Information Sciences 239 (2013) 226–240

Contents lists available at SciVerse ScienceDirect

Information Sciences

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

FastSIR algorithm: A fast algorithm for the simulation of the epidemic spread in large networks by using the susceptible–infected–recovered compartment model

Nino Antulov-Fantulin [a,*], Alen Lančić [b], Hrvoje Štefančić [c], Mile Šikić [d,e,*]

[a] Division of Electronics, Laboratory for Information Systems, Rudjer Bošković Institute, Zagreb, Croatia
[b] Faculty of Science, Department of Mathematics, University of Zagreb, Zagreb, Croatia
[c] Theoretical Physics Division, Rudjer Bošković Institute, Zagreb, Croatia
[d] Faculty of Electrical Engineering and Computing, Department of Electronic Systems and Information Processing, University of Zagreb, Croatia
[e] Bioinformatics Institute, A*STAR, Singapore, Republic of Singapore

ARTICLE INFO

Article history:
Received 26 August 2011
Received in revised form 13 March 2013
Accepted 18 March 2013
Available online 25 March 2013

Keywords:
SIR compartment model
Epidemic spreading algorithm
Computational epidemiology

ABSTRACT

We propose two efficient epidemic spreading algorithms (Naive SIR and FastSIR) for arbitrary network structures, based on the SIR (susceptible–infected–recovered) compartment model. The Naive SIR algorithm models full epidemic dynamics of the well-known SIR model and uses data structures efficiently to reduce running time. The FastSIR algorithm is based on the probability distribution over the number of infected nodes and uses the concept of generation time instead of explicit time in treating the spreading dynamics. Furthermore, we also propose an efficient recursive method for calculating probability distributions of the number of infected nodes. The average case running time of both algorithms has also been derived and an experimental analysis was made on five different empirical complex networks.

1. Introduction

Complex networks represent the structure of communication networks [3,7] or social contact interactions [23,27] very well. Therefore, it is reasonable to study computer virus propagation or epidemic spreading on complex networks [2,28,15]. Modeling the spread of an epidemic in a population is usually done by dividing individuals in the population into subdivisions with some common characteristic features called compartments. The SIR model is a good model for many infectious diseases where each individual in a population can be in one of three different compartments. Those who are susceptible to the disease are in the susceptible compartment, those who are infected and can transmit the disease to others are in the infected compartment and those who have recovered and are immune and those who are removed from the population are in the recovered compartment. Some infectious diseases are described with a different number of compartment models, e.g. SIS model (susceptible-infected-susceptible) where individuals cannot have long lasting immunity and therefore the recovered compartment does not exist.

Different mathematical frameworks have been used to study epidemic spreading. We can divide them into two big categories based upon assumptions they make: the homogeneous mixing framework and the heterogeneous mixing framework.

* Corresponding authors. Addresses: Ruder Bošković Institute, Bijenička cesta 54, 10000 Zagreb, Croatia (N. Antulov-Fantulin). Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb, Croatia Tel.: +385 16129781 (M. Šikić).

E-mail addresses: nino.antulov@irb.hr (N. Antulov-Fantulin), alen@student.math.hr (A. Lančić), shrvoje@thphys.irb.hr (H. Štefančić), mile.sikic@fer.hr (M. Šikić).

0020-0255/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2013.03.036

The homogeneous mixing framework assumes that all individuals in a population have equal probability of contact. This is a traditional mathematical framework [21,19,31], where differential equations can be applied to understand epidemic dynamics. The models in this framework predict the epidemic threshold which divides the healthy and the infective phase of the SIR and the SIS models.

In reality, each individual has contact with only a small fraction of individuals in a population. As the assumption of the homogeneous mixing fails to describe a realistic scenario of disease spreading, the heterogeneous mixing is described by using a network structure. The small world network property [39] and the scale-free network property [4,16] have a great impact on the outcome of an epidemic spreading. The SIR and the SIS models in the heterogeneous mixing framework imply that a disease will always spread for certain power-law degree distributions [30]. We can also make a further division in this framework by assumptions they make: the bond percolation, the mean-field and the particle network approach.

