IA/ML for network modeling
Practical work : Models for real mobile networks

Thomas MICHEL

February 3, 2022

1 Introduction

The present practical work consists in the analysis of different properties of mobile networks and the comparison of these properties with evolving graphs models. The two datasets that are used are RollerNet and Infocom06. These two datasets were collected using *iMotes* which, once fixed on people or at fixed locations, were able to detect other *iMote* devices in their neighborhood. The mobile networks were obtained by collecting the neighborhood information of each *iMote* at regular intervals corresponding to the time steps of the dynamic graph. Each device corresponds to a node of the graph, and two nodes are linked in the graph if the corresponding $iMotes$ are close at this time step.

2 Dynamic graph properties

The following section will describe the metrics, I implemented for this practical work.

2.1 Inter-contact duration time

Inter-contact duration time corresponds to the time separating two contacts between the same pair of nodes. One way to use this value to get insights on some properties of the network is by plotting the distribution of the durations of all the contacts happening during the experiments. As we can see on figure 1, several inter-contacts have a very short duration. The duration around 1 time steps can reflect contact creation and deletion due to the limited range of the device for nodes that stays close but not close enough for the contact to be continuous. However, we can observe that the most of the inter-contacts have a small duration compared to the duration of the experiment. This can be explained by the fact that the nodes have frequent contacts with the nodes they have already met, either due to the nodes being frequently mixed in a somewhat homogenous manner or the presence of smaller groups that are close to each others and meet often while never getting into contact with the other groups.

2.2 Average degree

We can measure the evolution of the average degree of the nodes in the graph as an indicator of the density of the connection of the nodes. In the context of mobile graphs, it can be interpreted as a measure of the average spacial density of the devices. As we can observe on figure 2, this value can vary greatly through time due to factors that depend directly on the context the data were collected in. For example, the Infocom06 data were collected over the course of four days, the arrival and depart of the people carrying the *iMote* devices can be clearly identified on the figure. Similarly, the Rollernet data may reflect some global movements of the group during the experiment, such as the individuals getting closer to each other during a difficult section of the itinerary to avoid road dangers.

The previous interpretation can be completely false in general since the distribution of degree can be very heterogeneous (large standard deviation), consequently the changes in average degree can be caused by changes affecting a minority of nodes. However, in the cases of data representing the physical distance of data, the way the data are collected mitigates the possibility of having

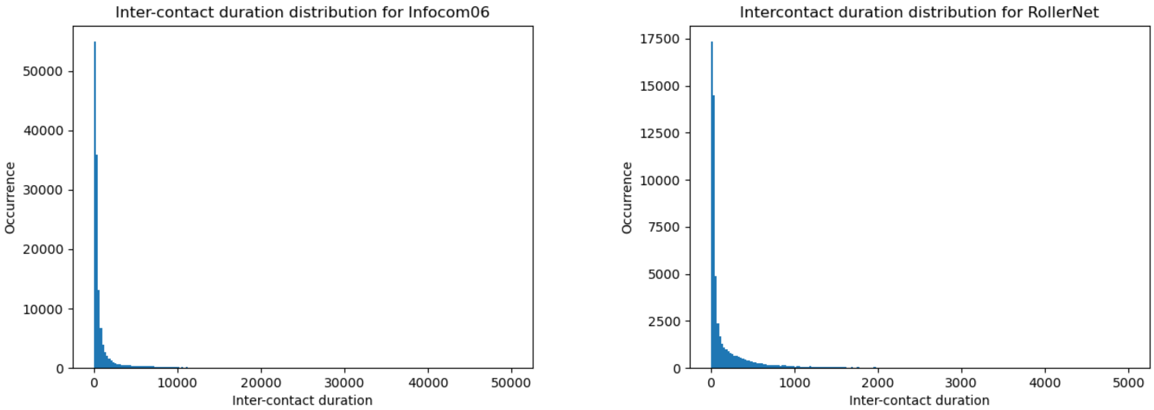

Figure 1: Distribution of inter-contact durations in the two experiments. Some of the biggest inter-contact duration times were very rare and were omitted for better visibility.

