Physica A 464 (2016) 198–210

Contents lists available at ScienceDirect

Physica A

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

A general stochastic model for studying time evolution of transition networks

Choujun Zhan [a,b,*], Chi K. Tse [b], Michael Small [c]

[a] Department of Electronics Communication and Software Engineering, Nanfang College of Sun Yat-Sen University, Guangdong 510970, China
[b] Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hunghom, Hong Kong
[c] School of Mathematics and Statistics, University of Western Australia, Crawley 6009, Australia

HIGHLIGHTS

- A general model describing the dynamics of transition networks is derived.
- A simulation algorithm for studying the network evolutionary behavior is proposed.
- The disease propagation dynamics in different networks generally have different properties but they do share some common features.
- The model provides a good prediction of user growth in the Facebook network.

ARTICLE INFO

Article history:
Received 3 April 2016
Received in revised form 11 June 2016
Available online 6 August 2016

Keywords:
Dynamic complex network
Stochastic process model
Simulation algorithm

ABSTRACT

We consider a class of complex networks whose nodes assume one of several possible states at any time and may change their states from time to time. Such networks represent practical networks of rumor spreading, disease spreading, language evolution, and so on. Here, we derive a model describing the dynamics of this kind of network and a simulation algorithm for studying the network evolutionary behavior. This model, derived at a microscopic level, can reveal the transition dynamics of every node. A numerical simulation is taken as an "experiment" or "realization" of the model. We use this model to study the disease propagation dynamics in four different prototypical networks, namely, the regular nearest-neighbor (RN) network, the classical Erdös–Renyí (ER) random graph, the Watts–Strogátz small-world (SW) network, and the Barabási–Albert (BA) scalefree network. We find that the disease propagation dynamics in these four networks generally have different properties but they do share some common features. Furthermore, we utilize the transition network model to predict user growth in the Facebook network. Simulation shows that our model agrees with the historical data. The study can provide a useful tool for a more thorough understanding of the dynamics networks.

1. Introduction

Many large-scale real-world networks, such as the Internet, the global aviation network and protein interaction networks, can be described as *complex networks* in which nodes represent individuals or organizations, and links mimic interactions among the nodes [1,2]. In many cases, each node of the network may assume several states and can change from one state

* Corresponding author at: Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hunghom, Hong Kong.
E-mail addresses: zchoujun2@gmail.com (C. Zhan), michael.tse@polyu.edu.hk (C.K. Tse), michael.small@uwa.edu.au (M. Small).

http://dx.doi.org/10.1016/j.physa.2016.07.034
0378-4371/

to another. For instance, in a rumor spreading network [3], each node of the network can be regarded as being in one of three possible states, namely, *spreader*, *ignorant* and *stifler*. Other examples can be found in epidemic spreading networks described by the susceptible–infected–recovered (SIR) model [4–6], the susceptible–infected–susceptible (SIS) model [7,8], or the susceptible–infected (SI) model [9–11]. Furthermore, complex networks of language evolution [12,13], the growth of membership-based websites [14], the spread of social behavior [15,16], spatial prisoner's dilemma game [17], and so on, also fall in this category. In this paper, we refer to this important class of networks as *transition networks*, which exist in many social, biological and communication systems.

For convenience of illustration, we discuss the main idea here using the familiar SIR epidemic spreading model [18,19]. Moreover, we stress that the proposed model is generally applicable to networks with any connection mechanism, like the SEIR, SIS, language evolution, human behavior spreading and so on, and in any network setting, such as the Erdös–Renyí (ER) random graph, Watts–Strogátz (WS) small world and Barabási–Albert (BA) scalefree networks. In an SIR model, individuals are categorized as Susceptible, Infected, and Recovered. For instance, under the homogeneous mixing hypothesis assumption [20,21], a standard deterministic SIR model can be described in terms of the densities of susceptible $S(t)$, infected $I(t)$, and recovered individuals $R(t)$. The ODEs provide a crude description revealing only the density dynamics, but fail to reveal disease propagation dynamics at any specific node. The homogeneous mixing hypothesis discards possible heterogeneities in the network, and inhomogeneous dynamics cannot be captured with this standard SIR model. Furthermore, the ODEs are deterministic while disease spreading is often subject to non-deterministic processes depending on the nature of transmission. Hence, in principle, a stochastic model is more realistic than a deterministic one. In addition, stochasticity introduces variances and co-variances which influence the behavior, and helps push the system away from the deterministic attractor so that transitions may play a significant role. Also, stochasticity may cause extinction of the epidemic [22].

In this paper, a stochastic process model is established for general complex transition networks with arbitrary connectivity distribution and transition probability. In most cases, the proposed model cannot be solved analytically, as it is the case with most models using ODEs. Hence, the proposed model has to be solved numerically. In our study, a simulation algorithm, which does not involve fixed time steps, is established. Specifically, our model aims to provide analytical expressions of

1. the probability that the next transition will occur at a certain time;
2. the probability that a particular transition will occur;
3. the probability that a particular transition will occur with a specific link.

Clearly, with these expressions, this model can expose the transition dynamics of every individual in continuous time, whereas the ODE-based model describes the dynamics of "density" and the standard Monte-Carlo simulation can only capture the dynamics of each node at pre-defined discrete time-steps. It should be noted that no mathematical model can fully predict transition dynamics of a natural system. However, the proposed model offers a more realistic physical meaning of the relationship between individuals in a system and is thus able to provide a better overall picture of the transition network behavior.

In addition, our study examines the spreading network in four representative complex network models, namely, the regular nearest-neighbor (RN) network, the ER random-graph network, the WS small-world network, and the BA scale-free network. A brief overview of epidemiological characteristics shows that the epidemic dynamics of the aforementioned four types of complex networks have their own particular characteristics but they also share some common features. One of the findings is that homogeneity or heterogeneity may not be a key condition controlling the rate of epidemic spreading, while the average path length may play an important role. Also, the transition network model is used to predict user growth of Facebook during 2004 to 2013. Simulation shows the proposed model fits well with the historical data. Details of the model derivation and experimental results will be shown in the subsequent sections.

In the next section, we describe the model in detail. In Section 3, a numerical simulation algorithm is presented. Some numerical simulations of disease propagation and a summary of the results are presented in Section 4, while the study of user growth of Facebook network is shown in Section 5. Finally, the conclusion is given in Section 6.

2. Formulation of general stochastic process model for transition networks

The standard deterministic ODE-based SIR model [23] is established to describe the dynamics of "density", which is similar to the concept of "molecular concentration" in modeling a chemical reaction process. Hence, while an ODE-based model exposes the dynamics of "molecular concentration" in continuous time, our model reaches to the "molecular" level and reveals the dynamics of every "molecule" in continuous time. To facilitate discussion, a numerical simulation of disease propagation is first provided to give a sense of some differences between the standard SIR model and the proposed model applied to the SIR case. A typical numerical simulation using the proposed model generates time series $\{X(t_i), Y(t_i), Z(t_i) \mid t_i \in \Re^+, i = 1, 2, \dots\}$, with the following properties:

1. One and only one infection or recovery occurs in the time interval $[t_{i-1}, t_i)$. Thus, we treat the infection/recovery as being completed at time t_i, i.e., at time t_i, the state of the network changes.
2. The numbers of S, I and R individuals are changing at t_i. Hence, in the time interval $[t_{i-1}, t_i)$, we can treat $\{X(t), Y(t), Z(t)\} = \{X(t_i), Y(t_i), Z(t_i)\}$.

