A Greedy Approach for Dynamic Control of Diffusion Processes in Networks

Kevin Scaman Argyris Kalogeratos Nicolas Vayatis

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

***Abstract*—This paper investigates the control of a diffusion process by utilizing real-time information. More specifically, we allow the network administrator to adjust the allocation of control resources, a set of treatments that increase the recovery rate of infected nodes, according to the evolution of the diffusion process. We first present a novel framework for describing a large class of dynamic control strategies. These strategies rely on sorting the nodes according to a priority score in order to treat more sensitive regions first. Then, we propose the *Largest Reduction in Infectious Edges* (LRIE) control strategy which is based on a greedy minimization of the cost associated to the undesired diffusion, and has the benefits of being efficient and easy to implement. Our simulations, which were conducted using a software package that we developed and made available to the community, show that the LRIE strategy substantially outperforms its competitors in a wide range of scenarios.**

I. Introduction

Network diffusion processes (DP) have received much attention recently from scientists of various fields, such as epidemiology and medicine, sociology, computer science, and marketing. Among other applications, a DP can model the spread of a disease or information (ideas, news, product penetration) in a network through node-to-node propagation. The main purpose of modeling a DP can be: its analytic description that would lead to accurate predictive models about its evolution given the current state, and the control of the process given a certain type of available control actions.

In this work, we study the *control of diffusion processes* and we use the epidemic control as reference. The analysis of this problem and the design of control strategies are complex and depend heavily on factors such as: i) the type of the diffusion process, e.g. each node can be prone to single or multiple infections, ii) the network structure, and iii) the type of control actions available to authorities. Our aim is to perform *dynamic epidemic control* using *real-time resource allocation*. At each instant in time, a certain budget of resources is available and the authorities need to decide which nodes should receive them based on the current state of the network. This setting is particularly representative for the control of undesired diffusion processes in a social network, such as the spread of particular interests, malicious behaviors (e.g. violence, racism) or even health related behaviors such as obesity (recently shown to be diffusive through a social network [1]) or smoking.

It is generally hard to collect real-time information about a DP. For epidemic control, it is currently out of reach to be able to know the exact state of every individual in a large network, or the complete underlying network of relations. However, the common practice of epidemiologists is to work on a simplified network where each node summarizes the state of large parts of the actual network. Transportation networks are good candidates for this purpose, not only because each node (e.g. an airport) corresponds to specific geographic areas, but also because control policies can be easily applied directly on those sites (e.g. cancellation of flights, thermal cameras, etc.). On the other hand, real-time information about a DP taking place on a digital network is quite feasible to collect even at the level of individuals, e.g. tracking the diffusion of memes and links among the users of a social network [2, 3], and could enable the application of dynamic control strategies to their full potential.

Our contributions are: First, we propose a model formulation for the dynamic control of diffusion processes as a *dynamic resource allocation* (DRA) problem. Second, we investigate DRA strategies and propose the novel *Largest Reduction in Infectious Edges* (LRIE) control strategy based on the minimization of a second-order approximation of the cost associated with a diffusion process. We explain that LRIE greedily minimizes the number of *infectious edges* that can transmit the infection from infected to healthy nodes of the network. This way, it reduces the scattering of the infection across the network and allows for efficient DP control. Third, our experimental study on randomly generated and real-world networks shows that LRIE outperforms well-known centrality-based strategies, whose performance is suboptimal for this particular problem. Fourth, our simulations were conducted using a software package that we developed and made available to the community[1].

[1] The Matlab code and supplementary technical material for this work are available at: `http://kalogeratos.com/material/lrie-dra/`.

II. Related work

Most studies focus on static control where the strategy is applied prior to the epidemic. A budget of resources is used to set up barriers in the network aiming to reduce the spread of a possible virus threat through the population [4, 5, 6, 7, 8, 9]. In essence, such *vaccination* approach provides vaccines to immunize some nodes, thus, render them uninvolved in any future spread of the virus. A slightly different approach consists in distributing *antidotes* that, instead of immunizing a node, increase its recovery rate [10]. The considered priority of a node to receive a vaccine or antidote largely depends on graph-theoretic node attributes, such as centrality and degree, that ignore the nodes' infection states when an epidemic is already present.

