ECOLE NORMALE SUPÉRIEURE PARIS-SACLAY
UNIVERSITÉ PARIS-SACLAY

PARCOURS IA
MACHINE LEARNING FOR NETWORK MODELING
PRACTICAL WORK / PROJECT

Models for real mobile networks

Practical study of the Edge-Markovian evolving model

Élève(s)
Augustin GODINOT

Enseignant(s)
Fabien TARISSAN
Argyris KALOGERATOS

Contents

List of Figures

I Introduction

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 datasets : RollerNet, and Infocom06.

In the first part we will discuss some properties of evolving graphs, then we will present the Edge-Markovian evolving graph model, and finally we will present an improvement of such model to better reflect real graphs properties.

This report and the code can be found on the github repository at `https://github.com/grodino/graphia` along with explanations on how to run it.

1. Datasets

RollerNet It was 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 were given to several dozens of participants (around 2 500 persons were attending the event). The iMotes used Bluetooth technology and monitored their neighbourhood every 15 second.

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 involved almost 100 devices, 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).

2. Evolving graph

In the two experiments, each node in the graph represented a person or a location. Both entities were always taking part in the experiment. Hence, we can consider that the nodes of the graph do not change over time. However, throughout the experiments, the beacons could come in contact or leave the range of other beacons.

Definition 1 (Evolving graph)**.** We define an evolving graph as a set V of fixed vertices with a set E_t of edges that can be created and deleted over time :

$$G_t = (V, E_t)$$

Remark 1. This definition is not the most general but suits well our case. However, as an example where this definition would not work, we can think about a communication network. In fact, on such network, servers can become unavailable, therefore forbidding any connection (edge). In this case, we would have to consider that node can be removed/added from the graph at any moment.

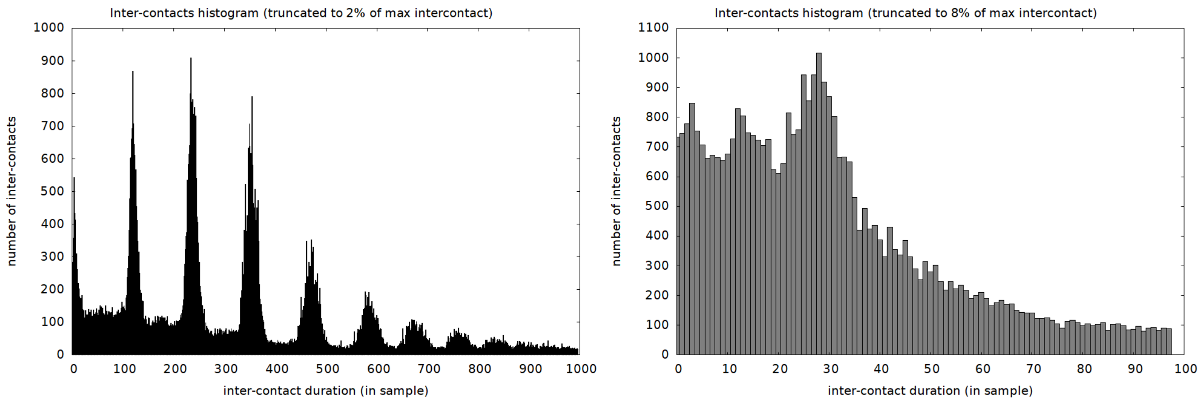
II Relevant properties of evolving graphs

In order to gather parameters for future graph models, we need to analyse and find relevant properties for evolving graphs. Furthermore, such properties would also be needed for the evaluation of our model.

Time evolving graphs have two main components to extract properties from : time dependence and topology. We will focus on three properties : inter-contact time, average degree and fraction of created/deleted nodes.

1. Inter-contacts

Definition 2 (Inter-contact)**.** We call inter-contact duration the time that separates two contacts involving the same pair of nodes.



(a) Infocom06 dataset

(b) RollerNet dataset

Figure 1: Inter-contact histogram, truncated

The inter-contact histograms are plotted on figure 1.

Remark 2 (histogram values)**.** The histograms are shifted to the right. This means that a value of 0 on the scale corresponds to a -1, a value on the scale of 1 corresponds to a 0 etc... During the computation of the histogram, if a pair of nodes is never reformed, the inter-contact value returned is -1.

On the Infocom06 dataset, we can observe some periodicity in the inter-contact : inter-contact values of 3h20, 8h20, 11h30 ... are much more represented. The dataset was produced during a conference and some beacons were located in the elevators. Therefore, a plausible hypothesis is that conferences lasted around 3h, after which, people would leave the room via the elevators.

