Network and Communication Technologies; Vol. 5, No. 1; 2020
ISSN 1927-064X E-ISSN 1927-0658
Published by Canadian Center of Science and Education

Modelling Malicious Attack in Social Networks

Amusan O.[1], Thompson A. F.[2], Aderinola T. B.[3] & Alese B. K.[4]

[1] Computer Science Department, The Federal University of Technology, Akure, Nigeria

[2] Cybersecurity Department, The Federal University of Technology, Akure, Nigeria

[3] Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia

[4] Computer Science Department University of Mines and Technology, Tarkwa, Ghana

Correspondence: Thompson A. F., Cybersecurity Department, The Federal University of Technology, Akure, Nigeria. Tel: 234-803-489-7857. E-mail: afthompson@futa.edu.ng

Received: April 18, 2019 Accepted: January 17, 2020 Online Published: February 6, 2020

doi:10.5539/nct.v5n1p37 URL: https://doi.org/10.5539/nct.v5n1p37

Abstract

Online Social Networks (OSNs) are based on actual trust relationships in environments which help people communicate with friends, family and acquaintances. Malicious individuals take advantage of this trust relationship to propagate malware through social networks. We study the dynamics of malware propagation among OSN users. Social networks users are referred to as nodes which is in two compartments: Healthy (H), or Infected (I). A H node could either be susceptible to infection (S) or removed (R). Simulations were carried out in R using the EpiModel network simulation package. Two networks were simulated thrice with different parameters to give better average values. Two categories of nodes, first category comprises of 3000 nodes with fewer connections and the second category comprising of 7000 nodes are the influential nodes with more connections. The larger network tends to have a higher fraction of nodes getting infected per unit time due to the high level of connectivity, as opposed to the small network where the number of connections is few. However, the infection tends to persist in the network as long as the birth rate is not equal to zero.

Keywords: Online Social Networks, nodes, EpiModel network, malware

1. Introduction

1.1 Background

A social network consists of individuals called nodes, connected by interdependencies such as friendship, kinship, common interest, dislike, etc. The structure of these networks are often represented using graphs (Kosorukoff, 2011). According to Faghani et al. (2012), there are three major types of OSN malware – Cross Site Scripting (XSS) worms, Trojans and ClickJacking worms. An XSS worm self-propagates among visitors of a website with the aim of progressively infecting other visitors. Trojans like Koobface use social engineering and spread themselves in social networks by broadcasting messages with fascinating topics.

1.2 Importance of Modeling Malicious Attacks in OSNs

Social network analysis has proved useful in epidemiology to help understand how patterns of human contact or aid inhibit the spread of infectious diseases in a population. A susceptible host is healthy but prone to infection. An infectious host had been infected and can infect other susceptible host. In cases where recovery from the infection confers *immunity*, hosts that recover are said to be *removed*, since they would not be susceptible to infection after recovery (Aderinola, Thompson, & Alese, 2017).

The dynamics of social networks can sometimes be modelled by using agent based models, which provide insight into the interplay between communication rules, rumour spreading and social structure. Social network analysis could also be used as a tool for mass surveillance (Kosorukoff, 2011).

Malware propagation in OSNs has become a major security threat to social networks users. More so, OSN malware propagate faster than traditional malware (Faghani et al., 2012).

Apart from being a platform for malware propagation, social networks have been disparaged for breaching the privacy of their users. This is such that Employees have been hired or fired, and University applicants have been screened based on their behaviour on social networks. There are various stories of user privacy breaches, with

unfortunate consequences (Mahmood, 2013). Malicious hackers often plan and strategize for cyber-attacks on social networks.

Analysts scrutinising these discussions can warn system administrators about attacker capabilities and intentions. This kind of analysis is often done manually. With recent machine learning technologies, Lincoln Laboratory has demonstrated the possibility of automatically detecting such malicious discussions from various OSNs (Lippmann *et al*., 2016).

1.3 Related Work

Lloyd, Valeika, and Cintr (2005) discussed the use of network models to describe the impact of local spatial structure on infection spread. They provided a detailed discussion on the implications of spatial structure on the dynamics of diseases infections in small-world networks. However, heterogeneity, which is an important aspect of the structure of many populations, was merely mentioned, but was not discussed into detail.