Other past studies have considered *static resource allocation* [10, 11], or the particular case of *contact tracing* [12, 13]. The latter, study the potential pathways of disease transmission between individuals, in order to take highly invasive, local, and predesignated actions (e.g. treatment or isolation), which is especially effective for suppressing a DP when there are still very few infected nodes.

However, when dealing with a real-world epidemic scenario, it is natural for authorities to take real-time actions. This is achieved by utilizing any available information about the infection state of the population and its connectivity. While vaccination can be characterized as *preventive* since vaccines are given to healthy nodes, resource allocation is essentially *corrective* since its purpose is to aid the infected nodes to recover more quickly.

Indeed, recent studies have elaborated the idea of taking into account real-time information. *Antidote distribution* or *palliative care* has been investigated using dynamic strategies and optimal control [14, 15, 16, 17]. These works examine how the size of the available budget of resources should vary through time -yet not how it should be distributed- to efficiently suppress an epidemic while meeting, at the same time, certain cost and efficiency constraints. For this purpose, a *mean-field assumption* is made to describe the macroscopic evolution of the epidemic in terms of several high-level differential equations.

Dynamic resource allocation was studied in [13, 18]. In the first, the authors considered curing rates that can depend on the current node infection state, but only the dynamic contact tracing was investigated. In the second, a dynamic optimization framework was designed to produce strategies that are robust to adversarial intervention, and then, they were used for epidemic control. The authors also derive an optimal dynamic control strategy based on a simplification of the DP. However, it is questionable if their approach is applicable for large networks (their simulations were on networks of less than ten nodes).

To the best of our knowledge, the question of how the available resources should be distributed in an arbitrary and large network in real-time remains unanswered. Our work is not opposed to static vaccination strategies, or macroscopic dynamic control strategies that adjust the budget size of the resources. It is rather *complementary* to those: even if real-time reaction during the DP is possible, preventive action would still be valuable for the overall effectiveness of the DP control. Moreover, while herein we focus on *how to distribute* resources in the network, we do not address the question of *how many* resources should be given at each time, nor we consider auxiliary constraints (e.g. cost).

III. Framework for epidemic diffusion and control

A. Modeling epidemic spread: The *Susceptible-Infected-Susceptible* (SIS) model [19] is used to model the DP of a disease. Each node is either healthy, hence *susceptible* to be infected, or already *infected*. A susceptible node can become infected only if at least one of its neighboring nodes is already infected, while an infected node returns to the susceptible state after a period of recovery time. Therefore, the SIS model is appropriate for diseases for which the nodes do not develop permanent immunity and remain prone to multiple infections. Interesting to note, such a model can also be valid for *information diffusion* in a social network.

We use the standard SIS model formulated as a *Continuous-time Markov Process* [19]. In the following, $t \in \mathbb{R}_+$ will be the (continuous) time variable. Let $A \in \mathbb{R}^{N \times N}$ be the adjacency matrix of an arbitrary directed and weighted network of N nodes, where $A_{ij} = w_{ij}$ is a non-zero weight only if an edge exists from node i to node j. Let also $X(t)$ be the infection state vector at time t, where $X_i(t) = 1$ if node i is infected at time t, and 0 otherwise. The control of a DP is achieved via a *dynamic resource allocation* (DRA) approach: a set of nodes is determined to receive a treatment in order to recover more quickly. Let $R(t)$ be the vector representing the distribution of resources in the network, hence $R_i(t) = 1$ if node i receives a treatment at time t, and 0 otherwise. Using the formalism of [20], we model the diffusion with a continuous-time Markov process with the following transition rates:

$$\begin{aligned} &X_i(t) : 0 \to 1 \text{ at rate } \beta \textstyle\sum_j A_{ji} X_j(t); \\ &X_i(t) : 1 \to 0 \text{ at rate } \delta + \rho R_i(t), \end{aligned} \tag{3.1}$$

where β, δ, ρ are parameters describing, respectively, the infection rate, the recovery rate without a treatment, and the increase in the recovery rate that a treatment induces. Roughly speaking, Eq. 3.1 indicates that a susceptible node gets infected at a rate proportional to the number of its infected neighbors. Conversely, an infected node recovers at a constant rate δ if it does not receive a treatment, and at $\delta + \rho$ if it is being treated at this particular time ($R_i(t) = 1$). This model is similar to the *heterogeneous N-intertwined SIS* [10], except that we restrict the node recovery rate in $\{\delta, \delta + \rho\}$, instead of a general $\delta_i \in \mathbb{R}_+$. This also implies

that multiple resources cannot be accumulated on the same node to increase the recovery rate. Finally, we define two dimensionless parameters: $r = \frac{\beta}{\delta}$ the *effective spreading rate* of the DP, and $e = \frac{\rho}{\delta}$ the *treatment efficiency*.

B. Dynamic resource allocation (DRA): A DRA strategy determines the resource allocation $R(t)$ aiming to suppress the diffusion process. For a resource allocation $R(t)$, let $N_I^R(t) = \sum_i X_i(t)$ be the number of infected nodes at time t. We consider that the undesired diffusion process induces a cost $\mathcal{C}(N_I^R)$ which depends on the number of infected nodes during the epidemic under a given strategy R. The purpose of a DRA strategy is to minimize this cost using a predefined budget of resources $b(t)$, i.e. $\min_R \mathcal{C}(N_I^R)$. While many cost functions can be used for different purposes (see examples in §V-A), we will focus on the following cumulative cost (see [16]):

$$\mathcal{C}(N_I^R) = \int_{t=0}^{+\infty} e^{-\gamma t} \mathbb{E}[N_I^R(t)] dt, \tag{3.2}$$

where R is a valid DRA strategy and $\gamma \geq 0$ is a parameter that reduces the impact of the long-term behavior of $N_I^R(t)$. To simplify what follows, we omit to mention the strategy R in $N_I^R(t)$ when there is no ambiguity on the strategy used to decide the resource allocation.

We now define valid DRA strategies: we consider strategies which can depend on the state vector $X(t)$. Note that, since $X(t)$ is a stochastic function (i.e. a random variable in a function space), $R(t)$ 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, $R(t)$ is adapted to the *natural filtration* associated to $X(t)$. Formally, a DRA strategy is defined as the stochastic process:

$$\begin{aligned} &R : \mathbb{R}_+ \to \{0,1\}^N \\ &\text{s.t. } \forall t \in \mathbb{R}_+, \textstyle\sum_i R_i(t) \leq b(t). \end{aligned} \tag{3.3}$$

At each time instance t, a limited *budget* $b(t) \ll N$ of resources is available for distribution. The formulated framework is quite generic but an extensive analysis of the kinds of problem it could model is beyond the scope of this work. Here we suggest a series of constraints to devise a tractable problem variation to further work with:

- *Unlimited resources, disposed at constant rate*. A fixed amount of b_{tot} resources is available at each time.
- *Inability to store resources for later use*.

Intractability of optimal strategies. In theory, optimal strategies can be found using the framework of Markov decision processes (MDP) and optimal control. However, such approaches are computationally intractable due to the very large state space: 2^N states for a network of N nodes. A basic MDP approach would require to store a parameter for each element of the state space, which would be prohibitive in practice for even a network of 50 nodes. For this reason, we investigate other strategies and, specifically, a greedy heuristic for solving the DRA problem presented in §IV-B.

