ATLANTIS PRESS

Advances in Computer Science Research, volume 87

3rd International Conference on Mechatronics Engineering and Information Technology (ICMEIT 2019)

The Application of Gillespie Algorithm in Spreading

Xiaomin Deng [a], Xiaomeng Wang [b, *]

College of computer and information science, Southwest University, Chongqing 400715, China.

[a]760565290@qq.com.com, [b, *] wxm1706@163.com

Abstract. The contagion models of disease-spread which predict the epidemics grow with time goes by have been widely researched in social networks. The discrete-time simulation method, Monte Carlo Simulation where time is discretized into uniform steps and transition rates between states are replaced by transition probabilities, are mostly applied when simulating the models. In this paper, we propose a continuous-time approach, the Gillespie algorithm, which can be used for fast simulation of stochastic processes, is event-driven rather than using equally-spaced time steps. We show how the method can be adapted to the epidemic models, mainly in the susceptible-infected model and susceptible-infected-susceptible model, and confirm the accuracy of the method with numerical simulations. Based on the accuracy of the method, we make some changes in epidemic models to make the models more applicable.

Keywords: contagion model; epidemic; discrete-time; continuous-time; Gillespie Algorithm.

1. Introduction

The complex networks whose nodes identify the individuals or organizations, and links between nodes represent the presence of relationship or interaction among them can describe many social, biological, and communication systems [1][2]. Such characterizations make a fundamental theory on dynamical spreading. In recently, the evolution spreading dynamics in complex networks has attracted much attention and which are ubiquitous in real life system, it can be theoretically described by spreading dynamics of financial [3], innovations [4], commercial products [5][6], rumors [7], opinions [8] and many others being treated as epidemics [9] [10] [11] [12].

The contagions, such as epidemic spreading, rumor spreading, and information spreading mentioned above are usually modeled as biology contagions, in which infection transmission can be sufficient by a single activated source [13] [14]. The most classical models for biology contagions is the SIS (susceptible-infected-susceptible) model and SIR (susceptible-infected-recovered) model in networks [15]. However, the simplest contagion model is SI (susceptible-infected) model [16] [17], which has been extensive and widely studied in complex networks. all of this disease models can be fall into “simple contagion” that is any process by which a node can easily become infected by a single contact with an infected neighbor. In other words, a susceptible individual only requires a single contact with an infected individual to allow a pathogen to propagate. Traditionally, the simple contagion models have been applied to sociological spreading behaviors in order to predict how behavior would diffuse across a network [18]. Without specifics statements, this paper focuses on SI and SIS models. At each time step, for reversible SIS model, each infected node transmits the infection to its susceptible neighbors at rate β, and then every infected node return to susceptible at rate μ. For the irreversible SI model, each infected node tries to infected every susceptible neighbor with rate β, nodes in SI model either in susceptible state or infected state, the critical difference from SIS model is the infected nodes cannot return to susceptible state.

When investigating the spreading dynamics of simple contagion models, some important information such as what are the dynamical mechanisms behind the phenomena, the method of simulation and so on will catch the researcher main attention. When simulating the contagion models, the first consideration is which simulation method is used to realize the process. The simulation methods apply to models will be our key point in this paper. The mostly used to modeling epidemic spreading dynamics is the discrete-time Monte Carlo simulations [19] [20] which often used as a gold standard. It is Such approaches can be either numerical or theoretical [21], in which time is discretized into time-steps of length Δt, events occur with certain probabilities. These probabilities are known as the state transition probabilities, which are simply the product of the corresponding rate and the

time-step Δt. Through using discrete-time approximations that in the limit time-steps $\Delta t \to 0$, we can simulate the continuous-time stochastic process. Nevertheless, this method leads to some limitations when simulating the continuous-time process [22]. Another generally used simulation approach is the mean-field solution [23] raised by Newman. In some cases, the mean-field theory can make a high-level accuracy, but it cannot predict well since the network contains very complex structures which cannot be characterized only by degree distribution [24]. Moreover, in mean-field theory, it considered the states of neighbors are independent, and neglect the dynamical correlation among the neighbors.

In this paper, we proposed a different method in simulating the epidemic models, the Gillespie algorithm, can simulate the continuous-time contagion dynamics in networks, In [25], it said that the easy use of a Gillespie algorithm in practice can lead to fast simulation of contagion processes on time-varying networks. Although it is not entirely direct to code, instead of checking at each time-step if each possible reaction takes place, we can exactly obtain the next stochastic process through the Gillespie algorithm simulation. fit the simulation results to the theory in order to make the accuracy of the new method.

