

Toward Understanding Spatial Dependence on Epidemic Thresholds in Networks

Zesheng Chen

Department of Computer Science

Indiana University - Purdue University Fort Wayne, Indiana 46805

Email: chenz@ipfw.edu

Abstract—Social influence in online social networks bears resemblance to epidemic spread in networks and has been studied through epidemiological models. The epidemic threshold is a fundamental metric used to evaluate epidemic spread in networks. Previous work has shown that the epidemic threshold of a network is exactly the inverse of the largest eigenvalue of its adjacency matrix. In this work, however, we indicate that such a threshold ignores spatial dependence among nodes and hence underestimates the actual epidemic threshold. Focusing on regular graphs, we analytically derive a more accurate epidemic threshold based on spatial Markov dependence. Our model shows that the epidemic threshold indeed depends on the spatial correlation coefficient between neighboring nodes and decreases with the death rate. Through both analysis and simulations, we show that our proposed epidemic threshold incorporates a certain spatial dependence and thus achieves a greater accuracy in characterizing the actual epidemic threshold in regular graphs. Moreover, we extend our study to irregular graphs by conjecturing a new epidemic threshold and show that such a threshold performs significantly better than previous work.

I. INTRODUCTION

Information diffusion and influence spread in online social networks (OSNs) bear resemblance to epidemic process in networks, which is an active interdisciplinary research area among physics, mathematics, epidemiology, social science, and computer science [3], [8], [13], [14]. Many scenarios of information cascading in OSNs have been modeled through epidemiological models. For example, the popular linear-cascade model on social influence, studied in [9], can be described as a special *susceptible-infected-recovered* (SIR) model [17]. The cascading behavior in large blog graphs has been characterized as the classic *susceptible-infected-susceptible* (SIS) model in [11]. Moreover, the *susceptible-infected-cured* (SIC) model is proposed in [18] to study the propagation of conflict information (*e.g.*, rumor and anti-rumor) in OSNs. It is therefore of great importance to study accurate mathematical models of epidemic process in networks, which can help us design network structures, protocols, and policies to facilitate the spread of good information (*e.g.*, products, news, and innovations) and counteract the propagation of unwanted information (*e.g.*, viruses, misinformation, and rumors).

The epidemic threshold is a fundamental metric used to evaluate epidemic spread in networks [1], [5], [6], [7], [10], [12], [17], [19]. Such a threshold reflects the condition on which an infection will either die out or become epidemic.

Specifically, in the classic SIS model, a node in a network can be either susceptible or infected. If the node is infected, it can be cured and become susceptible with a death rate; otherwise, it can be infected by one of its infected neighbors with a birth rate. When the ratio between the birth rate and the death rate is greater than the epidemic threshold, the infection will become epidemic; otherwise, it will die out.

An important discovery on epidemic spread is that in the SIS model, the epidemic threshold for a network can be shown to be exactly the inverse of the largest eigenvalue of its adjacency matrix [1], [6], [12]. The process of deriving this threshold, however, assumes that the status of nodes in the network are independent of each other. Such a spatial independence assumption can lead to overestimating the spreading ability of an infection [2], [12]. Intuitively, the status of nodes in a network are positively correlated, and two neighboring nodes tend to be either both infected or both susceptible. Moreover, through simulations, Givan *et al.* found that the epidemic threshold from [1], [6], [12] cannot accurately reflect the actual epidemic threshold in some types of networks [7].

The goal of this work is to find a more accurate epidemic threshold in networks. Specifically, we attempt to answer the following questions:

- Can spatial dependence among nodes affect the epidemic threshold? If so, how significantly?
- How can we derive a more accurate epidemic threshold, taking into consideration a certain spatial dependence?
- Can the birth rate and the death rate affect the spatial dependence and thus the epidemic threshold? If so, how?

To answer these questions, we apply a mathematical modeling method and focus on approximating the complex spatial dependence among nodes in a network. Specifically, we apply a spatial Markov dependence assumption and derive a closed-form expression for epidemic thresholds in regular graphs. We then use simulations to evaluate the performance of our proposed epidemic thresholds. We summarize our discoveries and contributions in the following:

- We find that spatial dependence among nodes affects the epidemic threshold significantly and show that the epidemic threshold in a network depends on not only the largest eigenvalue of its adjacency matrix, but also the spatial correlation coefficient between neighboring nodes.
- We derive a new epidemic threshold in regular graphs

based on the assumption of spatial Markov dependence. Through extensive simulation studies, we show that our proposed threshold better reflects the actual epidemic threshold than the threshold from [1], [6], [12].

- We show, through both analysis and simulations, that the epidemic threshold also depends on the death rate. It is noted that the death rate is often assumed to be 1 in previous work [5], [6]. However, we find that for regular graphs with a relative small average nodal degree, the epidemic threshold decreases when the death rate increases.

In this work, we focus our analysis on regular graphs for two reasons: (1) It has been shown that the epidemic threshold proposed in [1], [6], [12] does not work well in regular graphs [7]. (2) More importantly, due to the symmetric property of regular graphs, we can derive a closed-form expression for the epidemic threshold. However, we also extend our study to irregular graphs and conjecture a new epidemic threshold for arbitrary networks. We apply simulations to evaluate the performance of the new epidemic threshold in a ring, an ER random graph, and a power-law topology in Section V.

Although we focus only on the SIS model in this paper, our conclusions on epidemic thresholds can be well extended to other epidemiological models, such as arbitrary cascade models studied in [17].

The remainder of this paper is structured as follows. Section II introduces the system model. Section III derives a new epidemic threshold in regular graphs, and Section IV evaluates the performance of our proposed threshold and compares it with previous work and simulation results. Section V extends our study to arbitrary networks. Finally, Section VI concludes this paper.

