

# On the relevance of the edge-Markovian evolving graph model for real mobile networks

Aurelie Faure de Pebeyre  
LIP6-CNRS-UPMC  
4 place Jussieu  
75005 Paris,France  
Email: Aurelie.Faure@lip6.fr

Fabien Tarissan  
LIP6-CNRS-UPMC  
4 place Jussieu  
75005 Paris,France  
Email: Fabien.Tarissan@lip6.fr

Julien Sopena  
LIP6-CNRS-UPMC  
4 place Jussieu  
75005 Paris,France  
Email: Julien.Sopena@lip6.fr

**Abstract**—The development of wireless devices led the scientific community to focus more and more on systems of interaction composed of moving entities. In this context, different models have been proposed in an attempt to capture properties of the observed dynamics. Among those models, the edge-Markovian evolving graph model is appealing since it enables to highlight temporal dependencies in the evolution of the graphs. This model relies on two parameters accounting respectively for the creation and suppression of links in the graph. Thus it assumes that these two parameters are sufficient to characterise the dynamics during all the evolution of the graph. In this paper, we test this hypothesis by confronting the model to 6 datasets recording real traces of evolving networks. In particular, we study the proportion of created and deleted links over the time. The results show that 5 of the 6 case studies present an heterogeneous distribution of those fractions which contradicts the underlying hypothesis of the model. Besides, in order to understand the importance this might have as regard structural properties of real networks, we also study the impact the Markovian model has on the mean degree over the time. It turns out that even in the suitable case, the model fails to reproduce correctly this property which indicates its inadequacy for even more complex properties of real evolving networks.

## I. INTRODUCTION

July 15, 2013

The development of wireless devices led the scientific community to focus more and more on systems of interaction composed of moving entities. Those systems are characterised by the fact that the links, defined by the proximity between two different wireless equipments, appear and disappear continuously which makes the structure of the systems in constant evolution. A key question of the domain consists then in taking advantage of the properties of the observed dynamics in order to define efficient and robust communication protocols [1].

In this context, if studies have already shown interesting phenomena, such as the contact and inter-contact duration time following a power-law [2] or the accordion phenomenon [3], few of them have proposed models able to capture the properties of the dynamics of those systems. Among the random evolving graph models, the *edge-Markovian evolving graph* model [4]–[6] is appealing as it relies on a temporal dependency to determine the state of the graph at the next time according to its current state. This model was used in theoretical studies focusing on diffusion protocols, such as *flooding* [7], [8] or *push* [9], in order to prove their efficiency.

However, even fewer analyses have been made to validate

the model as regard the real dynamics observed in mobile networks. As far as we know, only two papers [10], [11] have performed such an analysis. But, first, they all study the behaviour of a particular dissemination algorithm and, second, they do not consider properties of the dynamics inherent to those evolving networks.

This work intends to answer this open question by analysing rigorously the properties of different traces of mobile networks and by studying the relevance of the Markovian model as regard those properties on a wide range of different datasets.

The remain of the article is organised as follow: Section II presents the necessary background for this study, Section III shows the main results related to the confrontation between real data and the underlying hypothesis made by the Markovian model while Section IV studies more precisely the impact the model has on structural properties of evolving networks. Finally Section V concludes the article and opens on new perspectives.

## II. BACKGROUND

In this section we provide details on the edge-Markovian evolving graph model (Section II-A) and the datasets on which we relies to perform our analysis of the properties of the dynamics of evolving networks and tune the model parameters (Section II-B).

### A. Markovian model

The growing interest for mobile networks and evolving graphs led the scientific community to recently propose models to account for the new characteristics of those networks. In this context, the proposition relying on random evolving graphs due to Ferreira [12] allows to cover a wide range of dynamics. Indeed, in this framework an evolving graph is simply described as a succession of distinct graphs at each time step. This allows for the reuse of all the models for random graph generation by simply considering at each time step independent random graphs. Although interesting for its simplicity, this model fails to account for inherent properties of evolving networks for which a state at a given time directly relies on its state at the previous time step.

Many attempts have been made in order to capture this property. J. Leskovec et. al. [13] for instance proposed a model based on the growth of vertices and degrees. However, this

approach does not take into account the instability of the links. In their proposition, Chaintreau et. al. [14] used sequences of graphs generated uniformly at random to analyse the diameter of ad-hoc mobile networks. Pellegrini et. al. [15] explored the notion of temporal connectivity.

