Evolving networks

Pierre BORGNAT [a] Eric FLEURY [b] Jean-Loup GUILLAUME [c,1]
Clémence MAGNIEN [c] Céline ROBARDET [d] Antoine SCHERRER [a]
[a] *Université de Lyon, ENS Lyon, Laboratoire de Physique (UMR 5672 CNRS)*
[b] *Université de Lyon, ENS Lyon, INRIA/ARES*
[c] *Université Pierre & Marie Curie, LIP6 (UMR 7606 CNRS)*
[d] *Université de Lyon, INSA Lyon, LIRIS (UMR 5205 CNRS)*

Abstract. Most real networks often evolve through time: changes of topology can occur if some nodes and/or edges appear and/or disappear, and the types or weights of nodes and edges can also change even if the topology stays static. Mobile devices with wireless capabilities (mobile phones, laptops, etc.) are a typical example of evolving networks where nodes or users are spread in the environment and connections between users can only occur if they are near each other. This who-is-near-whom network evolves every time users move and communication services (such as the spread of any information) will deeply rely on the mobility and on the characteristics of the underlying network. This paper presents some recent results concerning the characterization of the dynamics of complex networks through three different angles: evolution of some parameters on snapshots of the network, parameters describing the evolution itself, and intermediate approaches consisting in the study of specific phenomena or users of interest through time.

Keywords. Complex networks, evolving networks, social networks.

Introduction

Complex networks play an important role in several scientific contexts: computer science, social and interaction networks or epidemiology. Typical examples of such networks are the Internet, web graphs, E-mail, phone calls, P2P networks, etc. In these networks, *links* between entities generally represent some kind of interaction. Studied as a whole, these networks share some non trivial properties and some problems span over a large variety of networks. For instance, the spreading of information is studied in computer science but also in epidemiology and the detection of dense subnetworks (communities) is also a problem having strong implications in many domains. Last decade, this domain has proposed a large set of tools which can be used on any complex network to get a deep insight on its properties and to compare it to other networks (see for instance [1] for a review of parameters).

However, one fundamental property has until recently be understudied. Complex networks evolve: new nodes and edges appear while some old ones disappear. These evolutions are often playing a key role in all the scientific domains cited above: people get new acquaintances, web pages are created or modified on a daily basis, machines are

[1]Corresponding Author E-mail: jean-loup.guillaume@lip6.fr

added or removed on the Internet, etc. If some studies are dedicated to the dynamics of complex networks [2,3,4] they are still too few. It appears crucial to better understand the evolution of these networks first to get knowledge but also to be able to generate random evolutive networks which can be used for simulation purposes.

In this paper, we detail three distinct approaches which are currently used to study complex networks and we explicit these approaches using typical complex networks. First, it is possible to describe the evolution of a network as a sequence of static networks and since there exist many parameters to describe accurately a static network, one can study the evolution of the network through the evolution of these parameters (Sec. 1). Second, one can study the evolution itself and define parameters to capture it, such as the rate of appearance or disappearance of nodes and edges (Sec. 2). Third, an intermediate approach can be used which consists in studying specific phenomena or users of interest through time (Sec. 3). For all these approaches methods from graph theory, statistical physics, data mining and random processes can be used and in many cases new tools and parameters have to be introduced.

1. Evolution of static properties

The most natural way to describe the dynamics of a complex network is to study the evolution of static properties through time. Static networks have been widely studied and a lot a simple parameters are available to describe a network as a whole (number of nodes and edges, number of triangles, specific subgraphs, length of paths, connected components, etc.) or to describe specific nodes (number of neighbors, number of edges between the neighbors, clustering coefficient [5], etc.).

Therefore it is possible to consider an evolving network as a time sequence G_t of networks (snapshots) and to study each of these independently. This yields for each parameter a time series which can be studied using signal processing notions (see Fig. 1). Properties such as the mean, standard deviation and other statistical properties can be computed on these time series.