II. SYSTEM MODEL

We use $G(V, E)$ to represent a network, where V is the set of nodes and E is the set of edges (or links). Specifically, we consider undirected graphs, *i.e.*, if an edge $(i, j) \in E$, then $(j, i) \in E$. Let $N_i = \{j | (j, i) \in E\}$ be a neighborhood of node i . In this work, we focus our analysis on regular graphs, *i.e.*, the size of the neighborhood is the same for all nodes. That is, $|N_i| = k$, for $\forall i \in V$, where k denotes the average nodal degree. Typical regular graphs include ring, lattice, and complete graphs, where $k = 2, 4$, and $|V| - 1$, respectively.

We study the problem of epidemic spread in a network using the classic SIS model. Specifically, a node or a computer in a network can be either infected or susceptible. A susceptible node can be infected by one of its already infected neighbors with a birth rate (or infection rate) β , where $0 < \beta \leq 1$. On the other hand, an infected node can be cured and change back to be susceptible with a death rate (or curing rate) δ , where $0 < \delta \leq 1$. As applied in [1], [2], [6], [12], we assume that the birth rate (or the death rate) is the same for all nodes and does not change with time.

Let $\tau = \beta/\delta$, the ratio between the birth rate and the death rate. The epidemic threshold, τ_c , is defined as when $\tau \leq \tau_c$, the epidemic dies out, and no node is infected; and when $\tau > \tau_c$,

a nonzero fraction of nodes remain infected for a long time. In previous work [1], [6], [12], it has been shown that the epidemic threshold is

$$\tau_{c,ind} = \frac{1}{\lambda_{max}(A)}, \quad (1)$$

where $\lambda_{max}(A)$ is the largest eigenvalue of the adjacency matrix A of the network. Moreover, when a regular graph is considered, $\lambda_{max}(A) = k$, and thus

$$\tau_{c,ind} = \frac{1}{k}. \quad (2)$$

However, this threshold is derived based on the assumption of independence among nodes and has been shown to be unable to accurately capture the actual epidemic threshold [7].

III. EPIDEMIC THRESHOLDS IN REGULAR GRAPHS

In this section, we first present a general mathematical framework, and then derive the epidemic threshold in regular graphs with the assumption of spatial independence. Finally, we apply the spatial Markov assumption to obtain a new epidemic threshold in regular graphs.

A. Mathematical Framework

We consider a discrete-time system and refer to the model presented by Chen and Ji in [2] as the starting point. Let $X_i(t)$ be the status of node i at time t , where $X_i(t) = 1$ if node i is infected at time t and $X_i(t) = 0$ otherwise. If node i is infected at time t , it will become susceptible with probability (or death rate) δ at time $t + 1$, *i.e.*,

$$P(X_i(t+1) = 0 | X_i(t) = 1) = \delta. \quad (3)$$

On the other hand, if node i is susceptible at time t , it will be infected by its infected neighbors with probability

$$I_i(t) = P(X_i(t+1) = 1 | X_i(t) = 0). \quad (4)$$

Thus, the status of node i at time $t + 1$ can be derived based on its status at time t and $I_i(t)$, *i.e.*,

$$P(X_i(t+1) = 1) = P(X_i(t) = 1)(1 - \delta) + P(X_i(t) = 0)I_i(t). \quad (5)$$

Since an infected node will infect its susceptible neighbor with birth rate β , the probability that susceptible node i is *not* infected by its neighbor j at time $t + 1$ is $(1 - \beta)^{x_j(t)}$, where $x_j(t) \in \{0, 1\}$ is the realization of $X_j(t)$. Hence, given node i is susceptible at time t and the status of its neighbors, the probability that it becomes infected at time $t + 1$ is $1 - \prod_{j \in N_i} (1 - \beta)^{x_j(t)}$. Let $\mathbf{x}_{N_i} = \{x_j(t) | j \in N_i\}$. Then,

$$\begin{aligned} I_i(t) &= \sum_{\mathbf{x}_{N_i}(t)} P(X_i(t+1) = 1, \mathbf{X}_{N_i}(t) = \mathbf{x}_{N_i}(t) | X_i(t) = 0) \\ &= \sum_{\mathbf{x}_{N_i}(t)} P(\mathbf{X}_{N_i}(t) = \mathbf{x}_{N_i}(t) | X_i(t) = 0) \left[1 - \prod_{j \in N_i} (1 - \beta)^{x_j(t)} \right] \\ &= 1 - E \left[\prod_{j \in N_i} (1 - \beta)^{X_j(t)} \middle| X_i(t) = 0 \right]. \end{aligned} \quad (6)$$

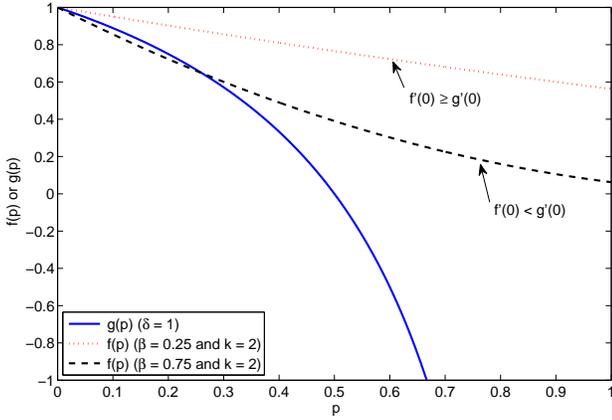


Fig. 1. Demonstration of the functions of $f(p)$ and $g(p)$.

Note that Equation (6) can be applied to arbitrary topologies, including both regular and irregular graphs. The difficulties of finding a closed-form expression to this equation lie in two aspects: (1) Spatial dependence between nodes, *i.e.*, $X_i(t)$ and $X_j(t)$'s are not independent. Intuitively, they are positively correlated [2], [12]. Moreover, given the status of node i , the status of its neighbors are also not independent. (2) Product form inside the expectation. When β is very small, $\prod_{j \in N_i} (1 - \beta)^{X_j(t)} \approx 1 - \beta \sum_{j \in N_i} X_j(t)$, which can simplify the derivation and has been applied in previous work [1], [6], [12]. However, when β is not small, such an approximation is obviously not accurate.

