IJCSS International Journal of Complex Systems in Science

Int. J. Complex Systems in Science
vol. **3(1)** (2013), pp. 97–105

Identifying critical nodes in multi-layered networks under multi-vector malware attack

Rafael Vida[1,2,3], Javier Galeano[3,†] and Sara Cuenda[4]

[1] *Dept. de Sistemas Informáticos, E. T. S. de Ingeniería (ICAI) , Univ. Pontificia de Comillas, Spain*

[2] *Seguridad en Desarrollo de Servicios y Sistemas, Telefónica, Spain*

[3] *Complex System Group (GSC), Universidad Politécnica de Madrid, Spain*

[4] *Dept. Economía Cuantitativa, Universidad Autónoma de Madrid, Spain*

Abstract.

Computer viruses are evolving by developing multiple spreading mechanisms that are simultaneously used during the infection process. The identification of the nodes that allow a better spreading efficiency of these kind of viruses is becoming a determinant part of the defensive strategy against malware. These multi-vector viruses can be modeled in multi-layered networks in which each node belongs simultaneously to different layers, adapting the spreading vector to the properties of the layer. This way, the same virus has different propagation rates in each layer and also in the multi-layered network considered as a whole. The set of nodes selected as initial group of infected subjects can determine the final propagation of the infection.

In this work, we analyze the spreading of a virus in a multi-layered network formed by M layers, given different sets of initial infected nodes, and, in particular, the effect of the initial selection on the efficiency of the infection. The initial group of infected nodes is selected according to properties of the nodes considered as part of a layer and also of the whole system. As an example, we apply this study to a multi-layered network formed by two layers: the social network of collaboration of the Spanish scientific community of Statistical Physics, and the telecommunication network of each institution.

Keywords: multi-layered networks, spreading models,
MSC 2000: 05C82

† **Corresponding author:** javier.galeano@upm.es
Received: December 5th, 2013
Published: December 31th, 2013

1. Introduction

Computer viruses are evolving by developing multiple spreading mechanisms that are simultaneously used during the infection process. These malware had a huge impact in the Internet network and use new methods of spreading. For these kind of viruses the propagation is easy and quick within a Local Area Network (LAN). However, effective security measures [1] limit the propagation of these viruses to other LANs. To overcome this limit, these viruses also make use of other secondary vector of propagation such as the social relations between humans. Due to the complexity of the virus propagation and infection, reinfection is quite common after virus removal, so it is technically complicated to clean a whole LAN quickly enough to stop the reinfection. The identification of the nodes that allow a better spreading efficiency of these kind of viruses is becoming a determinant part of the defensive strategy against malware.

The classic disease models describe individuals that are susceptible to, infected with and recovered from a particular disease. In the literature, two different models are commonly used: The susceptible-infectious-recovered (SIR) model and the susceptible-infectious-susceptible (SIS) model. The SIR model is more appropriated for infectious diseases that confer lifelong immunity, whereas the SIS model is used with diseases where repeat infectious are common [2], as in the case we are studying. This epidemic spreading models have been widely studied in the scientific literature, usually considering only one network [3, 4, 5, 6, 7, 8, 9] and, more recently, using several interconnected layers of networks [10, 11, 12, 13, 14]. However, none of these formalisms are convenient in the case of malware spreading, where nodes exist in different layers and the state of each node in every layer must be the same. In the case of virus propagation, the same computer that connects within a LAN is also sending and receiving e-mails and pen drives, following the social network of the users of the computers. Therefore the state of each node cannot depend on the layer.

In this work, we analyze the spreading of a virus in a multi-layered network formed by M layers, given different sets of initial infected nodes, and, in particular, the effect of the initial selection on the efficiency of the infection propagation. The initial group of infected nodes is selected according to properties of the nodes considered as part of a layer and also of the whole system. As an example, we apply this study to a multi-layered network formed by two layers: the social network of collaboration of the Spanish scientific community of Statistical Physics, and the telecommunication network of each institution.

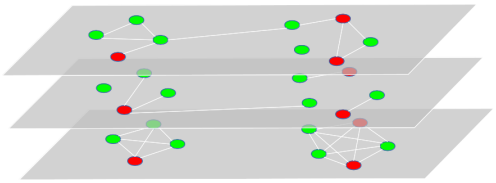

Figure 1: Multi-layer network where the state of each node in every layer is the same but the interactions within each layer may differ.

2. The Model

Let us consider M layers of networks formed each one by N nodes labeled in every layer with $i = 1, \ldots, N$. In this work we assume that nodes with the same label represent the same node, and that there are no links regarding the propagation of the epidemic connecting two nodes in different layers. Therefore, in our model, the usual adjacency matrix is replaced by a set of matrices, $A_{ij}^{(\alpha)}$ with $\alpha = 1, \ldots, M$, that specifies the links between nodes in each layer α. Notice that, in this *multi-layered networks*, the state of nodes with the same label must be the same, and the change in the state of one node in one layer changes automatically his state in all other layers. This is depicted in Figure 1.