Algorithm 1 Applying a score-based DRA strategy

Input : infection state vector $X(t)$, budget size b_{tot}, scoring function S.
Output: the resource allocation vector $R(t)$.

if $\sum_i X_i(t) < b_{tot}$ **then**
  return $X(t)$
end if
Let $R(t)$ a zero N-dimensional vector
Let $V \leftarrow \{S_i(X(t))\}_{i=1}^N$ a vector containing the node scores
Sort the elements of V in *descending* order
  and let I the node indexes of the ranking
for $i = 1$ **to** b_{tot} **do**
  $R_{I(i)}(t) \leftarrow 1$
end for
return $R(t)$

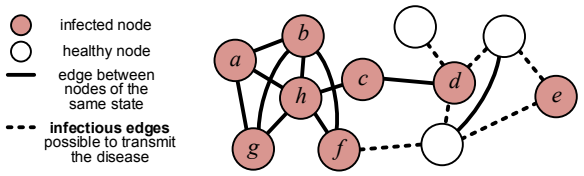


Fig. 1: Example network with healthy (*white*) and infected (*red*) nodes. *Dashed* edges denote *infectious edges* on which the disease might spread.

IV. DRA Control Strategies

A. Score-based strategies: A wide class of strategies can be described in terms of a scoring function S that takes as input the current infection state $X(t)$ and returns a priority order for the nodes of the network. More specifically, we define a *strategy based on score S* as a selection of the b_{tot} top ranked nodes according to $S(X(t))$:

$$R_i(t) = \begin{cases} 1 & \text{if } S_i(X(t)) \geq \theta_t; \\ 0 & \text{otherwise,} \end{cases} \tag{4.4}$$

where θ_t is a threshold value, set so that the distributed resources do not exceed the budget, i.e. $\sum_i R_i(t) = b_{tot}$. Note that, while the above formulation is general, simple scoring functions that rank the nodes based on their local properties, are not well-suited for planning coordinated actions, e.g. taking advantage of the position of other treatments when deciding where to allocate the current one. In the rest of the text, we will refer to the strategies that use scoring functions that are independent of the current state $X(t)$ as *static strategies*; they will form a baseline to assess the significance of real-time information for the DRA problem.

Alg. 1 presents a simple pseudocode for applying a score-based strategy. In general, its complexity is $O(E + N \log N)$ due to the sorting of N score values, where E is the number of edges and N the number of

nodes of the network. However, when the scoring function depends on *local* properties of the network and DP, the computational cost can be drastically reduced by partially updating the previous calculated node ranking, since only one node can change state at a time and only the scores of its neighboring nodes need to be updated. Any type of node attribute, or measurement associated to each node individually, can potentially be used to define a scoring function S. §IV-B presents the scoring functions considered in this article.

Fig. 1 shows the infection state of a network. Node h is the most connected, d has the highest diffusion rate (three healthy neighbors), e and h are the least and most probable to get reinfected if they recover. Scores that would give emphasis to properties like node *centrality* or *degree*, would tend to assign the highest priority to node h, while a strategy focusing on the most diffusive nodes would prefer to give a higher priority to node d.

B. LRIE – A greedy approach to DRA:

Optimization problem. In order to suppress a diffusion process as quickly and efficiently as possible, we consider the minimization of the integral of the number of infected nodes (see §V-A for a discussion on the quality metrics):

$$\min_R C_\gamma(R) = \int_{t=0}^{+\infty} e^{-\gamma t}\mathbb{E}[N_I(t)]dt, \tag{4.5}$$

where R is a valid DRA strategy (see §III-B) and $\gamma \geq 0$ is a parameter that reduces the impact of the long-term behavior of $N_I(t)$. Since the process is Markovian, it is straightforward that such an optimal strategy also minimizes $C_\gamma(R,t,X) = \int_{u=0}^{+\infty} e^{-\gamma u}\mathbb{E}[N_I(t+u)|X(t)=X]du$ for all time instances t and all state vectors $X \in \{0,1\}^N$. In the following, we achieve this optimization by approximating the short-term behavior of $\mathbb{E}[N_I(t+u)|X(t)=X]$.