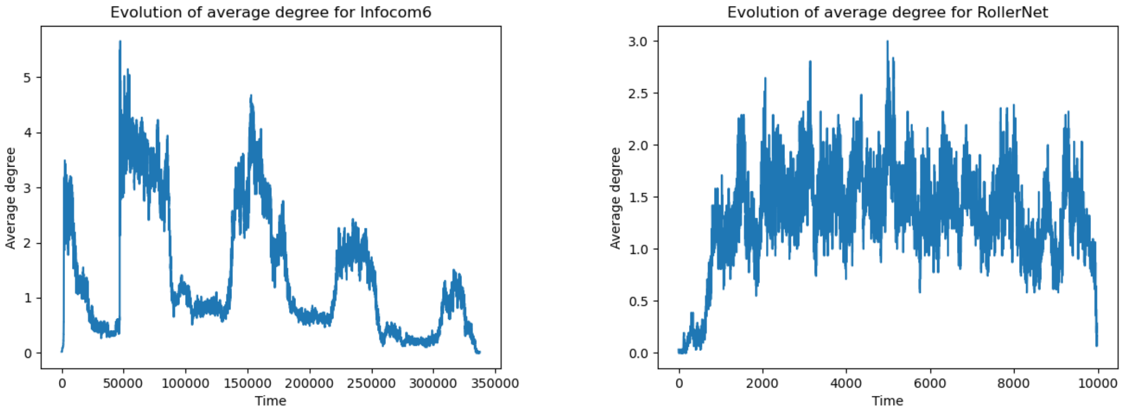

Figure 2: Evolution of average degree.

one high degree node and several small degree nodes. Indeed, when two nodes are closed from a third one, then they are likely close to each others. In particular, as it can be observed in our data on figure 3, the distribution of average degree of the nodes is not very spread out which supports the previous interpretation.

2.3 Creation and deletion of links

Another way to measure the evolution of the graph is to study the evolution of the quantity of links between the nodes through time. To capture this evolution, one can measure the fraction of created links at a given time step over the total number of links that could have been created. Similarly, the fraction of deleted links over the total number of links can be computed to complete the description of the link count changes at a given time step. This measure can help follow the link count and the trends of this count through time, but does not capture the actual positions of the links of the underlying graph. We can observe on figure 4 that the values of the fraction can change greatly from one time step to the other, leading to a poor visual representation. I chose to keep it this way to be able to observe this phenomenon, which may present difficulties when modeling the experiment.

2.4 Other metrics

Other metrics were implemented and used throughout this practical work, however they are not used in an in-depth analysis of the models in this report.

- *Moving nodes count* : This value corresponds to the number of unique nodes involved in the creation or the deletion of a link at each time step. It was intended to give a measure of