On the other hand, the RollerNet data set does not show such periodicity but rather a heavy-tailed distribution with a plateau phenomenon at the beginning.

In both cases, the distributions of the inter-contacts duration are very singular. This makes this property highly discrimitative, therefore useful, for future model comparison.

2. Average degree

Definition 3 (Degree of a node)**.** The degree d of a node is the number of nodes that are connected to it.

Definition 4 (Average degree of a graph)**.** The average degree D of a graph is the average degree of its nodes :

$$D = \frac{1}{card(V)} \sum_{v \in V} deg(v)$$

Proposition 1. The average degree of a graph at any given time t is :

$$D = \frac{2\, card(E_t)}{card(V)}$$

Proof. To prove this result, one has to see that an edge contributes by adding 1 to the degree of two nodes.

More formally, let's define the adjacency matrix A as such ($n = card(V)$) :

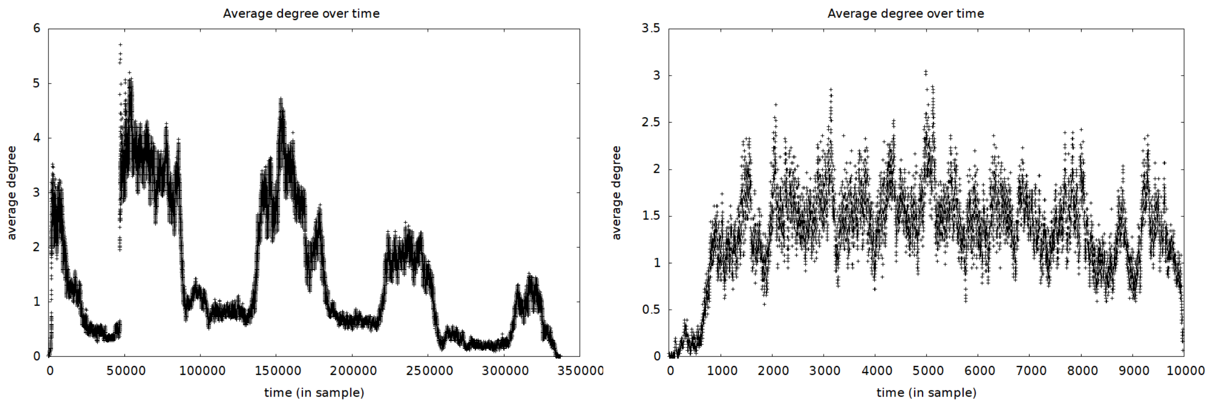
$$\forall (i,j) \in \{1,\ldots,n\}^2, A_{i,j} = \begin{cases} 1 & \text{if } v_i \text{ and } v_j \text{ are connected} \\ 0 & \text{else} \end{cases}$$

Because the graph is unoriented, A is symmetric. Moreover, the degree of a node is the sum of a line (or a column) :

$$\forall i \in 1,\ldots,n, d(v_i) = \sum_{j=1}^{n} A_{i,j}$$

$$\Rightarrow D = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} A_{i,j}}{n}$$

Or, this double sum is equal to twice the number of edges in the graph, hence the result. $\square$



(a) Infocom06 dataset

(b) RollerNet dataset

Figure 2: Average degree at each time

The average degrees over time are plotted on figure 2. The same observations as in the inter-contacts study can be made. The average degree of Infocom06 shows some periodicity whereas the rollernet data set seems to reach a steady state.

This property provides insights on the way beacons collectively move. If beacons tend to move in groups, the average degree will be high whereas if beacons tend to move independently, the average degree will be reduced.

3. Fraction of created and deleted probability

Definition 5 (Fraction of created links)**.** The fraction of created links f_c is the number of created links at a given time t over the number of links that could have been created.