The LRIE solution. Using a second-order approximation of $\Phi_{t,X}(u) = \mathbb{E}[N_I(t+u)|X(t)=X]$, we obtain the following approximation of $C_\gamma(R,t,X)$:

$$\begin{aligned} C_\gamma(R,t,X) = \tfrac{1}{\gamma}\textstyle\sum_i X_i + \tfrac{1}{\gamma^2}\Phi'_{t,X}(0) \\ + \tfrac{1}{\gamma^3}\Phi''_{t,X}(0) + O(\tfrac{1}{\gamma^4}), \end{aligned} \tag{4.6}$$

The minimization of the first and second order derivatives can be achieved simultaneously, and the resulting strategy, which we name *Largest Reduction in Infectious Edges* (LRIE), selects infected nodes according to the following scoring function:

$$\begin{aligned} S_{\text{LRIE}}(X(t)) &= A\overline{X}(t) - A^\top X(t) \\ &= \left[\textstyle\sum_j [A_{ij}\overline{X}_j(t) - A_{ji}X_j(t)]\right]_{i=1}^{N}, \end{aligned} \tag{4.7}$$

where $\overline{X}(t) = \mathbb{1} - X(t)$ is the vector indicating the healthy nodes, and $\mathbb{1}$ is the vector with ones for all coordinates.

This value can be seen as *the difference in the number of infectious edges* (i.e. edges that can transmit the disease from an infected to a healthy node) after healing a specific node. For the situation of Fig. 1, five infectious edges would be added if node h was healed. Respectively, one infectious edge would be added if node d was healed, while, if the infection was removed from node e, then the number of such edges would decrease by two. In essence, minimizing the number of infectious edges *reduces the scattering of the infection* across the network. Consequently, a smaller *front* is created separating the healthy region from the infected nodes, and enables better control over the DP. The pseudocode of Alg. 1 can be used for applying the LRIE strategy.

Note that, in theory, the method is also applicable to higher-order approximations of $C_\gamma(R,t,X)$. However, the mathematical complexity of the derivation of even third-order derivatives makes these improvements less practical, and better left to future investigation.

Scalability of LRIE. Eq. 4.7 can be easily computed for the needs of Alg. 1 by updating the score vector of the current state. When a node i changes state, only this node and its neighbors need an update of their scores. Then, sorting the scores will only cost $O(degree(i)\log N)$ provided the $(N - degree(i))$ other nodes are already sorted.

Technical details. We now provide the derivation of the S_{LRIE} score by computing the first and the second order derivatives of $\Phi_{t,X}(u)$. The Markov property of the process implies that $\Phi_{t,X}(u) = \Phi_{0,X}(u) = \mathbb{E}[N_I(u)]$ when the initial infection state is $X(0) = X$. From the formulation of Eq. 3.1, we derive the following formulas:

$$\begin{aligned} \tfrac{d}{dt}\mathbb{E}[X_i(t)] = &-\delta\mathbb{E}[X_i(t)] \\ &-\rho\mathbb{E}[X_i(t)R_i(t)] \\ &+\beta\textstyle\sum_j A_{ji}\mathbb{E}[\overline{X}_i(t)X_j(t)]. \end{aligned} \tag{4.8}$$

And, for two different nodes $i \neq j$:

$$\begin{aligned} \tfrac{d}{dt}\mathbb{E}[X_i(t)X_j(t)] = &-2\delta\mathbb{E}[X_i(t)X_j(t)] \\ &-\rho\mathbb{E}[X_i(t)X_j(t)(R_i(t)+R_j(t))] \\ &+\beta\textstyle\sum_k A_{ki}\mathbb{E}[\overline{X}_i(t)X_j(t)X_k(t)] \\ &+\beta\textstyle\sum_k A_{kj}\mathbb{E}[X_i(t)\overline{X}_j(t)X_k(t)]. \end{aligned} \tag{4.9}$$