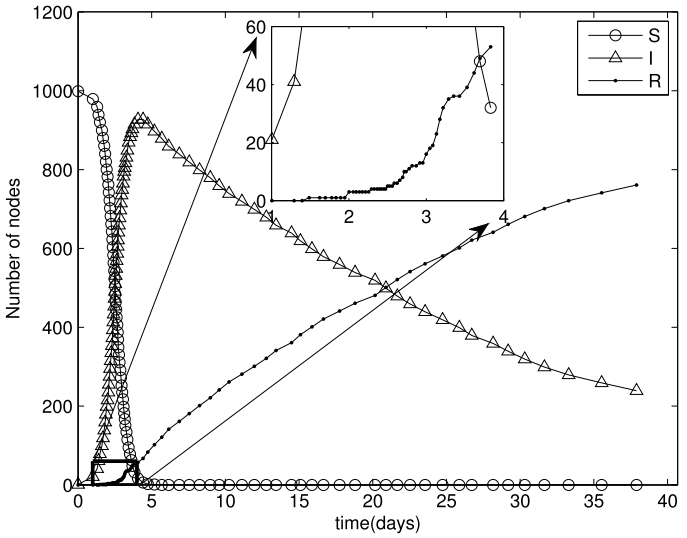

Fig. 1. A profile of disease propagation in a network with entirely susceptible population and a single infectious individual as initial condition.

As an example, consider the following simulated time series $\{X(t_0) = 999, Y(t_0) = 1, Z(t_0) = 0 \mid t_0 = 0\}$, $\{X(t_1) = 998, Y(t_1) = 2, Z(t_1) = 0 \mid t_1 = 0.541\text{d}\}$, $\{X(t_2) = 998, Y(t_2) = 1, Z(t_2) = 1 \mid t_2 = 2.241\text{d}\}$, ..., where d represents number of days. This data reveals that at $t_0 = 0$, there are 999 susceptible, 1 infected, and 0 recovered individuals, and in the interval $[t_0, t_1) = [0, 0.541\text{d})$, no infection has finished. The first infection completes at $t_1 = 0.541\text{d}$, and the numbers of S, I and R change to $X(t_1) = 998, Y(t_1) = 2, Z(t_1) = 0$. Then, there is no infection or recovery until $t_2 = 2.241\text{d}$, and an infected individual has recovered during the interval $[t_1, t_2) = [0.541\text{d}, 2.241\text{d})$. Hence, with the time series set $\{X(t_i), Y(t_i), Z(t_i) \mid t_i \in \Re^+, i = 1, 2, \ldots\}$, we are able to obtain almost every detail about the spreading of the disease, including when a transition (infection/recovery) occurs, how many transitions (infection/recovery) have occurred during a time interval, and when the state of a specific individual is transited (infected/recovery). Fig. 1 shows a snapshot of a single simulation of disease propagation in a complex network with 1000 individuals.

2.1. Preliminaries

Consider a general network $G = (V, E)$, where V and E denote the set of its nodes and edges, respectively. Suppose there are n nodes $\{v_1, v_2, \ldots, v_n\}$. Each node can be in one of k possible states $\{x_1, x_2, \ldots, x_k\}$ and can transit from one state to another [24]. This network is defined as *transition network*. The conceptual description of a transition is represented by the *transition channel* as shown below:

$$T_\mu : \left(x_{p_1} - x_{p_2} - \cdots - x_{p_{L_\mu}}\right) \xrightarrow{c_\mu} \left(x_{q_1} - x_{q_2} - \cdots - x_{q_{L_\mu}}\right), \tag{1}$$

where $\mu = 1, 2, \ldots, m$ is the index of the transition channel, m is the total number of types of transition channels, T_μ is the μth transition channel, and L_μ is the number of transition species in channel T_μ. Here, "$-$" represents a link. Hence, $x_{p_i} - x_{p_j}$ means a node in state x_{p_i} connects with another node in state x_{p_j}. The arrow indicates the transition direction, with c_μ being the rate of transition. Thus, $\left(x_{p_1} - x_{p_2} - \cdots - x_{p_{L_\mu}}\right)$ represents the set of prospective transition links before a transition occurs, while the resulting links are $\left(x_{q_1} - x_{q_2} - \cdots - x_{q_{L_\mu}}\right)$.

Here, we assume that all m transition channels are independent and all prospective transition links of the transition channel T_μ are homogeneous and exclusive. Then, each prospective transition link of the transition channel T_μ has the same transition probability and each transition occurs independently. The μth transition channel has a stochastic rate of transition c_μ. Thus, $c_\mu \Delta t$ is the probability that a prospective transition link of channel T_μ at time t will react in the next infinitesimal time interval $(t, t + \Delta t)$. For simplicity, we assume that c_μ is constant, but it can be made time-varying with no significant effect on the analysis.

As an example to illustrate the above definition, we consider a simple epidemiological model of disease spreading in a small network, as shown in Fig. 2. Each node of the network represents an individual in its corresponding state (susceptible, infected, or recovered), and each link is a connection along which the infection can spread. In this network, there are 6 nodes, i.e., $V = \{v_1, v_2, \ldots, v_6\}$, and 9 edges. Each node has 3 possible states, i.e., $k = 3$ and $\{x_1, x_2, x_3\} = \{S, I, R\}$. At time t_0, nodes $\{v_1, v_3, v_4, v_5\}$ are susceptible, $\{v_2\}$ is infected and $\{v_6\}$ is recovered. There are $m = 2$ transition channels, namely T_1

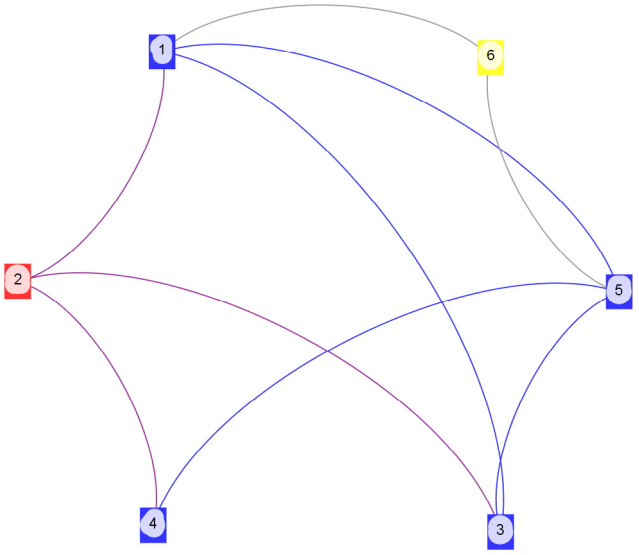

Fig. 2. A simple network with 6 nodes: red, blue and yellow nodes represent infected, suspected and recovered individuals, respectively. Disease can only be spread through the links. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and T_2:

$$\begin{aligned} T_1: &\quad (x_1 - x_2) \xrightarrow{c_1} (x_2 - x_2), \\ T_2: &\quad\quad (x_2) \xrightarrow{c_2} (x_3). \end{aligned} \tag{2}$$

Here, the first transition channel involves two transition species, and the second channel involves only one species. Thus, $L_1 = 2$ and $L_2 = 1$. For the first transition channel, the set of prospective transition links before transition occurs is $(x_1 - x_2) = \{e_{2,1}, e_{2,3}, e_{2,4}\}$, and for the second channel, the set of transition links only includes $(x_3) = \{v_2\}$. Here, we may set $c_1 = 0.04/\mathrm{d}$ and $c_2 = 0.01/\mathrm{d}$.

Similarly, the dynamics of language competition [12] in social networks can be model as

$$\begin{aligned} T_1: &\quad (x_1 - x_2) \xrightarrow{c_1} (x_1 - x_1), \\ T_2: &\quad (x_1 - x_2) \xrightarrow{c_2} (x_2 - x_2) \end{aligned} \tag{3}$$

where state x_1 represents the state of a speaker who selects language 1, while state x_2 corresponds to the state of a speaker who selects language 2. In this simple model, a speaker can either continue to speak its own language or switch to the other language.