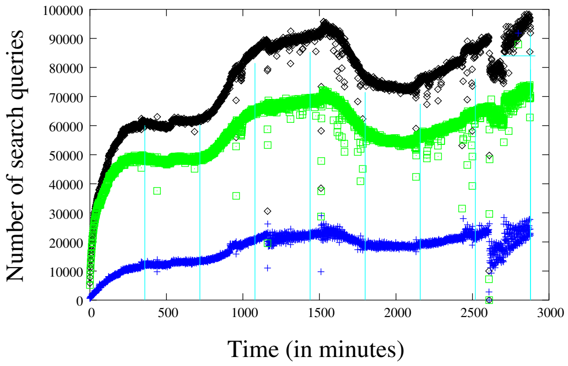

Figure 1. Number of queries per minute during 50 hours on a small size P2P Edonkey server. The three curves correspond to different types of queries. Day/night effects can be observed as well as the start of the measurement during which a lot of new peers connect which yields an increasing number of queries.

A more complex property is the autocorrelation function for a quantity X: $C_X(\tau) =< X(t+\tau)X(t) >_t - (< X(t) >_t)^2$, where $< \cdot >_t$ is the mean over time.

From this, we can extract a *correlation time* [6], defined as the first time were the function $C_X(\tau)$ equals zero (it always happens due to the summation rule of empirical C_X). The correlation time quantifies the "memory" of the property: the longer it is, the greater are the persistence of fluctuations in the data.

Note that very often data are not given as a sequence of snapshots but rather as a sequence of events: email networks for instance are defined by a set of triples (from,to,date), the date being the moment when the mail was sent. If one considers a snapshot every second, since it is likely that no two events (two emails sent exactly at the same time) happen simultaneously then the observed networks are very small. On the contrary, if the aggregation is done on a larger scale (every minute, hour or day) more events are to be observed on each snapshot but the temporal order of these events is going to be lost in each snapshot (mail replies or forwards for instance). In complex networks, different time scales can be used depending on the parameter or phenomenon observed. For instance when considering a typical P2P system, one can study the instantaneous throughput (one second or less), connection duration of a peer (minutes to days), download duration of a file (minutes to weeks) or even the duration during which a file is available on the network (up to years), etc.

2. Definition of dynamic properties

Considering the evolution as a sequence of snapshots is an efficient and simple approach in many cases but some properties cannot be directly observed in this framework. For instance it is natural to look at the duration of contacts or non-contacts between individuals in a network [7] or to study the evolution of communities. For these simple examples, one has to consider the evolution of the network from one time step to the next, or the whole evolution. Hereafter we detail the case of the evolution of communities in a network.

Communities are defined as dense subgraphs with few edges between them and can be found in many complex networks. The identification of such subgraphs is important in many contexts since such communities can correspond to groups of friends or people with similar interests, web pages with a similar content, etc. Moreover studies show that information (rumors for instance) spread more rapidly inside communities than between communities.

Many algorithms are available to find communities automatically on graphs, however theses methods are often time expensive and very sensitive to small modifications of the topology: the addition of one edge can have strong implications on the global community structure. Therefore it is likely that applying these methods will produce completely different decompositions for each snapshot. One approach has been used in [4] using a non classical definition of communities which allows to follow the evolution of community using a simple set of rules (birth, death, merge, split, growth and contraction). Similar ideas are presented in [8] by the identification of dense subgraphs in each snapshot, the subgraphs begin merged afterwards. In [9], the authors present an approach not specifically dedicated to the identification of communities but to the clustering problem in general which allows to cluster data in a timely fashion while keeping a good clustering and no strong variations from one snapshot to the next. Approaches using tools from data-mining are also available, which allow to compute dense sets of nodes with many interactions for a long period of time [10,11] (see Fig. 2).

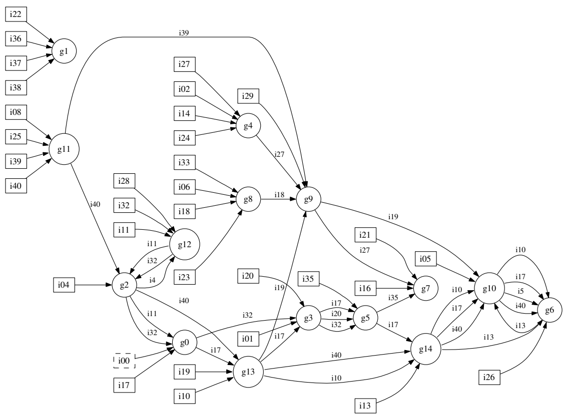

Figure 2. Time ordered trajectories of individual (square) in groups (circles) in a contact network. Groups are dense connected subgraphs which appear frequently in the evolving network.