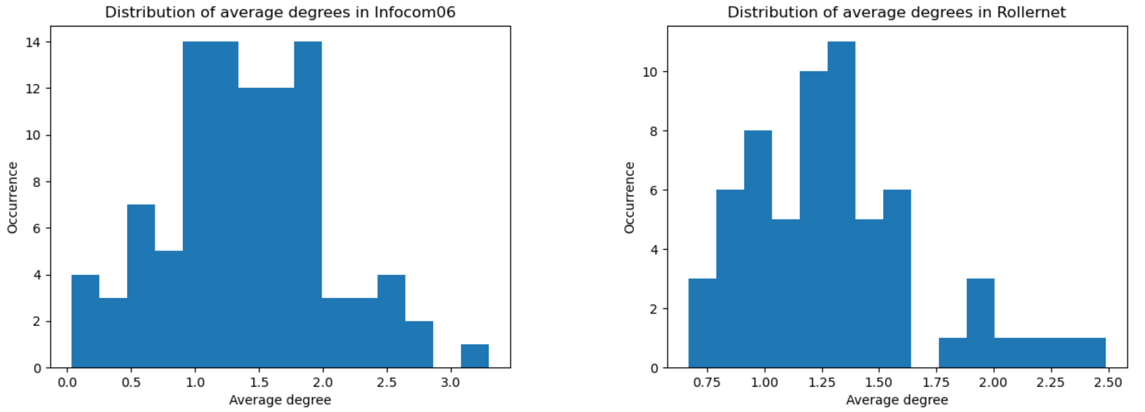

Figure 3: Distribution of the average degrees

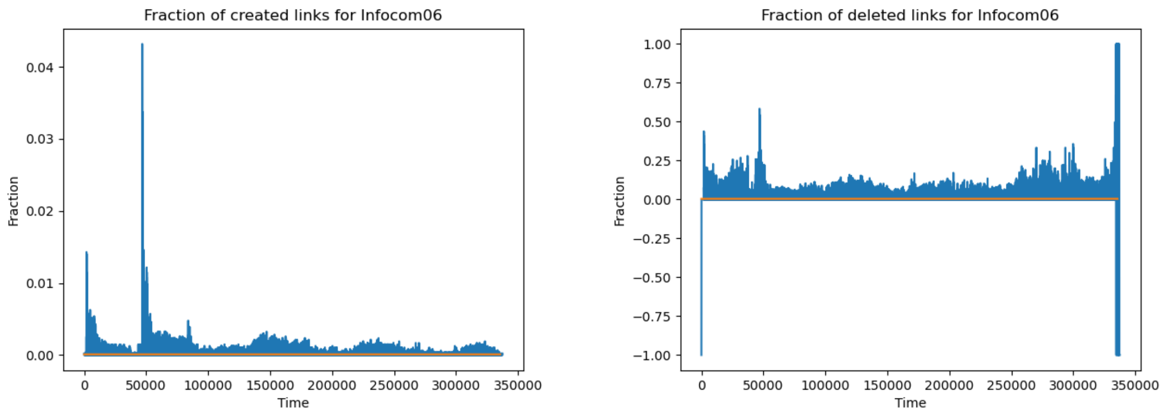

Figure 4: Evolution of the fraction of deleted links and the fraction of created links in the Infocom06 mobile graph. The horizontal orange line represents the average values of these fractions over the whole duration of the experiment.

the "mobility" of the nodes at each time steps (a small group is moving, or the movement is more global)

- *Meeting count* : This value records the number of unique nodes encountered throughout the experiment. It allows seeing if some nodes are more isolated than others during the experiment.
- *Average degree of each node* : A high value of this measure can mean that a node is more likely to stay in dense areas.
- *Average fraction of created and deleted links for each node*: Instead of computing these fractions for the whole graph, the values are computed locally by only considering the link around each node.

3 Modelling

In order to better understand the mechanisms leading to the properties observed before, we tried to find dynamic graph models able to reproduce these properties. In the following section we only consider RollerNet data. The results are similar for Infocom06.

3.1 Markovian model

The first model we consider is the edge-Markovian model. In this model, the transition between successive graphs is ruled by two independent parameters, which are the probability p that a new link is created and the probability d that an existing link is deleted. The presence or absence of a potential link is sampled according to a Bernoulli law, whose parameter depends only on the state of the edge at the previous time step (either p or d). The other parameters of this model are the number of nodes and the number of time steps. In order to try to reproduce the data at hand, we use the same number of nodes and the same duration. Additionally, the two probabilities are computed empirically by using the average of the fraction of created links and the fraction of deleted links over the course of the experiment.

To compare this model with the data, we first compute the evolution of the average degree of the nodes. We can observe on figure 5 that the range of the model average degrees is comparable to the one of the data, however its variations are very different. Indeed, the model is not able to capture the changes in the properties of the graph through time, such as the variation of the average degrees due to the day/night cycle in the case of Infocom06. We can observe with other metrics on figure 6 that the Markovian model can not capture the variation of the data through time and the distributions of the properties of the nodes.

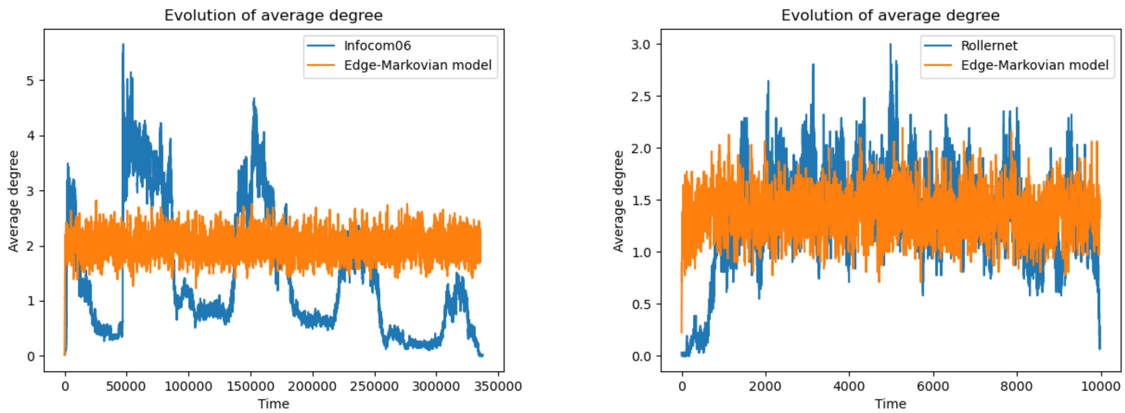

Figure 5: Comparison of the evolution of the average degree of the model with the degrees in the data

The behavior of the model may have been predicted to some extent because we chose to use the average fraction of created links and the average fraction of deleted links. However, this value does not reflect the actual variation of these fractions, as shown on figure 4. The number of creation and suppression of links vary a lot from one step to the next, and some large variations due to events during the experiment are not captured by the average values of the fractions.