Here, we focus on the simple epidemiological model (2). Suppose the network begins to evolve from $t = t_0$, and at time t, the network will be in a particular state. To determine the state at time t, we have to address the following three questions:

1. When will the next transition occur?
2. What transition will occur?
3. Which one of the prospective transition links will be selected?

Using the above SIR model as an example, the first question essentially means that a time interval Δt should be determined such that one and only one transition will complete at $t = t_0 + \Delta t$. Then, transition channel T_1 or T_2 will be selected for the next transition. Finally, if T_1 is selected (i.e., an individual will be infected), one prospective transition link among $\{e_{2,1}, e_{2,3}, e_{2,4}\}$ will be chosen. To answer the three questions stated above, a general model together with a simulation algorithm is developed in this paper.

2.2. General stochastic process model

The state matrix $\mathcal{S}(t)$ and the number of prospective transition links $h_\mu(\mathcal{S}(t))$ are required for establishing the model.

- $\mathcal{S}(t) \in \Re^{k\times N}$: at time t, if node v_j is in state x_i, $s_{i,j}(t) = 1$; otherwise $s_{i,j}(t) = 0$.
- $h_\mu(\mathcal{S}(t))$: number of distinct prospective transition links with state matrix $\mathcal{S}(t)$ for transition channel T_μ.

Referring to Fig. 2, at time t_0, we have

$$\mathcal{S}(t_0) = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{4}$$

Note that $\mathcal{S}(t)$ includes the state information of all the nodes. Transition channel T_1 includes 3 prospective transition links before transition, i.e., $\{e_{2,1}, e_{2,3}, e_{2,4}\}$. Thus, $h_1(\mathcal{S}(t_0)) = 3$, and likewise, $h_2(\mathcal{S}(t_0)) = 1$.

Let $P[\mathcal{S}(t) = M_s]$ denotes the probability that the transition network is in state $M_s \in \Re^{k\times N}$ at time t. The occurrence of the event $\mathcal{S}(t + \Delta t) = N_s$ can be thought of as an occurrence of the joint event: (1) $(\mathcal{S}(t + \Delta t) = N_s, \mathcal{S}(t) = M_s)$, which means that several transitions are completed in interval Δt; (2) $(\mathcal{S}(t + \Delta t) = N_s, \mathcal{S}(t) = N_s)$, which means in the time interval Δt, no transition takes place. Therefore, we can write

$$\begin{aligned} P[\mathcal{S}(t + \Delta t) = N_s] = &\sum_{j=1}^{\infty} \sum_{N_s^- \in \Omega_j^-} P[\mathcal{S}(t + \Delta t) = N_s, \mathcal{S}(t) = N_s^-] \\ &+ \left(P[\mathcal{S}(t) = N_s] - \sum_{j=1}^{\infty} \sum_{N_s^+ \in \Omega_j^+} P[\mathcal{S}(t + \Delta t) = N_s^+, \mathcal{S}(t) = N_s] \right) \end{aligned} \tag{5}$$

where Ω_j^- represents all the possible state matrices N_s^- which can transit into N_s after the jth transition (with $j = 1, 2, \ldots, \infty$); also Ω_j^+ represents all the possible state matrices N_s^+ which can transit from N_s after the jth transition (with $j = 1, 2, \ldots, \infty$). Here, the probability of $\mathcal{S}(t) = N_s$ at time $t + \Delta t$ is expressed in terms of the probabilities of all the possible states $\mathcal{S}(t)$. Manipulating conditional probabilities, we get

$$\begin{aligned} &P[\mathcal{S}(t + \Delta t) = N_s, \mathcal{S}(t) = N_s^-] = P[\mathcal{S}(t) = N_s^-] \times P[\mathcal{S}(t + \Delta t) = N_s \mid \mathcal{S}(t) = N_s^-], \\ &P[\mathcal{S}(t + \Delta t) = N_s^+, \mathcal{S}(t) = N_s] = P[\mathcal{S}(t) = N_s] \times P[\mathcal{S}(t + \Delta t) = N_s^+ \mid \mathcal{S}(t) = N_s]. \end{aligned} \tag{6}$$

Putting (6) in (5), we get

$$\begin{aligned} P[\mathcal{S}(t + \Delta t) = N_s] = &\sum_{j=1}^{\infty} \sum_{N_s^- \in \Omega_j^-} P[\mathcal{S}(t) = N_s^-] \times P[\mathcal{S}(t + \Delta t) = N_s \mid \mathcal{S}(t) = N_s^-] \\ &+ P[\mathcal{S}(t) = N_s] \times \left(1 - \sum_{j=1}^{\infty} \sum_{N_s^+ \in \Omega_j^+} P[\mathcal{S}(t + \Delta t) = N_s^+ \mid \mathcal{S}(t) = N_s] \right). \end{aligned} \tag{7}$$

Assume that at infinitesimal interval Δt, at most one prospective transition link of one transition channel undergoes a transition. Hence, $\Omega_1^- \neq \emptyset$ and $\Omega_1^+ \neq \emptyset$; otherwise $\Omega_j^- = \emptyset$ and $\Omega_j^+ = \emptyset$ $(j = 2, 3, \ldots, \infty)$. Thus, (7) can be simplified as

$$\begin{aligned} P[\mathcal{S}(t + \Delta t) = N_s] = &\sum_{N_s^- \in \Omega_1^-} P[\mathcal{S}(t) = N_s^-] \times P[\mathcal{S}(t + \Delta t) = N_s \mid \mathcal{S}(t) = N_s^-] \\ &+ P[\mathcal{S}(t) = N_s] \times \left(1 - \sum_{N_s^+ \in \Omega_1^+} P[\mathcal{S}(t + \Delta t) = N_s^+ \mid \mathcal{S}(t) = N_s] \right). \end{aligned} \tag{8}$$

Since there are m transition channels, (8) can be rewritten as

$$\begin{aligned} P[\mathcal{S}(t + \Delta t) = N_s] = &\sum_{\mu=1}^{m} \sum_{N_s^- \in \Omega_{1,\mu}^-} P[\mathcal{S}(t) = N_s^-] \times P[\mathcal{S}(t + \Delta t) = N_s \mid \mathcal{S}(t) = N_s^-] \\ &+ P[\mathcal{S}(t) = N_s] \times \left(1 - \sum_{\mu=1}^{m} \sum_{N_s^+ \in \Omega_{1,\mu}^+} P[\mathcal{S}(t + \Delta t) = N_s^+ \mid \mathcal{S}(t) = N_s] \right), \end{aligned} \tag{9}$$

where $\Omega_1^- = \Omega_{1,1}^- \cup \Omega_{1,2}^- \cdots \Omega_{1,m}^-$; $\Omega_1^+ = \Omega_{1,1}^+ \cup \Omega_{1,2}^+ \cdots \Omega_{1,m}^+$; $\Omega_{1,\mu}^-$ represents all the state matrices N_s^- which can transit into N_s after a μth transition; and $\Omega_{1,\mu}^+$ represents all the state matrices N_s^+ which can transit from N_s by a transition of the μth transition channel. We assume that the transition probabilities $P[\mathcal{S}(t + \Delta t) = N_s \mid \mathcal{S}(t) = N_s^-]$ and

$P[\mathcal{S}(t+\Delta t) = N_s^+ \,|\mathcal{S}(t) = N_s]$ do not explicitly depend upon time t at which the transition occurs. The fact that the transition depends only on the previous step or state is essentially the Markov assumption [25].

