Human Mobility Model Based on Time-Varying Bipartite Graph

Matteo Zignani
Dipartimento di Scienze dell'Informazione
Universitá di Milano
Milano, Italy
Email: matteo.zignani@unimi.it

***Abstract*—Nowadays human beings are surrounded by a heterogeneous networking environment consisting a growing number of portable computation and communication devices. As most devices are carried out by human beings, such a contact-based networks is highly influenced by human mobility. This fact implies that the presence of possible patterns in human movements can be exploited by wireless network applications in order to extract sensible informations on top of which novel mobile services can be deployed. Such information does not cover only the spatial or temporal dimension, but also concerns relational and social aspects of the involved people. In order to evaluate such applications we have to develop a mobility simulation sufficiently expressive and easily tunable. The most important goal in the mobility model research area is to provide a tool that can capture the most important and relevant features regarding both physical and social dimensions. For my PhD research I propose a new mobility model able to properly reproduce the spatial, temporal and social features that can be observed in real mobility datasets. In the model people move within a set of *geo-communities*, i.e. locations loosely shared among people, according to a bipartite time-varying graph; similarly, inside a geo-community, people move according to a modified version of classical random waypoint. We also derive social relationships from the bipartite graph representation by means of different types of projections on the node set. The purpose of this document is to briefly describe the state of the art in mobility model, and to outline my planned PhD research.**

***Keywords*-mobility model; time-varying graph;**

I. Introduction

In recent years we observed an exponential growth of mobile devices that have become widely popular. Most of these devices are equipped with short-range wireless network interface allowing the creation of an opportunistic connectivity for complementing both the wired and cellular, licensed network infrastructures. Such a contact-based infrastructure is highly influenced by human mobility.

By analyzing cell phone calls, Song *et al.* have shown that human mobility and the resulting contacts betweeen people have a high predictability and so it is not entirely random. The presence of possible patterns can be exploited by wireless network applications in order to infer relevant informations on top of which novel mobile applications can be deployed.

A fundamental issue in the design of new opportunistic services is the evaluation of the underlying protocols under practical and real mobility conditions. Two different evaluation methods can be used to the purpose: the first is based on traces taken from real environment, the second is based on synthetic traces generated from a mobility model simulation. Due to the limited number of datasets involving a large quantity of people, the second approach seems to be the best method to evaluate service performances in an environment with a variable high number of nodes. Thereby, it becomes necessary to develop models which can simulate different aspects of human mobility. Several mobility models, that overcome in expressivity and complexity the classical Random Waypoint, have been recently proposed. We can roughly divide these new models into two categories: location driven and social driven models.

- In a *location driven* approach, the main goal is to reproduce individual movement behaviour as observed in real mobility traces. Most of these models carefully reproduce many mobility patterns found in real traces; nevertheless they mainly focus on individual behaviours and do not consider any social interaction between nodes.
- By contrast, *social driven* models assume that sociality lead to the mobility of people. In general their goal is to simulate mobility by reproducing the social interactions.

All the above models suffer the same drawback: they focus only on a single out of many factors driving human mobility. In location driven models only spatial properties matter; whereas, in social driven models mobility is ruled only by social features and, as a consequence, the statistical distributions regarding spatial mobility, such as movement length, have not been evaluated. Moreover it has been shown that many models do not reproduce properties of the contact graph (graph modelling contact between mobile device or people) observed in real traces.

Starting from the actual scenario on the mobility model research area, the main goal is to realize a new mobility model that attempts to overcome this limitation by incorporating both social and spatial factors in the model. The development of such a mobility model requires to base its design on reliable and quantitative knowledge and predictions of some relevant information, such as which locations are visited by people in their daily life and how people join within communities. This advocates a realistic model able to describe both the human mobility throughout locations and the human attitude to socialize within communities. In our approach the combination of locations and communities is achieved by exploiting the concept of *geo-community*, informally defined as a location visited by a few people (the community) potentially in different time, and modelled by means of a bipartite graph.

The main features of our mobility model are: *(i)* all concepts and quantities involved in the model, together with their probability distributions, have to been inferred from a statistical analysis of real dataset, possibly obtained from GPS-based devices; *(ii)* by introducing the concept of geo-community, our models enables people sociality to emerge naturally from the sharing of locations, rather then being externally imposed on the basis of qualitative reasoning; *(iii)* we could deduce social relationships from traces by projections of the entire system node, geo-community on nodes; *(iv)* by considering a two-level framework, mobility inside geo-communities locations can be modelled by a non-uniform random waypoint.