This is why A. Clementi et. al. [4], [8] recently introduced the concept of edge-Markovian evolving graphs model (referred to further as *Markovian model*) which allows to point out this dependency. In this model, given a graph  $G_t$  at a given time  $t$ , the structure of the graph at the next time  $G_{t+1}$  directly depends on the one of  $G_t$  and is ruled by two independent parameters: the probability  $p$  of creation of an edge which does not exist and the probability  $d$  of deletion of an existing edge.

The interest of the model is twofold: first of all, it has been proven that whatever the values of  $p$  and  $d$ , and whatever the initial graph  $G_0$ , an evolving graph generated by this model converges towards a Erdős-Rényi random graph [16] with the probability for an edge to exist being  $\hat{p} = \frac{p}{p+d}$ . Secondly, it has only two parameters which enable to perform rigorous studies. Particularly, the latter have been performed on the efficiency of different diffusion protocols while remaining realistic as regard the temporal dependency principle.

However, the simplicity of this model may also be its weakness. Indeed it uses only two parameters to rule edges creations and deletions which supposes that the value of those parameters is representative of the **whole** evolution of the graphs over time. This is precisely what we intend to assess in this paper.

On a related notice, it is worth noticing that Scherrer et. al. [6] conducted an extensive analysis of mobile datasets in order to identify key properties of the dynamics and propose Markovian models able to reproduce it. In that sense their work is close to the one proposed in the present paper. But it actually differs on several aspects. Not only the datasets but also the chosen properties are different from the ones used in this study. Besides, they relied on properties identified in their analysis in order to propose several distinct models able to account for it. Our work however focuses on one simple unique model and confront it with characteristics observed on real datasets.

## B. Dataset

In order to test the relevance of the Markovian model, we have analysed 6 datasets which present traces of human-contact networks collected during different real-world experiments. In each of these experiments, people were carrying devices able to detect the presence of others similar devices nearby. The frequency of the measurements are different from one experiment to another and can vary from the second to the minute.

In our study, we used:

- **Rollernet [3]** : This dataset was collected during a rollerblading tour in Paris in August 2006. The tour lasted more or less 3 hours with a break of about 30 minutes; rollerbladers travelled about 20 miles; the tour took place in the streets which implies acceleration and speed reduction especially during passing traffic lights. *iMotes* have been given to 62 participants covering the entire assembly, knowing

that about 2500 people were participating in the tour. *iMotes* distributed used bluetooth technology and looked at their neighbourhood every 15s.

- **Infocom 05 and Infocom06 [2]** : Those two experiments were collected during the Infocom conferences respectively at Miami and Barcelona. Like RollerNet, they used the *iMote* technology with a scan every 120s. The first one (March 2005) lasted 4 days with 41 participants over the 800 participants to the conference. The second one (April 2006) had 98 devices among which 78 were participants to the conference, 17 were static and 3 have been placed in the elevators.
- **Unimi [17]** : This dataset was collected in the university of Milan by using *Pocket Mobile Trace Recorders* (PMTR) with 44 participants. These devices sent a signal every seconds.
- **hypertext09 [18]** : used *radio badges* distributed to 113 participants at the Hypertext conference (June 2009). These badges have recorded face to face contacts every 20s during 2,5 days.
- **Sociopattern [18]** : This dataset was collected during an exhibition on infectious propagation in Ireland. It also used radio badges distributed at the beginning of the exhibition. It is the only dataset in which every person participating to the exhibition had a badge. The experiment lasted 3 months with one record per day. The number of participants (88 to 410) is different from one day to another.

The table I below summarises some characteristics of the datasets.

Datasets	Duration	Participants	Contacts	Freq. (sec.)
RollerNet	3 hours	62	60 146	15
Infocom05	4 days	41	17 682	120
Infocom06	4 days	98	148 784	120
HT09	2,5 days	113	9 865	20
Socio	1 day	151	2 051	20
Unimi	10 days	44	11 895	1

TABLE I. DATASETS CHARACTERISTICS

In this study, we used all the datasets presented above. However due to space limitations, we have decided to present only 3 of them in the rest of the paper. We have chosen RollerNet, Infocom06 and SocioPattern because they stem from 3 distinct experiments and thus present different characteristics. Moreover, the duration of those experiences is totally different which entitles to draw general conclusions which do not depend of the duration time. Finally, the results obtained with the 3 other datasets are similar to what is presented here. More precisely, the results with Infocom05 are very similar to those of Infocom06 and the ones of Unimi and HT09 to SocioPattern. It is worth noticing that in SocioPattern, the number of participants is different every day. Here we only consider one particular day representative of a mean activity (in terms of number of participants).