Now, recall that $c_\mu \Delta t$ is the probability that a selected prospective transition link of T_μ at time t will make a transition in the next infinitesimal time interval Δt. Then, we have,

$$\begin{aligned} &P[\mathcal{S}(t+\Delta t) = N_s \,|\mathcal{S}(t) = N_s^-] = c_\mu \Delta t, \;\; N_s^- \in \Omega_{1,\mu}^- \\ &P[\mathcal{S}(t+\Delta t) = N_s^+ \,|\mathcal{S}(t) = N_s] = c_\mu \Delta t, \;\; N_s^+ \in \Omega_{1,\mu}^+. \end{aligned} \tag{10}$$

Putting (10) in (9), we have

$$P[\mathcal{S}(t+\Delta t) = N_s] = \sum_{\mu=1}^{m} \sum_{N_s^- \in \Omega_{1,\mu}^-} P[\mathcal{S}(t) = N_s^-] \times c_\mu \Delta t + P[\mathcal{S}(t) = N_s] \times \left(1 - \sum_{\mu=1}^{m} \sum_{N_s^+ \in \Omega_{1,\mu}^+} c_\mu \Delta t\right). \tag{11}$$

Moreover, since $\sum_{N_s^+ \in \Omega_{1,\mu}^+} 1 = h_\mu(N_s)$, we have

$$\sum_{N_s^+ \in \Omega_{1,\mu}^+} c_\mu \Delta t = c_\mu h_\mu(N_s) \Delta t = P_\mu(\Delta t), \tag{12}$$

where $P_\mu(\Delta t)$ is the probability that a T_μ transition will occur in the time interval $(t, t+\Delta t)$. Now, putting (12) in (11) yields

$$P[\mathcal{S}(t+\Delta t) = N_s] = \sum_{\mu=1}^{m} \sum_{N_s^- \in \Omega_{1,\mu}^-} P[\mathcal{S}(t) = N_s^-] c_\mu \Delta t + P[\mathcal{S}(t) = N_s] \times \left(1 - \sum_{\mu=1}^{m} c_\mu h_\mu(N_s) \Delta t\right). \tag{13}$$

Finally, re-arranging and taking the limit as $\Delta t \to 0$, we get the general transition equation as

$$\frac{\partial P[\mathcal{S}(t) = N_s]}{\partial t} = \sum_{\mu=1}^{m} \sum_{N_s^- \in \Omega_{1,\mu}^-} c_\mu P[\mathcal{S}(t) = N_s^-] - P[\mathcal{S}(t) = N_s] \times \sum_{\mu=1}^{m} c_\mu h_\mu(N_s). \tag{14}$$

3. Stochastic simulation algorithm

The foregoing derivation has provided a general model for studying the transition dynamics of very general transition networks at the microscopic level. However, as is obvious from (14), an exact analytical solution is unlikely to be achieved, and one has to resort to numerical solution. In the 1970s, Gillespie [26,27] developed a stochastic simulation algorithm to simulate or realize a chemical master equation (CME) model, which is similar to (14). In this section, we extend the Gillespie algorithm for simulating transition dynamics of a transition network.

At time t_0, the network state is $\mathcal{S}(t_0)$. The goal of stochastic simulation is to describe the evolution of state matrix $\mathcal{S}(t)$ from some given initial state $\mathcal{S}(t_0)$. Let $P_\mu(\tau, \Delta\tau)$ represents the probability that transition T_μ will occur in the infinitesimal time interval $(t+\tau, t+\tau+\Delta\tau)$, given the system is in state $\mathcal{S}(t) = N_s$ at time t. The probability $P_\mu(\Delta\tau)$ is calculated as

$$P_\mu(\tau, \Delta\tau) = P_0(\tau) P_\mu(\Delta\tau), \tag{15}$$

where $P_0(\tau)$ is the probability that, given state $\mathcal{S}(t) = N_s$, no transition will occur in the interval $(t, t+\tau)$. From (12), we have $P_\mu(\Delta\tau) = c_\mu h_\mu \Delta\tau$. Define $a_\mu = c_\mu h_\mu$, which represents the transition propensity of the T_μ transition. There are m transition channels, from the joint distribution. The probability that any of the m transitions occurs in the interval $\Delta\tau$ is

$$a^* \Delta\tau = \sum_{\mu=1}^{m} a_\mu \Delta\tau, \tag{16}$$

where $a^* = \sum_{\mu=1}^{m} a_\mu$. Note that the probability that no transitions occur in the interval $\Delta\tau$ is $1 - a^*\Delta\tau$ and therefore

$$P_0(\tau + \Delta\tau) = P_0(\tau)\left(1 - a^* \Delta\tau\right). \tag{17}$$

Re-arranging and taking the limit $\Delta\tau \to 0$, we get

$$\frac{\mathrm{d}P_0}{\mathrm{d}\tau} = -a^* P_0. \tag{18}$$

The probability that nothing happens in zero time is one, i.e., $P_0(0) = 1$. Then, the analytical solution of (18) is given by

$$P_0(\tau) = P_0(0)\, \mathrm{e}^{-a^*\tau} = \mathrm{e}^{-a^*\tau}. \tag{19}$$

Combining (12), (15) and (19), we get

$$P_\mu(\tau, \Delta t) = P_0(\tau)P_\mu(\Delta t) = a_\mu \, \mathrm{e}^{-a^*\tau}\Delta t, \tag{20}$$

which (since $\Delta t \to 0$) can be written as

$$P_\mu(\tau) = \lim_{\Delta t\to 0} \frac{P_\mu(\tau, \Delta t)}{\Delta t} = a_\mu \, \mathrm{e}^{-a^*\tau} \tag{21}$$

where $P\mu(\tau)$ is the probability density that transition T_μ will occur in the infinitesimal time interval $(t_0 + \tau, t_0 + \tau + \Delta\tau)$, namely, the probability that at time t, a transition T_μ occurs in interval $(t_0 + \tau, t_0 + \tau + \Delta\tau)$ and no other transition occurred in the previous interval.

Now, we obtain the probability that the next transition of any type will occur in the interval $(t + \tau, t + \tau + \Delta\tau)$ by integrating (21) over all transitions, i.e.,

$$P^{(1)}(\tau)\Delta\tau = \sum_{\mu=1}^{m} a_\mu \, \mathrm{e}^{-a^*\tau}\Delta\tau = a^* \, \mathrm{e}^{-a^*\tau}\Delta\tau \Rightarrow P^{(1)}(\tau) = a^* \, \mathrm{e}^{-a^*\tau}, \tag{22}$$

where $P^{(1)}(\tau)$ is probability density function (pdf) for the transition time interval, which determines when the next transition will occur. In order to perform simulation, this has to be implemented in an algorithm. At each time step the system is in one state and we can generate a simulation algorithm to answer all the three questions posed earlier based on the analytical expression of (21) and (22).

Question 1: When will the next transition occur?

According to $P^{(1)}(\tau) = a^* \, \mathrm{e}^{-a^*\tau}$, the corresponding cumulative distribution function is defined by

$$F(\tau) = \int_{-\infty}^{\tau} P^{(1)}(t)\,\mathrm{d}t = a^* \int_0^{\tau} \mathrm{e}^{-a^*t}\,\mathrm{d}t = 1 - \mathrm{e}^{-a^*\tau}. \tag{23}$$

We can use a random number generator to generate a number r_1 in the unit interval. If we choose a value τ such that $F(\tau) = r_1$, the pdf of τ will be the one corresponding to $P^{(1)}$. The random value of τ can thus be obtained as

$$\tau = F^{-1}(r_1) = \frac{1}{a^*} \ln\left(\frac{1}{1 - r_1}\right), \tag{24}$$

which is the time interval corresponding to the next transition.

Question 2: What transition will occur?