II. Related work

In recent years a lot of work has been done in the area of mobility models. In this section we present the principal models with which to compare results, resuming the categories presented in the introduction.

As regards location driven approach, the most important works are SLAW and TVC. SLAW is a model based on Levy walk and on the minimization of the distance covered, that captures the main statistical distributions on relevant quantities such as flight length, pause time, inter-contact time and speed observed in GPS-based traces.

TVC , instead, focuses on reproducing temporal (periodical reappearance) and spatial (location preference) regularities, two properties extracted from Wi-fi LAN-based traces. In order to capture preferential locations, it forces a node to visit its preferred location more often than other locations. In addition, TVC defines time periods during which a node moves towards its preferred location with higher probability. Thus, it reproduces periodical reappearances at the same location.

As concerns the second modeling approach, one of the first social driven model is the community-based mobility model (CMM). In CMM nodes belong to a primary community and by a rewiring process they may have links to external communities. The mobility of nodes is induced by probabilities defined by community attraction. Namely, in order to overcome a gregarious behavior, that is, all users tend to follow the first node that leaves the community where it is located, Boldrini *et al.* have proposed HCMM, which extends CMM by introducing a home-cell in which nodes tend to spend most of their time.

Gaito *et al.* proposed a two-level mobility model where nodes in communities share the same location preferences and inside each location they move according to a micro-mobility model.

In SIMPS nodes are continuously in motion and can assume two states. In the $socialize$ state, nodes move in the direction of social relationships, while in the $isolation$ state they try to escape from neighbors which they do not have social relations with. Switching process is controlled by a feedback loop so that each node needs for sociability to remain constant.

In Heterogeneous Human Walk (HHW) human mobility is based on heterogeneous centrality and on an overlapping community structure typical of social networks. HHW constructs a k-clique structure of overlapping communities built on common statistical properties, extracted by several real social networks.

Fischer *et al.* applied a similar approach requiring a conformance between the social graph taken in input by the model and the contacts induced by mobility traces generated by it.

III. Achieved Results

Unfortunately, so far neither the concept of location nor the concept of community has been univocally defined. Therefore, their specific features and, more importantly the relationship between the two concepts, are still not fully understood and exploited either. It is commonly accepted that people move during the day among locations and communities. Nevertheless, locations are just points in a simulation area and their characteristics, like size, distribution, number, population density, are still not known in full. Similarly, the concept of community has found different definitions in the literature depending on the particular context. There is the sociological approach, for example, which defines it as people sharing some interests; then there is the location aware definition where a community is a group of people meeting at a given location

Part of my work has been devoted to the formalization of the above concepts and to the development of a software tool for the extraction of geo-communities from real trace datasets. The first step in our methodology is to extract from GPS traces the points (locations) of interest - what we call *geo-locations* . Geo-locations are places that attract and interest a person for some reason. We then discriminate GPS points visited while moving - i.e. during the person's movement phase - from the points where the person has spent a relevant amount of time. In order to identify the real locations of a larger group of persons, what we call *geo-community*, we propose a hierarchical clustering method to aggregate individual geo-locations. The idea of geo-community we propose is strictly connected to the following assumption: people mobility is motivated by the need to go somewhere, more than by some social need and social interactions are created by sharing the same place: people who hang out at the same location have the potential to establish a social relationship or not, depending on whether or not they meet. But for the aims and purposes of deploying network connectivity and services, they belong to the same geo-community. So, a geo-community is both a point of attraction for people and the set of people who go there, maybe at different times.

After geo-community extraction, by statistically analyzing the available GPS traces, we show how the probability distributions of the main quantities involved in human movements and aggregations can be obtained. Our main results concern pause-times in geo-community, intra-distances, inter-distances and next geo-community choice. The latter point concerns how distance influences the choice of the next geo-community in the person's movement.