## III. ANALYSIS OF THE DYNAMICS

This section deals with the evolution of creation and deletion of links over time in the different datasets. This is

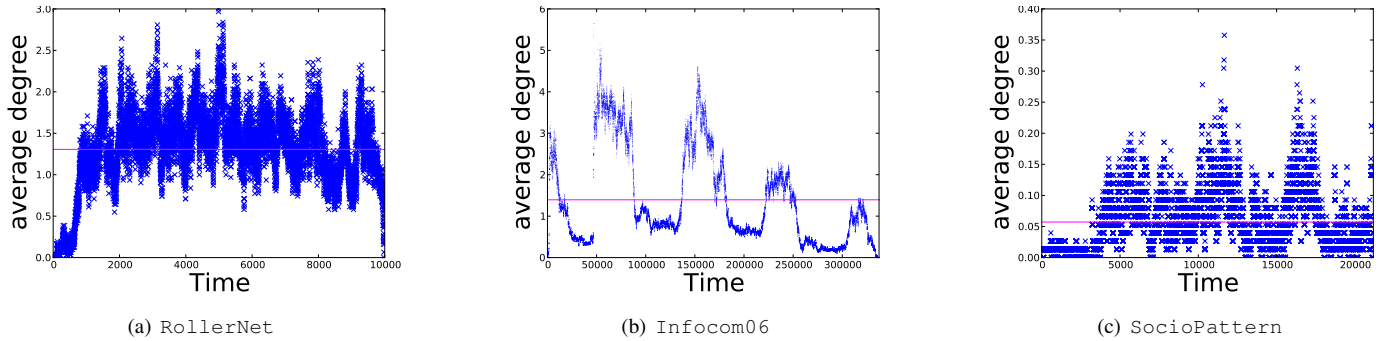


Fig. 1. Evolution of the mean degree over time

the main element on which is based the Markovian model.

We start by displaying the evolution of the mean degree over time in the different datasets. The results presented in Figure 1 show that all the graphs have a very low density, with a mean degree of 1.4 for RollerNet, 1.3 for Infocom06, and 0.06 for SocioPattern, with a density of  $2.2(10^{-2})$ ,  $1.3(10^{-2})$  and  $4(10^{-4})$  respectively. In addition, to the low density of connexions (particularly obvious for SocioPattern), the plots of Figure 1 show a very strong variation of the number of links in the 3 experiments. However, if those variations are quite regular for RollerNet, it is not the case for the two others traces. This is more emphasised for Infocom06, where we can distinguish relatively stable periods and sharp increases and decreases of the mean degree.

Since the Markovian model relies on mean values, we have studied the impact of those variations on the two parameters of the model. Thus we computed, for each time step  $t$ , the proportion of created (resp. deleted) links according to the number of non-existing (resp. existing) links at time  $(t - 1)$ . The results are shown in Figure 2 (resp. Figure 3).

As regard the creation of links, Figure 2 reveals that the range of values on all the plots is quite wide. This is particularly true for Infocom06 (Figure 2(b)) and SocioPattern (Figure 2(c)) on which the most extreme events are clearly observable. This is confirmed by the fact that the mean (see Table II below), the median, and the 75-percentile are very close to 0 indicating that low values are over-represented in those datasets. From this point of view, as regard RollerNet, the concept of mean value seems more relevant as it seems to position itself in the middle of the values taken over time.

Datasets	RollerNet	Infocom06	Socio
Created links	$3.2 (10^{-3})$	$9.5 (10^{-5})$	$9 (10^{-6})$
Deleted links	$1.4 (10^{-1})$	$4.5 (10^{-3})$	$1.6 (10^{-2})$

TABLE II. MEAN VALUE FOR THE FRACTIONS OF CREATED AND DELETED LINKS.