We will study a SIS epidemic spreading in a multi-layered network in which the contagion in every layer α, follows a different dynamics given by the contagion matrix $C^{(\alpha)}$ and where all layers have the same recovery rate μ [1].

Notice that, since the contagion matrices are different, the epidemic spreading may follow a contact process (CP), a reactive process (RP) or something in between. To this end we define the contagion matrix in each layer as

$$C_{ij}^{(\alpha)} = \beta^{(\alpha)} \left(\frac{w_{ij}^{(\alpha)}}{w_i^{(\alpha)}} \right)^{\gamma_\alpha}, \tag{1}$$

where $w_{ij}^{(\alpha)}$ is the weight of the links between nodes i and j in layer α, $w_i^{(\alpha)}$ is the total strengh [15] of node i in layer α, $\beta^{(\alpha)}$ is the contagion rate of layer α, and γ_α is the parameter that defines the contagious process in each layer α from a reactive process for $\gamma_\alpha = 0$ to a contact process for $\gamma_\alpha = 1$.

Notice that, since each layer may follow a different contagion process, time scales in each layer (or even in each node) may vary considerably. To account

[1] We consider that the propagation process depends on the layer in which it is embedded, but that the healing process depends on the mechanisms and services that he has available, although this specification can be generalized to a recovery rate layer-dependent.

for that issue we will model the system as a microscopic Markov process in continuous time, where the state of the system is defined by the state of every node, $\mathbf{x} = \{x_1, \ldots, x_N\}$, with $x_i = 0$ when node i is susceptible and $x_i = 1$ when it is infected. The transition rate for node i from infected to susceptible is

$$q_i^-(\mathbf{x}) = \mu x_i, \tag{2}$$

where μ is the (constant) recovery rate, which we assume layer-independent since the same healing mechanisms are available to all nodes (this assumption can, however, be easily relaxed). On the other hand, the transition rate from susceptible to infected is

$$q_i^+(\mathbf{x}) = (1 - x_i) \left[1 - \prod_{\alpha=1}^{M} \prod_{j=1}^{N} (1 - C_{ji}^{(\alpha)} x_j) \right]. \tag{3}$$

where (2) corresponds to the transition of node i from infected state to susceptible and (3) stands for the opposite. In the latter expression, all the possible contagions from the infected neighbors of node i in all the layers have been considered. With these transition rates, the Markov process in continuous time associated to the epidemic follows the master equation

$$\begin{aligned} \frac{\partial P(\mathbf{x}, t)}{\partial t} = \sum_{i=1}^{N} \Big\{ & \left[q_i^+(f_i(\mathbf{x})) + q_i^-(f_i(\mathbf{x})) \right] P(f_i(\mathbf{x}), t) \\ & - \left[q_i^+(\mathbf{x}) + q_i^-(\mathbf{x}) \right] P(\mathbf{x}, t) \Big\}, \end{aligned} \tag{4}$$

where $f_i(x_1, \ldots, x_i, \ldots, x_N) = (x_1, \ldots, 1 - x_i, \ldots, x_N)$ is the *flip* operator of the i-th component, that changes the state of node i from susceptible to infected and viceversa.

As in [**?**], the products in expression (3) can be grouped by layers and we can define the *effective contagion matrix*, $\bar{C}$,

$$\bar{C}_{ij} \equiv 1 - \prod_{\alpha=1}^{M} (1 - C_{ij}^{(\alpha)}), \tag{5}$$

and write the rate equation as

$$\begin{aligned} \frac{\partial P(\mathbf{x}, t)}{\partial t} = \sum_{i=1}^{N} \Bigg\{ & \left[x_i \left(1 - \prod_{j=1}^{N} (1 - \bar{C}_{ji} x_j) \right) + (1 - x_i)\mu \right] P(f_i(\mathbf{x}), t) \\ & \left[(1 - x_i) \left(1 - \prod_{j=1}^{N} (1 - \bar{C}_{ji} x_j) \right) + x_i \mu \right] P(\mathbf{x}, t) \Bigg\}, \end{aligned} \tag{6}$$

where we have used that there are no self-loops in these multi-layered network, and therefore $\bar{C}_{ii} = 0$.