The first improvement that could be made to the model would be to take into account the variation of the properties of the graph overtime. The second improvement is to introduce more diversity in the properties of the node, which can have different behavior when getting into contact with other nodes. These two improvements can either emerge from the creation of a more realistic model or be implemented by introducing more parameters to the model, such as indication of the evolution of a property of the network or a target distribution of node parameters.

3.2 Neighborhood Markovian model

A first attempt of improving the Markovian model consists in the introduction of new probabilities depending on the neighborhood of the link considered, to better capture the mechanism of link creation and deletion. The idea is to use the intuition that if two people are in contact of the third one, then the probability that they also come into contact is higher than in the case they do not have a neighbor in common. In order to verify this hypothesis, we computed new fractions of created and deleted links by splitting the cases depending on whether there exists a path of length 2 linking the two considered nodes or not. The average fractions are presented in table 1, we can observe that there exists a difference between the values of the previous edge-Markovian model and the actual data. We create a new model that takes these 4 average fractions into account.

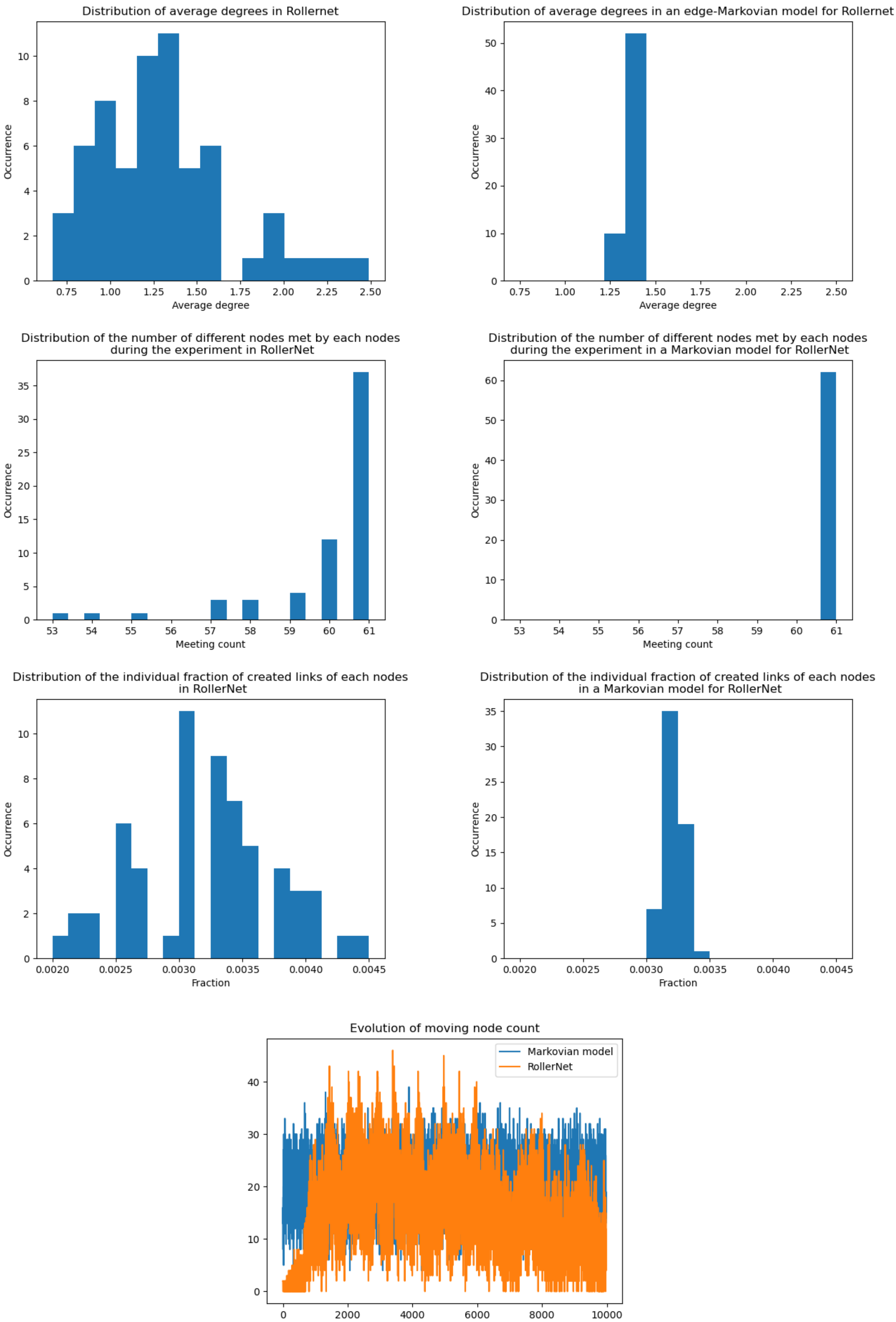

Figure 6: Comparison of RollerNet properties with its Markovian model.

	$F_{created}$(neighbors)	$F_{deleted}$(neighbors)	$F_{created}$(not neighbors)	$F_{deleted}$(not neighbors)
RollerNet	0.0149	0.1577	0.0027	0.1333
Edge-Markovian model	0.0021	0.1391	0.0032	0.0032

Table 1

