Dynamic Treatment Allocation for Epidemic Control in Arbitrary Networks

Kevin Scaman Argyris Kalogeratos Nicolas Vayatis

CMLA, École Normale Supérieure Cachan, CNRS
F-94230 Cachan, France
{scaman, kalogeratos, vayatis}@cmla.ens-cachan.fr

ABSTRACT

While static epidemic control, e.g. using *vaccination*, has been extensively studied for various network types, controlling epidemics dynamically remains an open issue. In this work, we first propose a general model formulation for the *dynamic treatment allocation problem* for the Susceptible-Infected-Susceptible diffusion model. Then, we investigate dynamic control strategies and further propose the novel *Largest Reduction in Infectious Edges* (LRIE) heuristic that gives priority to the treatment of nodes that have both a high dissemination rate of the infection to many healthy nodes, and low reinfection probability after recovery. Experiments on random and a real-world network show that the dynamic problem is significantly different from vaccination, since the latter strategies can lead to disastrous results, and that the proposed heuristic is an effective strategy under various initial infection conditions.

Categories and Subject Descriptors

G.5.2.2.5 [**Mathematics of computing**]: Stochastic control and optimization; L.5 [**Applied computing**]: Law, social and behavioral sciences

General Terms

Modeling, algorithms, experimentation

Keywords

Epidemic control, dynamic treatment allocation, virus models, diffusion processes, Markov processes, social networks

1. INTRODUCTION

Controlling diffusion processes in a network can have many applications in marketing (advertising campaigns), sociology (information diffusion in social networks), and epidemic control in medicine which is the problem we study in this work. Although the aim is simple to state, to control an epidemic, the type of virus diffusion (e.g. each node is prone to single, or multiple infections) and the type of actions that are available to authorities have a central role in the analysis of the problem and the design of strategies.

In the Diffusion Networks and Cascade Analytics Workshop of the 7th ACM WSDM, February 25-28, 2014, New York, USA.

Most studies focus on static control strategies that set up barriers in the network, prior to the emergence of an epidemic, aiming to reduce the virus spread through population [9, 11, 2]. In essence, this *vaccination* approach determines a set of nodes to receive a vaccine which would immunize them against any future spread of the virus. For this reason, the vaccination priority of a node largely depends on node attributes such as high centrality and degree.

In this paper, we focus on dynamic epidemic control. At each instant in time, a budget of treatments is given and the authorities need to decide which nodes should be treated according to current knowledge about the infection state of the network. Although some studies have considered static treatment allocation [7, 1] or the particular case of *contact tracing* [3], we believe that this problem is *inherently* dynamic. It is clear that any information about the infection state could be of great value in order to distribute the available resources right to the infection '*source*' in the network. In brief, vaccination could be characterized as *preventive*, as vaccine is given to healthy nodes, while a treatment is essentially *corrective*, trying to heal infected nodes.

Our contribution is: first, we propose a model formulation for the dynamic epidemic control as a *dynamic treatment allocation* (DTA) problem; second, we investigate DTA strategies and further propose the novel hybrid *Largest Reduction in Infectious Edges* (LRIE) heuristic that gives priority to the treatment of nodes that have both a high dissemination rate of the infection, and low reinfection probability after recovery. We also explain the connection between LRIE and the number of *infectious edges* in the network. Experimental results on random and a real-world network show that the DTA problem is significantly different from vaccination. Indeed, some of the latter strategies can lead to disastrous results whereas the proposed LRIE heuristic is an effective strategy under various initial infection conditions. Note that our work can be regarded as a generalization of [8] for the case of arbitrary networks.

In what follows, Sec. 2 and 3 provide the formulation and modeling details, Sec. 4 presents the experimental results and, finally, Sec. 5 provides concluding remarks and interesting ideas for future work.

2. MODELING EPIDEMIC SPREAD

The *Susceptible-Infected-Susceptible* (SIS) model [10] is used to model the diffusion of a disease. Each node in the network is either healthy, hence susceptible to be infected, or already infected (Fig. 1). Therefore, each node is prone to

Figure 1: Schematic view of the SIS model; nodes are prone to multiple infections.

multiple infections as it does not develop permanent immunity to the virus. A susceptible node can become infected if a neighboring node is also infected, while an infected node returns to the susceptible state after a certain period of recovery time.