The probability that, given a transition occurring at time τ, the transition is of type μ is equal to the conditional probability $P^{(2)}(\mu|\tau)$, which can be readily found as

$$P^{(2)}(\mu|\tau) = \frac{P(\mu, \tau)}{P^{(1)}(\tau)} = \frac{a_\mu \, \mathrm{e}^{-a^*\tau}}{a^* \, \mathrm{e}^{-a^*\tau}} = \frac{a_\mu}{a^*}. \tag{25}$$

Then, by generating a second random number r_2 in the unit interval, the type of transition that occurs at time τ will correspond to the value of μ^* that satisfies the inequality

$$\sum_{j=1}^{\mu^*-1} \frac{a_j}{a^*} \le r_2 < \sum_{j=1}^{\mu^*} \frac{a_j}{a^*}. \tag{26}$$

Question 3: Which one of the prospective transition links will be selected?

The question is answered by referring to the probability that the transition links of transition channel T_{μ^*} have the same transition probability. There are $h_{\mu^*}(N_s)$ prospective transition links. Then, the index of transition link is determined by generating a third random number r_3 in the unit interval. The selected transition link corresponds to the value of i^* ($i^* = 1, 2, \ldots, h_\mu(N_s)$), such that the following inequality is satisfied.

$$\frac{i^* - 1}{h_{\mu^*}(N_s)} \le r_3 < \frac{i^*}{h_{\mu^*}(N_s)}. \tag{27}$$

Finally, with the above information, we can update the state matrix $\mathcal{S}(t)$ to $\mathcal{S}(t + \tau)$ and generate the complete dynamical evolution of the transition network. For implementation of the algorithm, we need to compute $h_\mu(Ns)$ in the stochastic simulation. In the worst case (each node with degree N), the computation will become quite heavy, namely, the computation is roughly $O(N)$.

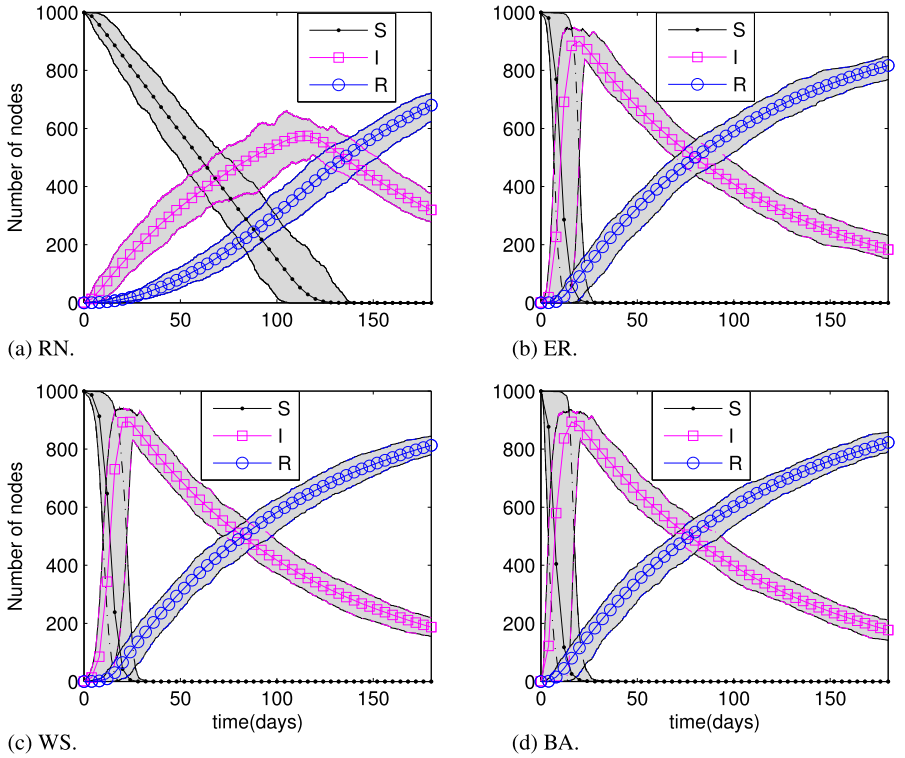

Fig. 3. (Color online) Disease propagation dynamics of the four different networks. Solid lines represent the number of S, I and R nodes, and the gray regions represent the deviation regions.

4. Experimental study of epidemic spreading

In this section, several epidemiological characteristics are studied by utilizing the proposed model to generate epidemic dynamics in several complex network settings. Among the many representative network models, the regular nearest-neighbor (RN) network (in which each node is connected to its $2m_0$ neighbors), the classical Erdös–Renyí (ER) random graph [28], the relatively new Watts–Strogaz (WS) small-world networks [29,30] and the Barabási–Albert (BA) scale-free network [31] are particularly significant and important. In this study, these networks are selected as illustrative examples. For comparison, the four types of networks all have 1000 nodes and about 10000 edges. For illustration, we set $c_1 = 0.04/\mathrm{d}$ and $c_2 = 0.01/\mathrm{d}$. But c_1 and c_2 can be any other constants larger than 0. All results are obtained by averaging over thousands independent runs for different realizations, based on the four different network models. For each simulation, the model is seeded with one randomly chosen initial infection.

4.1. Epidemic dynamics

Fig. 3 shows simulation results of the epidemic spreading dynamics for the four networks and the range of propagation dynamics in the RN network is wider than the others. The average epidemic spreading dynamics of the four different types of networks are plotted in Fig. 4(a) to (c), which reveal that the epidemic dynamics of the ER, WS and BA networks have similar behavior, while there is observable difference between the behavior of epidemic dynamics in the RN network and the other three networks.

One important aspect in epidemiology is the "epidemic curve", which is defined as the number of newly infected cases per time interval, namely, $c(t) = [S(t - \Delta t) - S(t)]/R(\infty)$, where Δt is a selected time interval and $R(\infty)$ represents the total number of infected nodes during the spreading. A typical example of the epidemic curve is provided in Fig. 4(d), which shows the rate of newly infected individuals per day (i.e., $\Delta t = 1$ day). Clearly, the epidemic curves provide information about the time of the peak of the spreading. The figure shows an interesting phenomenon. For epidemics spreading in the ER, WS and BA networks, there are "outbreak" peaks. Specifically, in a very short time, a lot of nodes are infected, while for the RN network, the infection seems more benign.

4.2. Spreading interval analysis

The simulation algorithm generates a time series $\{X(t_i), Y(t_i), Z(t_i)|i = 1, 2, \ldots\}$. At time t_i, a transition (infect or recover) process completes. From this time series, we can obtain $\{X(t_{S,1}), X(t_{S,2}), \ldots, X(t_{S,e}), t_{S,i} < t_{S,i+1}, t_i \in \Re^+\}$. At $t_{S,i}$, one more

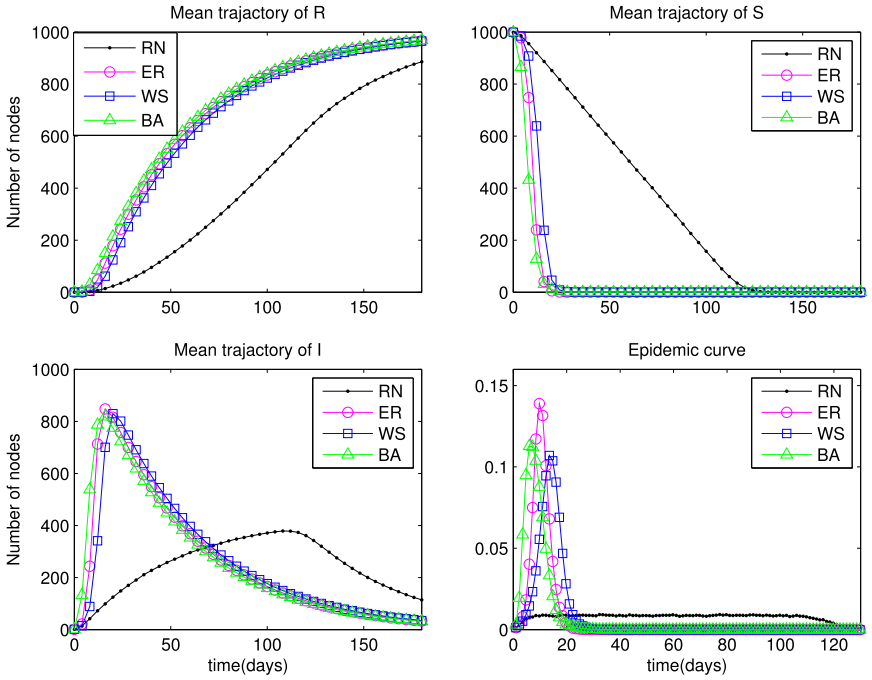

Fig. 4. (Color online) (a) Average of the R dynamics of the four different networks; (b) average of the S dynamics, (c) average of the I dynamics; (d) epidemic curves of the four different networks. y-axis is the normalized infection rate $c(t)$, and x-axis is the time.

