Investigating the Effect of Node Heterogeneity and Network Externality on Security Adoption

Zichao Yang John C.S. Lui
Computer Science & Engineering Department
The Chinese University of Hong Kong

1. INTRODUCTION

Network (positive) externality describes the phenomenon that when people align their behaviors with others, they can incur an explicit benefit [2]. A good example of network externality is the adoption of new technologies. For example, when companies introduce smartphone to the market, the benefit of smartphone to people is determined by *how many* people are using it. Network externality has been extensively studied. However, many previous results, e.g., [2] only focus on the impact of population size on network externality and only capture the influence of population size on people's willingness to adopt to a new technology. They did not consider the effect of heterogeneity, for example, different people can cast different influence on others and also different people can be influenced differently. In [3], the authors consider the influence of network externality on security measures deployment. We further enhance and generalize the work by considering the impact of node heterogeneity and differentiation.

2. MODEL

In this work, we present the mathematical model on how security protection can limit the spread of virus, in particular, we include *differential treatment* of different types of nodes and consider the impact of network externality in our evaluation. Our model includes two parts: the epidemic model and the economic model. The epidemic model is used to characterize the spread of virus or malware in a network. The economic model is used to evaluate the expected payoff of nodes. Based on the epidemic and economic model, nodes can determine whether to invest in security protection or not by evaluating their expected payoff.

Epidemic Model: Let $G = (V, E)$ be an undirected graph with vertex set V and edge set E. For $i, j \in V$, if $(i, j) \in E$, then nodes i and j are neighbors and we use $i \sim j$ to denote this relationship. Let $X = \{\text{healthy}, \text{infected}\}$ represent the set of states each node can be in. If node i is infected (healthy), then $X_i = 1$ ($X_i = 0$). Each infected node can contaminate its neighbors independently with probability q. Once a node is infected, it cannot recover to the healthy state. Note that this is similar to the *bond percolation process* [4] in which every edge is occupied with probability q. Each node has an initial state of being infected or not. Let us denote it by x_i where $x_i = 1$ if it is initially infected and $x_i = 0$ otherwise. Hence, at the steady state, a node is infected either because it is initially infected, or it contracted virus from its infected neighboring nodes. Hence, the final state of node i can be expressed in terms of the following recursive equation:

$$1 - X_i = (1 - x_i) \prod_{j \sim i} (1 - \theta_{ji} X_j) \quad \forall i \in V, \tag{1}$$

where θ_{ji} is a random variable indicating whether the edge (i, j) is occupied or not. According to previous discussion, θ_{ji} is a Bernoulli random variable with $\Pr(\theta_{ji} = 1) = q$. With Equation 1, the final probability that a node is infected can be derived by using the local mean field technique [3] given the initial probability of infection. Note that the infection can incur certain loss to a node. So a node needs to decide whether to invest in self-protection to decrease the infection probability or not by comparing the expected payoffs using the economic model, which we state below.

Economic Model: Every node i has an initial wealth $w_i \in \mathbb{R}_+$. A node's utility $u(y)$ is a function of wealth $y \in \mathbb{R}$. In our study, we consider nodes are *risk averse*, i.e., the utility function is strictly increasing and concave in y. However, for simplicity of illustration in this short paper, we assume that it is linear. If node i is infected, then it will incur a loss of $l_i \in \mathbb{R}_+$. In order to reduce the probability of being infected, the node can consider some self-protection measures, such as buying anti-virus software, installing firewall etc. For simplicity of analysis, we assume that the choice of a node i regarding self-protection is a binary decision: either the node invests with a cost of c_i, or it does not invest at all. If it decides to invest, it can still be infected with probability p^-. Else, it will be infected with p^+. Obviously we have $p^- < p^+$. We use S and N to denote the economic state of a node that it invests or does not invest in security protection respectively. A node makes the decision by maximizing its expected utility. In state N, the expected utility can be expressed as:

$$p^N u(w - l) + (1 - p^N) u(w), \tag{2}$$

where p^N is the final probability of a node being infected when it initially did not consider security protection, and l is the loss due to infection. The expected utility of a node which initially subscribed to security protection is:

$$p^S u(w - l - c) + p^S u(w - c), \tag{3}$$

where p^S is the final probability of a node being infected when it initially subscribed some self-protection measures with cost of c. Note that p^N and p^S are functions of p^- and p^+, as well as the infection probability q. They can be determined by the epidemic model and can be derived using Equation (1) with the local mean field technique [1, 3].