B. Independent Model

Assuming spatial independence between nodes, *i.e.*, $X_i(t)$ and $X_j(t)$'s are independent, we have

$$\begin{aligned} I_i(t) &= 1 - \prod_{j \in N_i} E \left[(1 - \beta)^{X_j(t)} \right] \\ &= 1 - \prod_{j \in N_i} [1 - \beta P(X_j(t) = 1)]. \end{aligned} \quad (7)$$

Set $p_{i,t} = P(X_i(t) = 1)$. Putting Equation (7) into Equation (5), we have

$$p_{i,t+1} = 1 - \delta p_{i,t} - (1 - p_{i,t}) \prod_{j \in N_i} (1 - \beta p_{j,t}). \quad (8)$$

We consider the steady state. Set $p_i = \lim_{t \rightarrow \infty} p_{i,t}$, so the above equation becomes

$$p_i = 1 - \delta p_i - (1 - p_i) \prod_{j \in N_i} (1 - \beta p_j). \quad (9)$$

When a regular graph is considered, due to its symmetric property, we have $p_i = p_j = p$, for $\forall i, j$. That is,

$$p = 1 - \delta p - (1 - p)(1 - \beta p)^k, \quad (10)$$

which leads to

$$(1 - \beta p)^k = \frac{1 - (\delta + 1)p}{1 - p}. \quad (11)$$

Set $f(p) = (1 - \beta p)^k$ and $g(p) = \frac{1 - (\delta + 1)p}{1 - p}$. Then the solutions to the above equation are the intersection points between curves $f(p)$ and $g(p)$. Note that $f(0) = g(0) = 1$, $f(1) \geq 0$, and $\lim_{p \rightarrow 1} g(p) \rightarrow -\infty$. Moreover, $f'(p) = -\beta k(1 - \beta p)^{k-1} < 0$, and $g'(p) = -\frac{\delta}{(1-p)^2} < 0$. Thus, whether Equation (11) has a non-zero solution depends on the slopes of $f(p)$ and $g(p)$ at $p = 0$. Figure 1 demonstrates two examples of $f(p)$ and how the slopes at $p = 0$ affect the intersection points between $f(p)$ and $g(p)$. Hence, $f'(0) < g'(0)$ for a non-zero solution in Equation (11). That is, $\frac{\beta}{\delta} > \frac{1}{k}$. Hence,

$$\tau_{c,ind} = \frac{1}{k}, \quad (12)$$

which is identical to Equation (2) and has been shown in [1], [6], [12] for the epidemic threshold in regular graphs. Here, we apply a different approach to obtain the same result.

C. Markov Model

Inspired by the local Markov property of Markov Random Field (MRF) [21], we assume spatial Markov dependence, *i.e.*, $X_j(t)$'s are independent given $X_i(t) = 0$. Then, we have

$$\begin{aligned} I_i(t) &= 1 - \prod_{j \in N_i} E \left[(1 - \beta)^{X_j(t)} \mid X_i(t) = 0 \right] \\ &= 1 - \prod_{j \in N_i} [1 - \beta P(X_j(t) = 1 \mid X_i(t) = 0)]. \end{aligned} \quad (13)$$

Set $p_{i,t} = P(X_i(t) = 1)$ and $p_{j|i,t} = P(X_j(t) = 1 \mid X_i(t) = 0)$. Putting Equation (13) into Equation (5), we have

$$p_{i,t+1} = 1 - \delta p_{i,t} - (1 - p_{i,t}) \prod_{j \in N_i} (1 - \beta p_{j|i,t}). \quad (14)$$

Similarly, we consider the steady state. Set $p_i = \lim_{t \rightarrow \infty} p_{i,t}$ and $p_{j|i} = \lim_{t \rightarrow \infty} p_{j|i,t}$, so the above equation becomes

$$p_i = 1 - \delta p_i - (1 - p_i) \prod_{j \in N_i} (1 - \beta p_{j|i}). \quad (15)$$

Consider the regular graph with an average nodal degree of k . Due to the symmetric property of regular graphs, we can set $p = p_i$ for $\forall i$, and $q = p_{j|i}$ for $\forall (i, j) \in E$. Then,

$$p = 1 - \delta p - (1 - p)(1 - \beta q)^k. \quad (16)$$

Define ρ as the spatial correction coefficient between neighboring nodes i and j (*i.e.*, $\forall (i, j) \in E$) in steady state. Setting $X_i = \lim_{t \rightarrow \infty} X_i(t)$, $\forall i \in V$, we have

$$\begin{aligned} \rho &= \frac{E[X_i X_j] - E[X_i]E[X_j]}{\sqrt{\text{Var}[X_i] \text{Var}[X_j]}} \\ &= \frac{P(X_i = 1, X_j = 1) - p^2}{p(1 - p)}. \end{aligned} \quad (17)$$

Thus,

$$E[X_i X_j] = P(X_i = 1, X_j = 1) = p(1 - p)\rho + p^2, \quad (18)$$

which leads to

$$\begin{aligned} q &= \frac{P(X_j = 1) - P(X_j = 1, X_i = 1)}{P(X_i = 0)} \\ &= (1 - \rho)p. \end{aligned} \quad (19)$$

Therefore, Equation (16) becomes

$$p = 1 - \delta p - (1 - p)[1 - \beta(1 - \rho)p]^k. \quad (20)$$

That is,

$$[1 - \beta(1 - \rho)p]^k = \frac{1 - (\delta + 1)p}{1 - p}. \quad (21)$$

Note that when $\rho = 0$, which means that neighboring nodes are independent, Equation (21) is reduced to Equation (11).