The same observation can be made with deleted links (Figure 3) for which it is even more obvious. Here, the range taken by the values is complete ( $[0..1]$ ) and some proportions can reach the maximum value, which means that all the existing edges have been removed. By investigating manually the datasets, it turned out that this behaviour is easily understandable since it is often due to the deletion of the *unique* edge present (proportion is then equal to 1) in the

graphs at that time, which appends usually at the beginning and at the end of the experiment. Finally, the RollerNet experiment (Figure 3(a)) is again more homogeneous than the two others which is validated by the values taken by mean, the median, and 75-percentile.

Thus, according to the plots presented above, it seems that characterising the dynamics of a network by two parameters is relevant for RollerNet but is inadequate for Infocom06 and SocioPattern since the observed values are not homogeneously distributed over the experiments.

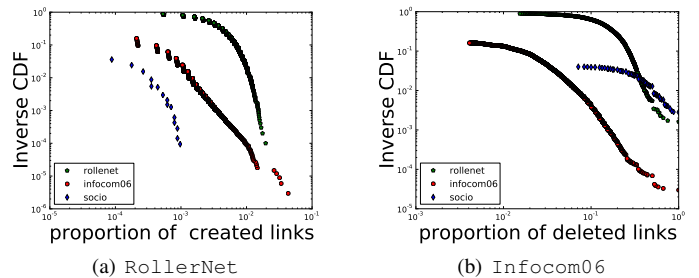


Fig. 4. Inverse cumulative distribution function of fractions of created and deleted links

To confirm this analysis, Figure 4 shows the inverse cumulative distribution function of those fractions on each datasets. It confirms the fact that the value of fractions of created links (left) is heterogeneous for two datasets (Infocom06 and SocioPattern). Indeed, we can see a straight line (in the log-log scale) over several orders of magnitude on the curve representing the Infocom06 experiment. On the contrary, the cumulative distribution for RollerNet sharply decreases around the mean value ( $3.2 \times 10^{-3}$ ), thus indicating that the set of values spread around the mean value. The same observation can be done for deleted links (right) even if it is less pronounced.

The figure confirms also an important observation for the three datasets in the case of deleted links: there is a non negligible number of fractions equal to 1. As said before, this observation is explained by the fact that when few links exist in the network, the value of the fraction is usually close to the extreme. This is especially notable when only one link exists, the fraction being either 0 or 1. Yet, this fact seems hardly compatible with the notion of a mean value made by the Markovian model (cf. table II).

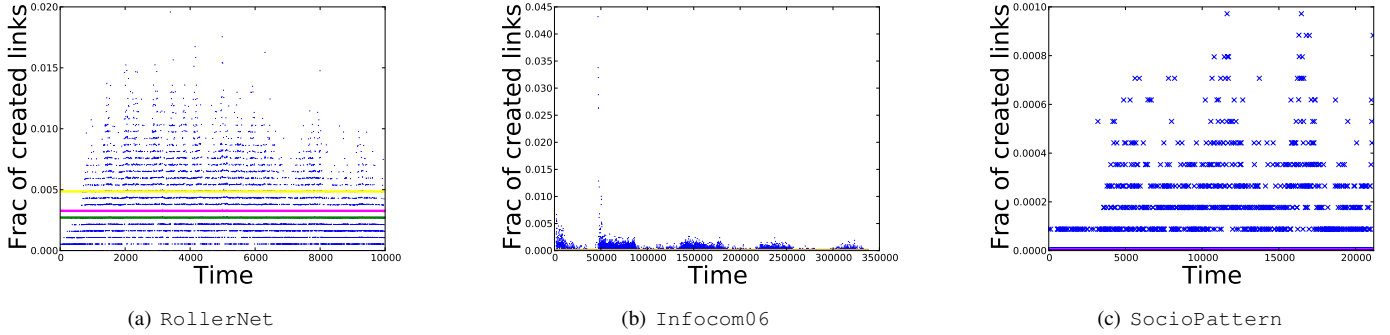


Fig. 2. Evolution of the proportion of created links with mean (pink), median (green) and 75-percentile (yellow).

