A basic toolbox for the analysis of dynamics of growing networks

Clémence Magnien,[†] Mahendra Mariadassou[‡] and Camille Roth[§]

CREA - CNRS - Ecole Polytechnique, 1 rue Descartes, 75005 Paris

Most complex networks evolve through time, yet up till now their analysis has mainly focused on static snapshots of these networks. Our goal is to provide a set of basic tools for the analysis of the dynamics of networks. These tools should be at the same time very general, in order be applied to all networks, whatever field they are originated from, and yet be relevent, in order to be able to capture meaningful behaviours. We restricted ourselves to the case of growing networks (*i.e.* networks in which nodes and links appear but are never removed), but our tools can easily adapted to general dynamic networks. Finally, we have applied our tools to two cases of growing networks, to illustrate their relevence.

Keywords: Complex Networks, Dynamic Networks, Network Analysis

1 Introduction

The analysis of complex networks has lead since a few years to the introduction of a set of statistical parameters providing information on networks, See for instance [BS03, Bar02]. Some of these parameters played a central role because they are very general: most real-world complex networks display the same behaviour concerning them, which is not captured by classical models (like regular graphs or random graphs).

In many cases, the networks under concern evolve during time: new nodes or links arrive, and some leave. This has been pointed as a central characteristics of real-world complex networks, but until now very little is known on these dynamics. Indeed, dynamic data are difficult to collect and represent, and even if one has such data there are very few, if any, statistical tools for their analysis. The small number of papers addressing this subject, are mostly concerned with answering precise questions, with no goal of generalisation [PHL04, OdC03, Hol03, Sni01, KMCB95, EL03, New01].

Our aim in this paper is to make a first step in this direction: we want to propose a set of simple statistical tools which may be used as a basic toolbox when one deals with a dynamic complex network. We want our tools to be: as simple and general as possible, in order to make them easily usable in various contexts; relevant, in the sense that they indeed capture some non-trivial behaviours of the dynamics; efficiently computable, or at least tractable in large practical cases (real-world networks can be very large, up to severla dozens of millions of nodes); and finally they should span well the variety of phenomena one may want to capture, at a basic and general level.

We actually restricted ourselves to a special case: we only considered *growing* networks, *i.e.* networks in which nodes and links arrive during time but are never removed afterwards. Many real-world cases actually fit in this special case. Moreover, most tools we will propose may easily be extended to the general case. New tools should however be introduced to deal with the general case, but this is out of the scope of this paper.

We will now present the formal framework we have considered as well as the two growing networks we have chosen to illustrate our approach, and then we will present the tools we have introduced for the analysis of the dynamics.

[†] magnien@shs.polytechnique.fr
[‡] mahendra.mariadassou@ens.fr
[§] roth@shs.polytechnique.fr

network	nodes	links	time-grain	density	av. distance	clus.
actor	869,986	51,520,496	1 year	1.361×10^{-4}	3.77	0.795
citations	29,555	352,807	30 days	9.147×10^{-4}	4.23	0.329
IP exch.	281,760	508,667	10 sec.	1.281×10^{-5}	5.29	5.914×10^{-5}

Table 1: Number of nodes and links, time grain, density, average distance and clustering coefficient of the growing networks under consideration.

2 Formal framework and methodology

There are several ways to describe a growing network. We will use the formalism of *coarse-grained* growing networks, which can be described as follows: a *coarse-grained* growing network is described as a series of sets $(\delta V_i, \delta E_i)$, $i = 1, 2, \ldots$ and a time grain denoted by δt. The network at time $t = \delta t \cdot i$ (*i.e.* the network at step i) is nothing but $G_i = (V_i, E_i)$ with $V_i = \cup_{j \leq i} \delta V_j$ and $E_i = \cup_{j \leq i} \delta E_j$. Notice that, even if a node or a link may appear several times (*i.e.* belong to several δV_i's or δE_i's), this information is not encoded in the graph at a given time. The coarse-grained formalism naturally induces several notions:

- $\delta V_i^{\top}$ is the set of nodes in δV_i which appear at step i, *i.e.* $\{v \in \delta V_i \text{ such that } \nexists j < i,\ v \in \delta V_j\}$. We will call these nodes the *external nodes of step i*.
- $\delta V_i^{\perp}$ is the set of nodes in δV_i already present before step i, *i.e.* $\{v \in \delta V_i \text{ such that } \exists j < i,\ v \in \delta V_j\}$. We will call these nodes the *internal nodes of step i*.
- $\delta E_i^{\top}$ is the set of links in δE_i between nodes which appear at step i, *i.e.* $\delta E_i^{\top} = \delta E_i \cap \delta V_i^{\top} \times \delta V_i^{\top}$. We will call these links the *fully external links of step i*.
- $\delta E_i^{\perp}$ is the set of links in δE_i between nodes which were already present before step i, *i.e.* $\delta E_i^{\perp} = \delta E_i \cap \delta V_i^{\perp} \times \delta V_i^{\perp}$. We will call these links the *internal links of step i*.
- $\delta E_i^{\overset{\top}{\perp}}$ is the set of links in δE_i between a node which appears at step i and a node which was already present, *i.e.* $\delta E_i^{\overset{\top}{\perp}} = \delta E_i \cap \delta V_i^{\perp} \times \delta V_i^{\top}$. We will call these links the *external links of step i*.