Each node needs to consider whether it should subscribe to some self-protection measures. The decision is based on the cost of investing in security measure, as well as the risk loss of being infected. The decision is non-trivial because one has to consider the *externality effect*. In general, a node needs to compare the cost on security investment and the risk. In particular, node i will choose to invest in security protection if and only if

$$c_i < (p^N - p^S)l_i. \tag{4}$$

In [3], the authors analyzed the case of *homogeneous* self-protection cost and risk loss, i.e., nodes, independent of their connectivity, will have the same investment cost and risk loss. However, this is not reasonable in practice, since many nodes have different risk loss. For example, nodes with low degree, representing individual users, have low risk loss, while nodes with high degree, representing large companies, have high cost and security risk if they are crippled by virus. Nodes with high degree are those who have high level of interaction with other nodes. Also, their decisions can have higher influence on others than those nodes with low degrees. Thus, nodes need to be *differentiated* according to their degree to represent different kinds of users in a network. It is important to incorporate node heterogeneity in the model and analyze the effect. It can also help us to understand the low level of self-protection measure adoption in real life and also different levels of adoption extent in different social classes.

In the following, we consider a Erdös–Rènyi random graph $G(n,p)$ with n nodes, where $p = \lambda/n$ is the probability that every possible node pair (i,j), $1 \le i < j \le n$, is connected. In the limit of large n, the degree of nodes in the random graph follows the Poisson distribution, i.e., $p_k = e^{-\lambda}\lambda^k/k!$. All results below can be extended to random graphs with general degree distribution [4].

3. DIFFERENTIAL TREATMENT: TWO TYPES CASE

Let us classify nodes into two types according to their degree. Let k_i denote the degree of node i. We define a degree threshold $\mathcal{K}$. If $k_i \le \mathcal{K}$, then the cost of self-protection is c_1 and the loss due to being infected is l_1. On the other hand, if $k_i > \mathcal{K}$, the cost of self-protection is c_2 and the loss due to being infected is l_2. It is reasonable to assume that $c_1 \le c_2$ and $l_1 \le l_2$. The initial probability of being infected is determined by economic state: p^- for S and p^+ for N. All edges have the same contraction probability q.

Assume that initially γ_1 (γ_2) fraction of the nodes with degree $k \le \mathcal{K}$ ($k > \mathcal{K}$) will invest in self-protection. Using the local mean field technique [1, 3], we can calculate the average final probability of nodes being infected , which we denote by h.

Proposition 1. *If γ_1 fraction of nodes with degree $k \le \mathcal{K}$ and γ_2 fraction of the nodes with degree $k > \mathcal{K}$ invest in self-protection, then h, the final average probability of nodes being infected, is given by the unique solution in $[0,1]$ of:*

$$\begin{aligned} h =& 1 - (1-p^+)e^{-\lambda qh} - (p^+ - p^-)[\gamma_1 \sum_{k \le \mathcal{K}} p(k)(1-hq)^k \\ &+ \gamma_2 \sum_{k > \mathcal{K}} p(k)(1-hq)^k], \end{aligned} \tag{5}$$

where $p(k) = e^{-\lambda}\lambda^k/k!$ is the probability mass function of the degree distribution of random graph.