As shown on figure 7, this model still has the same drawbacks as the edge-Markovian model. The range of its average degree is different, but still within the range of value of the data. We could not find a metric giving fundamentally different results from the edge-Markovian model other than the average fractions of created and deleted links when we consider different case depending on the neighborhood. In that last aspect, we have a model that one step closer to the data.

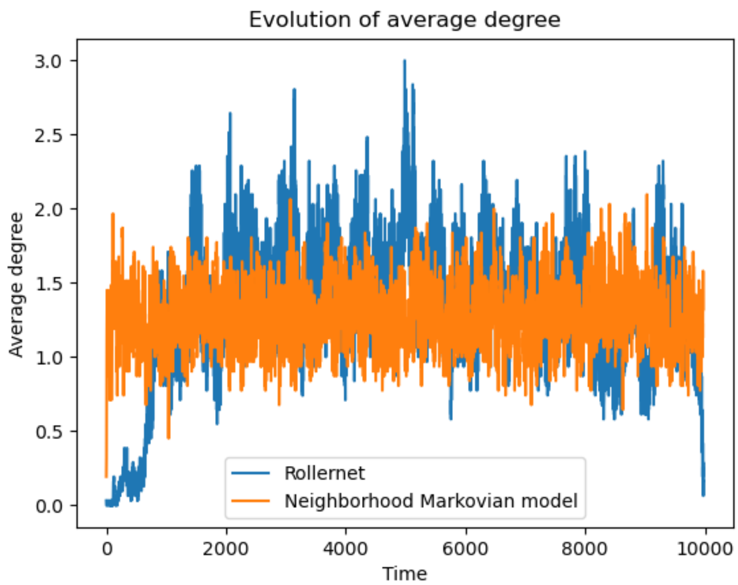

Figure 7: Comparison of the evolution of the average degree of the neighborhood dependent model with the degrees in the data

3.3 Time dependent model

In order to capture the variation of the degree overtime, we considered a model in which we give the fraction of deleted and created link at each time frame and use it as the values for the probabilities p and d at each simulation step. In some experiment, we smoothed this values by averaging them over a time window of a fixed size around each data point. As expected, this model provides convincing results when compared to the actual data with the average degree metric as shown in figure 8. The main drawback of this model is the amount of data used as parameter, which give a lot of control over the simulation and hide the underlying phenomena inside the parameters. We could actually reconstruct a dynamic graph with exactly the same evolution of the average degree as the data in a determinist manner.

Nevertheless, in this model we still sample the variations of the links independently of the others, as in the Markovian model. We can observe on figure 9 that as with the previous models, this new model can not capture the behaviors of the different nodes. They all have the same properties in the long run due to the independence of the creation and deletion of the links.