2. Related Work

2.1 Monte Carlo Simulation.

The straightforward way to simulate a stochastic process is to use a rejection sampling algorithm, which is widely used when constructed the epidemic models also named Monte Carlo simulation. In this mean, time-axis is discretized into many small and equal time-steps Δt , where the Δt should be chosen sufficiently small so that this discretization does not influence the outcome of process significantly. If events are described by rates γ, while the transition probabilities of events in the discrete-time interval Δt are $\gamma * \Delta t$, and must check at each time-step if each possible event take place.

2.2 Degree-based Mean Field Theory.

Supposed that node i has degree k, the variable $\rho_k(t)$ is defined as the fraction of nodes degree k that are infected at time t, we can also think as the probability is $\rho_k(t)$ when choosing a node whose degree is k and infected from all node. It makes the following approximation: using constant degree k which is defined for the entire class of nodes to replace the specific degree quantity of node i. that is to say, for all nodes in networks, the degree of each node is assumed the same k. the infected nodes of each time are obtained from

$$\gamma * k * (1 - \rho_k)\omega_k \tag{1}$$

where the $\omega_k(t)$ is the probability that a given neighbor of node i is infected.

3. Simulation Models

Based on the above consideration, the simulation algorithm for stochastic models are a well-reached area and which will be our main focus on. The Gillespie algorithm, which was originally introduced by Doob [25] [26] and popularized by Gillespie [27], are proposed as optimal simulation schemes both in terms of replicating the continuous-time process and computational speed [22]. Here, we apply the Gillespie algorithm to epidemic models, such as the SI (susceptible-infected) model, and SIS (susceptible-infected-susceptible) model.

3.1 Gillespie Algorithm on Static Networks.

The healthy nodes in the simple contagion model-SI model, which has two states susceptible and infected where healthy nodes are considered “susceptible”, can easily become infected by a single contact with an infected neighbor; If infected nodes transmit infection across their links at constant

rate β per unit time[29], susceptible nodes change it state and become infected at rate $k * \beta$, which k is the number of infected neighbors of the susceptible node, in other words, the rate that a susceptible node change to infected state is scaled linearly with the number of infected network neighbors.

As stated in section two, when simulating the SI model and SIS model, researchers were inclined to model it using Monte Carlo method. Here, we defined in this section the type of stochastic processes of epidemic models for which the Gillespie algorithm can be applied. The Gillespie algorithm allows us to perform exactly on stochastic simulation. When simulating, individuals are placed by nodes, the edges between two nodes represent the relationship of them. In general, dynamical processes can be described as a set of stochastic events. Similarly, the processes such as susceptible node become an infected node and the infected node recover to susceptible one can also be regarded as stochastic events in epidemic models. Making sure the next occurrence of the stochastic event is the critical thing while simulating spreading. There is only a class of stochastic event in SI spreading model, but in SIS model, we should be careful that there are two different classes of events in the same period. The advantage of adapting Gillespie algorithm is that we can clearly see the next event occurred and the time event occurred.

In an undirected connected network which is composed of N nodes. The spreading rate at which each susceptible node acquires infection from an infected neighbor is denoted by β, the recovery rate at which an infected node become a susceptible node is named by μ. For adapting Gillespie algorithm to simulations, the rate of possible stochastic events occurrence is denoted by k_i, as k_i is the product of the number of infected neighbors n and the transmission rate β when the random event is infection, with time step by, k_i is not always the same, as increasing of infected neighbors. However, when infected nodes recover and become susceptible, the rate k_i is a constant value equal to μ. It supposed that all nodes in networks are in a state of susceptible at time t=0, initialized a perturbation that making a node whose degree is non-zero in state infected, the set of possible stochastic events in general changes over time. At time t is not equal to zero, it can obtain the exact occurrence event and waiting time τ when the event will occur through the temporal Gillespie algorithm. We first sum up all rates of stochastic events that may occur, then generate a random variable which based on the sum of rates; the generated variable which is exponentially distributed is the waiting time expresses by τ for the occurrence of next event, we obtained τ from

$$\tau = -\frac{ln(r1)}{\sum_{i=1}^{n} k_i} \tag{2}$$

There are two random variables at each step of Gillespie algorithm, r1 and r2, which are uniformly distributed on the interval [0,1], one is used to draw the time interval τ via (2), the other is used to select the stochastic event occurring after the waiting time τ. On the basis of r2, we can determine the stochastic event occurs with probability

$$p = \frac{k_i}{\sum_{i=1}^{n} k_i} \tag{3}$$

With this information, the main steps used in the Gillespie algorithm when simulating the SI model and SIS model are the following:

1. At time t, there are various stochastic events may occur in the network.
2. Sum up the rates of stochastic events which may occur, there is only one.
3. Random generation of the waiting time τ from the cumulative distribution (2) based on the random variable r1 update the current time as $t \to t + \tau$.
4. On the basis of a random variable $r2$ generated form uniform distribution of [0,1], select a stochastic event to occur with the probability of (3).
5. Go back to the second step unless the number of possible events is zero or the total simulation time has been exceeded.

