Physica A 482 (2017) 375–391

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Exploring the evolution of node neighborhoods in Dynamic Networks

Günce Keziban Orman [a], Vincent Labatut [b], Ahmet Teoman Naskali [a]

[a] Computer Engineering Department, Galatasaray University, Ortaköy, Istanbul, Turkey

[b] Laboratoire Informatique d'Avignon, Université d'Avignon, Avignon, France

ARTICLE INFO

Article history:
Received 8 September 2016
Received in revised form 3 April 2017
Available online 28 April 2017

Keywords:
Dynamic networks
Network evolution
Network topology
Neighborhood events

ABSTRACT

Dynamic Networks are a popular way of modeling and studying the behavior of evolving systems. However, their analysis constitutes a relatively recent subfield of Network Science, and the number of available tools is consequently much smaller than for static networks. In this work, we propose a method specifically designed to take advantage of the longitudinal nature of dynamic networks. It characterizes each individual node by studying the evolution of its direct neighborhood, based on the assumption that the way this neighborhood changes reflects the role and position of the node in the whole network. For this purpose, we define the concept of *neighborhood event*, which corresponds to the various transformations such groups of nodes can undergo, and describe an algorithm for detecting such events. We demonstrate the interest of our method on three real-world networks: DBLP, LastFM and Enron. We apply frequent pattern mining to extract meaningful information from temporal sequences of neighborhood events. This results in the identification of behavioral trends emerging in the whole network, as well as the individual characterization of specific nodes. We also perform a cluster analysis, which reveals that, in all three networks, one can distinguish two types of nodes exhibiting different behaviors: a very small group of active nodes, whose neighborhood undergo diverse and frequent events, and a very large group of stable nodes.

1. Introduction

Dynamic Network Analysis is a subfield of Network Science aiming at representing and studying the behavior of systems constituted of interacting or related objects evolving through time. This domain gained attention lately, as attested by the recent publication of several surveys [1–5].

Authors generally distinguish two types of such dynamic systems. In the first, interactions or relationships are very short, or even punctual, and occur frequently. They are called *contact sequences* [4] and *streaming networks* [5] in the literature (among other names). In the second type, called *interval graphs* [4] and *slowly evolving networks* [5], interactions or relationships have a significant duration and are not so frequent. Dynamic network modeling can be applied to both types, although it is more suited to the second. It consists in representing the system evolution through a sequence of graphs. Each one of these graphs, which we call *time slices* in this article, represents the aggregation of the system changes over a given period of time. This representation allows studying the system properties with sequence-based tools such as time series analysis and sequential pattern mining.

E-mail addresses: korman@gsu.edu.tr (G.K. Orman), vincent.labatut@univ-avignon.fr (V. Labatut), tnaskali@gsu.edu.tr (A.T. Naskali).

http://dx.doi.org/10.1016/j.physa.2017.04.084
0378-4371/

According to Aggarwal & Subbian, one can distinguish two types of methods to analyze dynamic networks [5]. On the one hand, so-called *maintenance methods* consist in performing an initial analysis on the first time slice (e.g. detect the community structure), and then updating its output for each subsequent time slice (e.g. move certain nodes from one community to another). This is typically done with methods originally developed for static networks and later adapted to dynamic ones. On the other hand, *evolution methods* embrace the longitudinal nature of the data and focus on describing the changes caused by temporal evolution. They are useful to identify and understand the rules governing the evolution of the studied network, and allow the design of new models. Note that it is possible to simultaneously belong to both types of methods. Another important feature is the granularity used to perform the analysis. First, *macroscopic* methods deal with the whole network at once. They are the most widespread in the literature, e.g. network density, network diameter [6]. Second, *mesoscopic* methods consider the network at an intermediary level, generally that of the community structure, e.g. modularity measure [7], size of the communities [8]. Finally, *microscopic* methods focus on individual nodes and possibly their direct neighborhood, e.g. clustering coefficient [9].

By definition, methods specifically designed to handle a dynamic network are more likely to take advantage of the specificity of such data. Regarding the granularity, if macroscopic and mesoscopic methods are widespread in the literature, it is not the case for microscopic methods. Yet, the benefits of such approaches are numerous: by allowing the tracking of finer evolution processes, they complement macroscopic and/or mesoscopic results. They help identifying behavioral trends among the network nodes, and consequently outliers. These results can facilitate the description and understanding of processes observed at a higher level, and can be useful to define models of the studied system, especially agent-based ones.

In this work, we propose a method specifically designed to study dynamic slowly evolving networks, at the microscopic level. We characterize a node by the evolution of its neighborhood. More precisely, we detect specific events occurring among the groups of nodes constituting this neighborhood, between each pair of consecutive time slices, which we call *neighborhood evolution events*. This method constitutes our main contribution. To the best of our knowledge, it is the first attempt at describing the dynamics of a network based on such local events. For each node, our method outputs a sequence of categorical features corresponding to the different types of events it experienced. These features can then be analyzed with any tool capable of processing temporal categorical data, in order to extract meaningful information regarding the network dynamics. This knowledge can noticeably be used to categorize nodes, or identify trends and outliers. For illustration purposes, and as a second contribution, we analyze three real-world networks: DBLP (scientific collaborations), LastFM (social relations through musical tastes) and Enron (email exchanges). We first apply our event identification method, before using *Sequential Pattern Mining* and some complementary analysis to extract higher level information.

The rest of this article is organized as follows. In Section 2, we review in further detail the existing work the most related to our method. In Section 3, we formally define the concept of *neighborhood evolution event*, and describe the method we use to detect such events in dynamic networks. In Section 4, we apply our method to three real-world networks and discuss the obtained results. Finally, in Section 5, we comment on the limitations of our work, how they could be overcome and how our method could be extended.

2. Related work

This section does not aim at being exhaustive, but rather at presenting the various general families of methods existing to characterize dynamic graphs. A number of reviews exist which describe them in further detail, and list them in a more exhaustive way [4,10,5].

As mentioned in the introduction, the most straightforward way of characterizing a dynamic network is to use measures designed for static networks, by applying them to each time slice and considering the resulting time series. A number of studies adopted this approach for a variety of measures, at different granularities. To cite a few, at the macroscopic level: number of links [11], size of the giant component [12], diameter [6,12], principal Eigenvalue of the adjacency matrix [11]; at the mesoscopic level: modularity [7], size of the communities [8], number of communities [13], numbers of motifs (predefined small subgraphs) [14]; and at the microscopic level: degree [14,6,9], clustering coefficient [9].

The next step is to apply tools originally designed to process static networks at each time slice, like before, but with some additional updating mechanism allowing to smooth the results. This is particularly common in the domain of community detection, like for instance in [13]. A related incremental approach, consisting in performing an initial computation on the first time slice and updating it at each additional time slice, is also applied to the processing of centrality measures, such as the *incremental closeness* [15]. In contrast, more recent tools go farther in the adaptation of these static methods to the analysis of dynamic networks and better describe or detect the changes caused by temporal evolution. For instance, in [16], Gupta et al. describe a method allowing to detect the most significant changes in the distance between nodes, assuming such changes correspond to important events in the evolution of the graph.

With their method based on the identification of *role dynamics*, Rossi et al. use several such measures at once [17]. They consider each one of them as a feature characterizing the topological position of a node in the graph. They apply non-negative matrix factorization to identify a small number of node roles, corresponding to typical evolutions of the features for the considered period of time. Each role is characterized by its dependence on certain features, and each node is more or less related to certain roles. The prevalence of a role can change with time, which can be used to characterize the network