Tao et al., (2006) presented a detailed review of classical epidemic models and some interesting research problems in epidemiology. However, most studies on malware propagation in scale-free networks, small-world networks and online networks are often analysed with realistic network topologies, but with simple assumptions on user activities.

Yan, et al. (2011) studied the impact of initial infection, user activity, social structures on malware propagation by using trace-driven simulation. However, the work assumed that each user becomes active only after she performs an activity on the social network, and that each user has only one account on the network.

Ikhalia and Johnnes (2014) investigated common approaches to malware propagation on online social networks and provided insight to the operations and various types of malware that are proliferated in online social networks.

Xiao, Freeman, and Hwa (2015) described a scalable approach to finding groups of fake accounts registered by the same malicious user. They presented a scalable and time-sensitive machine learning approach to finding groups of fake accounts registered by the same actor.

Wen (2014) presented a *Susceptible-Infectious-Immunized* (SII). The model presented, however, assumed that online network users check messages at regular periods, which may not be true in the real scenario.

This study will model and simulate malware propagation in online social networks using the *Susceptible-Infected-Immunized* model.

The rest of this paper is organized as follows. In Section 2 the details of the network structure, malware propagation model and simulation are discussed. The results of the simulation and their interpretation are discussed in Section 3.

2. Method

2.1 Network Structure

Social networks are based on relationships between users of the network, referred to as nodes. The links between them are referred to edges. The social network will be modelled as an undirected graph. If two nodes are connected, information can flow both ways. The topology of a social network with *N* nodes can be represented as an *N* by *N* adjacency matrix:

$$\begin{bmatrix} p_{11} & \cdots & p_{1N} \\ \vdots & & \vdots \\ p_{N1} & \cdots & p_{NN} \end{bmatrix} \; p_{ij} \in [0,1] \tag{1}$$

where p_{ij} represents the probability of contact between nodes i and j.

When p_{ij} = 0, the nodes are not connected. When p_{ij} = 1, the nodes are connected, that is, they are "friends". Therefore, the node degree is the number of friends a user has. It is assumed that the nodes' degrees of users in the network exhibit the power-law distribution, which is described as

$$P(k) \sim k^{-\gamma}, \; \gamma > 0 \tag{2}$$

where $P(k)$ is the probability that a node connects to k nodes.

2.2 Malware Propagation

In a social network, the propagation of malware depends on the behaviour of users. The primary human behaviours affecting propagation are the message checking time of each user and the probability of clicking a malicious link.

In general, nodes will be in two compartments: Healthy (H), or Infected (I). A H node could either be susceptible to infection (S) or removed (R). The infected nodes could be active or dormant. Then we have

$$X_i(t) = \begin{cases} H,\ healthy & \begin{cases} \boldsymbol{S}, susceptible \\ \boldsymbol{R}, removed \end{cases} \\ \boldsymbol{I},\ \ infected & \begin{cases} A, active \\ D, dormant \end{cases} \end{cases} \tag{3}$$

A user is infectious only at the active state. The infection relies on contact between an infected and a susceptible node, with the susceptible node receiving and checking a malicious message from the infected node. The infection probability will be represented by $v(i,t)$. A node can move into the removed state from any of the other states with a recovery probability $r(t)$. It is assumed that once the user recovers, he is aware of such kinds of malicious links and cannot be re-infected. The classical SIR epidemic model is adapted to model the dynamics of the network:

$$\frac{dS}{dt} = -\lambda S + bN - \mu_s S \tag{4}$$

$$\frac{dI}{dt} = \lambda S - vI - \mu_i I \tag{5}$$

$$\frac{dR}{dt} = vI - \mu_r R \tag{6}$$

where b is the birth rate, μ are the mortality rates specific for each compartment, and ν is the recovery rate.

2.3 Network Simulation

The model simulation depends on three parameters. The *contact rate* between within a connection in each time unit. The *infection probability* is the risk of transmission given contact with an infected person. In an SIR model, the recovery rate is the reciprocal of the average duration of disease; likewise, the reciprocal of the death rates is the average life expectancy for persons in those compartments. However, in the case of the network being modeled, it can be assumed that there are no deaths in the network, but there are births. In other words, users are not removed, but more users can join the network.