Table 1
Average infection period and incubation period (days).

	RN	ER	WS	BA
Incubation period	5.0027 ± 1.6884	4.3943 ± 4.7191	5.0585 ± 1.7402	3.8711 ± 2.5340
Infection period	110.7929 ± 6.2407	12.0122 ± 0.6553	15.1274 ± 1.0332	12.3934 ± 0.7701

node is infected and $X(t_{S,1}) - X(t_{S,2}) = 1$. Similarly, time series $\{Y(t_{I,1}), Y(t_{I,2}), \ldots, Y(t_{I,e}), t_{I,i} < t_{I,i+1}, t_i \in \Re^+\}$ and $\{Z(t_{R,1}), Z(t_{R,2}), \ldots, Z(t_{R,e}), t_{R,i} < t_{R,i+1}, t_i \in \Re^+\}$ can be found. Here, the definition of spreading period is given as follows. Assume in the time interval $[t_0, t_e]$, a total of N susceptible nodes are infected, while in time interval $[t_{p1}, t_{pe}]$, there are $N(1-2\delta)$ susceptible nodes infected, where $0 < \delta \ll 1$. In intervals $[t_0, t_e)$ and $(t_{pe}, t_e]$, only $N\delta$ susceptible nodes are infected. In this study, we define $[t_{p1}, t_{pe}]$ as the *outbreak period*. Also, we treat interval $[t_0, t_{p1}]$ as the *incubation period* for the epidemic spreading. There is almost no infection during an incubation period. For instance, for $\delta = 0.005$, 10000 individuals are infected during the first 200 days, but 9900 individuals are infected during day 5 to day 30, while in the interval [0, 5] day and (30, 200] day, only 50 individuals are infected, respectively. Then, the outbreak period is [5, 30] day with a duration of 25 days and the incubation period is [0, 5) day with a duration of 5 days. More precisely, by plotting $\Delta t_{S,i}$ against $t_{S,i}$, we obtain the *spreading interval* profile of the four network models, as given in Fig. 5, where the x-axis is $\{t_{S,2}, t_{S,3}, \ldots, t_{S,e},\ t_{S,i} < t_{S,i+1}, t_i \in \Re^+\}$ and the y-axis is $\{\Delta t_{S,i} = t_{S,i+1} - t_{S,i},\ i = 1, 2, \ldots, e-1, t_i \in \Re^+\}$. Each stem reveals that a susceptible node is infected. The thicker the stem, the more susceptible nodes are infected. The gray region indicates the outbreak period whereas the green region indicates the incubation period.

From Fig. 5, one can find the outbreak period of the RN network, $T_{\text{per}}^{\text{RN}}$, is much longer than the others ($T_{\text{per}}^{\text{ER}}$, $T_{\text{per}}^{\text{BA}}$ and $T_{\text{per}}^{\text{EG}}$), which are shown in green. The incubation periods are shown in gray in the same figure. Table 1 shows the statistical results of the outbreak periods and the infection periods. It reveals that the WS network has the longest incubation period, while the others have a slightly shorter incubation period. However, the durations of the outbreak periods are very different. In the regular network, it requires about 110 days to infect 99% of the nodes, while in other networks, the time required is much shorter. Roughly, the outbreak period of the RN network is about 9 times longer than other three networks. Note that the outbreak period of the WS network is about 15 days, which is obviously larger than the outbreak periods of the ER and BA networks which are about 12 days. In conclusion, we have $T_{\text{per}}^{\text{BA}} \approx T_{\text{per}}^{\text{ER}} < T_{\text{per}}^{\text{WS}} \ll T_{\text{per}}^{\text{RN}}$.

From our results, the duration of the outbreak period is related to network structure. Note that RN, ER and WS networks are homogeneous, while the BA network is heterogeneous. But the outbreak period relationship is $T_{\text{per}}^{\text{RN}} \gg \{T_{\text{per}}^{\text{BA}}, T_{\text{per}}^{\text{ER}}, T_{\text{per}}^{\text{WS}}\}$. So, homogeneity or heterogeneity is not a key condition that controls the rate of epidemic spreading in a network. Moreover, the approximated average path lengths of the networks [28–30,32,33] are found as 1000, 2.4, 30 and 3.5, for the RN, ER, WS and BA networks, respectively. Comparing the relative durations of the outbreak periods, i.e., $T_{\text{per}}^{\text{BA}} \approx T_{\text{per}}^{\text{ER}} < T_{\text{per}}^{\text{WS}} \ll T_{\text{per}}^{\text{RN}}$, it

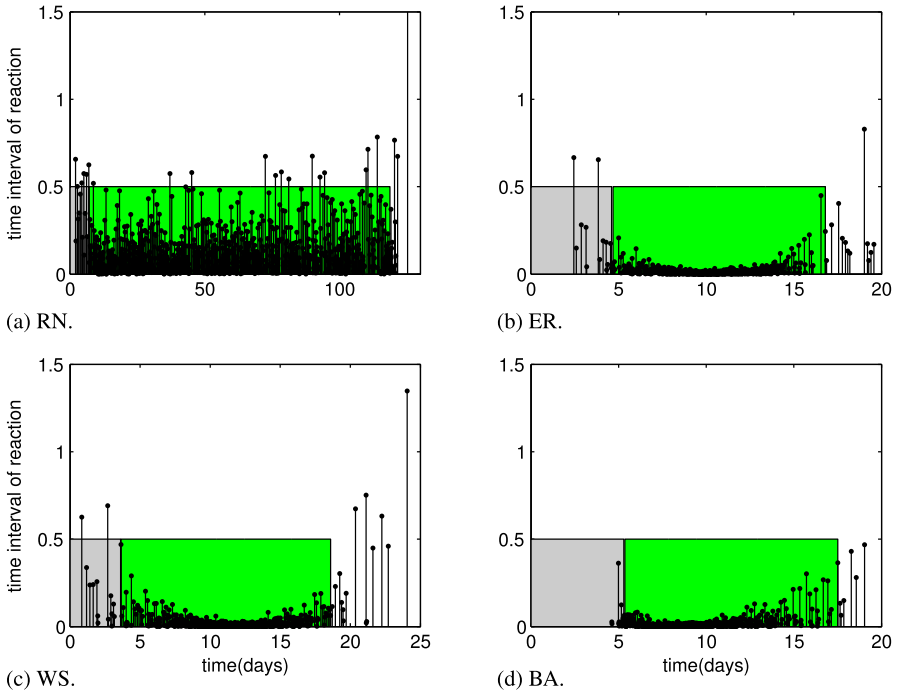

Fig. 5. Spreading period of the four types of networks. Gray region is outbreak period and green region is incubation period. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