The results obtained using any community detection algorithm for evolving network give some information on the communities (lifetime, rate of apparition and disappearance, probability of merging and splitting, etc.) [11]. The study of the evolution of the network can therefore be done at a different scale which is not local and not global.

3. Study of specific users and phenomenon

Another approach which can be used to study evolving networks is to study specific nodes or groups of nodes of particular interest (for instance communities). In many contexts algorithms and protocols are designed for average users and it is important to know the number of users who are significantly deviant from this average and how they behave in order to optimize protocols. For instance in most P2P systems, the load for a peer is somehow proportional to the number of files shared, users which share many files can therefore become bottlenecks if they are queried too often. In Fig. 3 we show results obtained on a typical P2P network when trying to identify the users who share many files. These users are more likely to be queried very often by other peers.

3.1. Transmission of information

A typical phenomenon on complex networks is the diffusion such as for viruses in epidemiology, routing in computer networks, innovation, etc. If recent studies have taken into account the real dynamics [12,13], for most of them the process of transmission is living on a static graph. In the static case the main issues are to find networks parameters which explain the persistence of viruses within a given graph. It has been shown for instance that there is a strong relation between the largest eigenvalue of the adjacency matrix of the network and the epidemic threshold [14].

Dynamics are also central in new communication services which are relying on mobile users spread in the environment. The routing of information in such a context depends on the connectivity between nodes and the mobility of these nodes. Understanding the characteristics of these networks (called Delay Tolerant Networks) is therefore cru-

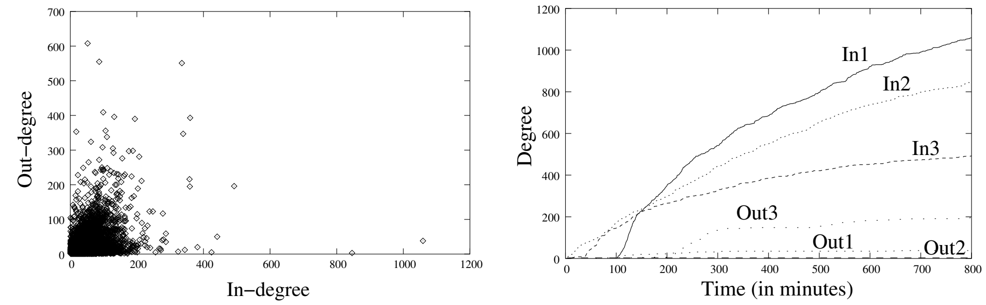

Figure 3. Left: joint distribution of the number of files offered by a peer (in-degree) versus the number of files he looks for (out-degree) in a P2P system. Peers offering many files and peers who do not offer any file but look for many (free-riders) can be easily identified and further studied. Right: evolution of the number of files requested for the three peers offering the more files (the three rightmost on the left figure). After 13 hours some of their files still have not been requested.

cial to propose protocols suitable for this context. The simplest ways to transmit data in this context are the opportunistic forwarding algorithms [7]: when a node needs to send some data to a destination, it uses its contacts to relay the data to the destination. Two naive algorithms belong to this class, the first one consists in waiting to be connected to the destination to send the data directly, the second consists in forwarding the data to all the neighbors which in turn are going to pass it to their neighbors. This is going to flood the network with the data which eventually should arrive at the destination (see Fig. 4).