The SIR model then becomes:

$$\frac{dS}{dt} = -\lambda S + bN \tag{7}$$

$$\frac{dI}{dt} = \lambda S - vI \tag{8}$$

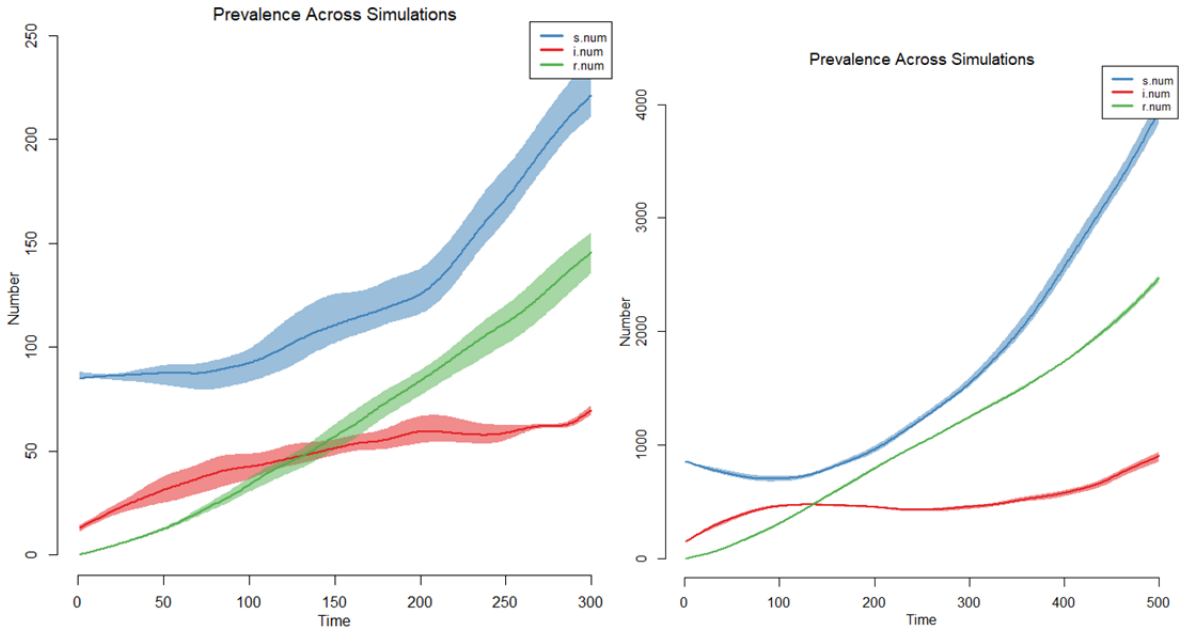
$$\frac{dR}{dt} = vI \tag{9}$$

Simulations were carried out in R using the EpiModel network simulation package (Jenness, Goodreau, & Morris, 2018). Two networks were simulated with different parameters. Each simulation was carried out three times with the same parameters to give better average values.