Our aim is to introduce a general formalism to describe real-world growth of complex networks. It is therefore essential to use real-world cases to illustrate and evaluate the relevance of our approach and of the parameters we will introduce. We will therefore use two large growing complex networks all along the paper: the *actor* networks, where the nodes are film actors and where a link exists between two actors if they acted together in a film, [Dat], and the *IP exchange* networks, in which the nodes are computers on the Internet and a link exists between two computers if the they echange a packet. For more details, see [Arc].

Table 1 gives the number of nodes and links of these networks, their time grain, as well as the basic statistics for these networks: density, average distance, clustering coefficient. For a definition of these statistics see for instance [GL05]. We have not represented here the degree distributions of these networks. They are, not surprisingly, well approximated by power-laws [GL05].

3 Analysis of the dynamics

The first quantity to consider when analysing growing networks is the evolution of the number of nodes and links of each type, presented in Figure 1. Notice that this already allows us to spot different behaviours in the networks we consider.

The statistics we will now present are based on the degrees of the nodes, and their evolution. Since the degree distributions of most real-world networks are heterogeneous, the mean degree is not representative of the network: most nodes' degree is smaller than the mean degree, and some node have a much larger degree.

We have therefore considered separately these two types of nodes. There is a large number of small degree nodes, and they form the bulk of the network. Therefore it makes sense to consider them as a whole, a natural quantity to study being then the overall fraction of small-degree nodes in the network. Figure 2

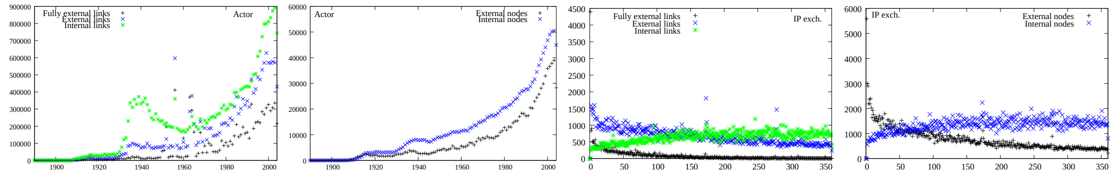

Figure 1: Time evolution of the number of fully-external, external and internal links and external and internal nodes for *actor* and *IP exch.* networks

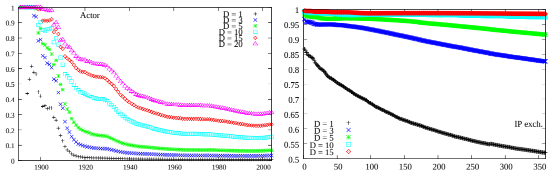

Figure 2: Time evolution of the number of small degree nodes for *actor* and *IP exch.* networks. The quantity plotted is the fraction of nodes with degree $\leq k$, for different values of k.

presents the time-evolution of the fraction of nodes of degree $\leq k$, for different values of k. We observe a difference between our two graphs: for the *actor* network, these values for the smallest values of k, seem to converge to a constant value. This is not the case for the *IP exch.* network.

High-degree nodes are the exceptional elements of a network, and therefore cannot be considered to form a homogeneous group. We have therefore considered separately the ten nodes that have the highest degree at the end of the network evolution. Once these nodes are identified, we can study the time evolution of their degree throughout the network's evolution. Figure 3 plots the time-evolution of a representative subset of these nodes, compared to the time-evolution of the maximal degree of the network (*i.e.* the degree of the highest degree node of the network at each time step).

The first conclusion is that, for each of the networks under study, the highest-degree nodes have a very *similar behaviour*. The degree of these nodes grows with an S-shape for the *actor* network, while the growth is linear for the *IP exch.* network. This seems to be a general characteristic of growing networks (confirmed by experiments not reproduced here). From the observation of the highest-degree nodes, we can also gather general information on the network: in the *actor* network, the time at which most of these nodes appear is almost the same, somewhere close to 1930. This is very different frow a case where the nodes degrees would grow with a similar shape, but at different times throughout the network's history, and therefore indicates a global property of the network.

