viruses or messages spread throughout the network. We refer to Section 2 for a more detailed overview of the literature on this subject. In the case of distances (also referred to as jump size), there is a general agreement on the fact that their distribution can be well approximated with a power law with exponential cut-off [6] [10] [20].

With regard to the second point, i.e., the definition of mobility models for network evaluations, there have been many proposals in the last years. One the most interesting category of mobility models are social-aware mobility models. This models have identified in the social relationships between the users the driver of users' movements [13][5]. Often these models employ concepts from social network theory to formalize the relations between nodes [18][3]. The opposite approach to social-aware mobility is location-aware mobility. Here the idea is that nodes move across specific locations, whose popularity exerts different attraction on nodes. In [3] we claimed that, in a realistic mobility model, social attraction and location attraction must go together. Therefore we extended an existing social-based model (CMM [18]) to include the attraction towards popular locations. The resulting mobility model (HCMM) is the object of the work presented in this paper. The evaluation of both CMM and HCMM has focused on the inter-contact times generated by the model. In this paper, we analyze the HCMM model from the point of view of the generated distance distribution. In particular, we show (Section 3) that the HCMM as-it-is is not able to reproduce important statistical features of the jump size distribution. Therefore, in Section 4 we modify the standard HCMM to include such a feature. In Section 5 we propose a mathematical model that describes the jump size distribution, for both the standard and the modified HCMM. Finally, in Section 6 we evaluate through simulation the modified HCMM, showing that realistic jump sizes are now produced.

The main contributions of this paper are:

- the definition of a new mobility model that merges sociality with realistic travelling patterns
- the definition of a mathematical model for the distribution of distances under this mobility model
- the analysis of the statistical features of the new mobility model.

2. RELATED WORK

Understanding how people move has been the goal of many works. Pedestrian motion and traffic dynamics have been described [11] in terms of the self-organizing behaviour shown by the participating entities, often referred to also as collective or swarm intelligence. The complexity of these systems is the result of the interaction among the single entities that behave according to very simple rules. Example of these applications are the emergency of phantom traffic jams on highways [11] [19] or trail formations in city parks [12]. Knowing traffic dynamics can allow more efficient transportation and prevention systems, with the result of less time wasted stuck in a traffic jams, increased driving safety (accidents are more likely to occur when sudden traffic jams creates), reduced pollution, and so on. Pedestrian mobility behavior is a valuable piece of information for urban planning or for coping with panic situations.

Recently the interest in understanding human mobility patterns has gained new popularity, triggered by studies on biological and mobile epidemics. At the beginning, air transportation and long-distance traffic were studied. However, human mobility is more complex that that, and can involve long and short distances, different types of transportation media, etc. This complexity has been a great obstacle to quantitatively assess the properties of human travelling. The first comprehensive study was presented in [6], where the path of banknotes has been used as an alias for the path of the humans carrying such banknotes. They found that the distance travelled by banknotes had a power law decay, thus implying that the corresponding trajectories are Lévy flights, whose superdiffusive spread is attenuated by long periods of rest. It has been argued, however, as banknotes go from one person to another, that the recorded paths were not directly related to actual human movements but were rather a convolution of some unknown human mobility pattern.

The dispute has been settled by [10], where traces related to mobile phone users are used. In this case, there is a more direct correspondence between the recorded paths and the user itself, because mobile phones are carried during the everyday routine. The observed distribution of distances between users' positions at consecutive calls are well described by a truncated power law, corresponding to a truncated Lévy flight.

The fact that human movements are associated with Lévy flights is confirmed also by [20]. Here authors analyse GPS traces related to different settings (college campuses, city,

theme park, state fair) and in all cases the power law tendency is detected. Very interesting also the observation about the reason for the power law: it is not due to geographical constraints, but its generated by human intentions, as already suggested in [1]. Along with the power law tendency, authors of [20] also highlight the effect of a limited area truncating the tail of the power law.

In [16], the same authors propose a mobility model, called SLAW (Self-similar Least Action Walk), that generates truncated power law jump sizes. To the best of our knowledge, SLAW is the only mobility model, together with the modified HCMM that we propose in this paper, that tries to map realistic travelling patterns in addition to realistic inter-contact times. But, unlike HCMM, SLAW does not include any concept of socialization between nodes. This model includes also other aspects of real human mobility, such as truncate power law inter-contact times and the fact that people are attracted by popular locations.

3. HCMM

The Home Cell Mobility Model (HCMM) [3] merges two fundamental aspects of human mobility: social and location attraction. The main idea behind this model is that users move towards other people with whom they share social relations, or, even better, towards places where these people are supposed to be Nodes (representing users) move in an $m \times n$ grid. They are organized into social communities, and each community is initially assigned to one of the cells of the grid. HCMM, as the name suggests, uses the concept of *home cell*. Each user is supposed to belong to a main social community (at any given point in time), and the user home cell is the cell within which the members of its social community preferentially move. Users sometimes also move to other communities in different cells, due to social relations with communities other than their main one.