3. Results

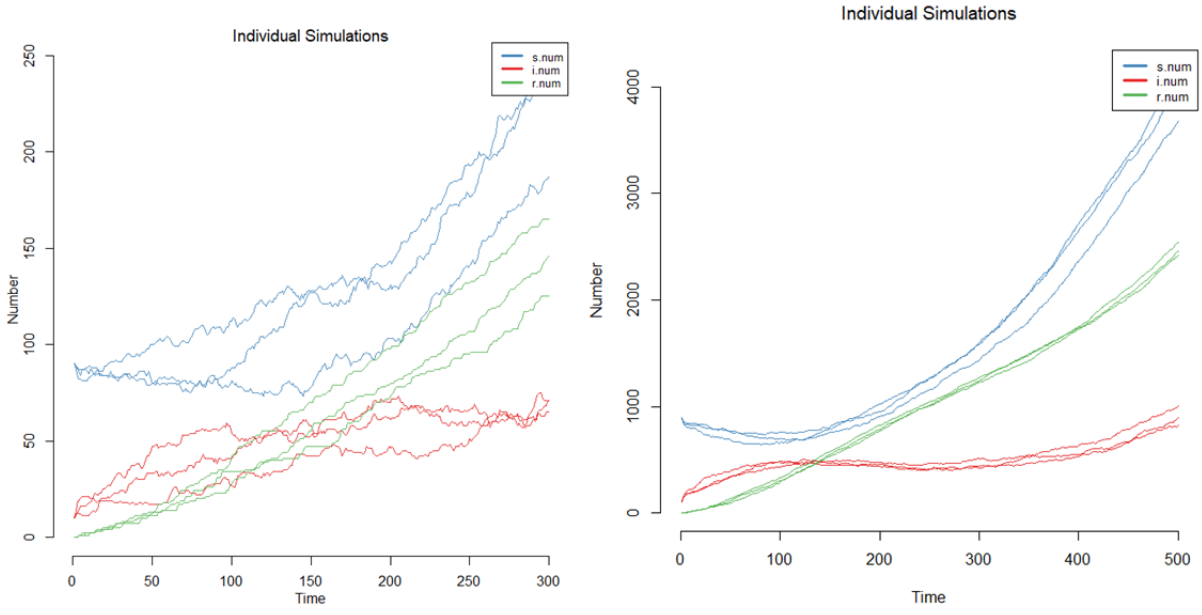
A total of 5 simulations were run, with T = 500. The output is shown in Figure 1. The plot displays the prevalence of the compartments in the model across the 3 simulations. The individual simulations are represented by the thin lines, the means across simulations at each time step are plotted with thicker lines, and the polygon band shows the inter-quartile range across simulations.

It can be seen from Figure 2 that given similar parameters and fraction of infected nodes, the infection trends are expected to follow the same pattern. Figure 3 shows the flows between the different compartments in the model across the 3 simulations: the births (b.flow), susceptible-infected (si.flow), and infected-recovered (ir.flow). The larger network would have a higher birth rate. In other words, a large social network will have a higher number of users joining than a small network. As shown by the SI flows, the larger network tends to have a higher fraction of nodes getting infected per unit time due to the high level of connectivity, as opposed to the small network where the number of connections is few. However, the infection tends to persist in the network as long as the birth rate is not equal to zero.



(a) (N = 100, N_f = 404; T = 300, I_0 = 10; b = 0.005) (b) (N = 1000, N_f = 7718; T = 500, I_0 = 100; b = 0.004)

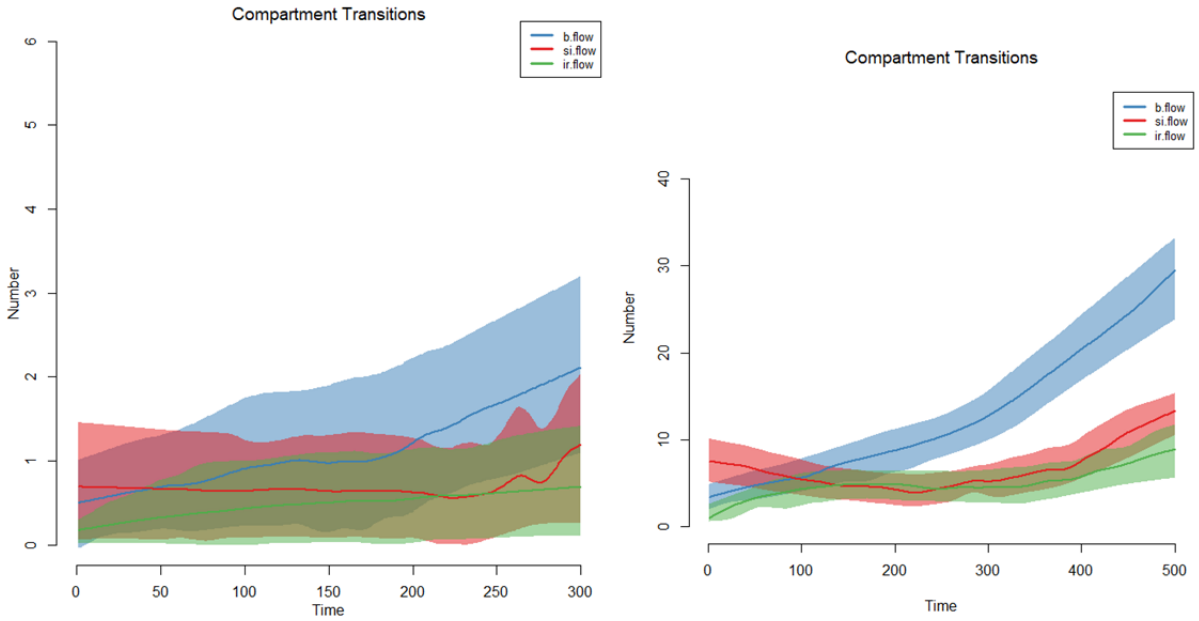
Figure 1. Infection Prevalence Across Compartments (c = 0.8; γ = 0.01; P_{inf} = 0.8)