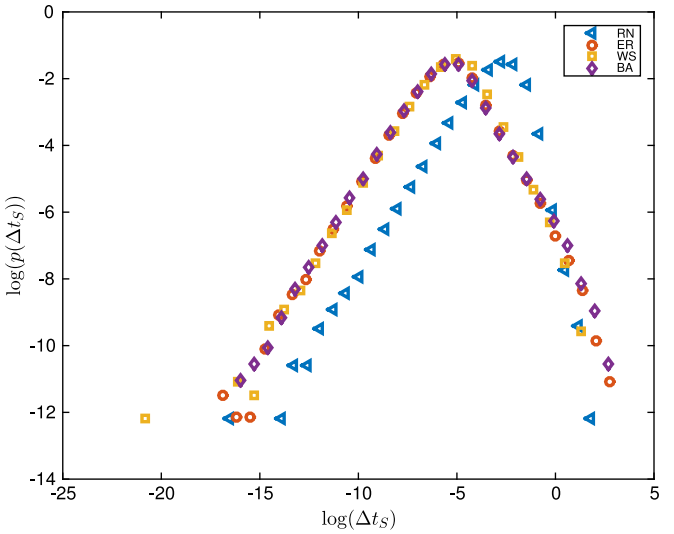

Fig. 6. (Color online) Probability density $p(\Delta t_S)$ of the time interval Δt_S of the four types of networks.

may be inferred that the average path length may be one of the key factors controlling the rate of epidemic spreading in a network. For a network with a shorter average path length, the time required for infecting the whole network is shorter. The distribution $p(\Delta t_S)$ of the time interval Δt_S is shown in Fig. 6. Results show that the four curves display a simple 'tent-shaped' form and the distribution is exponential, which has been reported in a lot of literatures [34,35].

5. Modeling Facebook user growth

Thousands of new products and services emerge every year, and most of them disappear within a few years while a small number of them make their way to the global market with overwhelming success. Facebook, for example, began in December 2004, and there were only about a million users, but after just about 8 years, in January 2013, Facebook users reached 1060 million. The rapid growth of Facebook is an interesting case relevant to the present study of transition networks. Fig. 7 shows

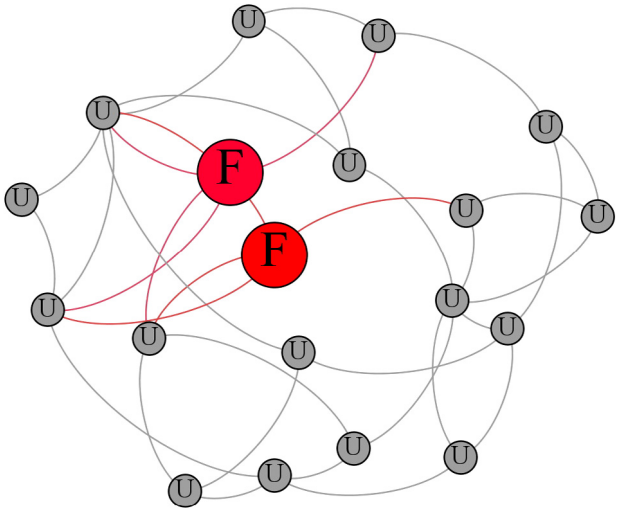

Fig. 7. (Color online) A simple social network: red nodes represent Facebook users, gray nodes are prospective users who do not join Facebook at the moment.

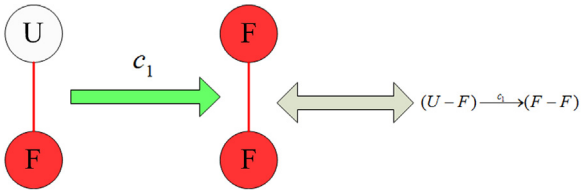

Fig. 8. (Color online) The transition channel of Facebook user growth model.

a simple local illustration, where the red nodes represent Facebook users, and the gray nodes are the prospective users who can become Facebook users. Links between two nodes imply that the two individuals have relationship, e.g., being friends, relatives, family members, etc. Each link connecting a Facebook user and a prospective user (red link in Fig. 7) is a connection along which the prospective user can "transit" into a Facebook user. This model has a simple practical resemblance as people can be "persuaded" or "infected" by friends and/or family members to become Facebook users. This transition process is shown in Fig. 8.

In this network, each node can assume either of two possible states, namely, U and F, which represent a prospective Facebook user and a Facebook user, respectively. Suppose an individual converts from U to F with a probability, per unit of time, of c_1. Then, there is $m = 1$ transition channel, i.e., T_1:

$$T_1 : (U - F) \xrightarrow{c_1} (F - F), \tag{28}$$

where $(U - F)$ is the set of prospective transition links and $(F - F)$ is the set of resulting links after transition.

To test our model, we collected data of Facebook user growth from 2004 to 2013, as shown in Fig. 9. We fit the model solutions to the data. In this study, for simplicity, one node represents 1 million people, which can be prospective users U or Facebook users F. The networks have 1200 nodes, which form a fully connected network. We elaborately select $c_1 = 2.86 \times 10^{-6}$/d. In this case, the model can provide a good prediction of the number of Facebook users. All results are obtained by averaging over thousands independent runs for different realizations. For each simulation, the model is seeded with one randomly chosen node representing one million Facebook users in December 2004. The solid line in Fig. 9(a) shows one realization of the model, which fits well with the historical data of Facebook user growth. Fig. 9(b) shows the average of 1000 simulations of Facebook user growth and the deviation. Real-world online social networks may contain millions of nodes. Based on the mean-field assumption, we can derive an ODE model that describes the user growth profile in continuous time from (28). The detailed steps of derivation are omitted for brevity, and we give the final first-order ODE model here:

$$\dot{x}(t) = c_1 N x - c_1 x^2, \tag{29}$$

where $x(t)$ is the expected number of Facebook users.

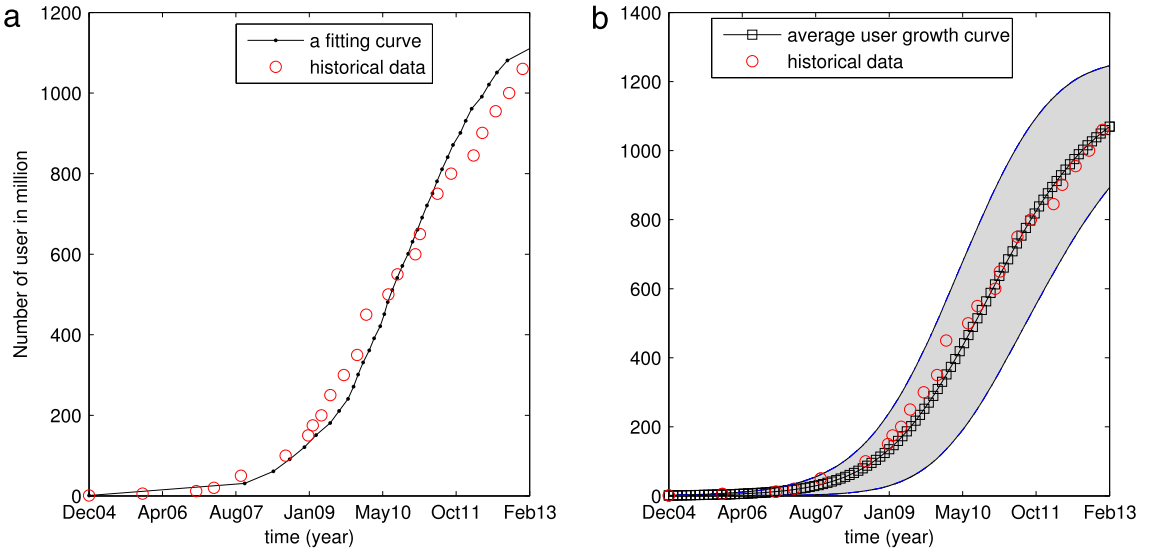

Fig. 9. (Color online) Facebook growth chart during December 2004 to December 2013. (a) Red circles are the historical data of Facebook users, while solid line represents an estimated number of Facebook users; (b) square line represents the average Facebook user growth curve, while the gray section represents the deviation region.