Let p_1^S (p_1^N) denote the final infection probability of a node which has degree $k \le \mathcal{K}$ and has initialled subscribed (not subscribed) to the self-protection mechanism. Similarly, let p_2^S (p_2^N) denote the final infection probability of a node which has degree $k > \mathcal{K}$ and has initially subscribed (not subscribed) to the self-protection mechanism. With Proposition 1, we can derive the following conditional probabilities:

Corollary 1. *For nodes with degree $k \le \mathcal{K}$,*

$$p_1^S = 1 - (1-p^+)\frac{\sum_{k \le \mathcal{K}} p(k)(1-qh)^k}{\sum_{k \le \mathcal{K}} p(k)}, \tag{6}$$

$$p_1^N = 1 - (1-p^-)\frac{\sum_{k \le \mathcal{K}} p(k)(1-qh)^k}{\sum_{k \le \mathcal{K}} p(k)}. \tag{7}$$

For nodes with degree $k > \mathcal{K}$, we have

$$p_2^S = 1 - (1-p^+)\frac{\sum_{k > \mathcal{K}} p(k)(1-qh)^k}{\sum_{k > \mathcal{K}} p(k)}, \tag{8}$$

$$p_2^N = 1 - (1-p^-)\frac{\sum_{k > \mathcal{K}} p(k)(1-qh)^k}{\sum_{k > \mathcal{K}} p(k)}. \tag{9}$$

Each node needs to make a decision to perform self-protection or not by maximizing the expected utility. Nodes will invest in self-protection if their utility with investment is greater than that without investment, so

$$\gamma_1 \quad = \Pr((p_1^N - p_1^S)l_1 \ge c_1), \tag{10}$$

$$\gamma_2 \quad = \Pr((p_2^N - p_2^S)l_2 \ge c_2). \tag{11}$$

Note that the conditional probabilities p_1^S, p_1^N and p_2^S, p_2^N are functions of γ_1 and γ_2. Equations (6) to (11) form fixed point equations. By Proposition 1 and Corollary 1, we can compare the utilities to determine the fraction of users that will invest in self-protection. For $k \le \mathcal{K}$, we have

$$\begin{aligned} p_1^N l_1 - (p_1^S l_1 + c_1) &= (p_1^N - p_1^S)l_1 - c_1 \\ &= (p^+ - p^-)\frac{\sum_{k \le \mathcal{K}} p(k)(1-qh)^k}{\sum_{k \le \mathcal{K}} p(k)} l_1 - c_1. \end{aligned} \tag{12}$$

Let $f_1(\gamma_1, \gamma_2) = (p^+ - p^-)\frac{\sum_{k \le \mathcal{K}} p(k)(1-qh)^k}{\sum_{k \le \mathcal{K}} p(k)}$ (because h is a function of γ_1 and γ_2), then Equation (12) becomes:

$$p_1^N l_1 - (p_1^S l_1 + c_1) = f_1(\gamma_1, \gamma_2)l_1 - c_1. \tag{13}$$

Here, $f_1(\gamma_1, \gamma_2) = (p_1^N - p_1^S)$ is the probability reduction for nodes being finally infected if they invest in self-protection. Similarly, for $k > \mathcal{K}$, we have

$$p_2^N l_2 - (p_2^S l_2 + c_2) = (p^+ - p^-)f_2(\gamma_1, \gamma_2)l_2 - c_2, \tag{14}$$

where $f_2(\gamma_1, \gamma_2) = (p^+ - p^-)\frac{\sum_{k > \mathcal{K}} p(k)(1-qh)^k}{\sum_{k > \mathcal{K}} p(k)}$. It is easy to verify that both $f_1(\gamma_1, \gamma_2)$ and $f_2(\gamma_1, \gamma_2)$ are increasing functions in γ_1 and γ_2, which indicates that γ_1 and γ_2 degenerate to indicator functions. In other words, either no nodes will invest in self-protection, or all of them will invest in self-protection. This also shows the effect of network externality: the value of investing in self-protection *increases* with the number of nodes doing the investment.

It can be shown that for any $0 \le \gamma_1 \le 1$ and $0 \le \gamma_2 \le 1$,

$$f_2(\gamma_1, \gamma_2) < f_1(\gamma_1, \gamma_2), \tag{15}$$

which indicates that nodes with higher degree are less sensitive to invest in self-protection. In other words, investing in self-protection will lead to lower reduction in the final infection probability for nodes with higher degree.