Second, we observe two different regimes when we compare the behaviour of the highest-degree nodes to the time evolution of the maximal degree of the network. For the *actor* network, the maximal degree is mainly carried by three different nodes: the first one apprears around 1910, and is not among the ten highest degree nodes. The maximal degree is then carried, beween approximatively 1935 and 1950, by the second-highest degree node, and by the highest degree node afterwards. This is very different from what happens for the *IP exch.* network, where the maximal degree is carried by the same node throughout the

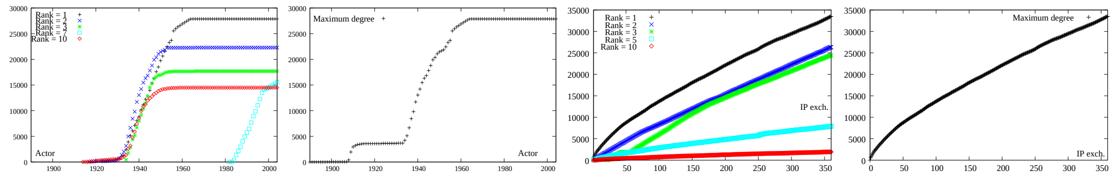

Figure 3: Time behaviour of a few high degree nodes for *actor* and *IP exch.* networks, as well as the evolution of the maximum degree for these networks. Rank is the rank of the node, based on his degree, in the final network.

evolution of the network.

The statistics we have introduced rely on the heterogeneous nature of the degree distribution of real-world networks, and seem to capture both properties common to all growing networks, and some specificities of the networks under study. They are therefore relevent for the analysis of networks' dynamics.

4 Conclusion

In this paper, we have undertaken the study of the dynamics of networks, restricting ourselves to the case of growing networks. We have introduced a formal framework for this study, then we have introduced some statistics to study the dynamics of networks, mainly concerning the degrees of nodes and their evolution. To illustrate the relevance of our statistics, we have measured them on two growing networks, without trying to give an interpretation of the observed behaviours.

The statistics we have introduced are relevent in the sense that they capture different behaviours for the networks we have studied. Another way to show that a statistic is relevent would be to compare our measurements with what happens for randomly growing networks: if the behaviour of a given network is similar to that of a random network, then the statistic does not capture something specific to this network. If, on the other hand, the two behaviours are different, then the satistics are relevent. We fully expect that this is the case for the statistics we have introduced, but this comparison has yet to be done.

Our tools, though quite general, do not address all simple characteristics of growing networks dynamics. There are in particular two simple, natural question that are not answered here. The first one concerns the degrees of the nodes between which new links appear: do these links tend to appear attached to high degree nodes, or independently of the nodes' degrees? The second question concerns the distance in the network between nodes joined by new links: are such two nodes closer in general than two randomly chosen nodes? These two questions should be addressed in order to yield tools for answering these questions.

Finally, our tools can be easily adapted to deal with the case of general dynamic networks. They do not, however, fully address the questions natural in this context, and new tools might be introduced to answer these questions.

Acknowledgments. This work was partly funded by the GAP (Graphes, Algorithmes et Probabilités) project, and by the PERSI (Programme d'Étude des Réseaux Sociaux de l'Internet, `http://www.liafa.jussieu.fr/ persi`) project.

References

[Arc] The Internet Traffic Archive. http://ita.ee.lbl.gov/.

[Bar02] A.-L. Barabási. *Linked: The New Science of Networks*. Perseus Publishing, 2002.

[BS03] S. Bornholdt and H.G. Schuster, editors. *Handbook of Graphs and Networks*. Wiley-VCH, 2003.

[Dat] The Internet Movie Database. http://www.imdb.com/.

[EL03] Eli Eisenberg and Erez Y. Levanon. Preferential attachment in the protein network evolution. *Physical review Letter*, 91(13), September 2003.

[GL05] J.-L. Guillaume and M. Latapy. Relevance of massively distributed explorations of the internet topology: Simulation results. In *Proceedings of the IEEE 24-th INFOCOM'05 conference*, 2005.

[Hol03] Petter Holme. Networks dynamics of ongoing social relationships. *Europhysics Letters*, 64(3), January 2003.

[KMCB95] David L. Banks Kathleen M. Carley, Ashish Sanil and Dean Behrens. Testing alternative models of network evolution. 1995.

[New01] M.E.J. Newman. Clustering and preferential attachment in growing networks. *Phys. Rev. E*, 64, 2001.

[OdC03] Roberto N. Onody and Paulo A. de Castro. Complex network study of brazilian soccer players. *Physical Review E*, 70, 2003.

[PHL04] Chistofer R. Edling Petter Holme and Fredrik Liljeros. Structure and time-evolution of an internet dating community. *Socical Networks*, 26(2), May 2004.

[Sni01] Tom A.B. Snijders. Statistical evaluation of social networks dynamics. *Social Methodology, Pp 361-395*, 2001.