In more detail, HCMM is inspired by the well known Caveman model [22], and works as follows. Given an initial number of communities C, each node is assigned to one of these communities, until all communities have the same number of members (referred to as n in the following of the paper). The social relationships between nodes are represented through a weighted graph. A link between two nodes x and y implies that the two nodes are "friend" (i.e., share a social relation) and the weight w_{xy} of the link gives the strength of the friendship. At the beginning, all the nodes belonging to the same community are connected with a social link. Each of these links is then uniformly rewired to the nodes of other communities with a probability p_r, called *rewiring probability*. This is the method for creating external relationships, i.e., relationships between nodes belonging to different communities and thus to different home cells. The friendship between nodes is then used to trigger nodes' movements. In HCMM the endpoint of a movement is called *goal*. The selection of the goal for the next movement consist of two 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 the case of a node x roaming withing its home cell. x will remain inside the home cell for the next movement with a probability proportional to the social attraction of the home cell. Otherwise x will move towards another cell c_j with a probability proportional to the social attraction exerted by cell c_j. 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:

$$SA_{c_i} = \sum_{\substack{y=1 \\ y \in c_i}}^{n} w_{xy}, \qquad \forall i = 1, \cdots, C \tag{1}$$

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 for node x with a probability CA_{c_i} equal to

$$CA_{c_i} = SA_{c_i} / \sum_{j=1}^{C} SA_{c_j}. \tag{2}$$

When x is located in an external cell, the selection of the next goal is performed as follows. With a probability p_e, x remains in the external cell for the next movement, 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, and the node starts to move towards the goal. The whole process is summarized in Figure 1 using a Markov Chain .

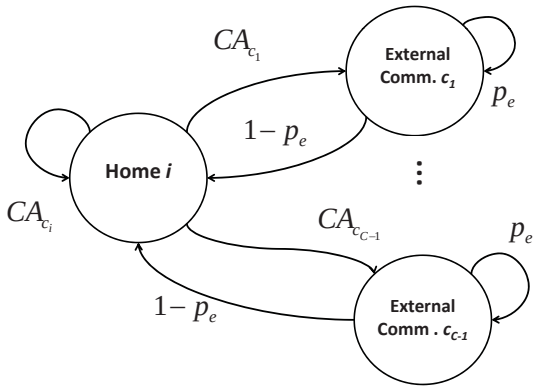

Figure 1: Markov chain for standard HCMM

Analyzing HCMM from the point of view of jump sizes, we can identify two main kinds of movements: movements within the same cell (or *local*) and movements from one cell to another (or *external*). External movements are generally longer than local 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 nodes towards which a link is rewired is selected uniformly at random. In addition, having communities 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 cells 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, as a consequence, no truncated power law jump size can appear. This argument will be formalized in Section 5 using an analytical model. However, this result is in contrast with what has been shown in [10] about real mobility. Therefore, in the next Section, we propose a modification to the standard HCMM to better reflect real mobility. We anticipate that this modification preserves the property of the original HCMM (shown in [3]) of producing truncated power law contact and inter-contact times. For reasons of space, this is shown in [4].

4. MODIFIED HCMM

In the standard HCMM the problem is that, being the rewiring process uniformly distributed between communities, there is no preferential selection of short distances. Instead, in [10] it is clearly shown that people tend to travel preferentially over short trips and this is at basis of the characteristic distribution of jump sizes. Therefore, we have modified the standard HCMM to include preferential selection of short distances.

After the completion of a trip within its home cell c_i, a node can perform either another trip within the home cell (with probability proportional to the social attraction of the home cell, that is computed as in standard HCMM) or a trip towards an external cell with the complementary probability. When an external trip is drawn, the external goal cell is selected differently from the standard HCMM. In fact, each cell c_j is selected with a probability inversely proportional to the power of the average distance $d_{c_i c_j}$ between the home cell of the node and the cell itself (Equation (3)). With this modification to the original HCMM algorithm, short distances are preferred over long ones. Note that, as in stardard HCMM, only cells for which the social attractivity is greater than zero are selected. For what concerns the other movement dynamics, they remain the same as in standard HCMM. The modified model is summarized in Figure 2.

$$CA_{c_j} = \begin{cases} (1 - CA_i)\frac{d_{c_i c_j}^{-\alpha}}{\sum_{z=1}^{C-1} d_{c_i c_z}^{-\alpha}} & \text{if } SA_j > 0 \\ 0 & \text{otherwise} \end{cases}, \qquad j \neq i \tag{3}$$

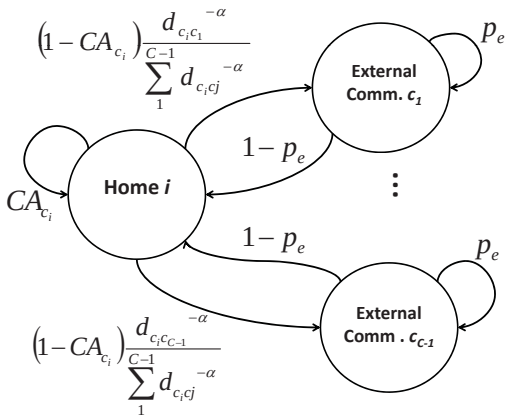

Figure 2: Markov chain representing the modified HCMM

5. MODELLING MOVEMENTS IN HCMM

In this Section we propose a model for the distribution of jump sizes, $P(d)$, that is applicable to both the standard and the modified HCMM. The assumption of the model is that the initial placement of communities on the grid is given and no more than one community can be assigned to the same cell. We consider a $m \times n$ grid scenario with C communities, each having assigned n nodes. The rewiring probability is 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 their home cell and the other cells of teh scenario. Given this fact, all nodes belonging to the same home cell "see" the other cells at the same distance. Separating the contributions $P_i(d)$ of the different home cells, we can write $P(d)$ as

$$P(d) = \frac{1}{C}\sum_{i=1}^{C} P_i(d) \tag{4}$$

Let us assume that the home cell of community i is cell c_i. All the formulas given hereafter refer to a tagged home cell c_i and are valid for all its nodes. In addition, as in HCMM nodes cannot move towards empty cells, we simply discard such cells from the model.