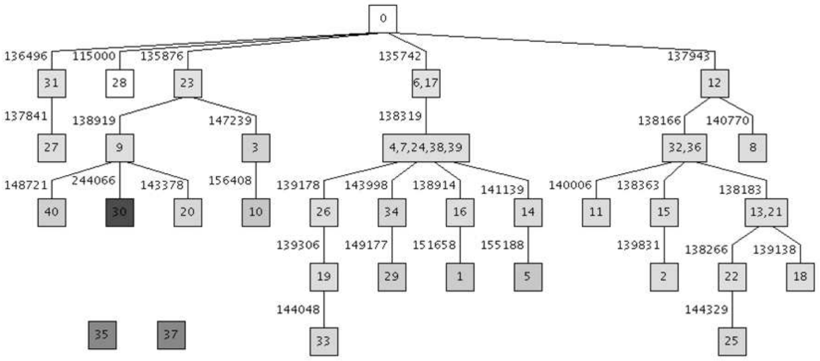

Figure 4. Naive instantaneous flooding in a mobile network, where node 0 is trying to send a message to all other nodes at time 115000. Time on the edges represents the earliest time for the message to be transmitted to the group of nodes below, *i*.e. the time when an edge is created between a node on top and a node at the bottom. Note that two users (35 and 37) cannot be reached. The data used for the flooding simulation are described in [11]

4. Conclusion

We presented here three different approaches which can be used to study an evolving network:

- the evolving network can be considered as a sequence of snapshots and each of these snapshots can be studied as a static network;

- properties can be defined on the evolution itself, for instance the duration of contacts in a network or the evolution of communities through time ;
- finally, specific users or phenomena can be studied, the more obvious being the diffusion of information in an evolving network.

Many studies have been focused on static networks, therefore the first approach is the more developed, however the definition of proper time scales is a unsolved problem and there is no warranty that such time scales can be defined in an automatic way given an evolving network. For both other approaches, much work has to be done in order to define new relevant parameters to describe the evolution as precisely as possible.

Finally, using all the parameters obtained with the previous approaches would allow the introduction of evolutionary models for dynamic complex networks. Such models could be used in formal contexts and for simulation purposes. Defining random models is not an easy task and even for static networks some simple parameters cannot be captured in a satisfactory way. A few models have already been introduced (see for instance [15]) which are modifying a given network by the addition of nodes and edges, however the aim is in general not to generate an evolving network but to eventually obtain a network with a given set of static properties. Much therefore remains to be done in this direction.

References

[1] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics 74, 47*, 2002.

[2] S.N. Dorogovtsev and J.F.F. Mendes. *Handbook of Graphs and Networks: From the Genome to the Internet*, chapter Accelerated growth of networks. Wiley-VCH, 2002.

[3] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: Densification laws, shrinking diameters and possible explanations. In *ACM SIGKDD*, 2005.

[4] G. Palla, A. Barabasi, and T. Vicsek. Quantifying social group evolution. *Nature*, 446:664–667, April 2007.

[5] J.D. Watts and S.H. Strogatz. Collective dynamics of'small-world'networks. *Nature*, 393(6684):409–10, 1998.

[6] H. Abarbanel. *Analysis of Observed Chaotic Data*. Springer, 1996.

[7] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of human mobility on the design of opportunistic forwarding algorithms. In *INFOCOM*, 2006.

[8] Y. Chi, S. Zhu, X. Song, J. Tatemura, and B.L. Tseng. Structural and temporal analysis of the blogosphere through community factorization. In *KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 163–172. ACM Press, 2007.

[9] D. Chakrabarti, R. Kumar, and A. Tomkins. Evolutionary clustering. In *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 554–560. ACM, 2006.

[10] J. Besson, C. Robardet, J.-F. Boulicaut, and S. Rome. Constraint-based concept mining and its application to microarray data analysis, 2005. Intelligent Data Analysis 9(1):59-82, 2005.

[11] E. Fleury, J.-L. Guillaume, C. Robardet, and A. Scherrer. Analysis of dynamic sensor networks: Power law then what? In *Second International Conference on COMmunication Systems softWAre and middlewaRE (COMSWARE 2007)*, Bangalore, India, 2007. IEEE.

[12] J. Leskovec, L.A. Adamic, and B.A. Huberman. The dynamics of viral marketing. In *EC '06: Proceedings of the 7th ACM conference on Electronic commerce*, pages 228–237. ACM Press, 2006.

[13] J. Leskovec, M. McGlohon, C. Faloutsos, N. Glance, and M. Hurst. Cascading behavior in large blog graphs, 2007.

[14] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos. Epidemic spreading in real networks: An eigenvalue viewpoint. In *22nd Symposium on Reliable Distributed Computing*, 2003.

[15] R. Albert, H. Jeong, and A.L. Barabasi. The diameter of the world wide web. *Nature*, 401:130–131, 1999.