Set $h(p) = [1 - \beta(1 - \rho)p]^k$ and $g(p) = \frac{1 - (\delta + 1)p}{1 - p}$. Then the solutions to Equation (21) are the intersection points between curves $h(p)$ and $g(p)$. Note that $h(0) = g(0) = 1$, $h(1) \geq 0$, and $\lim_{p \rightarrow 1} g(p) \rightarrow -\infty$. Moreover, $h'(p) = -\beta k(1 - \rho)[1 - \beta(1 - \rho)p]^{k-1} < 0$, and $g'(p) = -\frac{\delta}{(1-p)^2} < 0$. Thus, whether Equation (21) has a non-zero solution depends on the slopes of $h(p)$ and $g(p)$ at $p = 0$. That is, $h'(0) < g'(0)$ for a non-zero solution of Equation (21). That is,

$$\frac{\beta}{\delta} > \frac{1}{k(1 - \rho)}. \quad (22)$$

Hence, the epidemic threshold is

$$\tau_{c,mar} = \frac{1}{k(1 - \rho)}. \quad (23)$$

When $\rho > 0$, $\tau_{c,mar} > \tau_{c,ind}$ for the same regular graph. That is, ignoring the spatial dependence among nodes, $\tau_{c,ind}$ underestimates the actual epidemic threshold. On the other hand, $\tau_{c,mar}$ incorporates a certain spatial dependence and depends on the correlation coefficient between neighboring nodes.

D. Spatial Correlation

To find the epidemic threshold in Equation (23), we derive the correlation coefficient ρ using the spatial Markov assumption. Specifically, set $P_{uv}(t) = P(X_i(t) = u, X_j(t) = v)$, where $u, v \in \{0, 1\}$, and $q_{uv}(t) = P(X_i(t+1) = 1, X_j(t+1) = 1 | X_i(t) = u, X_j(t) = v)$ for simplifying the notation. Thus,

$$P_{11}(t+1) = \sum_{u,v \in \{0,1\}} [P_{uv}(t)q_{uv}(t)], \quad (24)$$

where

$$\begin{aligned} q_{11}(t) &= (1 - \delta)^2 \\ q_{10}(t) &= (1 - \delta) \left[1 - (1 - \beta) \prod_{l \in N_j - \{i\}} (1 - \beta p_{l|j,t}) \right] \\ q_{01}(t) &= (1 - \delta) \left[1 - (1 - \beta) \prod_{l \in N_i - \{j\}} (1 - \beta p_{l|i,t}) \right] \\ q_{00}(t) &= \left[1 - \prod_{l \in N_j - \{i\}} (1 - \beta p_{l|j,t}) \right] \left[1 - \prod_{l \in N_i - \{j\}} (1 - \beta p_{l|i,t}) \right] \end{aligned}$$

The derivation of $q_{uv}(t)$'s is based on the spatial Markov assumption [2] and is in the same spirit of deriving $I_i(t)$ in Equation (13). Setting $P_{uv} = \lim_{t \rightarrow \infty} P_{uv}(t)$, putting the

above equations into Equation (24), and considering $t \rightarrow \infty$ and regular graphs, we have

$$\begin{aligned} P_{11} &= P_{11}(1 - \delta)^2 + P_{00} \left[1 - (1 - \beta q)^{k-1} \right]^2 + \\ &\quad (P_{10} + P_{01})(1 - \delta) \left[1 - (1 - \beta)(1 - \beta q)^{k-1} \right]. \end{aligned} \quad (25)$$

Note that $P_{01} = P_{10} = p - P_{11}$, $P_{00} = 1 - 2p + P_{11}$, and $q = (1 - \rho)p$. Thus, setting $r = (1 - \beta q)^{k-1} = [1 - \beta(1 - \rho)p]^{k-1}$, we have

$$\begin{aligned} P_{11} &= P_{11}(1 - \delta)^2 + (1 - 2p + P_{11})(1 - r)^2 + \\ &\quad 2(p - P_{11})(1 - \delta)[1 - (1 - \beta)r]. \end{aligned} \quad (26)$$

Since we are interested in the epidemic threshold, we focus on the case when p is very small and approaches 0 from the right. When $p \rightarrow 0^+$, from Equation (18) $P_{11} \rightarrow \rho p$ by ignoring the higher order of p . Similarly, $r \rightarrow 1 - (k-1)(1 - \rho)\beta p$. Thus, when $p \rightarrow 0^+$, the item $(1 - 2p + P_{11})(1 - r)^2$ is in the order of p^2 and does not contain constant or p term, and we can safely ignore it. Moreover, the item $1 - (1 - \beta)r$ becomes β by ignoring the higher order of p . Therefore, Equation (26) becomes

$$(2\delta - \delta^2)P_{11} = 2\beta(1 - \delta)(p - P_{11}). \quad (27)$$

Putting $P_{11} = \rho p$ into the above equation, we have

$$\rho = \frac{2\beta(1 - \delta)}{(2\delta - \delta^2) + 2\beta(1 - \delta)}, \quad (28)$$

when $p \rightarrow 0^+$, *i.e.*, for the case of the epidemic threshold.

It is noted that from Equation (28), $0 \leq \rho < 1$. Specifically, when $\delta = 1$, $\rho = 0$. That is, if we assume that the death rate is 1 as proposed in previous work such as [5], [6], there is no spatial dependence from our analysis. On the other hand, however, if $\delta < 1$, $\rho > 0$, *i.e.*, there are spatial correlations between neighboring nodes, which can affect the epidemic threshold. Since $\frac{\partial \rho}{\partial \delta} < 0$ and $\frac{\partial \rho}{\partial \beta} > 0$, ρ increases when δ decrease or β increases. Another observation is that our derived ρ is independent of the number of neighbors, *i.e.*, k .

Putting Equation (28) into Equation (22), we find the epidemic threshold in regular graphs:

$$\tau_{c,mar} = \frac{2 - \delta}{2(k - 1) - (k - 2)\delta}. \quad (29)$$

It can be seen that if $\delta = 1$, $\tau_{c,mar} = \tau_{c,ind}$. If $\delta < 1$, however, $\tau_{c,mar} > \tau_{c,ind}$. Moreover, since $\frac{d\tau_{c,mar}}{d\delta} < 0$, $\tau_{c,mar}$ increases when δ decreases. Thus, $0 < \delta \leq 1$ leads to

$$\frac{1}{k} \leq \tau_{c,mar} < \frac{1}{k-1}. \quad (30)$$

That is, the epidemic threshold in regular graphs is in $[\frac{1}{k}, \frac{1}{k-1})$.

IV. SIMULATION RESULTS AND PERFORMANCE EVALUATION

In this section, we evaluate the performance of the estimation of the epidemic threshold in regular graphs through simulations.

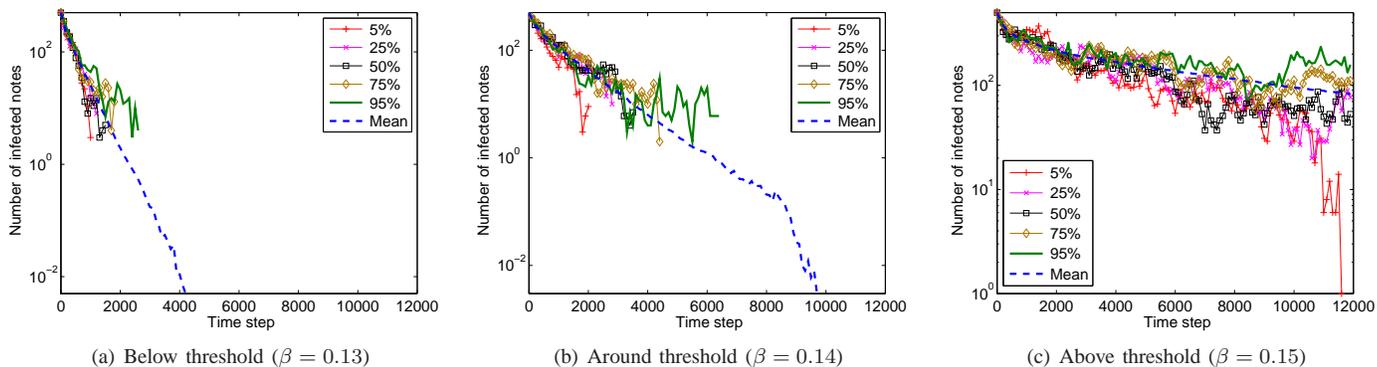


Fig. 2. Sample runs of epidemic spread in a ring graph ($\delta = 0.1$ and $|V| = 1000$).

A. Simulation Setup

We simulate the spread of epidemics with different birth rates and death rates in regular graphs such as ring, lattice, and complete graphs. The simulator is based on discrete time. In each time step, if node i is infected, it will become susceptible with the probability of δ at the next time step; otherwise, it will be infected by its infected neighbor j with the probability of β . Here, probabilities are created by a random number generator.

At the beginning of simulations, we assign half of nodes to be infected. Specifically, we assume that $|V|$ is even and node identifier is between 0 and $|V| - 1$. For ring graphs, node i 's neighbors are nodes $(i - 1) \bmod |V|$ and $(i + 1) \bmod |V|$. We assign nodes with even identifier (*i.e.*, nodes 0, 2, \dots , $V - 2$) to be infected initially. For lattice graphs, we use (i, j) to denote node's location, where $0 \leq i, j \leq m - 1$ and $m^2 = |V|$. Node (i, j) has four neighbors: nodes $((i - 1) \bmod m, j)$, $((i + 1) \bmod m, j)$, $(i, (j - 1) \bmod m)$, and $(i, (j + 1) \bmod m)$. We assign initially infected nodes in the following way: if i is even, choose nodes with j being even; otherwise, choose nodes with j being odd. For complete graphs, all nodes connect to each other. We select nodes 0, 1, \dots , $|V|/2 - 1$ to be infected initially.

We run each simulation long enough so that it reaches the steady state. Specifically, we use 12000 time steps. For each scenario, we run 1000 times using different seeds and average the number of final infections over these 1000 runs. To find the epidemic threshold (*i.e.*, τ_c) for a given death rate, we apply binary search as described in Algorithm 1, where ϵ is a very small number (*e.g.*, $\epsilon = 10^{-4}$). In the input of the algorithm, β_{low} is a case when the epidemic dies out, whereas β_{high} is a case when the epidemic survives.

Figure 2 shows sample runs of epidemic spread in a ring graph with a fixed death rate (*i.e.*, $\delta = 0.1$) and different birth rates (*i.e.*, $\beta = 0.13, 0.14, \text{ and } 0.15$). The figure plots how the number of infected nodes changes with time. In each subfigure, the “5%” curve indicates that the epidemic spreads no faster than this curve in 50 out 1000 simulation runs. The similar definition is applied to the “25%”, “50%”, “75%”, and “95%” curves. Moreover, the “mean” curve is the average over

Algorithm 1 Finding epidemic threshold τ_c

Input: $\delta, \beta_{low}, \beta_{high}$

Output: τ_c

while $\beta_{high} - \beta_{low} > \epsilon$ **do**

$\beta = (\beta_{high} + \beta_{low})/2$

Simulate epidemic spread using β and δ

Average the number of final infections over 1000 runs and get avg_inf_num

if $avg_inf_num > 0$ **then**

$\beta_{high} = \beta$

else

$\beta_{low} = \beta$

end if

end while

$\tau_c = \beta_{low}/\delta$

1000 runs. It can be seen that in the “mean” curve, the number of infected nodes can be under 1 for some time steps. This is because, in these time steps, the infection either has died out for some runs among 1000 runs or has a small number of infected nodes for other runs. Note that we use the log scale for the y-axis to make the spread process more visible. It can be seen that when β is small and thus the ratio between β and δ is below the epidemic threshold, the infection dies out quickly with an exponential rate. When the ratio between β and δ is around the epidemic threshold, the infection still dies out, but with a much slower rate. When the ratio between β and δ is above the epidemic threshold, a nonzero fraction of nodes are infected, and the infection becomes epidemic.

B. Performance Evaluation

We compare the performance of two estimators from Equations (12) and (29) (*i.e.*, $\tau_{c,ind}$ and $\tau_{c,mar}$) with simulation results. Figure 3 shows epidemic thresholds with different death rates ($0 < \delta \leq 1$) for ring, lattice, and complete graphs. It can be seen that $\tau_{c,mar}$ is a more accurate estimator than $\tau_{c,ind}$. For example, in ring graphs, when $\delta = 0.1$, the actual epidemic threshold is 1.4, whereas $\tau_{c,mar} = 0.95$ and $\tau_{c,ind} = 0.5$. There is about 50% performance improvement

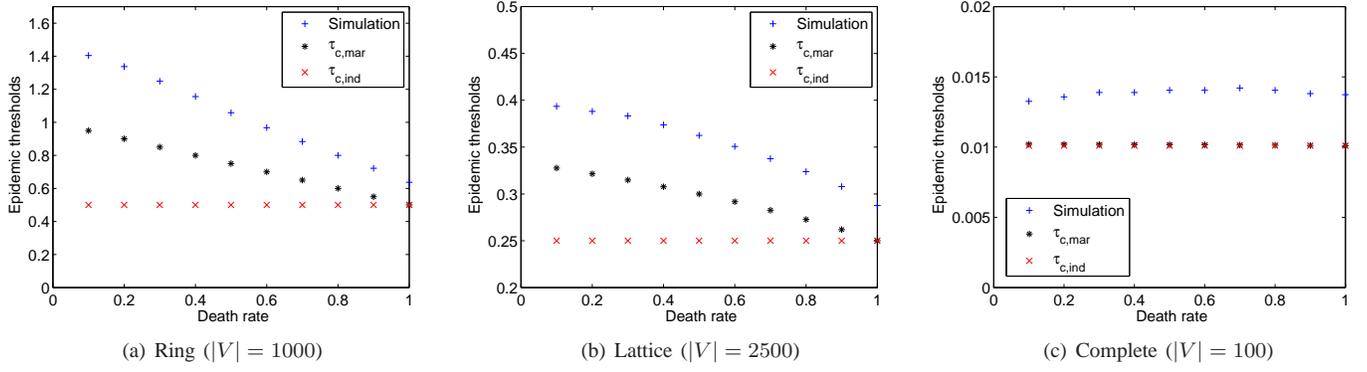


Fig. 3. Epidemic thresholds of regular graphs.

from the independence model to the Markov model. Therefore, the spatial independence assumption, which has been widely applied in previous work [1], [6], [12], significantly underestimates the epidemic threshold, whereas our proposed Markov model can incorporate a certain spatial dependence and predict the threshold more accurately. Moreover, while $\tau_{c,ind}$ is independent of δ , our model is able to catch the tendency changes with δ . For example, as indicated by Figures 3(a) and 3(b), both simulation results and our model show that the epidemic threshold decreases when the death rate increases.

We further define relative errors of the estimation for $\tau_{c,ind}$ as

$$\epsilon_{ind} = \frac{|\tau_{c,ind} - \tau_c|}{\tau_c}. \quad (31)$$

Similarly, we can define ϵ_{mar} for $\tau_{c,mar}$. Figure 4 shows how relative errors change with the death rate for three regular graphs. We can see that for all cases, $\epsilon_{mar} < \epsilon_{ind}$ when $\delta < 1$. When δ is small in ring and lattice graphs, ϵ_{mar} is about half of ϵ_{ind} . On the other hand, when δ is large or the graph is a complete graph, ϵ_{mar} is slightly better than ϵ_{ind} . Our evaluation shows that the epidemic threshold depends heavily on the spatial correlation between neighboring nodes.

V. EPIDEMIC THRESHOLDS IN ARBITRARY NETWORKS

We further extend our study on epidemic thresholds to irregular graphs, such as ER random graphs [4] and power-law topologies [15], [16]. Based on Equations (1) and (23), we conjecture that the epidemic threshold of an arbitrary network is

$$\tau_{c,cor} = \frac{1}{\lambda_{max}(A)(1 - \rho_e)}, \quad (32)$$

where $\lambda_{max}(A)$ is the largest eigenvalue of the adjacency matrix A of the network and ρ_e is the average spatial correlation coefficient between neighboring nodes in steady state at the epidemic threshold.

We verify our conjecture through simulations. It is noted that at the epidemic threshold, the epidemic dies out, and thus we cannot calculate the spatial correlation coefficient from simulations. Moreover, when β/δ is just above the epidemic

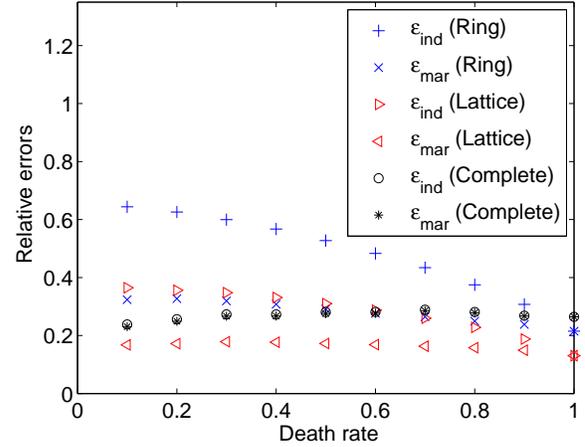


Fig. 4. Relative errors of epidemic thresholds in regular graphs.

threshold and the average number of final infections is close to 0, there are few samples of infected nodes, which makes the estimated correlation coefficient inaccurate. On the other hand, when the average number of final infections is large, the calculated coefficient is accurate, but may be very different from ρ_e in Equation (32). To obtain a reasonable estimate of ρ_e , we apply the value of the correlation coefficient when the average number of final infections is about 1. The details of finding spatial correlation coefficient ρ_e are given in Algorithm 2, where τ_c is the epidemic threshold when the death rate is δ , and β_s is a small value to increase the birth rate (e.g., $\beta_s = 0.0005$).

Figure 5 compares the analytical result of the spatial correlation from Equation (28) with the empirical spatial correlation from Algorithm 2 in ring graphs with 1000 nodes and different death rates ($0 < \delta \leq 1$). It can be seen that although there is a clear gap between theoretical and empirical results, the analytical results based on the Markov assumption are able to capture a certain spatial correlation and reflect the tendency of how correlation coefficients vary with death rates. In other words, while the correlation coefficient is always 0 in the inde-

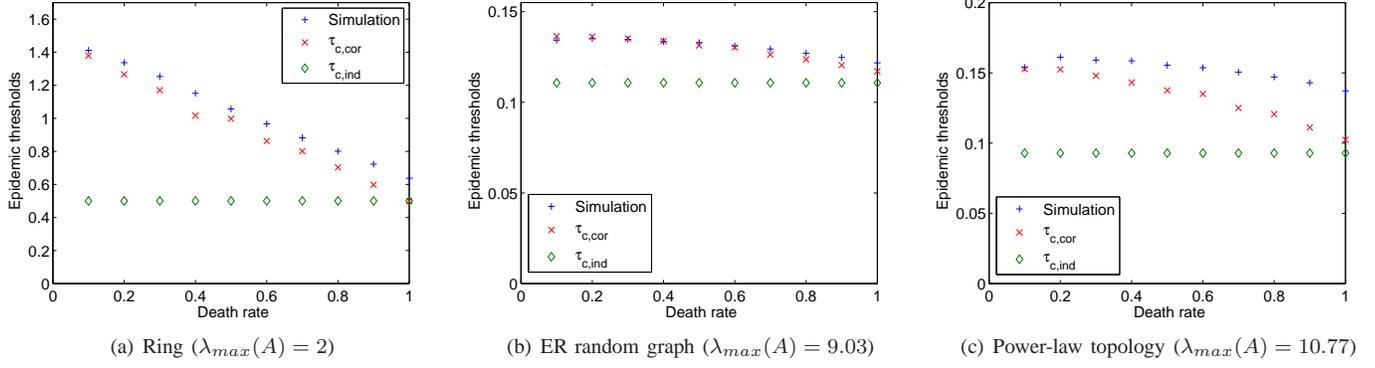


Fig. 6. Epidemic thresholds in an arbitrary network ($|V| = 1000$).

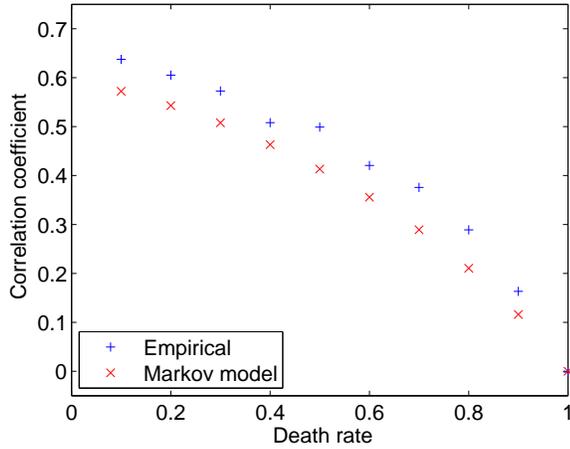


Fig. 5. Correlation coefficients at the epidemic threshold in ring graphs ($|V| = 1000$).

pendent model, the Markov model advances our understanding in spatial dependence by considering the interactions between neighboring nodes. Moreover, Figure 5 shows that when the death rate is less than 0.7, the empirical spatial correlation coefficient ρ_e is more than 0.4, which indicates the relatively strong correlations between neighboring nodes.

We compare the performance of two estimators from Equations (1) and (32) (*i.e.*, $\tau_{c,ind}$ and $\tau_{c,cor}$) with simulation results in Figure 6. The networks studied include a ring graph, an ER random graph, and a power-law topology, all with 1000 nodes. The ER random graph is with the average nodal degree of 7.91 and $\lambda_{max}(A) = 9.03$. The power-law topology is generated by BRITE [20] and is with the average nodal degree of 3.99 and $\lambda_{max}(A) = 10.77$. Similar to the simulation setup for regular graphs in Section IV, we run each scenario 1000 times with different seeds and apply the binary search as described in Algorithm 1 to find the epidemic threshold. At the beginning of simulations, we select randomly half of nodes to be infected. From Figure 6, it can be seen that $\tau_{c,cor}$ performs significantly better than $\tau_{c,ind}$ in estimating

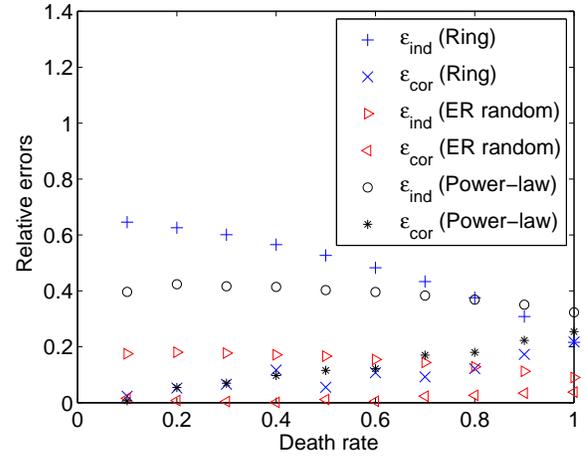


Fig. 7. Relative errors of epidemic thresholds in arbitrary networks.

Algorithm 2 Finding spatial correlation coefficient ρ_e

Input: δ, τ_c, β_s

Output: ρ_e

Set $\beta = \tau_c \times \delta$ and $found = 0$

while $found = 0$ **do**

 Simulate epidemic spread using β and δ

 Average the number of final infections over 1000 runs and get avg_inf_num

 Average the correlation coefficient over 1000 runs and get ρ

if $avg_inf_num \geq 1$ **then**

$\rho_e = \rho$

$found = 1$

end if

$\beta = \beta + \beta_s$

end while

the actual epidemic threshold in an arbitrary network. For example, in the power-law topology, when $\delta = 0.5$, the actual epidemic threshold is 0.156, whereas $\tau_{c,cor} = 0.138$ and $\tau_{c,ind} = 0.093$.