Using Eq. 4.8, we can write the derivative of $\mathbb{E}[N_I(t)]$ as:

$$\begin{aligned} \tfrac{d}{dt}\mathbb{E}[N_I(t)] = &-\delta\mathbb{E}[N_I(t)] \\ &-\rho\mathbb{E}[X(t)^\top R(t)] \\ &+\beta\mathbb{E}[X(t)^\top A\overline{X}(t)], \end{aligned} \tag{4.10}$$

$$\text{and } \Phi'_{t,X}(0) = -\delta N_I(X) - \rho X^\top R(t) - \beta X^\top A\overline{X}. \tag{4.11}$$

Minimizing this derivative w.r.t. $R(t)$ is thus equivalent to only selecting nodes which are infected. In the following, we consider that $X_i(t) = 0 \Rightarrow R_i(t) = 0$ and $\sum_i R_i(t) = \min(b_{tot}, \sum_i X_i(t))$ (i.e. $R(t)$ minimizes the derivative in Eq. 4.10). Using Eq. 4.9 and 4.10, the second order derivative

of $\mathbb{E}[N_I(t)]$ can be written as:

$$\begin{aligned}\frac{d^2}{dt^2}\mathbb{E}[N_I(t)] &= -\delta\frac{d}{dt}\mathbb{E}[N_I(t)] \\ &\quad -\rho\frac{d}{dt}\mathbb{E}[X(t)^\top R(t)] \\ &\quad +\beta\frac{d}{dt}\mathbb{E}[X(t)^\top A\overline{X}(t)] \\ &= -\beta\rho\mathbb{E}[\{A\overline{X}(t)-A^\top X(t)\}^\top R(t)] \\ &\quad +\Xi(t),\end{aligned} \tag{4.12}$$

where $\Xi(t)$ is independent of $R(t)$ [2]. We thus have:

$$\Phi''_{t,X}(0) = -\beta\rho\{A\overline{X} - A^\top X\}^\top R(t) + \Xi, \tag{4.13}$$

where Ξ is independent of $R(t)$. Minimizing the second order approximation of $C_\gamma(R,t,X)$ is thus equivalent to selecting infected nodes that maximize the following score: $S_{\text{LRIE},i} = \sum_j [A_{ij}\overline{X}_j(t) - A_{ji}X_j(t)]$.

V. Experimental results

The DRA strategies were compared using simulations on various random and two real-world 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 initializations of the nodes' infection state when infection is less than 100%. In all cases we set $\delta = 1$. In order to conduct our simulations we developed a software package in Matlab that we made publicly available for research use at: `http://kalogeratos.com/material/lrie-dra/`.

A. Quality assessment for DRA strategies: In literature, many quality metrics are available related to diffusion processes (DP). In our experiments, we used the metrics below, which are functions of the number of infected nodes at time t (i.e. $N_I(t)$):

- *Time to extinction* (↓): $T_{ext} = \min_t\{N_I(t) = 0\}$ (see [20]). Describes the speed of convergence to a healthy network (necessarily bounded since $X(t) = 0$ is the only absorbent state of the Markov Process). Note that this value can be very high, specifically exponential w.r.t. N [20], for an aggressive DP and inefficient treatment resources (i.e. r high and e low, see §III-A).
- *Area Under the Curve* (AUC) (↓): $\int_{t=0}^{+\infty} N_I(t)dt$. This is a special case of *cumulative metrics* (see §3.2 and [16]) representing the overall cost of an epidemic (i.e. small value is better). However, AUC uses equal temporal weighting for short- and long-term behavior, thus gives the *total number of infected nodes* during the DP.
- *Stable infection state* (↓): $N_I(T_\infty)$ for a sufficiently large T_∞ value (see e.g. [4]). For a particularly aggressive DP, the epidemic becomes a pandemic and $N_I(t)$ converges, for reasonably large time period, to a stable non-zero value. Of course, this is not a true stationary value since, as explained above, $\lim_{t\to+\infty} N_I(t) = 0$. However, in the case of a pandemic, this convergence happens after an exponentially long time period [20], while the DP reaches a non-zero stable state in reasonable time.