Given the above considerations, the concept of geo-community allows us to describe human mobility in terms of displacement between geo-communities and to model the entire process by means of an undirected bipartite graph $G = (U, V, E)$ where U is the set of the geo-communities and V is the set of nodes. Consequently, E can be define as $E = \{(v, u) | u \in U, v \in V \wedge u \in l(v)\}$, where by $l(v)$, we indicate the set of geo-communities visited by the node v. Starting from a static bipartite graph, we can introduce dynamic aspects by adding temporal information to edge connecting geo-communities and nodes. In particular each edge will be characterized by a list of pairs (t, d), where t represent the instant when the node reaches the geo-community and d indicates the amount of time spends in such geo-community.

The bipartite dynamic graph well fit with our approach where nodes move in shared areas (geo-communities) and the sharing of these areas by different nodes generates contacts and social interactions. It also represents a general framework by means of which to analyze movements between communities not necessarily georeferenced.

IV. Future Work

A set of challenge will be considered during the development and the analysis of the model:

Temporal list creation: in the setting of the arrival times t_i in geo-communities we have to face some constraints regarding distance and node behaviours, whereas in assigning durations we have to base our decision on statistics inferred from real traces.

Mobility inside geo-community: since it was shown in our previous work that Random Waypoint is unsuitable in reproducing the temporal features of the contacts inside geo-community, we have to find a more realistic mobility model. We intend to investigate the family of the non-uniform random waypoints in order to identify the best one in capturing the distributions involved in contacts.

Social relationship inference: Static bipartite graph representation enables us to make explicit the social relationships resulting from the sharing of common location between nodes. We could use the bipartite network to infer such connections, creating a one-mode projection from the two-mode bipartite form. By analysing such projection on the vertex set, we can deduce, for each node, the existence and the strength of its social interactions with all the other nodes.

A first step in the validation of our approach based on geo-communities has been to investigate GPS-based traces by extracting them. Our main purpose is to analyse the movement of the nodes among geo-communities and the structure of the contacts. From this analysis some indications about the assignment of the values to t and d emerge. In particular d should have to follow a power-law with exponential cut-off, whereas t values should depend on a criterion that involves the distance between geo-communities. Following this approach we will also analyze contact-based traces to confirm the probability distribution of d and to find what characteristics influence the choice of the next social community and consequently the shape of t.

In the first prototype of our model we will assume that the choosing process for a given node can be modelled by a finite time-homogeneous Markov chain where the states are the geo-communities that can be visited by the node and the probability of transition is a function of a rank among geo-communities depending by their distances.

Inside a geo-community we will focus on a particular variety of non-uniform random waypoint model similar to Levy walk. According to this model, the distances covered by a single node are distributed according to a power-law distribution. This simple model reflects the requirement of capturing some properties about distances that came out from our GPS trace analysis. In particular, we observed that transition lengths follow a Pareto distribution.

So far, our method has been driven by the spatial issues of human mobility. We proceeded backward, from locations to communities. However, we are aware of the relevance of social relationships in deploying most part of the emerging mobile computing services. To this purpose, we will introduce some methods to capture human social aspects from spatial data instead of imposing them as a priori knowledge or using the contact graphs. Social relationships naturally emerge by the undirected bipartite graph, as a matter of fact by analyzing its projection on the persons set, for each person we can deduce the existence and the strength of his social interactions with all the other people. This kind of representation enables us to make explicit the social relationships which develop when people spend time in the same location. We use the bipartite network to infer such connections, creating a one-mode projection from the two-mode bipartite form.

We will perform a projection on the nodes alone by constructing an undirected graph $G_{proj} = (V, E')$ where

$E' = \{(v_1, v_2)|v_1, v_2 \in V, \exists u \in U, (v_1, u), (v_2, u) \in E\}$. Each geo-community in the bipartite graph results in a cluster of vertices in the projection. We can also capture more information about the strength of the relationship by performing a weighted projection, assigning to each edge between two vertices a strength equal to the number of groups shared the vertices.

We will develop a way of projecting based on a preferential mechanism. This approach is justified by many results in literature showing the presence of a so called ”hub node”. Our way of projecting builds the subgraphs of their geo-communities by using two kinds of construction. If k, the number of users belonging to a geo-community, is less than a fixed threshold η, then we apply an unweighted projection creating a k-clique. Otherwise we build a Barabasi-Albert graph on the vertices inside the geo-community. After subgraph constructions we join them in an unique graph obtaining a *preferential projection*.