4. Results

According to the method of section three, in this section, presenting and analyzing a set of simulations and fitting it to theory will be our main focus.

Though normalizing the time and instantaneous cumulative transition rate, we constructing a temporary Gillespie algorithm that is applicable in Poisson processes. We use the pseudocode and c++ implementations for its application to simulate the epidemic spreading of SI and SIS models for homogeneous populations. In this section, we show the main results of the article. In part A: we verify the accuracy of the temporal Gillespie algorithm numerically by comparison with the theory [30]. For part B: we make some extend in epidemic spreading models which can make the Poisson models more widespread use. Notably, the infected rate β at which an infected node transmit infection through associated edge is the same theoretically for all nodes in networks. Yet, in realistic infectious diseases dynamics, there are also instances where someone who is more sensitive to the infections may lead to fast infection; someone who is immune to infections may make the infection slow. By making consideration of communication, in reality, we present the situation that different infected rate exists in the same networks.

4.1 Fitting to the Theory.

The simulations in this section are conducted to verify the accuracy of the Gillespie algorithm when adapted to SI and SIS models. The simulation of SI and SIS begin at a fixed fraction of initiators $p = 0.001$ on an ER graph with $N = 1000$, for both SI and SIS simulations. Because of very few parts of initiators at the beginning, and as the main purpose is fitting the simulation to verifying the accuracy of the Gillespie algorithm when applied in spreading models. In order to increase reliability of final results, the fraction of infections must cover as many as possible of all nodes when spreading is terminated. One avenue to realize the goal is by taking out nodes which do not consist of the giant component, such approach can make sure infections are disseminated in the giant component, however, when selecting initiators, it is necessary to avoid choosing the worked nodes before. The other way is set an appropriate average degree <k>. As soon as <k> becomes large enough for a giant component to arise [31]. Only if the network is fully connected, contagion can reach all network ultimately, the circumstance that choosing invalidate initiators has the low possibility to occur and reserve the initial scale of the network. On account of the easy implementation of the latter means, we use the second solution to simulate spreading dynamically.

We first look dynamic of SI model which spread the infection at infected rate $\beta=2$, the simulation results of SI model are shown in Fig.1. As the case that one node either in susceptible state or infected state in spreading processes, once a susceptible node become infected, it cannot return to susceptible again. Therefore, as represented in the Fig.1, at time $t=1.6$, almost all nodes are infected and the infection is terminated. Additionally, in terms of fitting with [30], the full line (simulation) and the dotted line basically coincide.

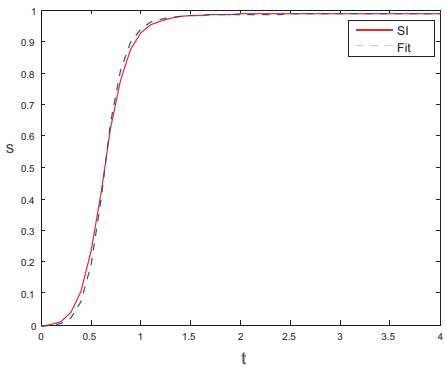

Fig. 1 SI dynamics with $\beta = 2$ on ER network with *N=1000* nodes the infected density S vs time with average degree $< k > = 6.0$, *p=0.001*, and the data are averaged over 100 independent runs. The numerical simulations are implemented and represented by the red solid line, the blue dashed line is obtained according to the theory from [30].

Let's consider SIS contagion process, which is somewhat more complicated than SI contagion process. As with SIS process, two rates define the stochastic dynamics. Each susceptible node switches to infected at rate $\beta = 2$, each infected node changes to susceptible state at a constant rate $\mu = 1$. The results in Fig.2 obtained from simulations indicate the SIS dynamical process, as can be seen in the graph, after t=1.5, the fraction of infected nodes in the metastable both in simulation and theory [30], but that does not mean there is no dynamical process occurring in networks.

