

M2 – NSD (Practical Work 6 - Sessions 12–14)

Models for real mobile networks.

Lionel Tabourier, Fabien Tarissan

The purpose of this practical work is to analyse the properties of different mobile networks and to compare the results to the one obtained on mobile networks generated by the edge-Markovian evolving graph model. To do so, we will rely on two dataset : ROLLERNET, and INFOCOM06.

1 Preliminaries

Here below, you will find some information related to the experiments leading to the datasets under study :

- **RollerNet** : it has been collected during a Rollerblade tour in Paris in August 2006. The tour lasted approximately 3 hours, with a break of 30 minutes, and covers around 30 km. *iMotes* have been given several dozens of participants (around 2500 persons were attending the event). The *iMotes* used Bluetooth technology and monitor their neighbourhood every 15 seconds.
- **Infocom06** : This experiment took place during the Infocom conference in Barcelona in April 2006. Like ROLLERNET, it relies on the *iMote* technology with a scan every 120 seconds. The experiment involve almost 100 devises, among which a majority were carried by persons attending the conference, while the other have been placed in various specific locations (17 were static, 3 have been placed in elevators).

Those dataset consist in lines of tuples describing contacts among participants. The format is :

`n1 n2 ts te`

where n_1 and n_2 are the identifiers of the two nodes involved in the contact, t_s stands for the time at which the contact started, and t_e the time at which the last contact between n_1 and n_2 has been recorded. It is worth noticing that the contacts are undirected and that, by convention, $n_1 < n_2$.

For the sake of exemplifying, here below is a small dataset involving 4 nodes during 10 units of time :

```
1 2 126 128
1 3 125 128
1 4 123 128
2 3 121 121
2 4 120 121
2 4 124 126
3 4 122 128
```

Exercise 1 Copy the files ROLLERNET and INFOCOM06 from the repository
/Enseignants-Public/tarissan/sdr/2015/

For each of those files, compute the number of nodes, the total number of contacts and the total duration of the experiment.

2 Normalisation

Those datasets don't start at time 0, which makes difficult the comparisons. We will then unify starting and ending times by making all the experiments start at time 0. Thus, on the former example, this would result in :

```
1 2 6 8
1 3 5 8
1 4 3 8
2 3 1 1
2 4 0 1
2 4 4 6
3 4 2 8
```

Exercise 2 Write a program that normalise all files such that they all start at time 0.

Those files have been sorted by identifier in increasing order. Sometimes, it is more useful to use a version of the dataset sorted by time and in which the information related to starting and ending a contact is splitted. Therefore, a line :

```
n1 n2 ts te
```

will generate two lines

```
ts n1 n2 C
```

and

```
tf+1 n1 n2 S
```

where the fourth column indicates is a character (C or S) that indicates whether the event is referring the the starting time or ending time of a contact (or, in other word, whether it is a creation or a suppression of a link).

Remark : The suppression of a link occurs at time $t_e + 1$, t_e begin the last time instant at which the contact has been recorded.

In our small example, this results in :

```
0 2 4 C
1 2 3 C
2 2 3 S
2 2 4 S
2 3 4 C
3 1 4 C
4 2 4 C
5 1 3 C
6 1 2 C
7 2 4 S
9 1 2 S
```

```
9 1 3 S
9 1 4 S
9 3 4 S
```

Exercise 3 Write a program that generate such a version

Let us recall that the command `sort -n -kx,x file` allows to sort the file `file` according to the numerical increasing order of the values contained in column `x`.

3 Analysis

This part is dedicated to the analysis of different properties of contact networks.

3.1 Inter-contact

We call *inter-contact duration time* the time that separate two contacts involving the same pair of nodes. For the second contact between 2 and 4 in the example, the inter-contact duration time is 2.

Exercise 4 Write a program that computes the inter-contact duration time associated to each observed contact.

Exercise 5 Compute and display the distribution of the inter-contact duration times.

3.2 Average degree

We focus now on the evolution of the average degree of the nodes in the network.

Exercise 6 Write a program that compute the average degree at each time instant. Is it mandatory to compute the degree of each nodes to answer the question ?

Exercise 7 Using *gnuplot*, display the evolution of the average degree over time for the datasets `ROLLERNET` and `INFOCOM06`. Comment the plots.

3.3 Creation and deletion of links

We focus now on the dynamics related to the creation and deletion of links over time. Let us recall that the fraction of created links at a given time is defined as the number of created links over the number of links *that could have been* created, that is according to the number of links that do not exist at preceding time. Similarly, the fraction of deleted links is defined as the number of deleted links over the number of existing links at preceding time.

The computation of those two quantities require then to memorise both the number of links that have been created (resp. deleted) at a given time but also the number of non existing links (resp. existing links) at previous time.

Exercise 8 Write a program that compute the fraction of created and deleted links at each time step and save the output in a file. When the fraction cannot be computed¹, the value will be `-1`.

1. For instance, if the graph is empty, one cannot compute the fraction of deleted links.

4 Modelling and simulating

The two quantities computed above will serve as a basis for the edge-Markovian model seen in class. Indeed in this model, given a graph G_t at a given time t , the structure of the graph at the next time step G_{t+1} depends directly on the one of G_t and is ruled by two independent parameters : la probability p that a new link is created and the probability d that an existing link is deleted. Thus, if $G_t = (V, E_t)$, then $G_{t+1} = (V, E_{t+1})$ with E_{t+1} determined as follows. For all pairs $(n_1, n_2) \in V \times V$ with $n_1 < n_2$:

- if $(n_1, n_2) \notin E_t$, then $(n_1, n_2) \in E_{t+1}$ with probability p .
- if $(n_1, n_2) \in E_t$, then $(n_1, n_2) \notin E_{t+1}$ with probability d .

The models needs 3 information for each time instants : the set E_t of existing links at time t , and the probabilities p and d . Those corresponds clearly in our study to the average value of the fractions computed previously.

We need also two other information in order to simulate the model :

- **The initial graph** G_0 : in order to reproduce the context of the different experiments, we will consider only the empty graph where $E_0 = \emptyset$
- **The number of time steps** T : in order to reproduce the context of the different experiments, we will take the duration time of the experiments.

Exercise 9 *Write a program that generates a random trace according to the edge-Markovian model. This program will take as input the number of nodes, the probability of creation of links, the probability of deletion of links and the number of time steps.*

Exercise 10 *Using previous programs, compare the evolution of the average degree over time on the randomly generated traces to the ones obtained on the real datasets. Comment.*

Exercise 11 *Write a conclusion that discusses the relevance of the markovian model in regards to the dataset studied here.*