3.4 Other models

A final model that I would have liked to implement is a model closer to the reality than the previous ones. In this model, a subset of the nodes would be chosen randomly and the links of these nodes would be modified to simulate a movement inside the whole group. The moving nodes would detach from some nodes in their neighborhood and attach to another group of

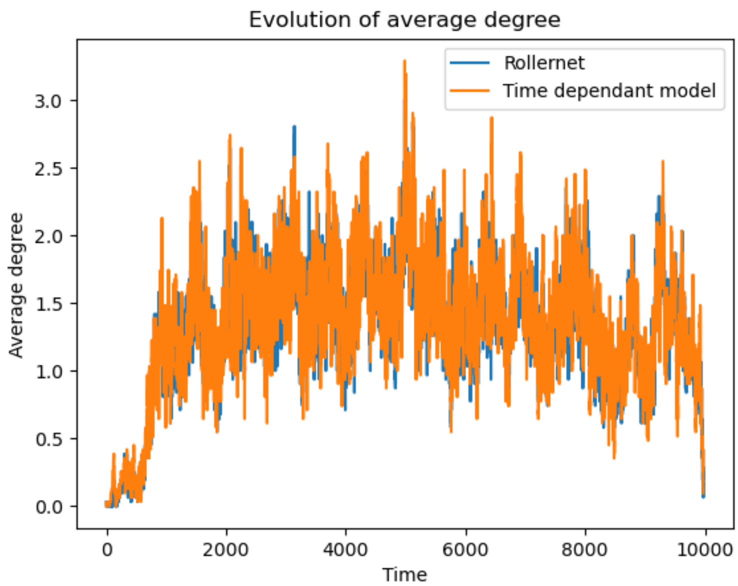

Figure 8: Comparison of the evolution of the average degree of the neighborhood dependent model with the degrees in the data

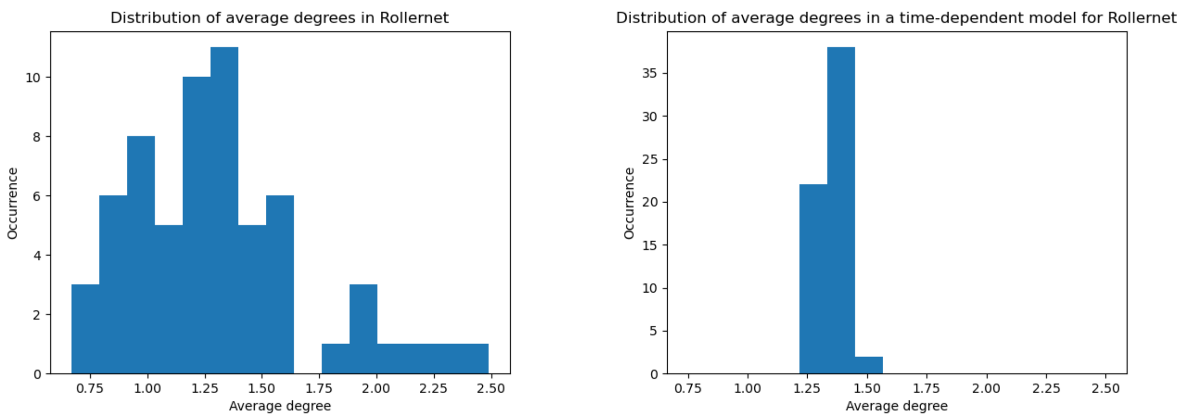

Figure 9: Distribution of average degrees

well-connected nodes. This model could be made even more realistic by collecting statistics on the behaviors of the nodes, such as the time between each movement. We could then assign a renewal process to each node to simulate its movement (the renewal process telling when a node decide to move).

However, I was not able to implement this model since I did not find how to compute the metrics required to define the parameters of the model. For example, when a node is involved in the deletion of a link, it may be hard to tell if it is this one that moves or the other node. A way to do it would be to look at the change of neighborhood of each node, or to chose one of the two nodes such that the number of moving nodes at a given time step is minimized. I was not able to find a satisfying way to define this metric. Another restraint was the computation of the probabilities involved in the renewal process. A simplified model to get around this would be to consider the probability of a node to move at each time step by computing the total number of moves over the duration of the experiment.

In any case, this model would not be able to capture the variation in the data due to external phenomena such as the day/night cycle of Infocom06. However, it would be interesting to compare its results to other Markovian models to see if a probably more realistic model could have explained the data better than the three previous models for some metrics.

4 Conclusion

In this practical work, we were able to explore different metrics and models for mobile networks. In particular, we highlighted the shortcomings of the simple edge-Markovian model in explaining mobile networks data by proposing multiple metrics for properties both at a global and a local scale. We tried two alternative models aimed at correcting some flaws of the edge-Markovian model, however these alternative models remain unsatisfying when we try to describe the real mobile networks.