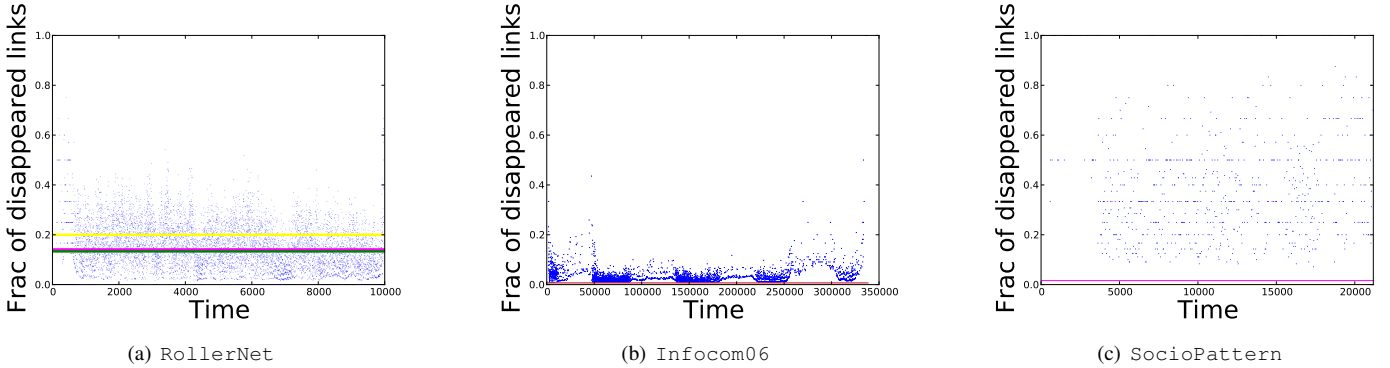


Fig. 3. Evolution of the proportion of deleted links with mean (pink), median (green) and 75-percentile (yellow).

As a partial conclusion, it seems that if the present results do not totally discredit the usefulness of the edge-Markovian evolving graph model, the analysis performed in the present section puts however the relevance of its underlying hypothesis into perspectives. As usual when dealing with properties of real networks, the dynamics of mobile networks present some heterogeneous aspects that is not captured by models relying on homogeneous parameters. In order to understand the impact of this result, we now study the relation of the Markovian model with other network properties.

#### IV. IMPACT OF THE MARKOVIAN MODEL

In this section, we study the evolution of the connexion in mobile networks and compare the properties observed in real data with the ones observed in the artificial traces generated by the Markovian model.

Thus, we compare each real trace with its equivalent in the Markovian model. To do so, we used the mean values of the fractions of created and deleted links presented in section III (see table II) and generated as much states as existing logs in experiments.

##### A. Impact on mean degree

First, we study the evolution of the mean degree over time. This characteristic is very simple (in structure) and quite close to the model's parameters since the degree is directly related with the creation and the deletion of the links. As such, this should be advantageous to the Markovian.

The obtained results are showed in figure 5 and are compared to those obtained on the real datasets studied (figure 1). As expected, the model is not suitable for datasets like Infocom06 and SocioPattern. Not only the variations observed over time on data do not exist in the trace generated by the model (which was foreseen given that the model is homogeneous), but, more important, the extreme events disappeared. We notice that in those two examples, the scale of the values on the y axes is modified (divided by 2) between real data and the model.

However, in the RollerNet experiment, the model reproduces quite well the evolution of the mean degree observed on the real trace. Except at the beginning, which is naturally a bias due to the measurement protocol, the behaviour is quite correct.

To be more precise in our analysis, we present figure 7 the inverse cumulative distribution function observed both in real data and in the generated traces. Again, the distribution is different for real data and for the model in the case of Infocom06 (middle) and SocioPattern (left).

Besides, and more surprisingly, the distribution is also different for RollerNet, whereas the comparison of figures 1(a) and 5(a) suggests that the Markovian model could be adapted. This shows that even if the distributions of the  $p$  and  $d$  values are homogeneous, the dynamics captured by those two parameters is not sufficient for reproducing structural properties of evolving networks. This raises a cautionary message. Indeed, if a property as simple as the mean degree is not well captured, it is very likely that more intricate properties, less directly related to those parameters, will also be poorly