For the SIS formulation we use continuous-time modeling of the disease spread using a *Continuous-time Markov Process* [10]. Thus, if t the time variable, then $t \in \mathbb{R}_+$. Let A be the $N \times N$ adjacency matrix of an undirected network of N nodes (A_{ij}=1 if there is an edge between nodes i and j, and 0 otherwise), and $X(t)$ the infection state vector at time t ($X_i(t)$=1 if node i is infected at time t, and 0 otherwise). Dynamic epidemic control is achieved via a *dynamic treatment allocation* (DTA) approach: a set of nodes is determined to receive medicine in order to recover more quickly. Let $M(t)$ be the vector representing the distribution of medicines in the network ($M_i(t)$=1 if node i is being treated at time t, and 0 otherwise). Then, the overall dynamics of the system are described as follows:

$$X_i(t) = \begin{cases} 0 \to 1 & \text{at rate } \beta \sum_j A_{ij} X_j(t) \\ 1 \to 0 & \text{at rate } \delta + \rho M_i(t) \end{cases}, \qquad (1)$$

where β, δ, ρ are parameters describing, respectively, the infection rate, the recovery rate without treatment, and the increase in the recovery rate when the node is being receiving medicine. Roughly speaking, Eq. 1 indicates that a susceptible node gets infected at a rate which is proportional to the number of its infected neighbors. Conversely, an infected node recovers (i.e. becomes susceptible again) at a constant rate δ if it does not receive any treatment, and at $\delta+\rho$ if is being treated at this particular time ($M_i(t)$=1). This model is similar to the *heterogeneous N-intertwined SIS* [7], but with a restriction on the possible values of the nodes' recovery rate (δ or $\delta+\rho$, instead of a general δ_i). Finally, we define the dimensionless parameters: $r=\frac{\beta}{\delta}$ the *effective spreading rate* of the disease, and $e=\frac{\rho}{\delta}$ the *treatment efficiency*.

3. DYNAMIC TREATMENT ALLOCATION

3.1 Proposed general DTA framework

A DTA strategy is a treatment allocation $M(t)$ aiming to suppress the epidemic. Here, we consider strategies which can depend on the state vector $X(t)$ and, since this is a stochastic function (i.e. a random variable in a function space), $M(t)$ is also a stochastic function. Nevertheless, we have to consider strategies which take into account only *past values* of $X(t)$, which are the observations up to time t. In mathematical terms, $M(t)$ will be adapted to the *natural filtration* associated to $X(t)$. Formally, a DTA strategy is a stochastic process:

$$M : \mathbb{R}_+ \to \{0,1\}^N \qquad (2)$$

$$\text{s.t. } \forall t \in \mathbb{R}_+, \quad \textstyle\sum_i M_i(t) \le b(t). \qquad (3)$$

At each instant t, a limited *budget* $b(t) \ll N$ of medicines is available for distribution. Without such a constraint the design of a strategy becomes trivial, because one could obtain an optimal strategy just by healing all the nodes of the network at once. The formulated framework is quite generic, but an extensive analysis of the kind of problem variations it could model is beyond the scope of this work. We suggest a series of constraints to devise a tractable problem variation to further work with, specifically:

- *Illimited resources, disposed at constant rate.* A fixed amount of b_{tot} medicines is provided to authorities at each time instance.
- *Inability to store resources for later use.*

3.2 Score-based strategies

The definition for DTA strategies can be simplified based on the above restrictions. Using the Markov property of the process ($X(t)$, $M(t)$), we can restrict ourselves to strategies that only depend on the *current infection state* $X(t)$:

$$M(t) = F(X(t)). \qquad (4)$$

Furthermore, according to Eq. 2 and since $M_i(t) \in \{0,1\}$, we can define the F function as a mechanism for selecting b_{tot} nodes from the network. This means that F can be implemented by a *scoring function* S that considers the current infection state $X(t)$ and returns a treatment priority score for each node. Finally, we can further restrict ourselves to strategies that distribute medicines only to infected nodes. This is due to the fact that a medicine is only active on infected nodes and has no preventive effect on healthy nodes against infection.

