Physica A 420 (2015) 98–107

Contents lists available at ScienceDirect

Physica A

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

Attack tolerance of correlated time-varying social networks with well-defined communities

Souvik Sur [a,c,*], Niloy Ganguly [b,c], Animesh Mukherjee [b,c]

[a] G. S. Sanyal School of Telecommunications, Kharagpur, 721302, India
[b] Department of Computer Science and Engineering, Kharagpur, 721302, India
[c] Indian Institute of Technology, Kharagpur, 721302, India

HIGHLIGHTS

- We identify the existence of short time correlation in temporal networks.
- The community based attack affects time correlated real-world networks most severely.
- We introduce a novel metric edge emergence factor to quantify short-time correlation.

ARTICLE INFO

Article history:
Received 28 October 2013
Received in revised form 18 August 2014
Available online 4 November 2014

Keywords:
Time-varying networks
Efficiency
Robustness
Attack tolerance
Community analysis

ABSTRACT

In this paper, we investigate the *efficiency* and the *robustness* of information transmission for real-world social networks, modeled as time-varying instances, under targeted attack in shorter time spans. We observe that these quantities are markedly higher than that of the randomized versions of the considered networks. An important factor that drives this efficiency or robustness is the presence of short-time correlations across the network instances which we quantify by a novel metric the—*edge emergence factor*, denoted as ξ. We find that standard targeted attacks are not effective in collapsing this network structure. Remarkably, if the hourly community structures of the temporal network instances are attacked with the largest size community attacked first, the second largest next and so on, the network soon collapses. This behavior, we show is an outcome of the fact that the *edge emergence factor* bears a strong positive correlation with the size ordered community structures.

1. Introduction

The seminal work [1] by Barabási et al. introduced the concept of error and attack tolerance of complex networks. They showed that scale-free networks are vulnerable to targeted node degree based attack due to the inherent inhomogeneity of the degree distribution whereas exponential networks are resilient from such attacks. The extent of vulnerability was measured in terms of change (before and after attack) in diameter, size of the individual clusters and average cluster size of the network. Following this, in later years, the resilience of static networks has been considerably investigated in Refs. [2–4]. However, since most of the real-world networks are time-varying [5] in nature, the attack strategies as well as the measurement tools effective for static networks may not be actually suitable to quantify the robustness of such networks, specially using epidemic dynamics [6].

* Corresponding author at: G. S. Sanyal School of Telecommunications, Kharagpur, 721302, India.
E-mail addresses: souviksur@gssst.iitkgp.ernet.in (S. Sur), niloy@cse.iitkgp.ernet.in (N. Ganguly), animeshm@cse.iitkgp.ernet.in (A. Mukherjee).

http://dx.doi.org/10.1016/j.physa.2014.08.074
0378-4371/

In order to better quantify the resilience of time-varying networks, the concept of temporal robustness has been introduced in Ref. [7], to measure the degree of tolerance against random failure in the network and has been subsequently used for targeted attack [8]. To effectively find the influential nodes in a time-varying network, researchers [8–10] have divided the network in two parts by considering the temporal network as a sequence of static networks taken at suitable time resolution [11]. They then identified important nodes from the initial data which can be used to launch attack on the remaining network. For example, [8] estimates important nodes using different metrics—average node degree, temporal closeness and number of node contacts-updates from the first half (50%) of the data and subsequently launch an attack by progressively removing those 'important' nodes. Considering such a framework, they study the effects of attacks on several real world networks (INFOCOM 2006 mobility trace and San Francisco Cab spotting data); however, since they consider two halves of the data they fail to observe any difference in effect between the real-world networks and random graphs.

On closer inspection, from the perspective of an attacker, it seems infeasible to split up the entire window of a sufficiently large network into two equal halves (or into 75%–25% as in Ref. [9]). This is because, to study the dynamics of a temporal network, one needs to observe it at the correct level of granularity. Therefore, for a sufficiently large network, one may investigate it at shorter time slices and study how the dynamics changes over the consecutive time slices. In other words, a more rational and practical way to launch an attack would be to do it in almost real time possibly by collecting evidences from the network instances within a shorter time-window (an hour) and attack the network structure in the following time-window (next hour) based on the evidence collected. For the purpose of our investigation, we choose the window of an hour since we observe that it is a representative choice among different others (e.g., 15 min, 30 min, 2 h and 6 h) for the dataset we considered. Intuitively, the scheme in [8], (a) is a less meaningful attack for dynamic networks and (b) leaves the shorter-time correlations in the network completely unseen. Note that by correlation we denote structural correlation, i.e., the existence of part of the network that recur over consecutive time-points.

We observe that there are certain attack strategies which work well for the network samples considered in Ref. [8], fail completely here. We see, in this shorter time-window, unlike [8], the temporal efficiency [7] of a real-world temporal network is significantly higher than that of its random counterpart. We find that if these networks are attacked based on the underlying hourly community structures with those nodes that appear in the largest size community targeted first for removal, then the attack seems to be successful in gradually collapsing the network thus allowing us to conclude that identification of the community structures as the target of the attack almost surely collapses the network even in the shorter-time window.

The different results obtained can be explained by considering the degree of time correlation present in subsequent (training and testing) networks. We quantify these shorter-time correlations in terms of a new metric called the *edge emergence factor* (ξ) that precisely computes how many edges branch out at some time instance from the end points of a single edge that existed in the immediate previous time instance. As we shall see, the ξ of a real time-varying social graph is significantly higher than uncorrelated random graphs. Note that the ξ is a manifestation of the dynamics of information spread that takes place through the social contacts in a temporal graph. We find that the key reason for the higher robustness in empirical networks arises from the fact that the ξ bears a strong correlation with the size of the hourly communities.

The paper is structured as follows. In Section 2, we describe the data that we have investigated, temporal network modeling of the data and different attack strategies. In Section 3, we present the results obtained from our investigation and attempt to connect the structure and the function of real-world temporal networks. Finally, we summarize our contributions in Section 4.

2. Materials and methods

2.1. Data

For the purpose of our investigation of robustness of time-varying networks, we consider three specific real-world face-to-face contact datasets and present our results for each of them.

2.1.1. Real-World networks

A detailed description of the datasets on which we conduct our experiments is as follows:

1. HYPERTEXT, 2009 ($HT^{09}_{Original}$): These data corresponds to face-to-face interactions of 113 attendees of ACM Hypertext 2009 conference held for 2.5 days between June 29th and July 1st, 2009 [12]. For data collection, active RFID devices were used to detect and record face-to-face proximity relations of persons wearing the RFID badges. These devices can detect face-to-face proximity (1–1.5 m) of another device with a temporal resolution of $\tau = 20$ s. Thus in a single hour there can be a maximum of $n = 3600/\tau = 180$ network snapshots.
2. INFOCOM, 2005 ($INF^{05}_{Original}$): The data were collected over 4 days at the IEEE INFOCOM 2005 conference [13]. Participants in the experiment were 50 students and researchers, equipped with mobile communication devices (i-Motes). The time resolution τ was again assumed to be 20 s. A link has been constructed at a certain time, if the two nodes were within the communication range.
3. INFOCOM, 2006 ($INF^{06}_{Original}$): These data were collected over 5 days at the IEEE INFOCOM 2006 conference in Barcelona [14]. In this case, number of participants were 78 students and researchers, equipped with i-Motes and an additional 20 stationary i-Motes were deployed as location anchors. The value of $\tau = 20$ s is also used here.