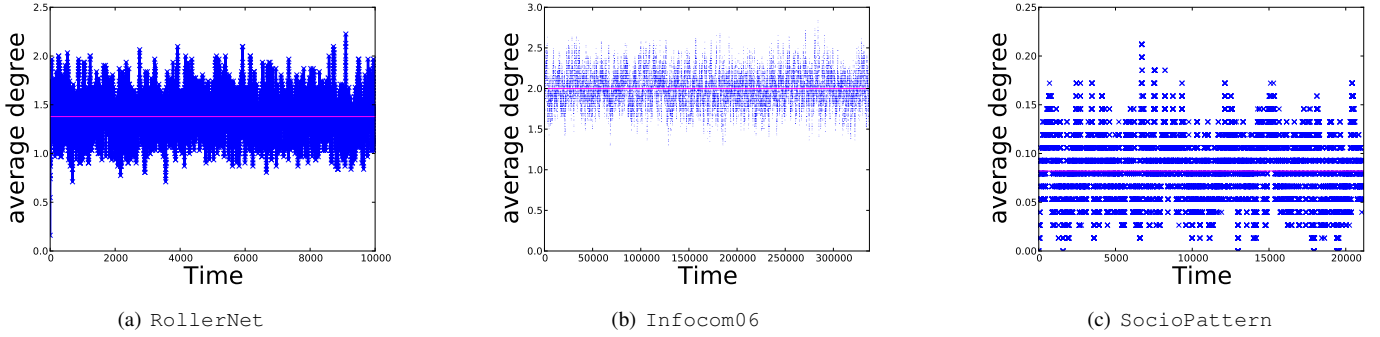


Fig. 5. Evolution of the mean degree over time on the Markovian model.

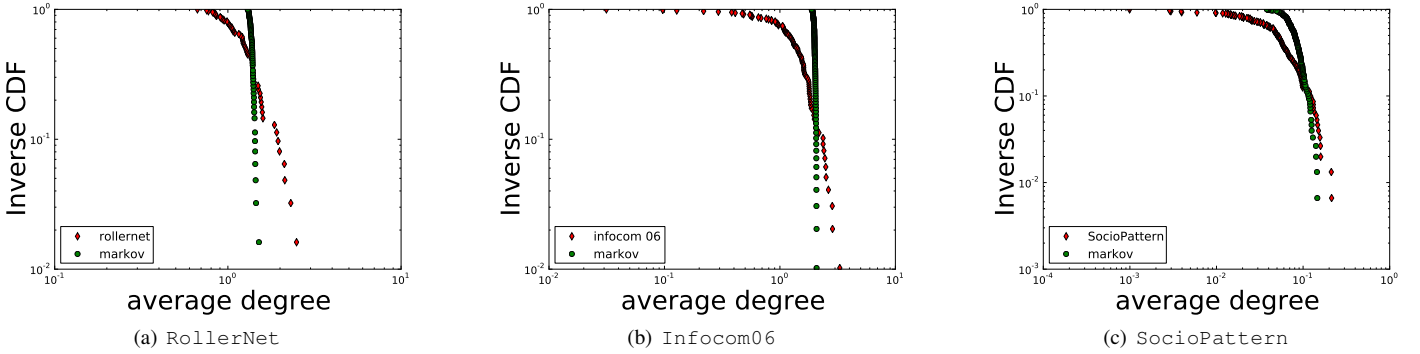


Fig. 6. Comparisons of cumulative distribution of mean degree.

reproduced.

### B. Impact on the degree distribution

In this section, we study the distribution degree, i.e., the probability that a node has a particular degree. This graph’s property is particularly interesting and has a strong influence on the efficiency of distributed algorithms.

The three histograms of figure 7 represent a comparison of the degree distribution on real datasets (in white) and their corresponding artificial Markovian trace (in grey). For each degree  $d$  (on the x axis) we plot the number of nodes connected to exactly  $d$  neighbours (y axes) as a mean value for all the graphs of a given trace.

If we focus first on the distribution of real data (white histogram), we notice that graphs corresponding to those experiments have an heterogeneous degree distribution. This is particularly clear for Infocom06 (figure 7(b)) but less observable for SocioPattern (figure 7(c)). For the last case, the very low number of connexions (density of  $4(10^{-4})$ , section III) explains that the phenomenon is less observable.

Now, if we compare the results of the real traces (white histogram) with the distributions obtained by the Markovian model (in grey), we can see that the model fails in reproducing the heterogeneous distribution. To understand this phenomenon, it is important to remember that the evolving graphs generated by the Markovian model converge toward an Erdős-Rényi [16] random graph. In such graphs, the degree distribution is known to follow a Poisson distribution (binomial law), which is confirmed by the figure.