We can thus define a *strategy based on score* S as a selection of the b_{tot} topmost infected nodes according to $S(X(t))$:

$$M_i(t) = \begin{cases} 1 & \text{if } X_i(t) = 1 \text{ and } S_i(X(t)) \ge \theta \\ 0 & \text{otherwise} \end{cases}, \qquad (5)$$

where θ is a threshold value set so that the distributed medicines do not exceed the available budget, or the number of infected nodes, i.e. $\sum_i M_i(t) = \min\left(\sum_i X_i(t), b_{tot}\right)$. Note that, while this formulation is general, simple scoring functions tend to order nodes based on their intrinsic properties and are not well suited for coordinated strategies.

In this way, we define several intuitive scoring functions and in Tab. 1 we provide the expressions to compute them.

- *Random* (RAND): selects nodes uniformly at random among infected nodes (without replacement).
- *Most Neighbors* (MN): selects infected nodes with the largest number of neighbors.
- *PageRank Centrality* (PRC): selects infected nodes that are central according to the PageRank algorithm [5].
- *Largest Reduction in Spectral Radius* (LRSR): selects infected nodes which lead to the largest drop in the first eigenvalue of the adjacency matrix of the network.
- *Most Susceptible Neighbors* (MSN): selects infected nodes with the most non-infected neighbors.
- *Least Infected Neighbors* (LIN): selects infected nodes with the lowest number of infected neighbors.

Strategy	Scoring function $S_i(X_i)$
RAND	R_i, where R_i is i.i.d. uniform in [0, 1]
MN	$\sum_j A_{ij}$
PRC	P_i, where P_i is the PageRank score for node i
LRSR	$\lambda_1 - \lambda_1^{G\setminus i}$, where λ_1 the largest eigenvalue of A, and $\lambda_1^{G\setminus i}$ the largest eigenvalue of the matrix $A^{G\setminus i}$ for the network without node i
MSN	$\sum_j A_{ij}(1-X_j)$
LIN	$-\sum_j A_{ij}X_j$
LRIE	$\sum_j A_{ij}(1-2X_j)$, sums equally MSN and LIN

Table 1: Derived DTA scoring functions.

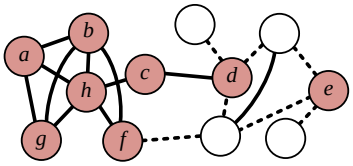

Figure 2: Example network with healthy (*white*) and infected (*red*) nodes. *Dashed* edges denote *infectious edges* on which the disease might spread.

- *Largest Reduction in Infectious Edges* (LRIE): a proposed novel hybrid heuristic that combines MSN and LIN to gain from both intuitive approaches.

While MN, PRC, and LRSR come from the static vaccination literature, MSN and LIN are intuitive strategies. MSN is reasonable because a node with numerous susceptible neighbors will spread the virus quickly, so healing it seems a good choice. On the contrary, nodes with many infected neighbors will get infected with high probability. Healing those nodes may not be a good choice since, most probably, they would be reinfected right after. The latter is exactly the intuition behind the LIN strategy.

We should notice the implied complementarity between MSN and LIN. In an indirect way, the former concentrates his efforts on '*central*' nodes with large degree, while the latter prefers to target nodes at network's '*periphery*'. In fact, MSN and LIN capture different aspects of how critical a node is for the diffusion. For this reason, we propose the combination of the two into the score-based strategy we named *Largest Reduction in Infectious Edges* (LRIE). This hybrid strategy seeks for nodes which are both *viral* for many healthy neighbors and, at the same time, *safe* in a neighborhood not heavily infected. The combination is achieved by adding up the scores, which proved to be better than other combinations we tested. A practical justification is that LRIE finds nodes whose healing would minimize the number of *infectious edges*, i.e. edges between infected and susceptible nodes.

Fig. 2 shows an example network and its infection state. We can observe that each scoring function would evaluate differently the treatment priority of these nodes given their current state. Specifically, node h is the most connected, d has the highest dissemination rate, e and h are the least and most probable to be reinfected if treated, respectively. The proposed LRIE strategy would choose node e to give the highest treatment priority, as it is viral but also the safest, and the order of the rest would be: d, f, c, $\{a, g\}$, b, h.

