Physica A 436 (2015) 272–285

Contents lists available at ScienceDirect

Physica A

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

Network topology inference from infection statistics

Igor Tomovski [a,*], Ljupčo Kocarev [a,b]

[a] Macedonian Academy of Sciences and Arts, bul. Krste Misirkov 2, 1000 Skopje, Macedonia

[b] Faculty of Computer Science and Engineering, University "Sv Kiril i Metodij", Skopje, Macedonia

HIGHLIGHTS

- Infectious SIS process may be used to determine the network topology.
- We derive mathematical relation between link weights and infection statistics.
- We test the result on both weighted and unweighted networks.

ARTICLE INFO

Article history:
Received 29 May 2014
Received in revised form 18 February 2015
Available online 22 April 2015

Keywords:
Network topology
SIS process

ABSTRACT

We introduce a mathematical framework for identification of network topology, based on data collected from infectious SIS process occurring on a network. An exact expression for the weight of each network link (existing or not) as a function of infectious statistics, is obtained. An algorithm for proper implementation of the analyzed concept is suggested and the validity of the obtained result is confirmed by numerical simulations performed on a number of synthetic (computer generated) networks.

1. Introduction

One of the key concepts of the modern world is the concept of networking. Networks form the core of almost any contemporary technical architecture (telecommunications, energy distribution, communication infrastructures), they give rise to new forms of social communications (social networks), etc. Recent discoveries in bio-medical sciences suggest that networks, and the specific topologies associated to them, are the basis of the biological life as we know it. Gene, protein and brain networks are just few examples of basic physiological networks that have been extensively studied in the last several decades.

When studying networks and especially dynamical processes occurring on networks, the usual problem definition is formulated as follows: given the network topology and knowing the individual dynamics of each entity (node), determine the collective dynamics arising from that specific topology. Number of different dynamical processes have been addressed using this approach: synchronization [1–3], blackouts and cascading failures [4,5], stochastic spreading processes (in particular virus spreading) [6–20], etc. Though not as frequent, the inverse problem is considered as well: given the individual dynamics of each node and the collective behavior of the networked system, to determine the topology of the network on which certain process occurs. This is a key issue in understanding systems in which, due to specific circumstances, the network architecture may not be determined using strait forward approaches. Good example are gene regulatory networks [21–24]

* Corresponding author.
E-mail address: igor@manu.edu.mk (I. Tomovski).

http://dx.doi.org/10.1016/j.physa.2015.03.090
0378-4371/

and functional brain networks [25–28]. Number of different theoretical dynamical models have been studied as well. Most of them have been focused on systems where both node dynamics and coupling are deterministic [29–34], although recently systems with some form of uncertainty or stochasticity have also been investigated [35–40].

In this article we suggest an answer to the question 'how can network topology be inferred from an infectious process which takes place on the network' by considering a SIS discrete time stochastic process. An exact expression for the weight of each network link, as a function of infectious statistics, is derived without making any approximations. To note, the results obtained for the SIS process may be further extended/generalized to other infectious, or more general to other Markov processes occurring on networks: we discuss this at the end of the article.

In the analysis we consider that the network structure, which describes how the graph is wired, may change in time due to the fact that edges are not continuously active. As an example, in networks of communication via e-mail, text messages, or phone calls, edges represent sequences of instantaneous or practically instantaneous contacts. The spreading of biological viruses depends on temporal patterns of contact, and the flow of information in social networks is influenced by the social activity patterns and rhythms of individuals. This has led to the concept of temporal networks [41–43]; we consider a SIS process occurring on temporal networks.

We would like to note that the results from numerical simulations presented in the paper indicate that practical implementation of the elaborated concept requires huge number of S–I–S (susceptible–infected–susceptible) transitions for proper assessment. This, in practice, yields for extensive observation/estimation period (time). In that sense one may question the practicality of the presented analysis. That however does not diminish the importance of the main result (Eq. (9)), due to several reasons. First, the obtained result is exact and may be considered in the future as a starting point for development of faster assessment algorithms, based on properly derived approximations from Eq. (9). The character of these approximations will be infection-specific and depends on the infection intensity. Next, the result may be further used as a base for development of methods for both technological network design and control. Infectious processes give a good insight into the network information capacity; on the other hand they are the core of dissemination processes occurring on networks. In that sense the result is essential when formulating problems like "to create networks or manipulate existing ones such that desired infection statistics (information flow, dissemination level) may be achieved". Finally, one should consider that technological progress will soon lead to development of ultra-fast networks, followed by dynamical processes occurring on similar time scales. In such conditions, sufficiently large number of transitions will occur, and therefore proper assessment may be performed, in reasonable time intervals.