The bond percolation approach applies the percolation theory to describe epidemic processes on networks [26,18,24]. The percolation theory predicts the mean epidemic size, but neglects epidemic dynamics. Analytical solutions of a mean outbreak size for the configuration network models have also been derived [9]. In order to show isomorphism of epidemic models to bond percolation processes [20], the epidemic percolation networks were introduced as a valuable tool for studying of the stochastic epidemic models.

The mean-field approach assumes that all nodes in a network with the same degree k with respect to an epidemic process are statistically equivalent [10,30]. This method enables us to write the epidemic time evolution equations for a network with an arbitrary degree distribution. By solving them, the relation of topology dependent features and the epidemic threshold have been discovered [38]. Recently, heterogeneous mean-field methodology [32] for epidemic spreading on interconnected complex networks has also been developed.

The particle network approach assumes that individuals are represented by particles which diffuse along edges on a network and each node contains some non-negative integer number of particles (reaction–diffusion processes). Some studies [11,12] used the contact network models between urban cities (cities are connected through airline transportation network) and the homogeneous mixing model inside urban cities and examined the influence of interventions (antiviral drugs and containments) to a worldwide spread of a pandemic.

Realistic epidemic simulations (GLEaMviz [8], EpiFast [6], EpiSims [5] and EpiSindemics [17]) have become a very important application of high-performance computing in epidemic predictions. Recently, predictions of the epidemic spread by using Global Epidemic and Mobility Model [34] has been validated on a public health study of 2009 H1N1 influenza. Optimal distribution of a vaccine supply can have a significant impact on an epidemic spreading outcome [33].

The large-scale fluctuations in network topology suggest that the role of each node in an epidemic process cannot be disregarded [35,37]. Our paper makes a contribution to the development in the field of SIR simulation algorithms on complex networks. These algorithms have an important role in the research of epidemic processes in the heterogeneous mixing framework.

In this paper, we propose the Naive SIR algorithm and the FastSIR algorithm for simulating spreading process with the SIR model on an arbitrary network structure. The Naive SIR algorithm simulates full epidemic dynamics of the SIR model on a network. The name "naive" suggests that the Naive SIR algorithm is the straightforward approach but nevertheless it uses data structures very efficiently. We use the Naive SIR algorithm as a baseline algorithm for comparison with the FastSIR algorithm. The main contribution of this paper is the FastSIR algorithm that uses probability distributions over the number of infected nodes which speeds up the process of recovery of infected nodes in a simulation, thus reducing the overall running time. Although the FastSIR algorithm does not follow epidemic dynamics in time, it still captures all infection transfers.

In Section 2 we formally define the epidemic simulation problem and other concepts used in this paper. Section 3 describes our implementation of the Naive SIR algorithm along with the running time and space complexity analysis. In Section 4 we describe the FastSIR algorithm along with the running time and space complexity analysis. By using a four-step proof, equivalence of the Naive SIR algorithm and the FastSIR algorithm regarding the number of infected nodes was shown. We also described how to efficiently implement probability distributions of the number of infected nodes by a recursive method. In Section 5 we described the results of the performance analysis of our algorithms on five empirical complex networks. In Sections 6 and 7 we described possible applications of our algorithms along with a discussion of results and a conclusion.

2. The epidemic simulation problem

We define the contact-network as an undirected and non-weighted graph $G(N,L)$ (N-set of nodes, L-set of links). A link (u,v) exists only if two nodes u and v are in contact during the epidemic time. We also assume that the contact-network during the epidemic process is a static one. To simulate epidemic propagation through a contact-network, we use the standard stochastic SIR model. In this model each node at some time can be in one of the following states: susceptible (S), infected (I) or recovered (R). Time is modeled in discrete time steps, and number of time steps necessary for one epidemic simulation is defined by the step at which epidemics stops spreading, i.e. when there are no infected nodes in the network.