(a) (N = 100, N_f = 404; T = 300, I_0 = 10; b = 0.005) (b) (N = 1000, N_f = 7718; T = 500, I_0 = 100; b = 0.004)

Figure 2. Infection Prevalence in Individual Simulations (c = 0.8; γ = 0.01; P_{inf} = 0.8)

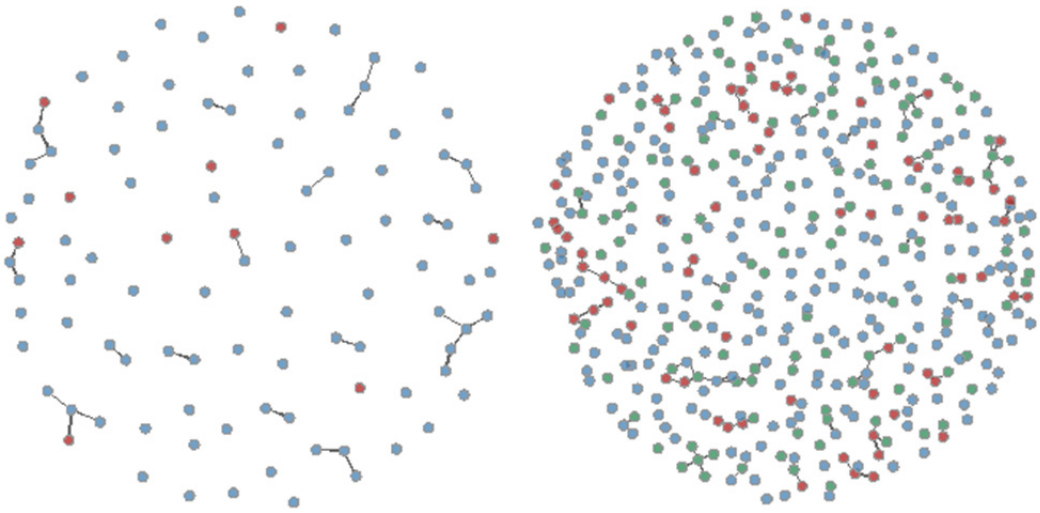
As long as new users keep joining the social network, the infections will persist, since every new user is susceptible, and will eventually be infected with a probability that depends on the user's awareness of such infection and the user activities. Figure 3 shows the flows between the different compartments in the model across the 3 simulations: the births (b.flow), susceptible-infected (si.flow), and infected-recovered (ir.flow). The larger network would have a higher birth rate. In other words, a large social network will have a higher number of users joining than a small network.



(a) (N = 100, N_f = 404; T = 300, I_0 = 10; b = 0.005) (b) (N = 1000, N_f = 7718; T = 500, I_0 = 100; b = 0.004)