Assumming the approximation that the marginal probability of node i being infected, p_i, does not depend on p_j, for $j \neq i$, and for small probabilities of infection [7], then from the master equation (6) we obtain that

$$\frac{dp_i(t)}{dt} \simeq -\mu p_i + \sum_{j=1}^{N} \bar{C}_{ji} p_j, \tag{7}$$

wich renders the general solution $\mathbf{p}(t) = \mathbf{p}(0)e^{(\bar{C}-\mu I)t}$. This solution, although approximated, can give us some insight of the epidemic dynamics on the multi-layered network depending on the initial condition of infected nodes $\mathbf{p}(0)$. Let $\lambda_{\max}$ be the highest eigenvalue of $\bar{C}$, then for $\mu > \lambda_{\max}$ $p_i(t) \to 0$ as t grows for any initial condition $\mathbf{p}(0)$ since all the eigenvalues of $\bar{C} - \mu I$ are negative: we are below the epidemic threshold. For $\mu < \lambda_{\max}$, as there is at least one positive eigenvalue in the exponent, we obtain that $p_i(t)$ diverges for large t for any initial condition $\mathbf{p}(0)$ that has a non zero component in the eigenspace formed by the eigenvectors with positive eigenvalues, and thus we are above the epidemic threshold. Therefore, the epidemic threshold of the network corresponds to case where $\lambda_{\max} = \mu$, in which the highest eigenvalue of the exponent $\bar{C} - \mu I$ is zero and the system converges to a non-null state for initial conditions with non zero components in the eigenspace of $\lambda_{\max}$ [2].

This analysis agrees with previous expressions of the epidemic threshold [7, 8], but also takes into account the importance of the initial set of infected nodes, $\mathbf{p}(0)$, in the spreading process, as it depends on the component of $\mathbf{p}(0)$ in the eigenspace of $\lambda_{\max}$, which may not always non-zero. This depency makes important the study of the final propagation in terms of the initial set nodes.

3. Case study: Scientific collaborations

Scientific collaborations usually take several researchers from the same or different institutions to gather. These meetings involve a wide scope of social interactions, such as e-mails, virtual or face-to-face meetings, research visits, invited talks, or conference attendances, among others. Some of these interactions include connecting a foreign laptop to a local network (by wire or Wi-fi connections) or connecting someone's pendrive to a computer.

Usually, universities or research institutions have one or more LANs, and each institution connects its own LANs using IP switches or routers. The

[2] We are considering left eigenvectors in this analysis from expression (7) and since the contagion matrix $\bar{C}$ is, in general, not symmetrical.

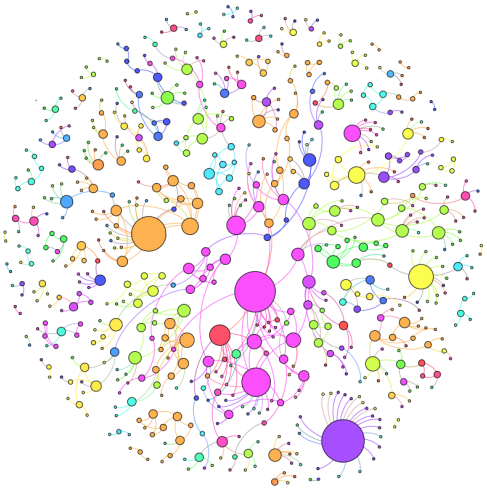

Figure 2: (color online) Social network of the Spanish Statistical Physics scientific community (FISES). Links represent collaborations between authors, the size of each node the number of links and the color indicates the affiliation. Therefore, links joining nodes of different colors show collaborations between research institutions.

internal IPs of each institution are considered trusted, whereas the external IPs are considered possibly dangerous. Therefore, the connections with external LANs use secured links with firewalls as a way to implement the perimetral security controls.

This isolation breaks down when we consider the social interactions of the scientific collaborations described above. For these reasons, and since the information of collaborations and affiliations is publicly available, the tandem formed by the LANs and scientific collaborations is a good example of real networks to include in our study. To this end, we have chosen the Spanish statistical mechanics (FisEs) research community, described in [**?**]. Both networks (the social network of the scientific collaborations and the physical network of the institutions LANs) are formed by 687 nodes distributed in 105 affiliations. Each institution is formed by a clique isolated from the rest (due to the security measures), and the largest institution has 39 researchers. In the social network there are 73 independent collaborations networks, the largest formed by 188 authors. The final network is depicted in Figure 2.

Strikingly, the number of connected components in the coupled network reduces dramatically to 8, the largest with 657 nodes, almost the total number of nodes in our networks. Note that, by coupling two poorly connected networks we have obtained a network that reduces in an order of magnitude the