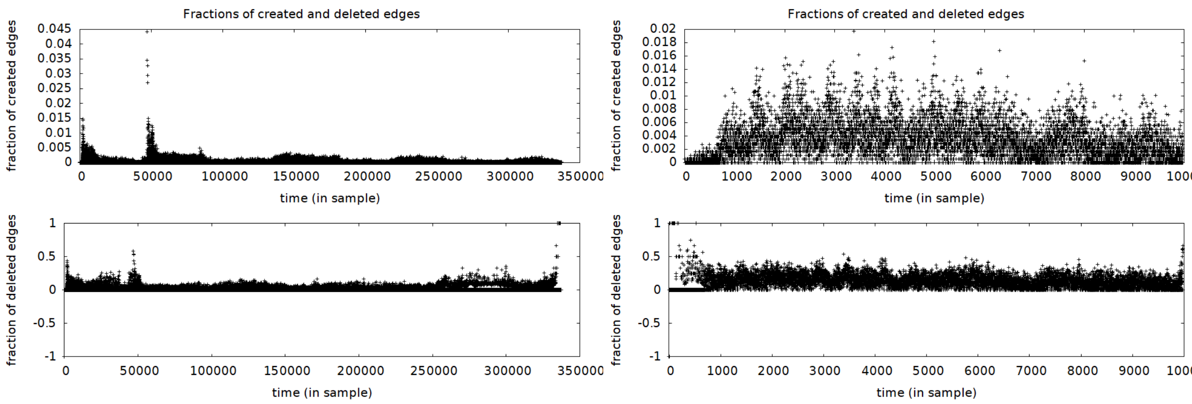
Proposition 2 (Fraction of created links)**.** Let Δ_t^+ be the number of created links between time steps $t-1$ and t, $n = card(V)$ the total number of vertices and e_t the number of edges at time t. The fraction of created links f_c can be expressed as such :

$$f_c = \frac{\Delta_t^+}{\binom{n}{2} - e_{t-1}}$$

Definition 6 (Fraction of deleted links)**.** The fraction of deleted links f_d is the number of deleted links at a given time t over the number of links that could have been deleted.

Proposition 3 (Fraction of deleted links)**.** Let Δ_t^- be the number of deleted links between time steps $t-1$ and t, $n = card(V)$ the total number of vertices and e_t the number of edges at time t. The fraction of deleted links f_d can be expressed as such :

$$f_d = \frac{\Delta_t^-}{e_{t-1}}$$



(a) Infocom06 dataset (b) RollerNet dataset

Figure 3: Creation and deleted fractions

The fractions of created and deleted edges are plotted on figure 3. On the two figures, we can see that this property is highly correlated to the average degree.

This property will be used in the next section as a parameter of the evolving edge-Markovian model.

III Modelling and simulating

1. The evolving edge-Markovian model

The objective of this section is to provide a simple model to explain some of the properties seen in the two datasets.

Definition 7 (Evolving edge-Markovian model)**.** Let p_c (resp. d_c) be the creation (resp. deletion) probability. The graph is initialised with zero edge. Then, for any given time t, the transition is defined as such :

For all $(v_1, v_2) \in V^2$,

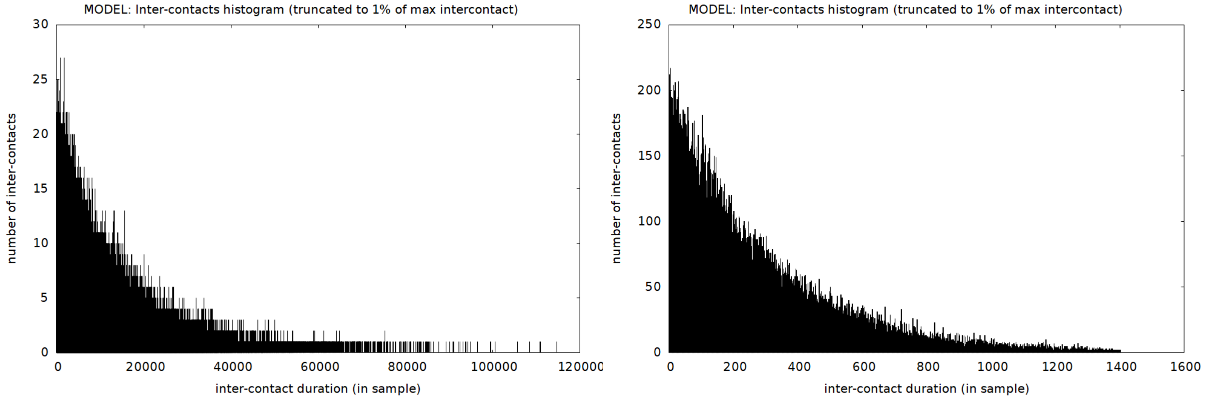
(deletion) if $(v_1, v_2) \in E_{t-1}$, then $(v_1, v_2) \notin E_t$ with probability p_d

(creation) if $(v_1, v_2) \notin E_{t-1}$, then $(v_1, v_2) \in E_t$ with probability p_c

2. Results

Figures 4, 5, and 6 present the results of the simulation. On all measurements taken from the model, we can observe that only the mean value of the property corresponds to the real data. The periodicity that we observed in the Infocom06 dataset is not picked by the model, nor the setup time of RollerNet's average degree.