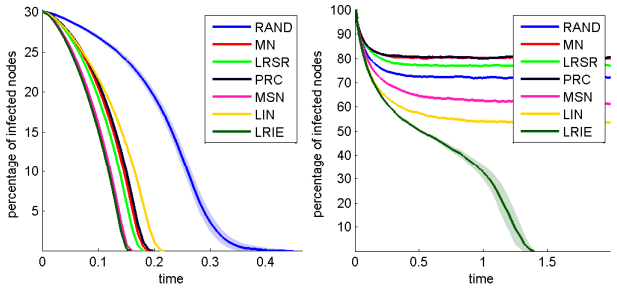

Figure 3: Results for Erdös-Rényi networks: N=10^4 nodes, p=0.001, r=2, b_{tot}=10 medicines. Left: e=4000, right: e=3000.

4. EXPERIMENTAL RESULTS

The DTA strategies were compared using simulations on one real-world and various random networks. To measure the performance of a strategy on a network, 10 to 100 simulations were performed starting from the same fixed overall infection level of the network (%), but with different random initialization of the nodes' infection state. In all cases we set δ=1. The results are illustrated in line plots where solid lines represent the expected number of infected nodes for each strategy, and their surrounding area is the 95% confidence interval under Gaussian hypothesis[1].

4.1 Experiments on simulated networks

The random networks we used are of two types: i) *Erdös-Rényi networks*, and ii) more realistic *scale-free networks* generated by the Barabási-Albert preferential attachment approach [5]. We generated a different network for each simulation using the same generation parameters (i.e. the edge probability p for type (i) and the number of added edges with each node m for type (ii)).

Fig. 3 and 4 present the results for Erdös-Rényi networks. In all simulations LRIE performs better than all its competing strategies. We observe two different behaviors depending on the percentage of initially infected population. If this is low (30% in Fig. 3 left), then centrality-based strategies (MN, PRC and LRSR) perform well and are able to suppress the epidemic. However, when the percentage of initially infected nodes is high (100% in Fig. 3 right) and the budget b_{tot} is low, only LRIE is able to suppress the epidemic. More importantly, MN, PRC and LRSR are counter-effective, as they give worse results than the random strategy. This comes from the fact that, in this case, central nodes have too many infected neighbors and, thus, are prone to quick reinfection. Fig. 4 presents a more realistic scenario where the treatment is only moderately effective (e=5). Scale-free networks are highly prone to epidemics, due to the existence of extremely highly connected nodes. The behavior of the compared DTA strategies are similar to the Erdös-Rényi case (see Fig. 5), except that the epidemic is more aggressive and some strategies do not manage to suppress it, even when initiated with a low percentage of infected nodes. We may also note that MN is more efficient in this case (Fig. 5 left) compared to PRC and LRSR, which is what one would expect since node degree is more informative in a scale-free network than in a uniformly random one.

[1]For N_{tests} simulations, this is $2\frac{\sigma_{N_{tests}}}{\sqrt{N_{tests}}}$, where $\sigma_{N_{tests}}$ the standard deviation of the measurements.

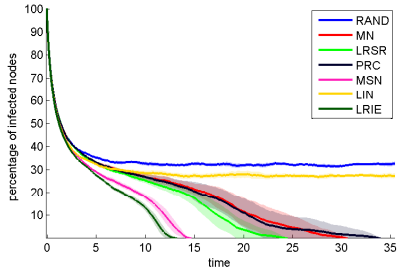

Figure 4: Results for Erdös-Rényi networks: N=10^4 nodes, p=0.001, r=0.2, e=5, b_{tot}=200 medicines.

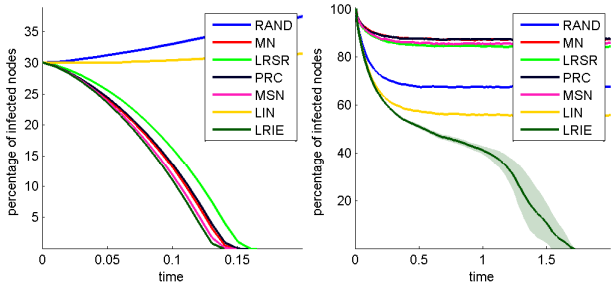

Figure 5: Results for random scale-free networks: N=10^4 nodes, m=5, r=2, b_{tot}=10 medicines. Left: e=4000, right: e=3000.