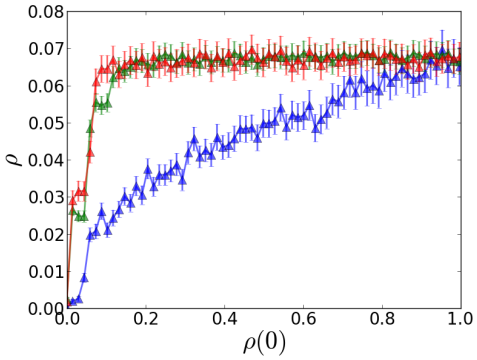

Figure 3: (color online) Density of infected nodes in the stationary state, ρ, vs. density of infected nodes in the initial seed, $\rho(0)$. The infection strategy is based on the strength of nodes in the contagion matrix $\bar{C}$. The contagion rates are $\beta_1 = 1.0$, $\beta_2 = 0.0$ (blue); $\beta_1 = 0.0$, $\beta_2 = 0.027$ (green) and $\beta_1 = 0.18$, $\beta_2 = 0.02$ (red).

number of connected components and with a largest component of the order of the number of total nodes. In figure are depicted the interconnections that the social network adds to the physical network. These interconnections affect deeply the virus spreading, as we will show below.

4. Virus propagation depending on the initial condition

In order to study the dependence of the initial condition of infected nodes in the virus propagation we performed numerical simulations of the stochastic contagion process described in section 2. using the network described in section 3.. We used the following procedure. First, we order the nodes by some intrinsic measurable magnitude π; if there is a set of nodes with the same value of π, we re-order them randomly keeping the total order of π in the list. Then we start the stochastic contagion with the first node in the list and take the average density of contagion in the stationary state after 200 simulations, ρ. We repeat the process with the folling nodes of the list and, finally, we repeat the overall process several times.

The first results of this procedure using the strength of nodes as the ordering magnitude is depicted in Figure 3, which shows the dependency on the density of initial infected nodes, $\rho(0)$, of the density in the stationary state ρ. We pursue to study this dependency with other properties such as the eigenvector centrality, the degree, etc., and compare them to a random picking of nodes with no specific order.

5. Conclusions

We study the propagation process in a multi-layered network in terms of the initial condition of infected nodes, $\mathbf{x}(0)$. We use the effective matrix $\bar{C}$ obtained in [16] in order to study this dependency in terms of different intrinsic magnitudes of the nodes of the coupled network. The results when the strength of nodes is chosen as the ordering magnitude are shown in Figure 3.

This procedure can be extended to other magnitudes, such as the eigenvector centrality, the degree, etc., in order to seek the optimal initial seed needed to percolate the network. The results can be of great interest in social contagion processes such as, for instance, viral marketing, meme spreading and bankruptcy of financial institutions.

Acknowledgements

We want to thank the financial support of MINECO through grants MTM2012-39101 for J.G. and FIS2011-22449 (PRODIEVO) for S.C., and of CM through grant S2009/ESP-1691 (MODELICO) for J.G. and S.C.

References

[1] V. Antoine, R. Bongiorni, A. Borza, P. Bosmajian, D. Duesterhaus, M. Dransfield, B. Eppinger, K. Gallicchio, J. Houser, A. Kim, et al., *Router security configuration guide, version 1.1 b. Technical Report C4-040R-02, System and Network Attack Center (SNAC), National Security Agency (NSA)* (2003).

[2] . M. J. Keeling and K. T.D. Eames, *J. R. Soc. Interface* **2**, 295 (2005).

[3] R. Pastor-Satorras and A. Vespignani, *Phys. Rev. Lett.*, **86**, p. 3200 (2001).

[4] R. Pastor-Satorras and A. Vespignani, *Phys. Rev. E*, **63**, p.066117 (2001).

[5] Robert M. May and Alun L. Lloyd, *Phys. Rev. E*, **64**, p.066112 (2001).

[6] Y. Moreno, J. Gómez, and A. Pacheco, *Phys. Rev. E*, **68**, p.035103 (2003).

[7] S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, and Y. Moreno, *Europhys. Lett.*, **89**, p.38009 (2013).

[8] P. Van Mieghem, J. Omic, and R. Kooij, *IEEE/ACM Trans. Net.*, **17**, p.1 (2009).

[9] K. Klemm, M. A. Serrano, V. M. Eguíluz, and M. San Miguel, *Sci. Rep.*, **2**, p.292 (2012).

[10] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, *Nature*, **464**, p.1025 (2010).

[11] S-W. Son, G. Bizhani, C. Christensen, P. Grassberger, and M. Paczuski, *EPL*, **97**, p.16006 (2012).

[12] A. Saumell-Mendiola, M. A. Serrano, and M. Boguñá, *Phys. Rev. E*, **86**, p.026106 (2012).

[13] C. Granell, S. Gómez, and A. Arenas, *Phys. Rev. Lett.*, **111**, p.128701 (2013).

[14] E. Cozzo, R. A. Baños, S. Meloni, and Y. Moreno, *arXiv:1307.1656* [physics.soc-ph].

[15] M. Barthélemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani *Physica A*, **346**, p.34–43 (2005).

[16] R. Vida, J. Galeano and S. Cuenda, *arXiv:1310.0741* [physics.soc-ph].