In HCMM, as in many grid-based mobility models, 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 $P_{c_i c_j}(d)$ that can be computed exactly but not in a closed form (see Appendix A for more details). When the starting and destination points are within the same cell, the distribution followed by the distance is the same as that of a node moving according to the Random Waypoint mobility model [14] and has already been computed in [2]. Being all cells statistically equivalent, the distribution of the movements within a cell is the same for all cells and equal to $P_{int}(d)$. The distributions for external and internal movements taken together can describe exhaustively all the possible paths that a node can travel. We can then express the distribution of movements of nodes belonging to community c_i as the composition of the distributions $P_{c_i c_j}(d)$ 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 to which friend nodes belong can be selected as the goal of a movement. In addition, in the modified version of HCMM, short paths are preferred over longer ones. Therefore we have to weight $P_{c_i c_j}(d)$ and $P_{int}(d)$ with the probabilities p_{ij} and p_{int} of actually having that component according to HCMM rules. Equation (5) shows the formula for $P_i(d)$.

$$P_i(d) = \sum_{\substack{j=1 \\ j \neq i}}^{C} p_{ij} * P_{c_i c_j}(d) + p_{int} * P_{int}(d) \tag{5}$$

In order to compute the probabilities p_{ij}, first we need to compute the social attraction SA and the cell attraction CA. 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 cells. 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 nodes in the same community. Each set of rewired links for a node x is a realization of $B(n-1,p)$. 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

$$n_e = (n-1) * p_r \tag{6}$$

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

$$n_e^{c_j} = n_e^c = \frac{(n-1) * p_r}{C-1}, \tag{7}$$

where we dropped the index j because the formula does not depend on the community j chosen. Please note that so far there have been no differences between the standard and the modified HCMM. If all weights are either 0 or 1, using Equation (7) we can write the social attraction SA exerted by a cell c_j on a generic node x as

$$SA_{c_j} = \sum_{\substack{y=1 \\ y \in c_i}}^{n} w_{xy} \sim n_e^c. \tag{8}$$

Following the same line of reasoning, the social attraction of the home cell c_i, that is determined by the links with nodes sharing the same home cell, can be written as

$$SA_{c_i} \sim n - n_e. \tag{9}$$

The cell attraction of cell c_j is equal to $SA_{c_j} / \sum_{z=1}^{C} SA_{c_z}$ (Equation (2)). After substituting Equations (8) and (9) in the formula, we obtain

$$CA_{c_j} = \frac{n_e^c}{(C-1)n_e^c + (n - ne)} = \frac{p_r(n-1)}{n(C-1)}. \tag{10}$$

Similarly, for the home cell the cell attraction is given by

$$CA_{c_i} = \frac{n - ne}{(C-1)n_e^c + (n - ne)} = \frac{n - p_r(n-1)}{n}. \tag{11}$$

We can now go on completing the analysis of Equation (5). Let us firstly focus on external movements. i.e., on p_{ij}. In Equation (5) p_{ij} gives the probability of a movement from the home cell to another cell c_j or vice versa. It can be expressed in terms of the transition probabilities of the Markov chain corresponding to the standard or modified 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:

$$p_{ij} = P(c_j|c_i) * P(c_i) + P(c_i|c_j) * P(c_j) \tag{12}$$

For both standard and modified HCMM, $P(c_i|c_j) = 1 - p_e$. $P(c_i)$ and $P(c_j)$ corresponds to the steady state probabilities for the standard and modified HCMM Markov chains, i.e.,

$$P(c_i) = \frac{1 - p_e}{(1 - p_e) + (1 - P(c_i|c_i))} \tag{13}$$

$$P(c_j) = \frac{P(c_j|c_i)}{(1 - p_e) + (1 - P(c_i|c_i))}. \tag{14}$$

For both models $P(c_i|c_i) = CA_{c_i}$. If we substitute Equation (11) into this formula, we can rewrite Equations (13) and (14) as

$$P(c_i) = \frac{1 - p_e}{(1 - p_e) + \frac{(n-1)p_r}{n}} \tag{15}$$

$$P(c_j) = \frac{P(c_j|c_i)}{(1 - p_e) + \frac{(n-1)p_r}{n}} \tag{16}$$

Now the only component missing is $P(c_j|c_i)$. In stardard HCMM $P(c_j|c_i)$ is equal to the cell attraction of c_j. By simply plugging Equation (10), we obtain

$$P_{std}(c_j|c_i) = \frac{p_r(n-1)}{n(C-1)} \tag{17}$$

Equation (17) is of absolute importance for our analysis. It says that $P_{std}(c_j|c_i)$ does not depend on the distance: i.e., in the average case, movements towards near cells are as likely as movements towards distant cells. This is in contrast with what has been shown about human mobility in [10]. It confirms the argument given in 3 about the reason why we decided to modify the original HCMM to include preferential selection of distances.