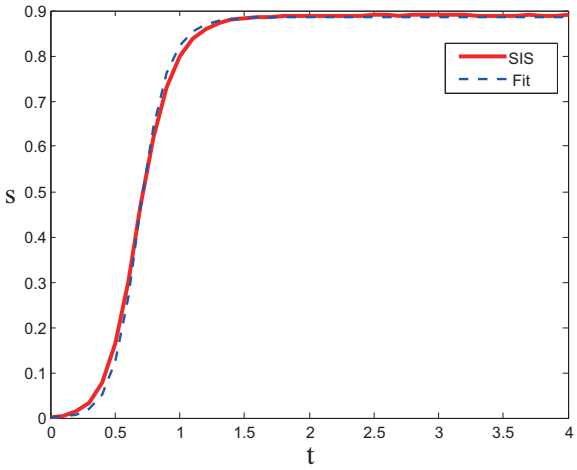

Fig. 2 Time evolution of the expected fraction of infected nodes S for SIS dynamics with $\beta = 2, \mu = 1$ on an ER network with 1000 nodes and average degree $<k>=6.0$, $p=0.001$. the trajectory which is shown as a red solid line is averaged over 100 realizations.

In accordance with the fitting results of SI and SIS, we arrive at a simple conclusion: the way of adapting the Gillespie algorithm to epidemic processes is the success, which can accurately describe the dynamical processes.

4.2 Make Different Infected Rate.

Regardless of the information spreading in social network, information forwarding in Weibo, dissemination in circle of friends and epidemic diffusing in community, the simplest propagative mechanism is that spreading is transmitted from infected one (the sender) to susceptible one (the receiver), the related mechanism of contagion conforms to SI contagion process. In real life of this propagation, there must be many factors influence the spreading, however, which cannot be fully considered and described when using the existing models. It is our main purpose to make the model more suitable for the phenomenon of communication in real life by changing some small details in the existing model. In reality, each individual has different ability to accept new things, and different friends will have a different influence on them. When forwarding news, people with greater influence, such as stars, are flying faster than ordinary people. When infectious diseases spread, people with poor health are infected more quickly than healthy people. Therefore, it is realistic to have multiple propagation rates in a network.

Different from the previous works, here we investigate the SI process on networks with different infectivity. Based on the above considerations and the difficulty of implementation, we simulate the propagation process in which there is only two transmission rate to different infectious rate of everyone in the network, and compare the disparate results with the original one (only one rate).

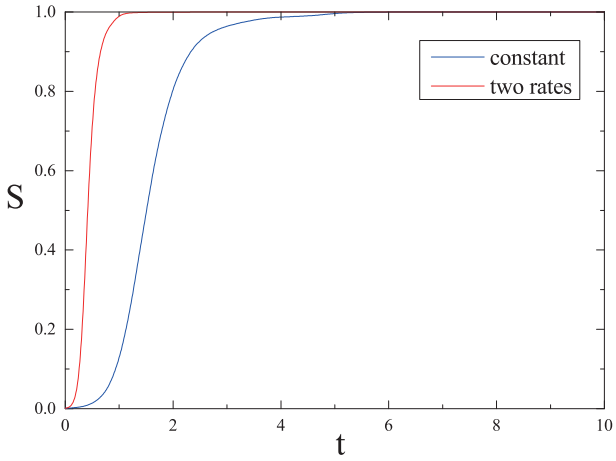

Fig .3 Time evolution of the expected fraction of infected nodes S for SI dynamics with $\beta_1 = 4, \beta_2 = 1$ shown by red line, and $\beta = 1$ represented by blue one on a constant ER network with 1000 nodes and average degree <k>=6.0, p=0.001. the trajectory which is shown as a red solid line is averaged over 100 realizations.

As is showed in Fig.3, Based on the results and conclusions of part A, we apply our approach to a stochastic system of spreading with two transmission rates. When simulating, the infectious speed of node is randomly assigned to one of two rates. Comparing the result of different infection rates to the same infection rate, we can see that the different infection rate accelerates the spreading process.

5. Conclusion

In the study of infectious disease models, researchers are more inclined to use discrete methods (Monte Carlo Simulation) to build models, fewer investigators use continuous-time to study the epidemic spreading, there are fewer ways to achieve continuous-time epidemic spreading and there are some limits of this mean. In this paper, we adapt the Gillespie algorithm in epidemic models, such as SI and SIS models for the sake of arriving at continuous-time spreading. In order to prove the validity and accuracy of the method, by fitting the simulated experiment and theory, we successfully proved the validity and feasibility of the method. Further, on the fundamental of continuous-time spreading models that have been implemented, we have made some extension. By setting a variety of propagation rate make the diffusion process in models more in line with communication phenomena in the existing life.

References

[1]. Newman MEJ. Properties of highly clustered networks. Phys Rev E(2003) 68:026121. Doi:10.1103/PhysRevE.68.026121.

[2]. S. Wasserman and K. Faust, See the special section on complex systems in science 284, 79 (1999).

