Notes on human mobility

Davide Cellai

September 10, 2010

1 A mobility model

1.1 Gonzalez paper

The distribution of displacements over "all" users is well approximated by the law:

$$P(\Delta r) = \frac{exp(-\Delta r/\kappa)}{(\Delta r + \Delta r_0)^\beta} \quad (1)$$

with $\beta = 1.75$; $\Delta r_0 = 1.5km$; $\kappa = 400km$ for D1 and $\kappa = 80km$ for D2. This is the total superimposition, and we'd like to use it as a check.

Equation (1) suggests that human motion follows a truncated Lévy flight. However, the observed shape of $P(\Delta r)$ could be explained by three distinct hypotheses:

A. each individual follows a Levy trajectory with jump size distribution given by equation (1);

B. the observed distribution captures a population-based heterogeneity, corresponding to the inherent differences between individuals;

C. a population-based heterogeneity coexists with individual Levy trajectories;

hence, equation (1) represents a convolution of hypotheses A and B.

1.1.1 Results

Distribution of giration radii r_g:

$$P(r_g) = \frac{exp(-r_g/\kappa)}{(r_g + r_g^0)^{\beta_r}} \quad (2)$$

with $\beta_r = 1.65$; $r_g^0 = 5.8km$; $\kappa = 350km$.

Distribution of displacements for a single user:

$$P(\Delta r|r_g) \sim F\left(\frac{\Delta r}{r_g}\right) \quad (3)$$

where

$$F(x) \sim \begin{cases} x^{-\alpha} & x < 1 \\ exp(-x) & x \gg 1 \end{cases} \tag{4}$$

and $\alpha = 1.2$.

1.2 Types of walks

According to Gonzalez et al., we represent people movements as agents whose displacements are taken from a truncated Levy flight, i.e. a truncated power law distribution.

$f : [\gamma x_0, x_0] \rightarrow \mathcal{R}$

$$f(x) = \frac{A}{(x+b)^\alpha}\theta(x_0 - x)\theta(x - \gamma x_0) \tag{5}$$

with $\gamma < 1$; $x_0 = higherCutoff$; and normalization:

$$A = \frac{(1-\alpha)}{\frac{1}{(x_0+b)^{\alpha-1}} - \frac{1}{(\gamma x_0+b)^{\alpha-1}}} \tag{6}$$

Cumulative distribution:

$$F(x) = \begin{cases} \frac{A}{1-\alpha}\left[\frac{1}{(x+b)^{\alpha-1}} - \frac{1}{(\gamma x_0+b)^{\alpha-1}}\right] & \text{for} \gamma x_0 < x < x_0 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Inverse cumulative distribution:

$$F^{-1}(y) = \left[\frac{1-\alpha}{A}y + \frac{1}{(\gamma x_0 + b)^{\alpha-1}}\right]^{-\frac{1}{\alpha-1}} - b \tag{8}$$

1.3 Choice of the parameters

The two distributions that we want to implement are $P(r_g)$ and $P(\Delta r|r_g)$, because $P(\Delta r)$ is the total superimposition, and we want to use it as a check.

1.3.1 The r_g distribution

As a first approximation, we use a sharp cutoff in dealing with the exponentials.

According to the experiment, b cannot be zero. b sets the unit of r_g and thus we can set it to one. Maybe, b can also be seen as the parameter which controls the spread of $P(r_g)$, i.e. the importance of the "small" r_g respect to the "large" ones.

So, our distribution is

$$P(r_g) = \frac{A}{(r_g + b)^\alpha} \tag{9}$$

which is defined in the domain $[0, x_0]$. The parameters are based on the experimental values. Since $\kappa/r_g^0 \simeq 60$, we set $x_0 = 60$. Then $b = 1$ and $\alpha = 1.65$.

1.3.2 The Δr distribution

Here we also replace a rapid decreasing function with a θ function.

Our implementation is:

$$P(\Delta r|r_g) = \begin{cases} \frac{A}{\left(\frac{\Delta r}{r_g}\right)^\beta} = \frac{A}{x^\beta} & \text{for} x \in [\gamma, 1] \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where γ is a parameter to avoid machine errors and the exponent is $\beta = 1.2$.

1.3.3 The transmission radius

This is a delicate point. If we want to consider bluetooth connections, we should assume a transmission radius of about $100m$. This implies an equivalent parameter $r_t = 0.017$. However, such a small radius will cause no connectivity in the network at any time of the simulation. Perhaps is more realistic to assume a transmission radius roughly similar to the mean cell size, which is about 2km.

2 Analysis of real traces

2.1 First approach

We are setting up a project for detecting real human mobility traces. It is sensible to prepare a plan of the analysis it will be carried out on the trace. A first set of issues is based on questions related to the experiments already present in the literature and on the insights given by Graham's paper. An incomplete list is:

1. General mobility analysis
2. Periodicity hunting
3. Spatial dependence of the degree
4. Buddy discovery

2.1.1 General mobility analysis

The traditional analysis of mobility should be carried out. We divide the time in equal intervals Δt. Then, we study the probability that an individual makes a movement of length r during the all duration of the experiment. We denote this quantity as $p(r)$ and we plot it against r. This function is very sensitive to the choice of the time interval Δt. When Δt is small, we detect the means of transport used, when it is larger, we give a coarse grained picture as in Brockmann's papers.

Another important information is the popularity of a location. This can be studied by plotting the probability of being in $\mathbf{x}$ ($v(\mathbf{x})$) against $\mathbf{x}$.

2.1.2 Periodicity hunting

The interest for the periodicity of the walk is rooted in the findings of Graham's paper. In principle, for every location $\mathbf{x}$, we want to generate the plot $\Pi(t)$, t, where $\Pi(t)$ is the probability that an individual that was is in $\mathbf{x}$ at time $t = 0$ is still there after time t. In practice, only the most attended locations will give interesting informations.

2.1.3 Spatial dependence of the degree

We want to study the correlation between degree (number of neighbours) and location. In order to do that the degree d can be plot against the position $\mathbf{x}$.

2.1.4 Buddy discovery

Another interesting question, potentially related to some applications, is whether the same individual is frequently met in different places. It is interesting, then, to determine for each individual the time spent together with Tom and the number of locations in which this happens.

2.2 A more decent program

Some of the most interesting questions about a HPN can be categorized under the question: **Who is influential?** In addition to that, other interesting questions are:

- correlations between local and global properties;
- routing;
- spatial sensing and energy management;
- P2P on HPN.

Regarding the purpose of discovering influential individuals, an interesting topic is to investigate if some nodes are trend-makers in discovering a popular location. To understand this, one has to study the time evolution of popularity of a particular location. If a boost is noticed, it would be interesting to understand whether some individuals are strarting frequenting a given location always before the others. Of course, we always have to bear in mind that correlation does not imply consequentiality (but it's better than nothing).

Another interesting analysis is to work out social relationships from the mobility trace. Inside this topic, the patterns of co-locations are particularly useful. Let $x_A(t)$ the position of the user A at time t. Then we can define the co-location of user A and B at time t as

$$c_{AB}(t) = 1 - \theta(|x_A(t) - x_B(t)| - \lambda) \tag{11}$$

where λ is the range within which we consider two users co-located. At this stage, $c_{AB}(t)$ can be averaged over different time intervals according to what it is interesting. The time average over the total duration of the experiment allows us to draw the network of co-locations, having users as nodes and co-location as the weight of the link between them. With different time averages, the evolution of this network can be studied. An interesting option is to compare the co-location networks of different times of the day: morning, afternoon, evening and night.