Nodes can determine whether to make investment or not by comparing the expected profit of investment $f_1(\gamma_1, \gamma_2)l_1$ with the cost c_1 for nodes with lower degrees and $f_2(\gamma_1, \gamma_2)l_2$ with c_2 for nodes with higher degrees. We proceed to compare $f_1(\gamma_1, \gamma_2)$ with c_1/l_1 and $f_2(\gamma_1, \gamma_2)$ with c_2/l_2. We have four cases to consider:

Case 1: If $f_1(0,0) > c_1/l_1, f_2(0,0) > c_2/l_2$, then there is a unique Nash equilibrium where all the nodes invest in self-protection. Even if initially none of the nodes invest in self-protection, the profit of investment exceeds the cost regardless of the degree of nodes and eventually, all nodes will purchase self-protection tools.

Case 2: If $f_1(0,0) > c_1/l_1, f_2(0,0) < c_2/l_2$, then all nodes with degree $k \le \mathcal{K}$ will invest in self-protection. This is because the profit of investment for nodes with lower degree exceeds the cost while the profit is smaller than the cost for nodes with higher degree.

- If $f_2(1,0) > c_2/l_2$, then all nodes with degree higher than $\mathcal{K}$ will invest in self-protection. In this case, the profit of investment for nodes with higher degree increases since nodes with lower degrees will do the investment. Hence, the investment in security by nodes with with degree $k \le \mathcal{K}$ will incentivize nodes with degree $k > \mathcal{K}$ to invest in self-protection.

- If $f_2(1,0) < c_2/l_2 < f_2(1,1)$, there exists a *tipping point* γ_2^*, such that $f_2(1, \gamma_2^*) = \frac{c_2}{l_2}$. This implies that if we can offer self-protection to γ_2^* fraction of nodes with degree higher than $\mathcal{K}$ for free, then this will incentivize all nodes with higher degrees to do the investment. The price of anarchy can be expressed as $\frac{\sum_{k\le\mathcal{K}} p_k p_1^S(1,0)l_1 + \sum_{k>\mathcal{K}} p_k p_2^N(1,0)l_2}{\sum_{k\le\mathcal{K}} p_k(p_1^S(1,1)l_1+c_1) + \sum_{k>\mathcal{K}} p_k(p_2^S(1,1)l_2+c_2)}$.

- If $c_2/l_2 > f_2(1,1)$, all nodes with degree $k > \mathcal{K}$ will not perform self-protection.

Case 3: If $f_1(0,0) < c_1/l_1, f_2(0,0) > c_2/l_2$, then all nodes with degree $k > \mathcal{K}$ will take self-protection.

- If $f_1(0,1) > c_1/l_1$, then all nodes with degree lower than $\mathcal{K}$ will take self-protection. In this case, the investment in security by nodes with degree $k > \mathcal{K}$ will incentivize nodes with degree $k \le \mathcal{K}$ to invest in self-protection.

- If $f_1(0,1) < c_1/l_1 < f_1(1,1)$, there exists a *tipping point* γ_1^*, such that $f_1(\gamma_1^*, 1) = \frac{c_1}{l_1}$. The price of anarchy can be expressed as $\frac{\sum_{k\le\mathcal{K}} p_k p_1^N(0,1)l_1 + \sum_{k>\mathcal{K}} p_k p_2^S(0,1)l_2}{\sum_{k\le\mathcal{K}} p_k(p_1^S(1,1)l_1+c_1) + \sum_{k>\mathcal{K}} p_k(p_2^S(1,1)l_2+c_2)}$.