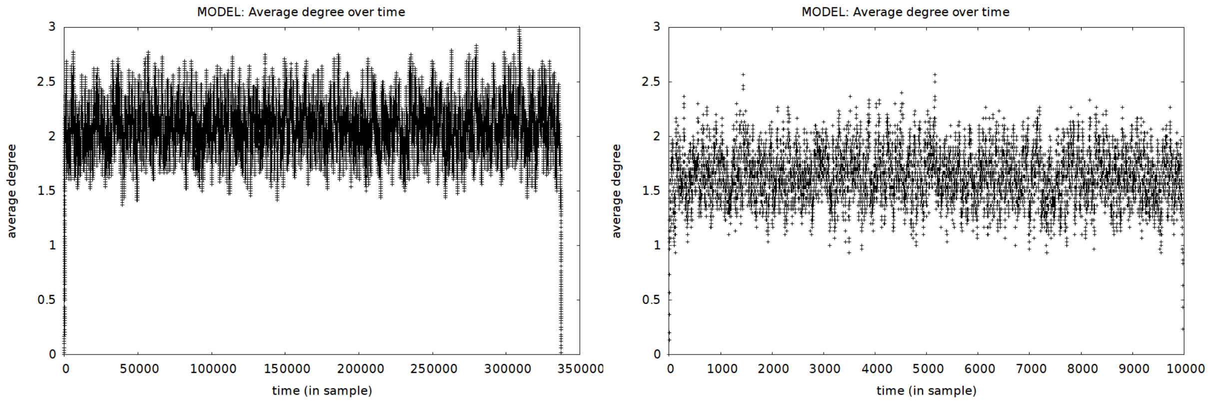
However, those results were expected : we chose to use the mean of the creation and deletion probability, therefore the variations of the property cannot be reflected in the model. Furthermore, the Markovian model used makes the implicit hypothesis that the process modifying the graph is stationary, which is false.



(a) Infocom06 dataset

(b) RollerNet dataset

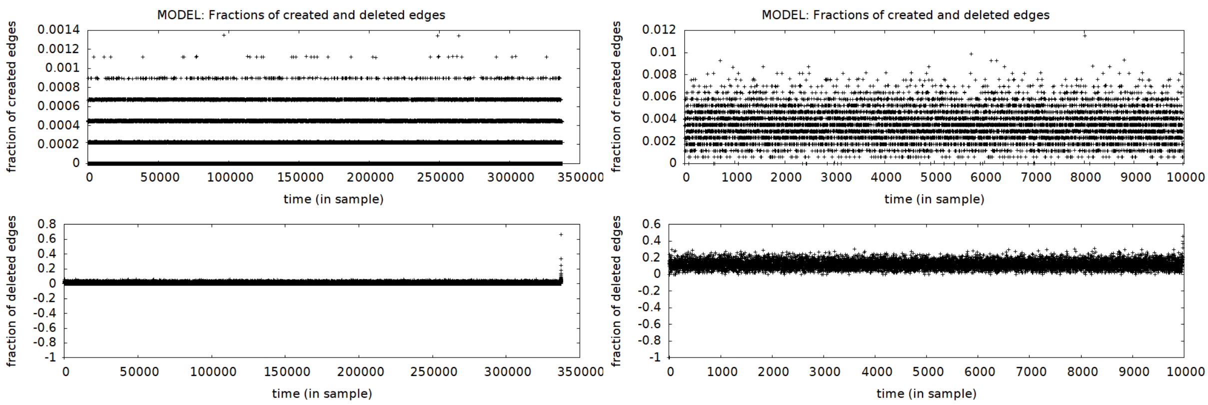
Figure 4: Modelled Inter-contact histogram, truncated



(a) Infocom06 dataset

(b) RollerNet dataset

Figure 5: Modelled Average degree at each time



(a) Infocom06 dataset

(b) RollerNet dataset

Figure 6: Modelled creation and deletion fractions

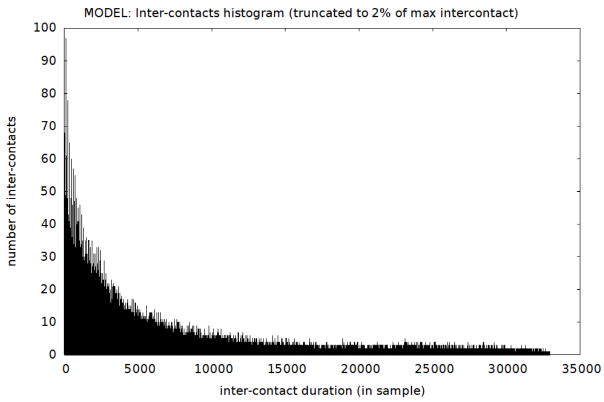
IV Improvements

1. Time dependent Edge-Markovian model

In this section, we explore the relevance of using time dependent creation and deletion probabilities as parameters for the Edge-Markovian model. Figures 7, 8, and 9 present the results of the simulation.