In the case of modified 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:

$$P_{mod}(c_j|c_i) = (1 - CA_i)\frac{d_{c_i c_j}^{-\alpha}}{\sum_{z=1}^{C-1} d_{c_i c_z}^{-\alpha}} = n_e \frac{d_{c_i c_j}^{-\alpha}}{\sum_{z=1}^{C-1} d_{c_i c_z}^{-\alpha}}. \tag{18}$$

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 $w_{c_i c_j}^{\alpha}$. By simple substitutions into Equation (12) we get the probability of having a movement between c_i and c_j:

$$p_{ij}^{std} = \frac{2(n-1)\,(1-p_e)\,p_r}{(C-1)\,(n\,(1-p_e) + (n-1)p_r)} \tag{19}$$

$$p_{ij}^{mod} = \frac{2(n-1)n\,(1-p_e)\,p_r w_{c_i c_j}^{\alpha}}{n\,(1-p_e) + (n-1)p_r} \tag{20}$$

Note again that p_{ij}^{std} does not depend on j (the target community), while p_{ij}^{mod} does. With regard to Equation (5), now that we have computed p_{ij}, only p_{int} is left. The procedure is similar to what we have described above. 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 (Equation (21)).

$$p_{int} = P(c_i|c_i) * P(c_i) + \sum_{i=1}^{C-1} P(c_j|c_j) * P(c_j) \quad (21)$$

$P(c_i)$ and $P(c_j)$ are given by Equations (15) - (16). $P(c_i|c_i)$ and $P(c_j|c_j)$ are the same for standard and modified HCMM and 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

$$\begin{aligned} p_{int} &= \frac{n-p_r(n-1)}{n} * \frac{1-p_e}{(1-p_e)+\frac{(n-1)p_r}{n}} + \\ &+ \sum_{i=1}^{C-1} p_e * \frac{P(c_j|c_i)}{(1-p_e)+\frac{(n-1)p_r}{n}} \end{aligned} \quad (22)$$

Being $P(c_j|c_i)$ different for the standard and modified HCMM, we can rewrite p_{int} as

$$p_{int}^{std} = -1 + 2p_e + \frac{2n\,(1-p_e)^{\,2}}{n(1-p_e)+(n-1)p_r} \quad (23)$$

$$\begin{aligned} p_{int}^{mod} &= \frac{(1-p_e)\,(-n+(n-1)p_r)}{n\,(1+p_e)+(-1+n)p_r} + \\ &+ \sum_{i=1}^{C-1} \frac{(n-1)p_e p_r w_{c_i c_j}^{\alpha}}{1-p_e+\frac{(n-1)p_r}{n}} \end{aligned} \quad (24)$$

Finally we can compute the expression for $P(d_i)$ for standard HCMM, as

$$\begin{aligned} P_{std}(d_i) &= \sum_{\substack{j=1 \\ j\neq i}}^{C} \frac{2(n-1)\,(1-p_e)\,p_r}{(C-1)\,(n\,(1-p_e)+(n-1)p_r)} * P_{c_i c_j}(d) + \\ &+ -1 + 2p_e + \frac{2n\,(1-p_e)^{\,2}}{n(1-p_e)+(n-1)p_r} * P_{int}(d) \end{aligned} \quad (25)$$

and for modified HCMM as

$$\begin{aligned} P_{mod}(d_i) &= \sum_{\substack{j=1 \\ j\neq i}}^{C} \frac{2(n-1)n\,(1-p_e)\,p_r w_{c_i c_j}^{\alpha}}{n\,(1-p_e)+(n-1)p_r} * P_{c_i c_j}(d) + \\ &+ \left(\frac{(1-p_e)\,(-n+(n-1)p_r)}{n\,(1+p_e)+(-1+n)p_r} + \right. \\ &+ \left. \sum_{i=1}^{C-1} \frac{(n-1)p_e p_r w_{c_i c_j}^{\alpha}}{1-p_e+\frac{(n-1)p_r}{n}} \right) * P_{int}(d) \end{aligned} \quad (26)$$

For $P_{int}(d)$ the formula given in [2] can be used, assuming that each cell has size $a \times b$:

$$P_{int}(d) = \frac{4l}{a^2 b^2} * \begin{cases} \frac{\pi}{2}ab - ad - bd + \frac{1}{2}d^2 & \text{for } 0 \le d \le b \\ ab\arcsin\frac{b}{d} + a * \sqrt{d^2-b^2} + & \\ -\frac{1}{2}b^2 - al & \text{for } b \le d \le a \\ ab\arcsin\frac{b}{d} + a * \sqrt{d^2-b^2} + & \\ -\frac{1}{2}b^2 - ab\arccos\frac{a}{d} + & \\ +b\sqrt{d^2-a^2} - \frac{1}{2}a^2 - \frac{1}{2}d^2 & \text{for } a \le d \le \sqrt{a^2+b^2} \\ 0 & \text{otherwise.} \end{cases} \quad (27)$$

Instead, there is no closed form for $P_{c_i c_j}(d)$ for generic c_i and c_j. This implies that also for $P_i(d)$ only numerical solutions are available.