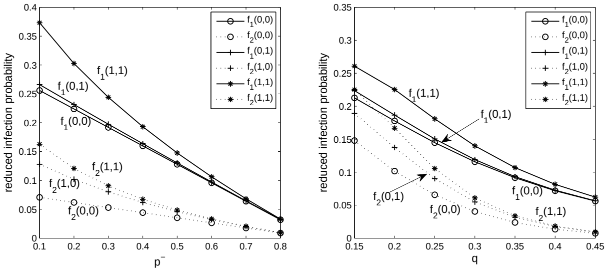
- If $c_1/l_1 > f_1(1,1)$, all nodes with degree lower than $\mathcal{K}$ will not take self-protection.

Case 4: If $f_1(0,0) < c_1/l_1 < f_1(1,1), f_2(0,0) < c_2/l_2 < f_2(1,1)$, then there exists a *tipping point* γ_1^* and γ_2^*. The price of anarchy can be expressed as
$\frac{\sum_{k\le\mathcal{K}} p_k p_1^N(0,0)l_1 + \sum_{k>\mathcal{K}} p_k p_2^N(0,0)l_2}{\sum_{k\le\mathcal{K}} p_k(p_1^S(1,1)l_1+c_1) + \sum_{k>\mathcal{K}} p_k(p_2^S(1,1)l_2+c_2)}$.

4. NUMERICAL RESULTS & CONCLUSION

In this section, we present numerical results to show that how various parameters may affect the adoption of security protection measures. In our experiments, we set the average degree of nodes $\lambda = 5$ and the degree threshold $\mathcal{K} = 5$.

First, we fix the initial probability of infection p^+ without secure measure and study the effect of p^-, the validity of self-protection, on the adoptability of self-protection measures. The result is shown in Figure 1(a). We set $p^+ = 0.9$, contagion probability $q = 0.3$ and vary p^- from 0.1 to 0.8. The figure shows how the reduced infection probability (thresholds) $f_1(\gamma_1, \gamma_2)$ and $f_2(\gamma_1, \gamma_2)$ change with p^-. From the figure, we can see $f_1(\gamma_1, \gamma_2) > f_2(\gamma_1, \gamma_2)$, which verifies previous claim. $f_1(\gamma_1, \gamma_2)$ and $f_2(\gamma_1, \gamma_2)$ decrease as p^- grows, which indicates that self-protection measures with higher quality imply more nodes to take the self-protection measures. This is because the network externality effect plays a small role if the self-protection quality is high. Notice that $f_1(0,1) - f_1(0,0)$, i.e., the gap between $f_1(0,1)$ and $f_1(0,0)$, is greater than $f_2(1,0) - f_2(0,0)$. It means that the adoption of self-protection for nodes with lower degree can incentivize higher degree nodes to invest in self-protection more than that higher degree nodes can influence those lower degree nodes. It is somewhat counter intuitive since we expect that nodes with higher degree can inflict more influence. One possible explanation is that nodes with lower degree takes a larger percentage of all the nodes, i.e., Pr($k \le \mathcal{K}$) > Pr($k > \mathcal{K}$).



(a) influence of p^- on thresholds (b) influence of q on thresholds

Figure 1: Influence of parameters on thresholds

In Figure 1(b), we investigate the effect of contagion probability q on the the thresholds. We set $p^+ = 0.4$, $p^- = 0.1$ and vary q from 0.15 to 0.45. As the figure shows, the thresholds decrease as q grows, i.e., a high contagion probability implies a greater network externality. When the contagion probability is high, taking self-protection will not lead to significant reduction in the final probability of being infected if no one decides to take self-protection. When contagion probability is high, people will decide to invest only if their cost and loss ratio c/l is low enough. Hence, high contagion probability will inhibit nodes to take self-protection.

5. REFERENCES

[1] D. Aldous and A. Bandyopadhyay. A survey of max-type recursive distributional equations. *The Annals of Applied Probability*, 15(2):1047–1110, 2005.

[2] D. Easley and J. Kleinberg. *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge Univ Pr, 2010.

[3] M. Lelarge and J. Bolot. Network externalities and the deployment of security features and protocols in the internet. In *Proceedings of the 2008 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, pages 37–48. ACM, 2008.

[4] M. Newman. *Networks: an introduction*. Oxford Univ Pr, 2010.