Remark 3 (simulation algorithm)**.** Following the Erdos-Renyi model generation trick, we could have randomly selected edge to create instead of nodes to connect. This would have reduced the complexity of the algorithm. However, because the graph creation is iterative, this would have implied a far more complex code than needed.

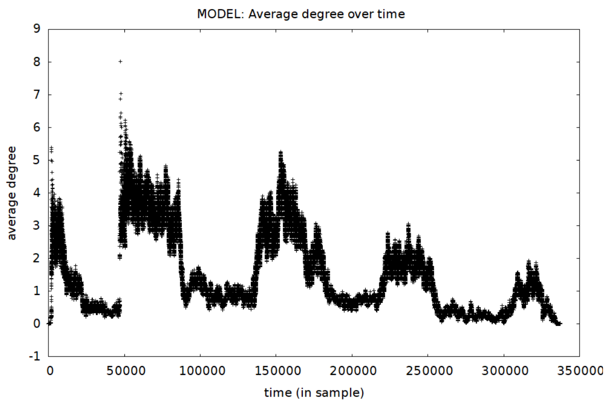
As expected, the creation/deletion fractions and the average degree are much closer to the dataset. However, we can observe that this new model still fails at capturing the periodicity seen in the inter-contacts histogram. Indeed, in the simulation, when a contact between two nodes ends, they are directly sent back to the available nodes for new connection, therefore not controlling the inter-contact duration.



(a) Infocom06 dataset

(b) RollerNet dataset

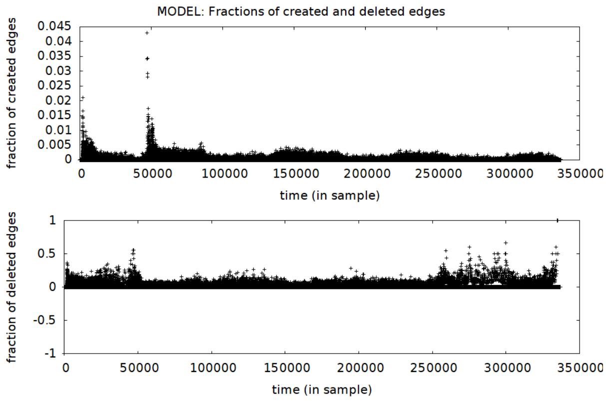
Figure 7: Modelled Inter-contact histogram, truncated



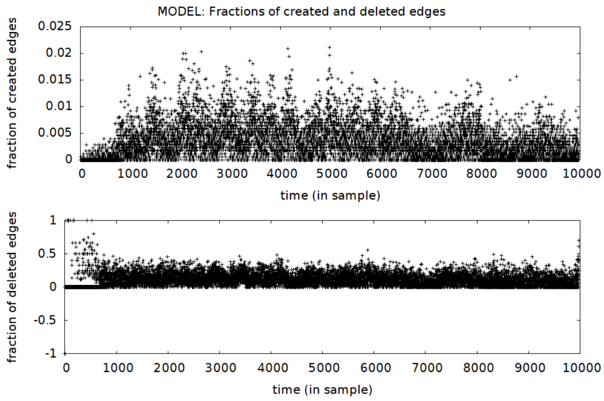
(a) Infocom06 dataset

(b) RollerNet dataset

Figure 8: Modelled Average degree at each time



(a) Infocom06 dataset



(b) RollerNet dataset

Figure 9: Modelled creation and deletion fractions

2. Further ideas

In order to reflect the singularities found in the real inter-contacts histograms, we could introduce a new mechanism in the simulation : When two nodes are disconnected, we draw a duration from the measured inter-contacts distribution, during which the two nodes cannot be connected again.

So far, this study focused on capturing three properties of evolving graphs via derivatives of the Edge-Markovian model. These properties however do not reflect a key aspect of real-world graphs : the community structure. Introducing a measure of how the communities are structured in the graph could help better understand how close from reality is our model.

V Conclusion

This practical work was a great introduction to graph theory, especially to evolving graphs. We discussed about the relevance of Edge-Markovian models to simulate real world evolving graphs. We observed that its simplicity is its strength (ease of understanding and implementation) but also its weakness (not adapted to time dependent processes on graphs).