Physica A 392 (2013) 3476–3483

Contents lists available at SciVerse ScienceDirect

Physica A

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

Threshold model of cascades in empirical temporal networks

Fariba Karimi [a], Petter Holme [a,b,c,*]

[a] IceLab, Department of Physics, Umeå University, 90187 Umeå, Sweden
[b] Department of Energy Science, Sungkyunkwan University, Suwon 440-746, Republic of Korea
[c] Department of Sociology, Stockholm University, 10691 Stockholm, Sweden

HIGHLIGHTS

- We study threshold models of social influence on empirical temporal datasets.
- The cascade dynamics of the model is sensitive to the temporal-network structure.
- For most datasets and parameter values, temporal patterns facilitate cascades.

ARTICLE INFO

Article history:
Received 20 October 2012
Received in revised form 26 December 2012
Available online 12 April 2013

Keywords:
Complex networks
Network theory
Threshold models
Temporal networks

ABSTRACT

Threshold models try to explain the consequences of social influence like the spread of fads and opinions. Along with models of epidemics, they constitute a major theoretical framework of social spreading processes. In threshold models on static networks, an individual changes her state if a certain fraction of her neighbors has done the same. When there are strong correlations in the temporal aspects of contact patterns, it is useful to represent the system as a temporal network. In such a system, not only contacts but also the time of the contacts are represented explicitly. In many cases, bursty temporal patterns slow down disease spreading. However, as we will see, this is not a universal truth for threshold models. In this work we propose an extension of Watts's classic threshold model to temporal networks. We do this by assuming that an agent is influenced by contacts which lie a certain time into the past. I.e., the individuals are affected by contacts within a time window. In addition to thresholds in the fraction of contacts, we also investigate the number of contacts within the time window as a basis for influence. To elucidate the model's behavior, we run the model on real and randomized empirical contact datasets.

1. Introduction

An important socio-economic mechanism is social influence—the spread of opinions and beliefs from the social surrounding to an individual. This type of process can be modeled as a threshold process—if the fraction of neighbors in a static network exceeds a threshold, then the focal vertex changes state. In his pioneering work from the 1970's, Mark Granovetter [1] pointed out that social collective behavior could be divided into processes depending on credulity (including threshold models) or vulnerability (like disease-spreading models). A recent, theoretically important development was made in Ref. [2], where Duncan Watts proposed a threshold model that is analytically tractable on networks. The study highlights effect of network structure on cascade sizes. Some other studies have tried to bridge the two classes—compartmental models and threshold models [3]. One of the key findings of Watts was that network structure affects cascade processes [2,4]. Another dimension that has been shown to be important for social spreading phenomena is the timing of

* Corresponding author at: IceLab, Department of Physics, Umeå University, 90187 Umeå, Sweden. Tel.: +46 702298396.
E-mail addresses: petter.holme@physics.umu.se, pttrhlm@gmail.com (P. Holme).

0378-4371/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physa.2013.03.050

contacts [5]. For this reason it makes sense to model social contact patterns as temporal networks [6]. This is a term for network representations where the explicit time of contacts is included. In this paper we extend Watts's cascade model to temporal networks.

Watts's cascade model assumes that an actor can switch between two states. It is a deterministic, non-equilibrium model of a cascade where actors can change to a new state but not back to the previous one. In the original model an actor is continuously influenced by its surrounding. In contrast, in a temporal network, there is no fixed surrounding in time. One has to decide what time into the past an actor can be influenced by its contacts. In our model, we integrate the influence from contacts over a time window. Our model is thus both taking into account the chronological order of events and the period a contact that can influence an actor. Outside the time window, communication does not influence individuals. This is to say that for social influence, communications in the past can be forgotten or become unimportant as time passes by.

We will use empirical temporal-network data as a substrate to run our model on. We compare our results to those of randomized versions of the original data. The results are compared to static and temporal structural measures characterizing the temporal network.

2. Methods

The type of data we consider in this paper are sets of triples (i, j, t), or *contacts*, which means vertices i and j have been in contact at time t. Let V be the set of N vertices, E the set of M edges (pairs of vertices that occur in at least one contact), and let C be the set of all Γ contacts.

We assume a system of vertices with one-to-one communication over edges (pairs of vertices connected at least at one point in time). The interaction is considered bidirectional in the sense that both vertices in contact can influence one another. Each edge has a list of time stamps representing the times of communication events. We assign a binary value to each vertex representing its current state with respect to whatever spreads through the network. The population is initialized in the state 0 except one random vertex that is assigned state 1. We then study the spread of state 1 in the network. We call state-1 vertices *adopters* and state-0 vertices *non-adopters*. The term adopters comes from that early threshold models often sought to capture the spread of new technology [7].

When we simulate the model, we follow the set of contacts in time order and let each contact be an opportunity for the vertices to change state. Vertices are influenced by their contacts within a finite *time window* from time θ in the past to the present. In other words, at time t, a node base its decision to change state, or not, on contacts within the interval

$$[t - \theta, t). \tag{1}$$

Let f_i be the fraction of contacts between i neighbors of state 1 within the time window. In our *fractional-threshold model*, if $f_i \geq \phi$ the vertex i will stay in state 1 for the remainder of the simulation. The reason that the agents remain in state 1 indefinitely is to conform to Watts' model. Watts motivate this choice by it making the model maximally simple (and analytically tractable) that is still capturing threshold behavior as a mechanism behind cascades in social networks. A similar setup is seen in disease spreading models where in e.g. the SI and SIR models agents cannot return to their original states. In addition to the fractional-threshold model, we consider another version where the agents respond to an absolute number F_i of interactions with state-1 neighbors. The state changes if $F_i \geq \Phi$ [8,9] (we call this version the *absolute-threshold model*).

In Fig. 1, we give an illustrative example of the model. We use a time-line representation of a temporal network. Contacts between individuals are illustrated by arcs. Red circles indicate adopters, white or gray circles indicate non-adopters. The threshold here is assumed to be $\phi = 0.5$ and the time window $\theta = 10$. As time progresses, the time window slides through each contact and updates vertices with respect to the contacts within the time window. In panel (a), vertex d does not change its state even though it is in contact with an infected vertex. At another time—see panel (b)—vertex d changes its color because inside the time window the fraction of red neighbors exceeds the threshold.

We simulate the model on six empirical temporal networks. The statistics we will present are averaged over at least 100 independent simulation runs.

3. Empirical datasets

We test our model on six empirical datasets generated by different types of human interactions. The datasets were obtained with all individuals anonymized to protect their identity. The first dataset consists of self-reported sexual contacts from a Brazilian online forum where sex-buyers rate and discuss female sex-sellers [10]. The second dataset comes from email exchange at a university [11]. It was used in Ref. [12] to argue that human behavior often comes in bursts. The third dataset was collected at a three-days conference from face-to-face interactions between conference attendees [13]. The fourth dataset comes from a Swedish Internet dating site where the interaction ranges from partner seeking to friendship oriented [14]. The fifth and sixth datasets come from a Swedish forum for rating and discussing films [15]. One of these datasets represents comments in a forum that is organized so that one can see who comments on whom. The other datasets comes from email-like messages. Table 1 summarizes details of the datasets such as number of vertices, number of contacts, sampling time and time resolution. Some of the datasets like the movie forum, the email and the conference contacts can be an underlying structure for social influence, spread of fads and ideas, or computer viruses. The Sexual-contact datasets and