6. Conclusion

Based on a stochastic process, we have derived a general model that captures the dynamics of transition networks. A practical simulation algorithm is provided, allowing convenient retrieval of meaningful time profiles for the key quantities that characterize the dynamic behavior of the network. Specifically, we study the epidemiological characteristics of the disease propagation in four representative complex networks, namely, the regular nearest-neighbor network, the random graph, the small-world network, and the scale-free network. By using the proposed model, we avoid making one of the assumptions underlying standard ODE-based models, which is a homogeneous fully connected population, namely, all individuals are susceptible to the disease and all suffer an equal, small, positive probability of contracting the disease. Here, we propose an alternative to the standard SIR model. Unlike the standard SIR model, the proposed model is more realistic and capable of describing dynamics at a microscopic level, the price to pay being the increased complexity. The results suggest that the topology of the underlying transition networks profoundly affects the performance of epidemic-spreading dynamics. Results may shed light on improving control and even to eradicate the infection from a network, for instance, by carefully selecting and quarantining individuals thereby shortening the average path length which has been found to be related to infection period. Furthermore, we model the growth (spreading) of Facebook with our transition model, and show that our model well fits the historical data of Facebook user growth. Here, the model we developed assumes that the total number of potential users is fixed, and no other competing products come into play. However, in the real world, there are other social networks competing with Facebook. Our transition model can also be modified to accommodate competing transitions, and we will leave this interesting topic to a future study.

In conclusion, the proposed model is expected to find applications in a variety of real-life transition systems, such as rumor spreading, mobile communication, language evolution and grid computing. This model may be applied to communication network for data dissemination, reliable group communication or replicated database maintenance. It will also be helpful in understanding social phenomena such as the spreading of new ideas in population or the efficiency of marketing campaigns. Finally, we point out that the general model studied here can be extended to time-variant transition networks, where the network topology changes with time. We defer a more detailed report of this issue to a future work.

Acknowledgment

This work was supported by Hong Kong Polytechnic University Grant G-YBAT.

References

[1] S.H. Strogatz, Exploring complex networks, Nature 410 (6825) (2001) 268–276.
[2] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Rev. Modern Phys. 74 (1) (2002) 47.
[3] D.J. Daley, D.G. Kendall, Epidemics and rumours, Nature 204 (1964) 1118.
[4] M.E. Newman, Spread of epidemic disease on networks, Phys. Rev. E 66 (1) (2002) 016128.
[5] R.M. May, A.L. Lloyd, Infection dynamics on scale-free networks, Phys. Rev. E 64 (6) (2001) 066112.
[6] Y. Moreno, R. Pastor-Satorras, A. Vespignani, Epidemic outbreaks in complex heterogeneous networks, Eur. Phys. J. B 26 (4) (2002) 521–529.

[7] R. Pastor-Satorras, A. Vespignani, Epidemic spreading in scale-free networks, Phys. Rev. Lett. 86 (14) (2001) 3200.
[8] R. Pastor-Satorras, A. Vespignani, Epidemic dynamics and endemic states in complex networks, Phys. Rev. E 63 (6) (2001) 066117.
[9] M. Barthélemy, A. Barrat, R. Pastor-Satorras, A. Vespignani, Velocity and hierarchical spread of epidemic outbreaks in scale-free networks, Phys. Rev. Lett. 92 (17) (2004) 178701.
[10] T. Zhou, G. Yan, B.-H. Wang, Maximal planar networks with large clustering coefficient and power-law degree distribution, Phys. Rev. E 71 (4) (2005) 046141.
[11] A. Vazquez, Polynomial growth in branching processes with diverging reproductive number, Phys. Rev. Lett. 96 (3) (2006) 038702.
[12] W.S. Wang, J.W. Minett, The invasion of language: emergence, change and death, Trends Ecol. Evol. 20 (5) (2005) 263–269.
[13] J. Ke, T. Gong, W.S. Wang, Language change and social networks, Commun. Comput. Phys. 3 (4) (2008) 935–949.
[14] B. Ribeiro, Modeling and predicting the growth and death of membership-based websites, in: Proc. 23rd Int. Conf. World wide web, International World Wide Web Conferences Steering Committee, 2014, pp. 653–664.
[15] R.P. Mann, J. Faria, D.J. Sumpter, J. Krause, The dynamics of audience applause, J. R. Soc. Interface 10 (85) (2013) 20130466.
[16] D. Centola, The spread of behavior in an online social network experiment, Science 329 (5996) (2010) 1194–1197.
[17] Z. Rong, Z.-X. Wu, G. Chen, Coevolution of strategy-selection time scale and cooperation in spatial prisoner's dilemma game, Europhys. Lett. 102 (6) (2013) 68005.
[18] R.M. Anderson, R.M. May, B. Anderson, Infectious Diseases of Humans: Dynamics and Control, Vol. 28, Wiley Online Library, 1992.
[19] J.D. Murray, Mathematical Biology, Vol. 3, Springer Berlin, Berlin Heidelberg, 1993.
[20] M. Small, C.K. Tse, Small world and scale free model of transmission of sars, Int. J. Bifurc. Chaos 15 (05) (2005) 1745–1755.
[21] M. Small, C.K. Tse, D.M. Walker, Super-spreaders and the rate of transmission of the sars virus, Physica D 215 (2) (2006) 146–158.
[22] M.J. Keeling, P. Rohani, Modeling Infectious Diseases in Humans and Animals, Princeton University Press, 2008.
[23] Y. Moreno, M. Nekovee, A.F. Pacheco, Dynamics of rumor spreading in complex networks, Phys. Rev. E 69 (6) (2004) 066130.
[24] A. Barrat, M. Barthelemy, A. Vespignani, Dynamical Processes on Complex Networks, Vol. 1, Cambridge University Press, Cambridge, 2008.
[25] B. Øksendal, Stochastic Differential Equations, Springer, Berlin Heidelberg, 2003.
[26] D.T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, J. Comput. Phys. 22 (4) (1976) 403–434.
[27] D.T. Gillespie, Exact stochastic simulation of coupled chemical reactions, J. Phys. Chem. 81 (25) (1977) 2340–2361.
[28] P. Erdös, A. Rényi, On the evolution of random graphs, Publ. Math. Inst. Hungar. Acad. Sci. 5 (1960) 17–61.
[29] D.J. Watts, S.H. Strogatz, Collective dynamics of small-world networks, Nature 393 (6684) (1998) 440–442.
[30] M.E. Newman, D.J. Watts, Renormalization group analysis of the small-world network model, Phys. Lett. A 263 (4) (1999) 341–346.
[31] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.
[32] B. Bollobás, O. Riordan, Mathematical results on scale-free random graphs, in: Handbook of Graphs and Networks, Vol. 1, 2003, p. 34.
[33] R. Cohen, S. Havlin, Scale-free networks are ultrasmall, Phys. Rev. Lett. 90 (5) (2003) 058701.
[34] M.H. Stanley, L.A. Amaral, S.V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M.A. Salinger, H.E. Stanley, Scaling behaviour in the growth of companies, Nature 379 (6568) (1996) 804–806.
[35] V. Plerou, L.A.N. Amaral, P. Gopikrishnan, M. Meyer, H.E. Stanley, Similarities between the growth dynamics of university research and of competitive economic activities, Nature 400 (6743) (1999) 433–437.