2. Model description

The SIS model is well known [8–18], however no formal mathematical definition of its raw (fundamental) form may be found in the literature. In this section, we conduct a thrall analysis of the descriptive form of the SIS process, that leads to strict derivation of the governing equation of the model.

To start with, following [41–43], we give a brief definition of a temporal network. Let $G = G(V, E)$ be a bidirectional and connected graph, where $V = \{0, \ldots, N-1\}$ is the set of nodes, $\|V\| = N$, and E is the set of edges or links, $\|E\| = L$. A *temporal network* is a sequence of graphs $G(t) = G(V, E(t)),\ t \in \mathbb{N}$, such that $(i, j) \in E(t)$ if $(i, j) \in E$ and the link between i and j is active in the direction $j \to i$ during $[t, t+1]$. We describe the network with *temporal network matrix* $\mathbf{A}(t) = [a_{ij}(t)]$. Its members $a_{ij}(t)$ are binary random variables, such that $a_{ij}(t) = 1$ if nodes $(i, j) \in E$ and the link is active in the direction from j to i during $[t, t+1]$ and $a_{ij}(t) = 0$ otherwise. For the statistical expectation $E[\mathbf{A}(t)] = \mathbf{W} = [w_{ij}]$ we use the term *network adjacency matrix* and for its members w_{ij} the term *weights*.

Let $s_i(t)$ be a random variable that denotes the infectious status of node i at time t, with $s_i(t) = 1$ if node i is infected, and $s_i(t) = 0$ otherwise. Let $\beta_{ij}(t)$ be a binary random variable that denotes that an infected node j will infect the susceptible node i, providing that the link $j \to i$ is active in the time frame $[t, t+1]$. Similarly, let $\delta_i(t)$ be a binary random variable that indicates whether the process of healing of node i will occur at time frame $[t, t+1]$. For the expectation values $\beta = E[\beta_{ij}(t)]$ and $\delta = E[\delta_i(t)]$ we use the terms *infection probability* and *curing probability*, respectively. Similar or identical terminology for these parameters may be found in the existing literature for the SIS model [8–10,18,11–17].

The analysis presented in the following text is based on several assumptions. To start with, the dynamics of $a_{ij}(t)$ does not depend on either $s_i(t)$ or $s_j(t)$, only on the standard operating procedures or behavioral patterns of nodes i and j, or stochastic influences from aside. In that sense $a_{ij}(t)$ are considered as stochastically independent variables. Next, a note should be taken on variables $\beta_{ij}(t)$. We say that they represent specific factors that at time interval $[t, t+1]$ favor or disfavor both infection transition of a specific agent from infected node j ($s_j(t) = 1$) to a susceptible node i ($s_i(t) = 0$), and its manifestation in node i at time $t+1$ ($s_i(t+1) = 1$), when all other contributing factors ($a_{ij}(t) = 1$) are fulfilled. It is assumed, that in reality, factors influencing $\beta_{ij}(t)$'s are unknown or may not be measured on an individual level. What may be determined is the statistical average β of the statistical assembly $\{\beta_{ij}(t)\},\ i, j \in [0, N-1],\ i \neq j,\ t \in Z^+$. Therefore, in the model, we treat $\beta_{ij}(t)$'s as statistically independent variables, as well. Though not explicitly stated in the existing literature, this argument is implicitly considered as valid. Similar discussion holds for variables $\delta_i(t)$.

Let $s_i^b(t)$, $a_{ij}^b(t)$, $\beta_{ij}^b(t)$ and $\delta_i^b(t)$ be logical (Boolean) '0'-'1' variables, analogues of binary random variables $s_i(t)$, $a_{ij}(t)$, $\beta_{ij}(t)$ and $\delta_i(t)$ respectively. The infection pattern of the SIS model, with no instantaneous curing and re-infection, occurring on a temporal network, may be in detail described as follows: a node i will be infected at time $t+1$ if: