Phase Transition of Multi-state Diffusion Process in Networks

Sungsu Lim
Dept. of Math. Sci.
KAIST
ssungssu@kaist.ac.kr

Kyomin Jung
Dept. of Computer Science
KAIST
kyomin@kaist.edu

John C. S. Lui
CSE Department
Chinese Univ. of Hong Kong
cslui@cse.cuhk.edu.hk

1. INTRODUCTION

The spread of viruses, ideas, technologies or behaviors in networks has been widely studied using mathematical models of contagion [7, 5, 8]. Understanding these dynamical processes is important to control and prevent the spread of diseases, and to maximize the influence of a product in online social networks. One of the most studied contagion models so far is the Susceptible-Infected-Susceptible (SIS) model. In this model, there are $k = 2$ states and each node in the graph is in one of these two states: *Susceptible* or *Infected*. This model describes the spread of contagions like flu (without immunity) or idea.

In practice, the SIS model can be quite restrictive since the degrees of interest in a contagion among individuals are different. For example, consider the case in which the diffusion process is designed to describe a product adoption [1, 6]. At some point of time after the product release, some people may have purchased the product while some may have not. For example, the consumer purchase decision process theory [2] suggests that there are five stages until a consumer purchases a product and influences others. The states include produce recognition, information search, alternative evaluation, purchase decision, and post-purchase behavior. This implies that one needs to further divide the *susceptible* state into several states according to the degree of interest. This is also intuitive because the adoption of a new product may need exposure from more than one customers.

In this work, we propose a generalization of the SIS model by allowing the number of states of adoption (or infection) to be more than two ($k \geq 2$). In particular, the states can range from 0 to $k-1$, where the state $k-1$ is the *active* state, that the node is infected and can influence other neighboring nodes. Nodes whose state is in 0 to $k-2$ can be promoted to a higher state if they are exposed to their infected neighbors (whose state is in $k-1$). We analyze the epidemic threshold dynamics, according to which initial condition leads to or prevents a disease outbreak. However, the traditional branching process approaches (that deal with a single initial spreader) [7] cannot be applied directly to this setup since we allow any fraction of initial infective nodes. Specifically, we use the *multidimensional mean-field method* to analyze our model and determine the condition of phase transition.

We believe that our work is a step towards elucidating the complex interactions between nodes in the epidemic spreading. The key result of our research is that our method predicts the behavior of the diffusion process accurately, and allows us to design a simple and effective vaccination or advertisement strategy. The analysis presented in this paper focuses on the case of a fully connected graph, but we also show experimentally that our method predicts the dynamics of the multi-state diffusion well for various types of networks if the initial node states are assigned in an i.i.d. manner. In the future work, we are planning to extend the analysis to arbitrary networks, under mild assumptions.

MAMA 2013 Pittsburgh, PA, USA

2. MODEL AND ANALYSIS

In our multi-state SIS model, we have a fully connected undirected graph $G = (V, E)$. Any node $a \in V$ can be in one of $k \geq 2$ states: $\{0, 1, \ldots, k-1\}$. Only nodes in state $k-1$ (infected or active state) can increase the state value of its neighbors, say node s, which is in state $j \in \{0, 1, \ldots, k-2\}$, to state value $j+1$ with the rate β_{j+1} (infection rate). Each node can be independently recovered with a rate γ (recovery rate). Figure 1 depicts our multi-state model.

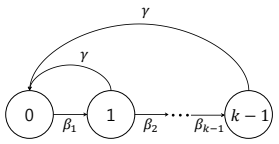


Figure 1: The multi-statemodel with k states.

First, we briefly review the analysis of the 2-state SIS model. For this SIS model, state 0 corresponds to *susceptible* (S) while state 1 corresponds to *infected* (I). The infection rate and the recovery rate are assumed to be β and γ, respectively. Let $x_0(t)$ and $x_1(t)$ be the fraction of nodes in state S and in state I, at time $t \geq 0$, respectively. Let $(x_0,\ x_1)$ be an equilibrium for this model. Given that G is a fully connected graph, we have $\frac{dx_1}{dt} = \beta x_0 x_1 - \gamma x_1$ and $x_0(t) + x_1(t) = 1$. For this model, there are two possible equilibria, one is $(x_0, x_1) = (1, 0)$ and the other is $(x_0, x_1) = (\frac{\gamma}{\beta}, 1 - \frac{\gamma}{\beta})$.

For an arbitrary graph G, let A be the adjacency matrix of G. Let $x_0^{(i)}(t)$ and $x_1^{(i)}(t)$ be the fraction of nodes in state S and state I, for each node $i \in V$ at time t, respectively. Let $(x_0^{(i)},\ x_1^{(i)})_{i \in V}$ be an equilibrium for this model. Then, $\frac{dx_1^{(i)}}{dt} = \beta x_0^{(i)} \sum_j A_{ij} x_1^{(j)} - \gamma x_1^{(i)} = 0$. In the beginning, $x_0^{(i)}(0) = 1 - c/n \approx 1$ if there are $c = o(1)$ initial infected nodes selected uniformly at random from G. Therefore, for each node $i \in V$, $\frac{dx_1^{(i)}}{dt} = \beta(1 - x_1^{(i)}) \sum_j A_{ij} x_1^{(j)} - \gamma x_1^{(i)} \approx$

$\beta \sum_j A_{ij} x_1^{(j)} - \gamma x_1^{(i)}$. At equilibrium, we have $\frac{dx_1^{(i)}}{dt} = 0$, which implies $\frac{\beta}{\gamma} A\mathbf{x} \approx \mathbf{x}$. Let λ_1 be the largest eigenvalue of A. It can be shown that the condition for the infection to die out over time is $\frac{\beta}{\gamma} < \frac{1}{\lambda_1}$, and infection survives and becomes an epidemic if $\frac{\beta}{\gamma} > \frac{1}{\lambda_1}$ [3, 5].

2.1 Ternary model

For the clarity of presentation, let us first consider a multi-state model with $k = 3$ states. Here, the state 1 represents that a node is exposed but *not* infected yet. For each state $s \in \{0, 1, 2\}$, let $x_s(t)$ be the fraction of nodes with state s at time t. Let (x_0, x_1, x_2) be an equilibrium for the model. Note that $x_0(t) + x_1(t) + x_2(t) = 1\ \forall t$. By the mean-field analysis, we derive a system of differential equation that describes the system dynamics:

$$\frac{dx_2}{dt} = \beta_2 x_2 x_1 - \gamma x_2, \quad (1)$$

$$\frac{dx_1}{dt} = -\beta_2 x_2 x_1 + \beta_1 x_2 x_0 - \gamma x_1. \quad (2)$$

Setting $\frac{dx_2}{dt} = 0$, we have $(x_2 = 0)$ or $(x_2 \neq 0$ and $x_1 = \frac{\gamma}{\beta_2})$. We are interested in identifying the condition for the non-trivial equilibrium which is the second case. Setting $\frac{dx_1}{dt} = 0$, we have $-\beta_2 x_2 x_1 + \beta_1 x_2 x_0 - \gamma x_1 = 0$. This implies that

$$x_0 = \left(\frac{\beta_2 x_2 + \gamma}{\beta_1 x_2}\right) x_1. \quad (3)$$

Thus, if $x_2 \neq 0$, then $x_1 = \frac{\gamma}{\beta_2}$ and $x_0 = (\frac{\beta_2 x_2 + \gamma}{\beta_1 x_2}) x_1$. Since $\sum_{i=0}^{2} x_i = 1$, we have $1 = x_2 + \frac{\gamma}{\beta_2} + (\frac{\beta_2 x_2 + \gamma}{\beta_1 x_2})\frac{\gamma}{\beta_2}$, or

$$\beta_1 \beta_2 x_2 = (\beta_1 x_2 + \gamma)(\beta_2 x_2 + \gamma) \quad (4)$$

The discriminant of this quadratic equation is $D = (\gamma(\beta_1 + \beta_2) - \beta_1\beta_2)^2 - 4\gamma^2\beta_1\beta_2$. The condition that $D \geq 0$ is equivalent to

$$\beta_2 \geq \left(\frac{\beta_1 + \gamma\sqrt{\beta_1}}{\beta_1 - \gamma}\right)^2, \quad (5)$$

yielding a rational solution if $D \geq 0$. Using this, we can determine the region for the phase transition, which is shown in Figure 2. The asymptotic lines are $\beta_1 = \gamma$ and $\beta_2 = \gamma$. Therefore, the infection survives if both β_1 and β_2 are large enough. Interestingly, we also discover that the condition is *symmetric* for β_1 and β_2. In Section 2.2, we extend this condition to a more general case with any arbitrary $k \geq 2$.

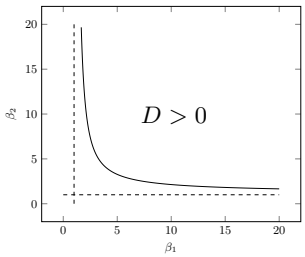


Figure 2: The phase transition diagram for the ternary model when $\gamma = 1$.

Stability analysis: Let us present the stability condition of the ternary model. Let $f_1(x_1, x_2) = \frac{dx_1}{dt} = -\beta_2 x_2 x_1 + \beta_1 x_2(1 - x_1 - x_2) - \gamma x_1$ and $f_2(x_1, x_2) = \frac{dx_2}{dt} = \beta_2 x_2 x_1 - \gamma x_2$. Define the Jacobian matrix $J = (\frac{df_i}{dx_j})_{i,j=1,2}$. Then, it can be shown that a fixed point (or an equilibria) of the system of differential equation defined by (1) and (2) is stable if the determinant (Det) of J is positive and the trace (Tr) of J is negative. We prove that $Tr < 0$ if $x_1 = \frac{\gamma}{\beta_2}$ and $Det > 0$ if $x_2 > \frac{1}{2} + \frac{\gamma}{\beta_1} + \frac{\gamma}{\beta_2}$. Therefore, a non-trivial equilibrium (x_1, x_2) is stable if $x_2 > \frac{1}{2} + \frac{\gamma}{\beta_1} + \frac{\gamma}{\beta_2}$, or a saddle point elsewhere. Since the sum of solutions of Eq. (4) is $1 - \frac{\gamma}{\beta_1} - \frac{\gamma}{\beta_1}$, there can be a stable non-trivial equilibrium where β_1 and β_2 are not so small with respect to γ. Note that $(x_1, x_2) = (0, 0)$ is a stable equilibrium since the eigenvalues of J are less than or equal to zero and the multiplicity of each is 1.

Examples: For instance, let $\beta_1 = \beta_2 = 5$ and $\gamma = 1$. Then $x_1 = 0.2$ and $25x_2 = (5x_2 + 1)^2$ from Eq. (4). Solving this, we have $x_2 = 0.076$ or 0.523 where $x_0 = 1 - x_1 - x_2$, respectively. Since $\frac{1}{2} + \frac{\gamma}{\beta_1} + \frac{\gamma}{\beta_2} = 0.9$, there is no stable non-trivial equilibrium. On the other hand, when $\beta_1 = \beta_2 = 20$ and $\gamma = 1$, we have $x_2 = 0.003$ or 0.897. Then, $0.897 > \frac{1}{2} + \frac{\gamma}{\beta_1} + \frac{\gamma}{\beta_2} = 0.6$ assures that there is a stable non-trivial equilibrium for this case.

Applications: By using our results, we can determine γ values to prevent large-scale disease spreading for any given (β_1, β_2). This leads to the following simple vaccination strategy. For the vaccination with a uniform rate γ, we set γ value so that $D < 0$. Note that $D < 0$ is equivalent to $(\beta_1 - \gamma)(\beta_2 - \gamma) < \gamma^2 + 2\gamma\sqrt{\beta_1\beta_2}$, and $\gamma > \frac{\beta_1\beta_2}{(\sqrt{\beta_1} + \sqrt{\beta_1})^2}$. When γ satisfies the above condition, the fraction of nodes at the final state converges to zero for large t.

2.2 Multi-state model

Let us proceed to analyze the multi-state model for any $k \geq 2$. For each $s \in \{0, 1, \ldots, k-1\}$, let $x_s(t)$ be the fraction of nodes of in state s at time t. Let $(x_0, x_1, \ldots, x_{k-1})$ be an equilibrium for the model. Then, we obtain the following system of differential equation.

$$\frac{dx_{k-1}}{dt} = \beta_{k-1} x_{k-1} x_{k-2} - \gamma x_{k-1}, \quad (6)$$

$$\frac{dx_s}{dt} = -\beta_{s+1} x_{k-1} x_s + \beta_s x_{k-1} x_{s-1} - \gamma x_s \quad \forall s \in \{1, \ldots, k-2\}, \quad (7)$$

$$\frac{dx_0}{dt} = -\beta_1 x_{k-1} x_0 + \gamma(1 - x_0). \quad (8)$$

Setting $\frac{dx_{k-1}}{dt} = 0$, we have $(x_{k-1} = 0)$ or $(x_{k-1} \neq 0$ and $x_{k-2} = \frac{\gamma}{\beta_{k-1}})$. We also set $\frac{dx_s}{dt} = 0$ for any $1 \leq s \leq k-2$, and $-\beta_{s+1} x_{k-1} x_s + \beta_s x_{k-1} x_{s-1} - \gamma x_s = 0$. This implies that

$$x_{s-1} = \left(\frac{\beta_{s+1} x_{k-1} + \gamma}{\beta_s x_{k-1}}\right) x_s = \left(\prod_{j=s}^{k-1} \frac{\beta_{j+1} x_{k-1} + \gamma}{\beta_j x_{k-1}}\right) x_{k-1}. \quad (9)$$

Thus, if $x_{k-1} \neq 0$, then the condition $\sum_{i=0}^{k-1} x_i = 1$ is equivalent to $1 = x_{k-1} + \frac{\gamma}{\beta_{k-1}} + \sum_{s=1}^{k-1}(\prod_{j=s}^{k-1} \frac{\beta_{j+1} x_{k-1} + \gamma}{\beta_j x_{k-1}}) x_{k-1}$, which is a $(k-1)$-dimension equation of x_{k-1}. Multiplying $\beta_1\beta_2 \ldots \beta_{k-1} x_{k-1}^{k-2}$ on both sides, we have

$$\beta_1\beta_2 \ldots \beta_{k-1} x_{k-1}^{k-2} = (\beta_1 x_{k-1} + \gamma) \ldots (\beta_{k-1} x_{k-1} + \gamma). \quad (10)$$

This holds for any $k \geq 2$, and one can check that this argument holds via mathematical induction on k.

Now let us consider the case in which the infection rates are increasing geometrically with a growth rate α so that $\beta_{k-1} = \alpha\beta_{k-2} = \ldots = \alpha^{k-2}\beta_1$. Then, for nonzero $\beta_1, .., \beta_{k-1}$, Eq. (1) is equivalent to $\frac{1}{x} = (1+\frac{\gamma}{\beta_1 x_{k-1}})\ldots(1+\frac{\gamma}{\beta_{k-1}x_{k-1}})$. For simplicity, we take $\beta = \beta_1$. Substituting y with $1/\beta x_{k-1}$, we have $\beta y = (1+\gamma y)(1+\frac{\gamma}{\alpha}y)\ldots(1+\frac{\gamma}{\alpha^{k-2}}y)$.

Let $g_1(y) = \beta y$ and $g_2(y) = (1+\gamma y)(1+\frac{\gamma}{\alpha}y)\ldots(1+\frac{\gamma}{\alpha^{k-2}}y)$. Then, these two functions of y are positive, monotone increasing, and convex for $y > 0$ since $g_2(y) = 0$ has only negative solutions $y = -\gamma, -\gamma\alpha, \ldots, -\gamma\alpha^{k-2}$. Thus, $g_1(y) = g_2(y)$ has at most two solutions. Moreover, for a fixed α, there is a *tipping point* β_t (or equivalently epidemic threshold) so that the equation has no solution if $\beta < \beta_t$ and has two solutions if $\beta > \beta_t$.

For instance, for the case that $\alpha = 1$ (i.e., the infection rates are homogeneous), $g_1(y) = g_2(y)$ is equivalent to $\beta y = (1+\gamma y)^{k-1}$. Note that the slopes of $g_1(y)$ and $g_2(y)$ are the same at the tipping point with $\beta = \beta_t$. Since $\frac{d}{dy}\beta y = \beta$ and $\frac{d}{dy}(1+\gamma y)^{k-1} = \gamma(k-1)(1+\gamma y)^{k-2}$, $y = \frac{1}{\gamma}\{(\frac{\beta}{\gamma(k-1)})^{1/(k-2)} - 1\}$ at the intersecting point. Substituting y with $\frac{1}{\gamma}\left((\frac{\beta}{\gamma(k-1)})^{1/(k-2)} - 1\right)$, we have $\frac{\beta}{\gamma}\{(\frac{\beta}{k-1})^{1/(k-2)} - 1\} = (\frac{\beta}{\gamma(k-1)})^{(k-1)/(k-2)}$. When β is nonzero, the solution of the above equation is

$$\beta = \gamma^{(k-1)}\frac{(k-1)^{(k-1)}}{(k-2)^{(k-2)}}. \quad (11)$$

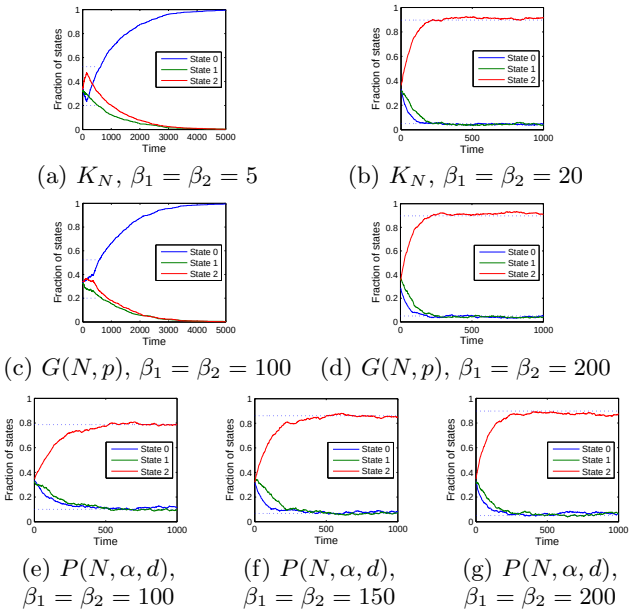
For instance, the threshold of β is $1, 2, 6.75$ if $\gamma = 1$ and $k = 2, 3, 4$, respectively.

Applications: Consider a computer virus outbreak in a network which is represented by G. We can devise an effective vaccination strategy from Eq. (6). For the vaccination with a uniform rate γ, we can prevent a large-scale virus spreading by using a suitable γ value. When $\gamma^{(k-1)} > \frac{(k-2)^{(k-2)}}{\beta(k-1)^{(k-1)}}$, the fraction of nodes at the final state converges to zero for large t. Furthermore, when k increases (i.e., it takes more phases for a computer virus to activate), the threshold for γ decreases *exponentially*. Thus it is easy to control the disease spread when the number of steps to reach the infection state is large. On the other hand, if the application of this model is to capture the dynamics of product's influence in an online social network, then decreasing k is more crucial.

3. NUMERICAL RESULTS

We conducted a set of experiments under the multi-state model. We study the dynamics of fractions of states for different network datasets. Our network datasets include (i) a *complete graph* K_N with N nodes, (ii) a *Erdös-Rényi random graph* $G(N,p)$ with N nodes and the probability of having an edge p, and (iii) a *random power law graph* $P(N,\alpha,d)$ with N nodes, the exponent α and the expected average degree d [4]. In our experiments, the initial state value of each node is chosen independently and uniformly at random from $\{0, 1, \ldots, k-1\}$ according to a given initial rate.

Note that although our analysis focuses on the case (i), we can also derive threshold values for the case (ii) by scaling the previous values by $1/p$. Figure 3 compares the dynamics for the ternary model with different β_1 and β_2 values. Figure 3(a) shows that if there is no stationary non-trivial



(a) K_N, $\beta_1 = \beta_2 = 5$ (b) K_N, $\beta_1 = \beta_2 = 20$

(c) $G(N,p)$, $\beta_1 = \beta_2 = 100$ (d) $G(N,p)$, $\beta_1 = \beta_2 = 200$

(e) $P(N,\alpha,d)$, $\beta_1 = \beta_2 = 100$ (f) $P(N,\alpha,d)$, $\beta_1 = \beta_2 = 150$ (g) $P(N,\alpha,d)$, $\beta_1 = \beta_2 = 200$

Figure 3: The dynamics of fractions x_0, x_1 and x_2 of states over time where $N = 10,000$ and $\gamma = 1$. For the cases (ii) and (iii), $p = 0.05$, $\alpha = 3$ and $d = 500$.

equilibrium, then (x_0, x_1, x_2) converges to $(1, 0, 0)$ for large t. However, if there is a stationary equilibrium, then there is a possibility that (x_1, x_2) converges to another point. In Figure 3(b), the dotted lines represent a stationary non-trivial equilibrium $(x_1, x_2) = (0.05, 0.897)$. In this case, (x_1, x_2) converges to either $(0, 0)$ or $(0.05, 0.897)$. Figure 3(c) and 3(d) show that the analysis holds for β_1/p and β_2/p where Np is not too small.

It is interesting to note that for the case (iii), our method still succeeds to predict the equilibrium condition with $N\beta_1/d$ and $N\beta_2/d$, as shown in Figures 3(e), 3(f), and 3(g). This is because the mean-field approach gives a good approximation for locally tree-like networks. Our current work is to extend our analysis to determine the condition of phase transition for general graphs with any $k \geq 2$.

4. REFERENCES

[1] S. Bhagat, A. Goyal, and L. V. S. Lakshmanan. Maximizing product adoption in social networks. In *WSDM*, 2012.

[2] R. D. Blackwell, P. W. Miniard, and J. F. Engel. *Consumer Behavior*. South-Western College Publishing, 10th ed., 2005.

[3] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos. Epidemic thresholds in real networks. *ACM TISSEC*, 10(4), 2008.

[4] F. Chung and L. Lu. Connected components in random graphs with given expected degree sequence. *Ann. Combin.*, 6:125–145, 2002.

[5] A. Ganesh, L. Massoulie, and D. Towsley. The effect of network topology on the spread of epidemics. In *INFOCOM*, 2005.

[6] K. Jung, B. Y. Kim, and M. Vojnović. Distributed ranking in networks with limited memory and communication link. In *ISIT*, 2012.

[7] M. E. J. Newman. The spread of epidemic disease on networks. *Phys. Rev. E*, 66:016128, 2002.

[8] B. A. Prakash, D. Chakrabarti, M. Faloutsos, N. Valler, and C. Faloutsos. Threshold conditions for arbitrary cascade models on arbitrary networks. In *ICDM*, 2011.