Similar to ϵ_{ind} in Equation (31), we define relative errors of the estimation for $\tau_{c,cor}$ as

$$\epsilon_{cor} = \frac{|\tau_{c,cor} - \tau_c|}{\tau_c}. \quad (33)$$

Figure 7 shows how relative errors vary with the death rate for three networks. We can see that $\epsilon_{cor} < \epsilon_{ind}$ for all cases when $\delta < 1$. Moreover, when δ is small, ϵ_{cor} is close to 0. Therefore, the actual epidemic threshold depends on not only $\lambda_{max}(A)$, but also the spatial correlations coefficient ρ_e .

VI. CONCLUSIONS

In this work, we have proposed a new epidemic threshold by taking into consideration a certain spatial dependence. Specifically, we have exploited the assumption of Markov spatial dependence and shown analytically that the epidemic threshold in regular graphs indeed depends on the correlation coefficient between neighboring nodes. Through extensive simulations, we have demonstrated that our proposed epidemic threshold better characterizes the actual threshold than the threshold from [1], [6], [12] in arbitrary networks, such as ER random graphs and power-law topologies. To the best of our knowledge, this is the first attempt in quantitatively understanding the effect of spatial dependence on epidemic thresholds in networks.

Our discoveries on epidemic thresholds have important implications and applications for predicting and controlling the dynamics of the epidemic spreading process. Compared with the previous work, our proposed epidemic threshold provides a more accurate prediction on whether an infection will die out or become epidemic. Especially, when β/δ is between $1/\lambda_{max}(A)$ and $1/\lambda_{max}(A)(1 - \rho_e)$ in an arbitrary network, it is predicted in previous work [1], [6], [12] that the infection will become epidemic; however, we show that the infection actually dies out. Moreover, an objective function for controlling epidemic spread should consider both the largest eigenvalue of the topology and the spatial correction between neighboring nodes [13].

Note that $\tau_{c,cor}$ in Equation (32) depends on the empirical results of the spatial correlation ρ_e . Thus, it is not an analytical estimator. As our on-going work, we plan to derive a closed-form expression for ρ_e in an arbitrary network.

REFERENCES

- [1] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, "Epidemic thresholds in real networks," *ACM Transactions on Information and System Security (TISSEC)*, vol. 10, no. 4, Jan. 2008.
- [2] Z. Chen and C. Ji, "Spatial-temporal modeling of malware propagation in networks," *IEEE Transactions on Neural Networks: Special Issue on Adaptive Learning Systems in Communication Networks*, vol. 16, no. 5, Sept. 2005.
- [3] D. J. Daley and J. Gani, *Epidemic Modeling: An Introduction*. Cambridge University Press, 2001.
- [4] P. Erdős and A. Rényi, "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci.*, vol. 5, 1960, pp. 17-61.
- [5] S. C. Ferreira, C. Castellano, and R. Pastor-Satorras, "Epidemic thresholds of the susceptible-infected-susceptible model on networks: A comparison of numerical and theoretical results," *Phys. Rev. E* 86, 041125, Oct. 2012.
- [6] A. Ganesh, L. Massoulie, and D. Towsley, "The effect of network topology on the spread of epidemics," in *Proc. of IEEE INFOCOM'05*, vol. 2, Miami, FL, March 2005, pp. 1455-1466.
- [7] O. Givan, N. Schwartz, A. Cygelberg, and L. Stone, "Predicting epidemic thresholds on complex networks: Limitations of mean-field approaches," *Journal of Theoretical Biology*, vol. 288, no. 1, Nov. 2011, pp. 21-28.
- [8] A. Guille, H. Hacid, C. Favre, and D. A. Zighed, "Information diffusion in online social networks: a survey," *ACM SIGMOD Record*, vol. 42, no. 2, May 2013, pp. 17-28.
- [9] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proc. of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2003, pp. 137-146.
- [10] J. O. Kephart and S. R. White, "Directed-graph epidemiological models of computer viruses," in *Proc. of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy*, May 1991, pp. 343-359.
- [11] J. Leskovec, M. McGlohon, C. Faloutsos, N. Glance, and M. Hurst, "Cascading behavior in large blog graphs: Patterns and a model," *SIAM International Conference on Data Mining (SDM)*, 2007.
- [12] P. V. Mieghem, J. Omic, and R. Kooij, "Virus spread in networks," *IEEE/ACM Transaction on Networking*, vol. 17, no. 1, Feb. 2009, pp. 1-14.
- [13] C. Nowzari, V. M. Preciado, and G. J. Pappas, "Analysis and control of epidemics: A survey of spreading processes on complex networks," *IEEE Control Systems Magazine*, vol. 36, no. 1, Feb. 2016, pp. 26-46.
- [14] R. Pastor-Satorras, C. Castellano, P. V. Mieghem, and A. Vespignani, "Epidemic processes in complex networks," *Rev. Mod. Phys.*, vol. 87, no. 3, Aug. 2015, pp. 925.
- [15] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Physical review letters*, vol. 86, no. 14, Apr. 2001, pp. 3200-3203.
- [16] R. Pastor-Satorras and A. Vespignani, "Epidemic dynamics in finite size scale-free networks," *Physical Review E*, vol. 65, no. 3, Mar. 2002, pp. 035108.
- [17] B. A. Prakash, D. Chakrabarti, M. Faloutsos, N. Valler, and C. Faloutsos, "Threshold conditions for arbitrary cascade models on arbitrary networks," in *Proc. of the 2011 IEEE 11th International Conference on Data Mining*, Vancouver, pp. 537-546.
- [18] J. Wang and W. Wang, "To live or to die: Encountering conflict information dissemination over simple networks," in *Proc. of the IEEE INFOCOM*, San Francisco, CA, Apr. 2016.
- [19] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, "Epidemic spreading: An eigenvalue viewpoint," in *Proc. 2003 Symposium of Reliable and Distributed Systems*, Florence, Italy, Oct. 2003.
- [20] BRITE, Power-Law Topology Generator [Online]. Available: <http://www.cs.bu.edu/brite/> (May/2016 accessed).
- [21] Markov Random Field, Wikipedia [Online]. Available: https://en.wikipedia.org/wiki/Markov_random_field (May/2016 accessed).