This last comparison shows that, although the Markovian model relies on the dynamics of creation and deletion of links, it fails in accounting for the repartition of those links over the nodes as observed in the different experiments. Indeed, the model distributes homogeneously the links in the network while in real networks, some nodes tend to gather the connexions and somehow “attract” other nodes – phenomenon also referred to as *preferential attachment*. This indicates that the model will face the same problem with other classical properties such as the *clustering coefficient*, generated by sets of pairwise connected devices. As regard this last property, a first global analysis confirms the statement. Table III shows the mean value of the clustering coefficient observed over the time both for the real datasets and the synthetic data generated by the model. In each case, the value of the clustering coefficient is at least one order of magnitude lower for the model than for the real dataset.

Datasets	RollerNet	Infocom06	Socio
real	$1.7 (10^{-1})$	$2.9 (10^{-1})$	$4 (10^{-1})$
model	$2 (10^{-2})$	$1.9 (10^{-2})$	$0 (10^0)$

TABLE III. MEAN VALUE FOR CLUSTERING COEFFICIENT.

## V. CONCLUSION AND PERSPECTIVES

The work presented in this article intended to confront the hypothesis of homogeneity made by the Markovian model with the dynamics observed in real datasets recording the activity of real evolving networks. The study showed that in most of the cases (5 of the 6 studies), this hypothesis was contradicted by the analysis.

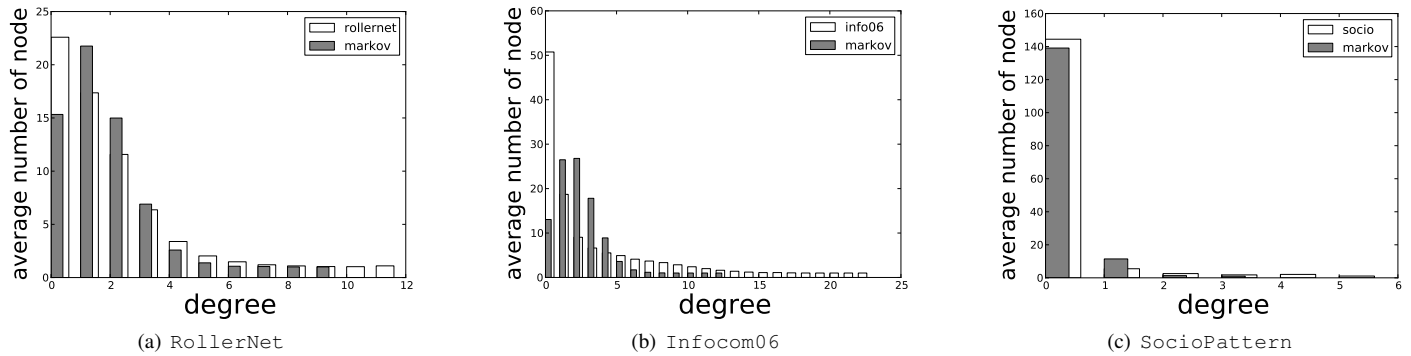


Fig. 7. Comparison of the distributions of the mean degree.

Moreover, we have shown that, even with RollerNet, which has an homogeneous repartition of the model's parameter, the model fails to reproduce the simple property of the mean degree over the time. Finally, the study highlights the fact that the model can not simulate degree distribution as observed in the real traces. This limit, resulting from the nature of the model, can be generalised to other properties like clustering coefficient. However, those negative results should be moderate as regard the use of the model. Particularly, the works [10], [11] have shown that this model could be adapted for studying flooding diffusion protocols. Thus it is not the purpose here to discredit the usefulness of the model but to show that the question of its representativeness as an evolving graph model accounting for the whole dynamics of networks is subject to caution.

As regard the perspectives, it would be interesting to see if a simple modification taking into account the fact that the fractions of created and deleted links are heterogeneous would be enough to improve the representativeness of the model. For example, we could base the model not on the mean values of  $p$  and  $d$  but on sets of values for  $p$  and  $d$  whose distributions follow an heterogeneous law.

It would also be interesting to look at some more intricate properties than the simple evolution of the mean degree. This one has been chosen here for its proximity with the model's parameters but it would also be relevant to look at the repartition of the connexion in the networks as well as the correlations existing between the link's creation and deletion over time. It is very likely that the periods with many links creations are related to periods of few links deletions (and reciprocally) like suggested by the accordion effect observed on RollerNet [3].