5.1 Model validation

In this Section we compare the distribution given by Equations (25) - (26) with the empirical distributions obtained through simulation. We consider a 10×10 grid on which 4 communities are placed, each having 50 nodes. 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 500000 seconds to ensure stationarity. As the model requires as input the initial positions of communities, let us consider the scenario in Figure 3 (similar results have been obtained in other scenario).

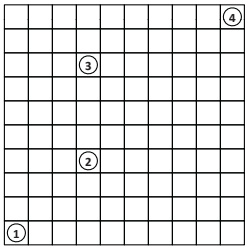

Figure 3: Communities' locations on the grid

First we focus on standard HCMM. Analytical and simulative results are shown in Figure 4, where the empirical probability distribution is obtained using the Kernel density estimation method. The two densities 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 travel distances on a $m \times n$ 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.

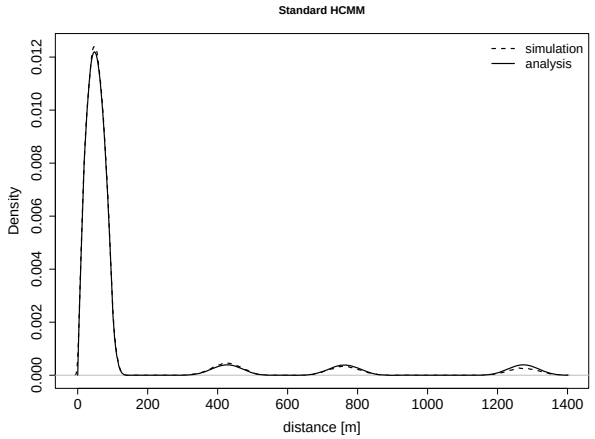

Figure 4: Kernel density vs analytical results

Figure 5 shows analytical against simulative results for the modified HCMM. Again, the two curves overlap almost perfectly. Note that in this case, long distances are rarely selected and short ones are preferred.

6. MOBILITY PATTERN EVALUATION

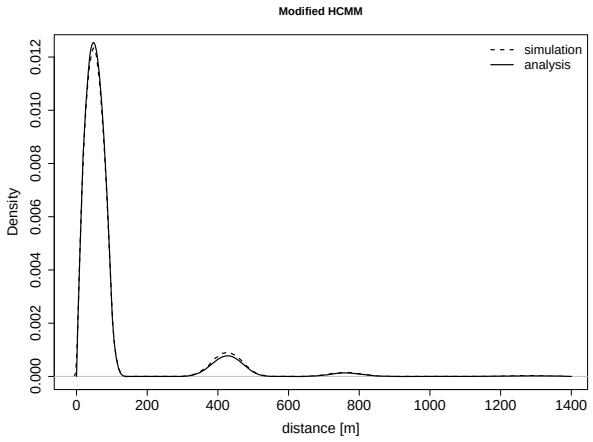

Figure 5: Simulation vs analytical results - Standard HCMM

In Section 4 we have extended the HCMM protocol to account for realistic travelling patterns. In this Section we evaluate the proposed model through simulation. In particular, we are interested in evaluating if prioritizing the distances chosen by the nodes is enough for obtaining that the jump size follows a truncated power law distribution. 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 other 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 sub-sampling: 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 particular, there is a power law behaviour when the density of the scenario, both in terms of communities and external social links, are sufficient to explore "enough" distances.

Our reference scenario is a 1000×1000 m^2 square, 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 Equation (3) 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.

6.1 Everybody is a friend

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 Figure 6, the continuos black line shows the empirical CCDF of the jump size obtained, the others are the Maximum Likelihood Estimantion 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 dicreases 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 [7], 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 $100\sqrt{2}$) and this is a good explanation for the rapid decay of the tail. This results are also consistent with [10], where a power law with an exponential decay was suggested as a very good approximation of human travelling patterns.

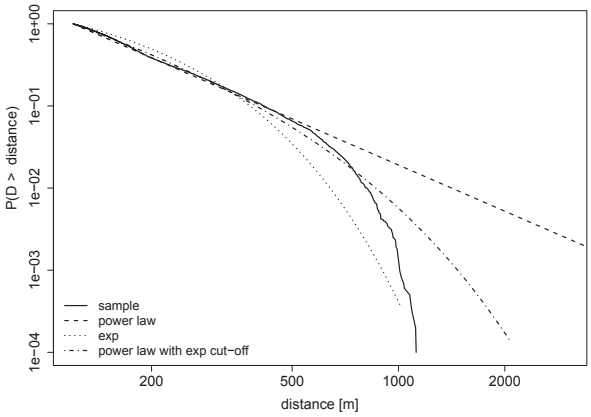

Figure 6: MLE fitting - 100 communities

Now we evaluate the effect of removing some groups from the grid. In this case, some distances related to unoccupied cells are not selected. Therefore, the power law selection of the modified HCMM model will consider only a subset of all the possible distances, thus creating what we have called subsampling effect. And this effect indeed appears in Figure 8, where we can see the evolution from a straight line plus cut-off to a non-straight, bumpy curve. The bumps are due to the distances that are not present, and therefore create suddent, short jumps on the distribution. The absence of some distances is more evident in the probability density function Figure 7. While with many groups, there is a constantly descreasing curve that is the result of the mixture in Equation (4), with few groups we can clearly distinguish the single bell-shaped curves corresponding to the available distances.