[3]. J. P. Glesson, T. Hurd, S. Melnik, and A. Hackett, in Advances in network analysis and its Applications (Springer, 2012) pp. 27-56.

[4]. Aral S, Muchnik L, Sundararajan A. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. Proc Natl Acad Sci USA. (2009) 106:21544-9.

[5]. Yanqing Hu, Shlomo Havlin, and Hernan Makse. The rise and fall of social communities: Cascades of followers triggered by innovators. In APS Meeting Abstracts, 2013.

[6]. P. Resnick, H. R. Varian, Recommender systems, Communications of the ACM 40 (1997) 56-58.

[7]. Yamir Moreno, Maziar Nekovee, and Amalio F Pacheco. Dynamics of rumor dpreading in complex networks. Physical Review E, 69(6):066130, 2004.

[8]. Damon Centola, Victor M Eguiluz, and Michael W Macy. Cascade dynamics of complex propagation . Physica A: Statistical Mechanics and its Applications, 374(1):449-456, 2007.

[9]. M. E. J. Newman, Networks: An intorduction, Oxford University Press, 2010.

[10]. M. J. Keeling, P.Rohani, Modeling infectious diseases in humans and animals, Princeton University Press, 2008.

[11]. R. Cohen, S. Havlin, Complex networks: Structure, robustness and function, Cambridge University Press, 2018.

[12]. V. Karyotis, M. Khouzani, Malware diffusion models for modern complex networks: theory and applications, Morgan Kaufmann, 2016.

[13]. R. M. Anderson, R. M. May, Infectious deiseases of humans: Dynamics ans control, Oxford university press, 1992.

[14]. D. Guilbeault, J. Becker, D. Centola, Complex contagion: A decade in review, in: Complex spreading phenomena in social systems, Springer, 2018,3-25.

[15]. H. W. Hethcote, The mathematious of infectious diseases, SIAM Review 42 (2000) 599-653.

[16]. M. Barth´elemy, A. Barrat, R. Pastor-Satorras, A. Vespignani, Velocity and hierarchical spread of epidemic outbreaks in scale-free networks, Physical Review Letters 92 (2004) 178701.

[17]. T. Zhou, J.-G. Liu, W.-J. Bai, G. Chen, B.-H. Wang, Behaviors of susceptible-infected epidemics on scale-free networks with identical infectivity, Physical Review E 74 (2006) 056109.

[18]. Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, et al. Identification of influential spreaders in complex networks. Nat Phys. (2010) 6:888-93. doi: 10.1038/nphys1746.

[19]. R. Pastor-Satorras and A. Vespignani, Physical Review Letters 86,3200 (2001).

[20]. Y. Wang, D.Chakrabarti, C. Wang, and C. Faloutsos, in Reliable Distributed Systems, 2003. Proceedings. 22nd International Symposium on (IEEE, 2003)pp.25-34.

[21]. D. Balcan, B. Goncalves, H. Hu, J. J. Ramasco, V. Colizza, and A. Vespignami, Journal of Computational science 1, 132 (2010).

[22]. Peter G Fennell, Sergey Melnik, and James P Gleesom. Limitations of discrete-time approaches to continuous-time contagion dynamics. Physical Review E, 94(5):052125, 2016.

[23]. M. E. J. Newman, C. Moore, D. J. Watts, Mean-Field Solution of the Small-World Network Model, Physical Review. Letter. 96 (2006) 034101.

[24]. A. L. Barab´asi, R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509-512.

[25]. Christian L Vestetgaard and Mathieu Genois. Temporal gillespie algorithm: Fast simulation of contagion processes on time-varying networks. PLoS computational biology, 11 (10): e1004579, 2015.

[26]. Joseph L Doob. Topics in the theory of markoff chains. Transactions of the American Mathematical Society, 52(1): 37-64,1942.

[27]. Joseph L Doob. Markoff chains-denumerable case. Transactions of the American Mathematical Society, 58 (3): 455-473, 1945.

[28]. Daniel T Gillespie. Exact stochastic simulation of coupled chemical reactions. The journal of physical chemistry, 81(25): 2340-2361, 1977.

[29]. Poeter MA, Gleeson JP. Dynamical systems on networks: a tutorial (2014). ArXiv preprint arXiv: 1403.7663.

[30]. Gleeson J. P.High-accuracy approximation of binary-state dynamics on networks, Physical. Review. Letter. 2011, vol.107 pg. 068701.

[31]. Pramesh Singh,Sameet Sreenivasan,Boleslaw K Szymanski, and Gyorgy Korniss. Threshold-limited spreading in social networks with multiple initiators. Scientific reports, 3:2330, 2013.