4.2 A real network: air traffic data

We further tested the DTA strategies on a real network of US air traffic for the year 2010[2], which contains 1574 nodes corresponding to the US airports that serviced domestic and international flights, and those non-US airports that serviced flights to US, during that year. The directed edges are weighted with the aggregated capacity of all flights on each direction. However, we symmetrized the adjacency matrix as A_s=A+$A^\top$ which finally contains 17215 weighted undirected edges. It is known that air transportation networks are scale-free small-world networks with multicommunity structure [4], properties that makes epidemic control difficult. Fig. 6 presents the simulation results for a scenario where the treatment is not strong enough to completely remove the disease. There is still a large difference between the stationary values of the strategies and LRIE outperforms all other methods. The observed persistence of the disease at a low infection level is due to the existence of very few high-degree nodes, a finding that has been widely reported for scale-free networks in literature [6].

5. CONCLUSION AND FUTURE WORK

In this paper we investigated the problem of dynamic epidemic control for the Susceptible-Infected-Susceptible diffusion model. First, we proposed a general model formulation as a *dynamic treatment allocation problem* (DTA). Then we investigated dynamic control strategies, starting with simple intuitive ones, and we further proposed the novel hybrid *Largest Reduction in Infectious Edges* (LRIE) heuristic that gives priority to the treatment of nodes that have both a high dissemination rate of the infection to many healthy nodes, and low reinfection probability after recovery.

[2]Source: US Bureau of Transportation Statistics (BTS). Available: `http://toreopsahl.com/datasets/#usairports`.

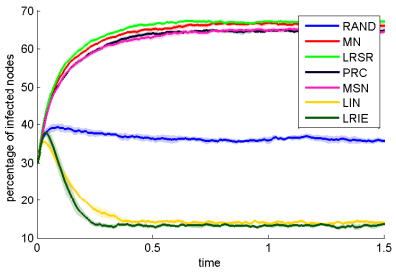

Figure 6: Results for the US air traffic network: N=1574 nodes, r=2, e=600, b_{tot}=10 medicines.

The experiments on one real-world and various random networks, with different initial infection conditions, show that: i) the DTA problem is fundamentally different to vaccination, since some of those approaches may lead to disastrous results, ii) the proposed LRIE strategy is the most effective and robust among the compared strategies. Our plans for future work include a deeper theoretical study of LRIE heuristic and extensive experimentation on various other network data.

6. ACKNOWLEDGMENT

This research is part of the SODATECH project funded by the French Government within the program of "*Investments for the Future – Big Data*".

References

[1] F. Chung, P. Horn, and A. Tsiatas. Distributing antidote using pagerank vectors. *Int. Math.*, 6(2):237–254, 2009.

[2] R. Cohen, S. Havlin, and D. Ben-Avraham. Efficient immunization strategies for computer networks and populations. *Ph. Rev. Let.*, 91(24):247901, 2003.

[3] K. Eames and M. Keeling. Contact tracing and disease control. *Proc. of the Royal Society of London. Series B: Biol. Sc.*, 270(1533):2565–2571, 2003.

[4] R. Guimerà, S. Mossa, A. Turtschi, and L. A. N. Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proc.of the NAS of the USA*, 102:77–94, 2005.

[5] M. Newman. *Networks: An Introduction*. Oxford University Press, Inc., New York, NY, USA, 2010.

[6] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Ph. Rev. Let.*, 86(14):3200–3203, 2001.

[7] V. M. Preciado, M. Zargham, C. Enyioha, A. Jadbabaie, and G. J. Pappas. Optimal resource allocation for network protection: A geometric programming approach. *arXiv preprint arXiv:1309.6270*, 2013.

[8] R. E. Rowthorn, R. Laxminarayan, and C. A. Gilligan. Optimal control of epidemics in metapopulations. *J. of the Royal Society Interface*, 6(41):1135–1144, 2009.

[9] H. Tong, B. A. Prakash, T. Eliassi-Rad, M. Faloutsos, and C. Faloutsos. Gelling, and melting, large graphs by edge manipulation. In *Proc. of the 21st ACM Int. Conf. on Inf. and Knowl. Management*, pages 245–254. ACM, 2012.

[10] P. Van Mieghem, J. Omic, and R. Kooij. Virus spread in networks. *IEEE/ACM Trans. on Net.*, 17(1):1–14, 2009.

[11] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos. Epidemic spreading in real networks: An eigenvalue viewpoint. In *Proc. of the 22nd Int. Symp. on Reliable Distributed Systems*, pages 25–34. IEEE, 2003.