6.2 Reduced number of friends

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. This results in the presence of bumps in the CCDF of jump sizes. In the next set of simulations we reduce the average number of external friends

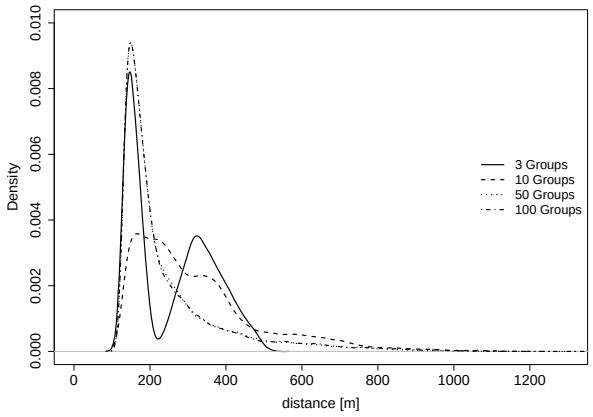

Figure 7: Kernel density estimate with varying number of communities

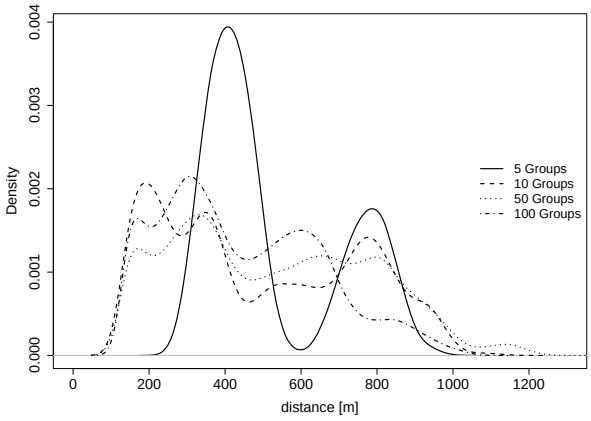

Figure 9: Kernerl density estimate of jump size with rewiring 0.1

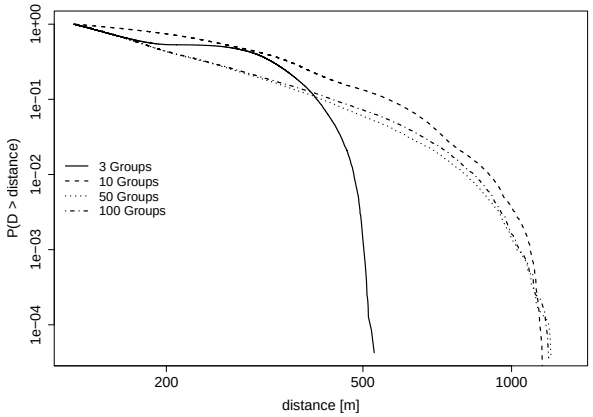

Figure 8: Jump size CCDF - Varying the number of communities

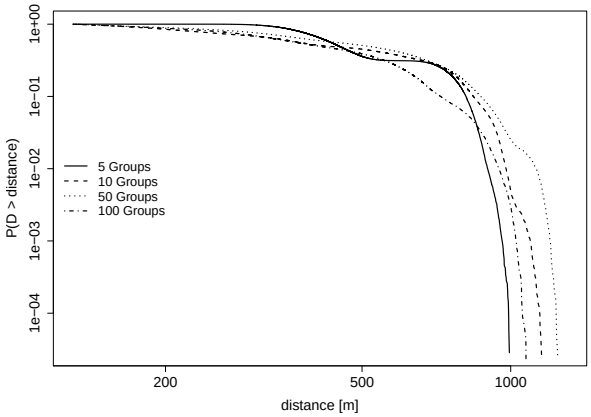

Figure 10: CCDF for rewiring 0.1

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 Figure 9, we can see that even in the case of 100 communities, the single humps are clearly visible. Each hump corresponds to the bell-shaped curve associated with a characteristic distance for which there is at least one friend. The CCDF of the jump size (Figure 10) is obviously affected by the subsampling as well. The straight lines in the head of the distributions are not visible anymore, neither with many or few groups, and bumps are present. 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 control the subsampling that is performed. This is shown in Figure 11, where we go from a non power law behaviour with low rewiring (high subsampling) to clear power law heads 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 Figure 12.

Concluding, in this Section we have shown the 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.