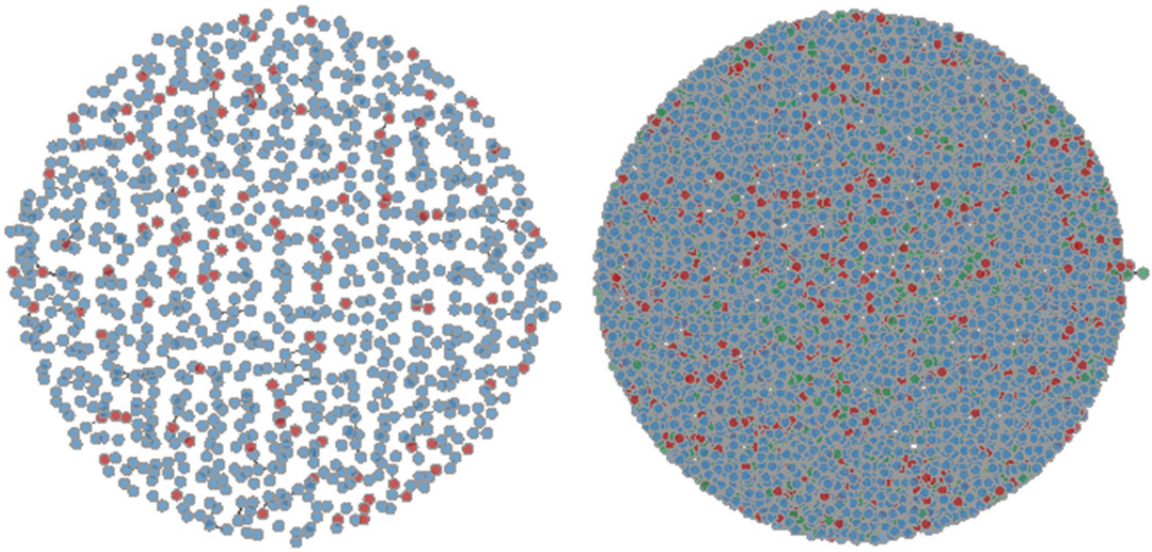
Figure 3. Nodes Transitions between Compartments (c = 0.8; γ = 0.01; P_{inf} = 0.8)

As shown by the SI flows, the larger network tends to have a higher fraction of nodes getting infected per unit time due to the high level of connectivity, as opposed to the small network where the number of connections is few. However, the infection tends to persist in the network as long as the birth rate is not equal to zero. As long as new users keep joining the social network, the infections will persist, since every new user is susceptible, and will eventually be infected with a probability that depends on the user's awareness of such infection and the user activities.



(N = 100, N_f = 404; T = 300, I_0 = 10; b = 0.005)

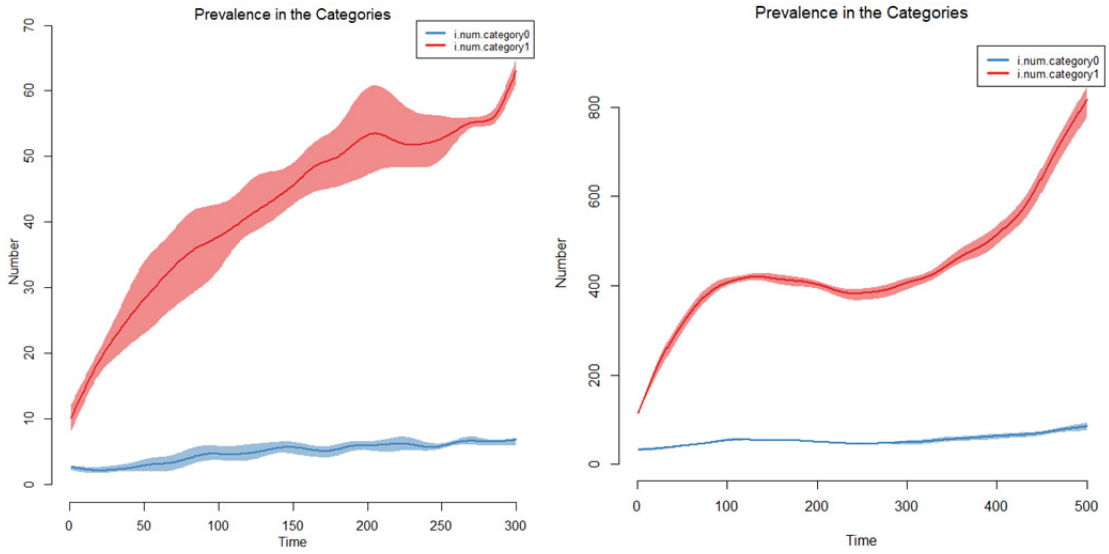
Figure 4. Network Snapshot at simulation start and end (c = 0.8; γ = 0.01; P_{inf} = 0.8)



(N = 1000, N_f = 7718; T = 500, I_0 = 100; b = 0.004)

Figure 5. Network Snapshot at simulation start and end (c = 0.8; γ = 0.01; P_{inf} = 0.8)

Figures 4 and 5 show the connectivity and disease prevalence for both networks at the start and end of the simulations. The infected nodes are indicated as red, the susceptible nodes as blue, and the recovered nodes as green. The lines indicate connection between nodes.