Finally, to follow the usefulness of the model in the domain of diffusion, it would be appealing to see how diffusion like the *gossip* protocols (such as the push diffusion studied in [9]) will perform on real data and confront it with the results obtained with the Markovian model. This could reaffirm the interest of the Markovian model for the community.

**Acknowledgement.** This work is partly supported by the DynGraph grant from the ANR (n 10-JCJC-0202).

## REFERENCES

- [1] M. Boc, A. Fladenmuller, and M. D. de Amorim, "Otiy: locators tracking nodes," in *Proceedings of the ACM CoNEXT conference (CoNEXT '07)*. ACM, 2007, pp. 14:1–14:12.
- [2] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 606–620, 2007.
- [3] P.-U. Tournoux, J. Leguay, F. Benbadis, V. Conan, M. D. De Amorim, and J. Whitbeck, "The accordion phenomenon: Analysis, characterization, and impact on dtn routing," in *Proceedings of INFOCOM'09*. IEEE, 2009, pp. 1116–1124.
- [4] A. E. Clementi, C. Macci, A. Monti, F. Pasquale, and R. Silvestri, "Flooding time in edge-markovian dynamic graphs," in *Proceedings of the Symposium on Principles Of Distributed Computing (PODC '08)*, 2008, pp. 213–222.
- [5] A. E. F. Clementi, F. Pasquale, A. Monti, and R. Silvestri, "Information spreading in stationary markovian evolving graphs," in *Proceedings of IPDPS'09*. IEEE, 2009, pp. 1–12.
- [6] A. Scherrer, P. Borgnat, E. Fleury, J.-L. Guillaume, and C. Robardet, "Description and simulation of dynamic mobility networks," *Computer Networks*, vol. 52, no. 15, pp. 2842–2858, 2008.
- [7] H. Baumann, P. Crescenzi, and P. Fraigniaud, "Parsimonious flooding in dynamic graphs," in *Proceedings of the 28th Annual ACM Symposium on Principles of Distributed Computing (PODC 2009)*. ACM, 2009, pp. 260–269.
- [8] A. E. Clementi, C. Macci, A. Monti, F. Pasquale, and R. Silvestri, "Flooding time of edge-markovian evolving graphs," *SIAM journal on discrete mathematics*, vol. 24, no. 4, pp. 1694–1712, 2010.
- [9] A. Clementi, P. Crescenzi, C. Doerr, P. Fraigniaud, M. Isopi, A. Panconesi, F. Pasquale, and R. Silvestri, "Rumor spreading in random evolving graphs," *arXiv preprint arXiv:1302.3828*, 2013.
- [10] M. Vojnovic and A. Proutiere, "Hop limited flooding over dynamic networks," in *Proceedings of INFOCOM'11*, 2011, pp. 685–693.
- [11] J. Whitbeck, V. Conan, and M. D. de Amorim, "Performance of opportunistic epidemic routing on edge-markovian dynamic graphs," *IEEE Trans. on Communications*, vol. 59, no. 5, pp. 1259–1263, 2011.
- [12] A. Ferreira, "On models and algorithms for dynamic communication networks: The case for evolving graphs," in *AlgoTel'02*. INRIA Press, 2002, pp. 155–161.
- [13] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over time: densification laws, shrinking diameters and possible explanations," in *Proceedings of the International Conference on Knowledge Discovery in Data mining*. New York, NY, USA: ACM, 2005, pp. 177–187.
- [14] A. Chaintreau, A. Mtibaa, L. Massoulie, and C. Diot, "The diameter of opportunistic mobile networks," in *Proceedings of the 2007 ACM CoNEXT conference*, New York, USA, 2007, pp. 12:1–12:12.
- [15] F. D. Pellegrini, D. Miorandi, I. Carreras, and I. Chlamtac, "A graph-based model for disconnected ad hoc networks," in *Proceedings of INFOCOM'07*. IEEE, 2007, pp. 373–381.
- [16] P. Erdos and A. Rényi, "On the evolution of random graphs," *Math. Inst. Hung. Acad. Sci.*, vol. 5, pp. 17–61, 1960.
- [17] S. Gaito, E. Pagani, and G. P. Rossi, "Opportunistic forwarding in workplaces," in *ACM Workshop on Online Social Networks (WOSN)*, 2009, pp. 55–60.
- [18] SocioPatterns, "<http://www.sociopatterns.org/>."