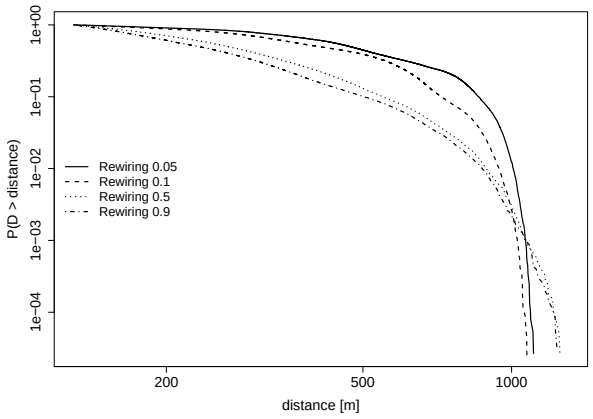

Figure 11: CCDF of jump size with varying rewiring

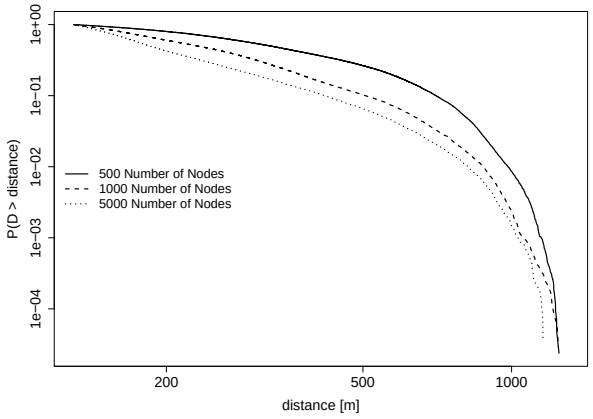

Figure 12: CCDF of jump size with varying number of nodes

7. CONCLUSIONS AND FURTHER WORK

In this paper we have extended an existing social-aware mobility model to include realistic jump sizes. We have evaluated the generated distance distribution of the proposed model, showing that, under some conditions on the mobility setting, it is able to reproduce the realistic distribution of distances. In addition, we have proposed a model for the distance distribution, that we have validated against simulative results.

Still there are some open research directions. We plan to explore other ways for generating power law distributed jump size, e.g., using a non uniform rewiring process. In addition, the application of the proposed mathematical model to mobility models other than HCMM or the use of this model for computing other properties of grid-framed mobility seem a very promising research direction.

8. ACKNOWLEDGMENTS

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

9. REFERENCES

[1] A. Barabàsi. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039):207–211, 2005.

[2] C. Bettstetter, H. Hartenstein, and X. Pérez-Costa. Stochastic properties of the random waypoint mobility model. *Wireless Networks*, 10(5):555–567, 2004.

[3] C. Boldrini, M. Conti, and A. Passarella. Users mobility models for opportunistic networks: the role of physical locations. *Proc. of IEEE WRECOM*, 2007.

[4] C. Boldrini, M. Conti, and A. Passarella. The sociable traveller: human travelling patterns in social-based mobility. Technical report, IIT-CNR 2009, `http://bruno1.iit.cnr.it/~chiara/pubs/mobiwac09tr.pdf`.

[5] V. Borrel, F. Legendre, M. de Amorim, and S. Fdida. Simps: Using sociology for personal mobility. *Arxiv preprint cs.NI/0612045*, 2006.

[6] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, 2006.

[7] H. Cai. Crossing over the bounded domain: from exponential to power-law inter-meeting time in manet. pages 159–170, 2007.

[8] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of human mobility on opportunistic forwarding algorithms. *IEEE Transactions on Mobile Computing*, 6(6):606, 2007.

[9] V. Conan, J. Leguay, and T. Friedman. Characterizing pairwise inter-contact patterns in delay tolerant networks. 2007.

[10] M. González, C. Hidalgo, and A. Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[11] D. Helbing. Traffic and related self-driven many-particle systems. *Reviews of modern physics*, 73(4):1067–1141, 2001.

[12] D. Helbing, F. Schweitzer, J. Keltsch, and P. Molnar. Active walker model for the formation of human and animal trail systems. *Physical Review E*, 56(3):2527–2539, 1997.

[13] 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*, pages 128–129. ACM New York, NY, USA, 2003.

[14] D. Johnson and D. Maltz. Dynamic source routing in ad hoc wireless networks. *KLUWER INTERNATIONAL SERIES IN ENGINEERING AND COMPUTER SCIENCE*, pages 153–179, 1996.

[15] T. Karagiannis and M. Vojnović. Power law and exponential decay of inter contact times between mobile devices. pages 183–194, 2007.

[16] K. Lee, S. Hong, S. Kim, I. Rhee, and S. Chong. Slaw: A mobility model for human walks. In *IEEE INFOCOM 2009*, 2009.

[17] A. Mathai, P. Moschopoulos, and G. Pederzoli. Random points associated with rectangles. *Rendiconti del Circolo Matematico di Palermo*, 48(1):163–190, 1999.

[18] M. Musolesi and C. Mascolo. Designing mobility models based on social network theory. *ACM SIGMOBILE Mobile Computing and Communications Review*, 2007.

[19] K. Nagel and M. Paczuski. Emergent traffic jams. *Phys. Rev. E*, 1995.

[20] I. Rhee, M. Shin, S. Hong, K. Lee, and S. Chong. On the levy-walk nature of human mobility. In *IEEE INFOCOM 2008. The 27th Conference on Computer Communications*, pages 924–932, 2008.

[21] M. Seshadri, S. Machiraju, A. Sridharan, J. Bolot, C. Faloutsos, and J. Leskove. Mobile call graphs: beyond power-law and lognormal distributions. 2008.