(a) (N = 100, N_f = 404; T = 300, I_0 = 10; b = 0.005 (b) (N = 1000, N_f = 7718; T = 500, I_0 = 100; b = 0.004)

Figure 5. Infection Prevalence in the Two Node Categories (c = 0.8; γ = 0.01; P_{inf} = 0.8)

The degree of nodes also affects the infection prevalence in the two categories as shown in Figure 5. It can be observed that in both simulations, the infection is more prevalent in the second category of nodes (category 1), since they have higher degrees. Hence, social network users that have more friends are likely to receive more messages containing malicious links and possibly forward same to their friends.

References

Aderinola, T., Thompson, A., & Alese, B. K. (2017). Epidemic Response Model for Malware Defense on Computer Networks. *International Journal on Cyber Situational Awareness, 2*(1). https://doi.org/10.22619/IJCSA.2017.100115

Faghani, M. R., Matrawy, A., & Lung, C. H. (2012). A study of Trojan propagation in online social networks. *2012 5th International Conference on New Technologies, Mobility and Security - Proceedings of NTMS 2012 Conference and Workshops*, 6-10. https://doi.org/10.1109/NTMS.2012.6208767

Fire, M., Goldschmidt, R., & Elovici, Y. (2014). Online Social Networks : Threats and Solutions. *IEEE Communications Surveys & Tutorials, 16*(4), 2019-2036. https://doi.org/10.1109/COMST.2014.2321628

Ikhalia, E. J., & Johnnes, A. (2014). Online Social Networks: A Vehicle for Malware Propagation. *Proceedings of the 13th European Conference on Cyber Warfare and Security (ECCWS-2014)*, (July), 95-101. https://doi.org/10.13140/RG.2.1.4331.5281

Jenness, S., Goodreau, S., & Morris, M. (2018). EpiModel: An R Package for Mathematical Modeling of Infectious Disease over Networks. *Journal of Statistical Software, 84*(8), 1-47. https://doi.org/10.18637/jss.v084.i08

Kosorukoff, A. (2011). *Social Network Analysis: Theory and Applications* (D. L. Passmore, Ed.). Passmore, D. L, 2011.

Lloyd, A. L., Valeika, S., & Cintr, A. (2005). Infection Dynamics on Small-World Networks. *Mathematical Studies on Human Disease Dynamics: Emerging Paradigms and Challanges*, 209-234. https://doi.org/10.1090/conm/410/07729

Pravallika, K., & Reddy, B. S. (2014). XSS Worm Propagation and Detection in Online Social Network. *International Journal of Science and Research (IJSR), 3*(7), 458-460. Retrieved from http://www.ijsr.net/archive/v3i7/MDIwMTQxMDE4.pdf

Tao, Z., Tao, Z., Zhongqian, F., Zhongqian, F., Binghong, W., & Binghong, W. (2006). Epidemic dynamics on complex networks. *Progress in Natural Science, 16*(70471033), 452-457. https://doi.org/10.1080/10020070612330019.

Wen, S. (2014). *Modeling the propagation and defense study of internet malicious information*, (April). Retrieved from http://dro.deakin.edu.au/view/DU:30067480

Xiao, C., Freeman, D. M., & Hwa, T. (2015). Detecting Clusters of Fake Accounts in Online Social Networks. *Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security - AISec '15*, 91-101. https://doi.org/10.1145/2808769.2808779

Yan, G., Chen, G., Eidenbenz, S., & Li, N. (2011). Malware propagation in online social networks: nature, dynamics, and defense implications. *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, 196-206. https://doi.org/10.1145/1966913.1966939