[22] D. Watts. *Small worlds: the dynamics of networks between order and randomness*. Princeton University Press, 1999.

APPENDIX

A. DISTANCES BETWEEN RANDOM POINTS IN DIFFERENT CELLS

Authors in [2] provide a formula for the distribution of distances between random points within the same cell. Here we perform the same analysis but for the case of random points belonging to different cells. The approach that we use is the same as in [17]. For the sake of simplicity we assume $a \times a$ wide cells (square cells). We consider a generic pair of cells belonging to the grid scenario. Our analysis uses k_1 and k_2 as defined in Figure 13. First we consider the unidimensional case (random point associated with different segments) and then we use the obtained results to solve the bidimensional case.

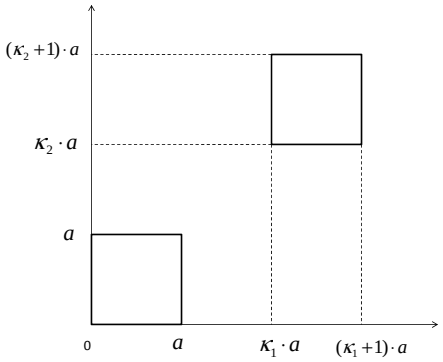

Figure 13: Squares over a grid

A.1 Random points in different segments

First we consider the case of two points P_{x_1} and P_{x_2}, uniformly distributed on two unidimensional segments: $[0, a]$ for I y

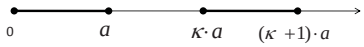

Figure 14: Points on different segments

density functions defining the positions of the two point are:

$$f_{P_{x_1}}(x_1) = \begin{cases} \frac{1}{a} & \text{if } 0 \le x_1 \le a \\ 0 & \text{else} \end{cases} \quad (28)$$

$$f_{P_{x_2}}(x_2) = \begin{cases} \frac{1}{a} & \text{if } ka \le x_2 \le (k+1)a \\ 0 & \text{else} \end{cases} \quad (29)$$

From Equations 28 - 29 we can compute the joint probability distribution:

$$\begin{aligned} f_{P_{x_1}P_{x_2}}(x_1, x_2) &= f_{P_{x_1}}(x_1) f_{P_{x_2}}(x_2) = \\ &= \begin{cases} \frac{1}{a^2} & \text{if } 0 \le x_1 \le a \\ & \text{and } ka \le x_2 \le (k+1)a \\ 0 & \text{else} \end{cases} \end{aligned} \quad (30)$$

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 $P(D \le d)$ can be obtained by integrating the joint probability in Equation 30 over the domain $\mathcal{D} = |x_2 - x_1| \le d$. This implies solving Equation (31).

$$P(D \le d) = \begin{cases} 0 & \text{if } d < 0 \\ \frac{1}{a^2} \int_{ka-d}^{a} \int_{ka}^{x_1+d} dx_2 dx_1 & (k-1)a \le d \le ka \\ \frac{1}{a^2} (\int_{0}^{(k+1)a-d} \int_{ka}^{x_1+d} dx_2 dx_1 + & \\ \quad + \int_{(k+1)a-d}^{a} \int_{ka}^{(k+1)a} dx_2 dx_1) & ka < d \le (k+1)a \\ 1 & d > (k+1)a \end{cases} \quad (31)$$

The result is given in Equation (32):

$$P(D \le d) = \begin{cases} 0 & \text{if } d < 0 \\ \frac{(a-ak+d)^2}{2a^2} & (k-1)a \le d \le ka \\ -\frac{(a^2(-1+k(2+k))-2a(1+k)d+d^2}{2a^2} & ka < d \le (k+1)a \\ 1 & d > (k+1)a \end{cases} \quad (32)$$

Differentiating (32) with respect to d, we get the probability density function for D in the unidimensional case:

$$f_D(d) = \begin{cases} \frac{a-|d-ak|}{a^2} & \text{if } (k-1)a \le d \le (k+1)a \\ 0 & \text{else} \end{cases} \quad (33)$$

A.2 Random points in different cells

In this Section we consider the bidimensional case shown in Figure 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:

$$D = ||P_2 - P_1|| = \sqrt{|y_2 - y_1| + |x_2 - x_1|} = \sqrt{D_x^2 + D_y^2} \quad (34)$$

From Section A.1 we know that the distribution of D_x and D_y is given by Equation (33). Thus, we can compute the joint distribution

$$\begin{aligned} & f_{D_x D_y}(d_x, d_y) = f_{D_x}(d_x) f_{D_y}(d_y) = \\ & = \begin{cases} \frac{(a-|d_x-ak_1|)(a-|d_y-ak_2|)}{a^4} & a(-1+k_1) \le d_x \le a(1+k_1) \\ & a(-1+k_1) \le d_y \le a(1+k_2) \\ 0 & \text{else} \end{cases} \end{aligned} \quad (35)$$

In order to compute $P(D \le d)$ we integrate Equation 35 over the domain $\sqrt{D_x^2 + D_y^2} \le d = D_x^2 + D_y^2 \le d^2$. 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 $P(D \le d)$. The only solution available is numerical, and it is the one that we used in Section 5.