Strategy	Scoring function $S^i(X)$ for node i
RAND	$\sigma(X_i)+R_i$, where R_i is i.i.d. uniform in [0, 1]
MN	$\sigma(X_i)+\sum_j A_{ij}$
PRC	$\sigma(X_i)+P_i$, where P_i is the PageRank score for node i
LRSR	$\sigma(X_i)+(\lambda_1-\lambda_1^{G\setminus i})$, where λ_1 is 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	$\sigma(X_i)+\sum_j A_{ij}\overline{X}_j$
LIN	$\sigma(X_i)-\sum_j A_{ji}X_j$
LRIE	$\sigma(X_i)+\sum_j[A_{ij}\overline{X}_j-A_{ji}X_j]$, sums MSN and LIN

TABLE I: Various derived DRA scoring functions. In all strategies, $\sigma(1) = 0$ and $\sigma(0) = -\infty$. Also, recall that $X(t)$ is the infection state vector and $\overline{X}(t)$ is the vector indicating the healthy nodes (see §IV-B).

Regarding the simulation results, they are being illustrated using the following figure types:

- *Line plots* represent, as solid lines, the expected number of infected nodes for each strategy, and their surrounding area is the 95% confidence interval under Gaussian hypothesis[3] (e.g. Fig. 2).
- *Heat maps* compare two strategies for a wide range of parameter values for the DP's effective spreading rate r and treatment efficiency e (see §III-A). In these simulations, we consider a total infection at the initial stage. The color of each point (e.g. Fig. 3) depicts the ratio $R_{(r,e)}$ of an employed quality metric on the performance of two strategies, for a set of values (r, e). Here AUC is used as quality metric: $\int_{t=0}^{T_\infty} N_I(t)dt$, where T_∞ denotes a sufficiently long time period. Contrary to the general AUC definition of the previous paragraph, here we stop the integration after a relatively long simulation period T_∞ in order to compare the quality of stable non-zero behaviors, as well. In this case, the ratio is equal to the ratio of *stable infection state* metrics.

***Technical details*.** In practice, the duration of the simulation plays a fundamental role in the quality of our results due to the integral definition of AUC. If the behavior is convergent to 0, then the simulation should run until convergence to the absorbent state. Otherwise, $N_I(t)$ is stationary for relatively long time, and the simulation can be terminated when stable behavior is reached. We used a statistical test to assess, with a 99% confidence, the non-zero value of the slope of $N_I(t)$ under a linear regression assumption. When the slope is

[2] We omit the complete derivation which is rather simple and straightforward. A technical appendix is available at: `http://kalogeratos.com/material/lrie-dra/`.

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

sufficiently small, we consider $N_I(t)$ to be stationary and we terminate the simulation. In such a case, the AUC is approximately $T_\infty N_{I\infty}$, where $N_{I\infty}$ is the stable value of the number of infected nodes.

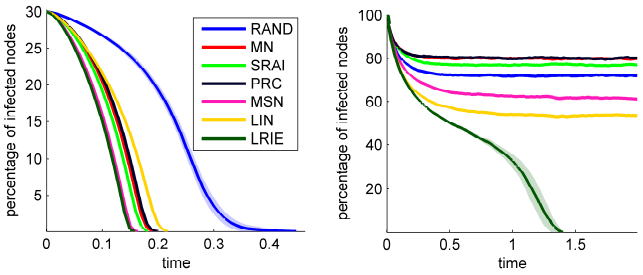
B. Competing strategies: To the best of our knowledge, there are no methods in the current literature that deal with the specific problem of diffusion control we introduced in §III-A. Consequently, the proposed *Largest Reduction in Infectious Edges* (LRIE) method is compared to several other heuristic scoring functions:

- *Random* (RAND): selects nodes uniformly at random, without replacement, among infected nodes.
- *Most Neighbors* (MN): selects the infected nodes with the largest number of neighbors.
- *PageRank Centrality* (PRC): selects the most central infected nodes according to PageRank algorithm [21].
- *Largest Reduction in Spectral Radius* (LRSR): selects the infected nodes which lead to the largest drop in the first eigenvalue of the adjacency matrix of the network.
- *Most Susceptible Neighbors* (MSN): selects the infected nodes with the most non-infected neighbors.
- *Least Infected Neighbors* (LIN): selects the infected nodes with the lowest number of infected neighbors.

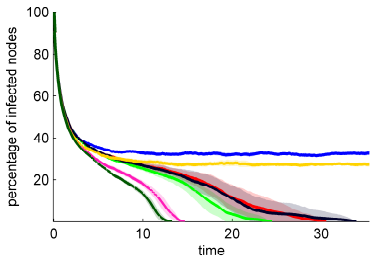
Tab. I provides the expressions to compute the scoring function that are experimentally compared. MN, PRC, and LRSR come from the static vaccination literature, and we will later refer to them as *centrality-based strategies* since they focus on nodes that are central in the network topology.

MSN and LIN are intuitive heuristics based on the assumptions that a node with many susceptible neighbors will spread the virus quickly, while a node with many infected neighbors will get infected with high probability. Notably, MSN and LIN are complementary to each other. Indirectly, the former focuses on '*central*' nodes with large degree, while the latter tends to target nodes at the network '*periphery*'. Thus, MSN and LIN capture different aspects of how critical a node is for the diffusion. The proposed LRIE strategy can also be seen as a combination of MSN and LIN, since it essentially seeks for nodes which are both *diffusive* for many healthy neighbors and, at the same time, *safe* in a mildly infected neighborhood.

C. Experiments on simulated networks: We used two types of random networks: i) *Erdös-Rényi networks* [22] (parameter: the edge probability p), ii) *scale-free networks* generated by the Barabási-Albert *preferential attachment* approach [23] (parameter: the number of added edges with each node m). We generated a different network for each simulation using the same values for model parameters).

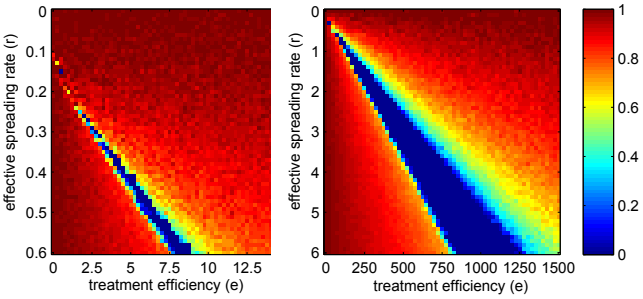


(a) $r=2$, $e=4000$, $b_{tot}=10$ (b) $r=2$, $e=3000$, $b_{tot}=10$



(c) $r=0.2$, $e=5$, and $b_{tot}=200$

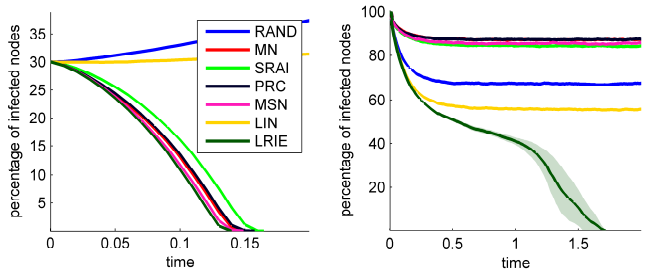
Fig. 2: Results for Erdös-Rényi networks: $N{=}10^4$ nodes, $p=0.001$.



(a) $b_{tot}=100$; small r, e (b) $b_{tot}=10$; large r, e values

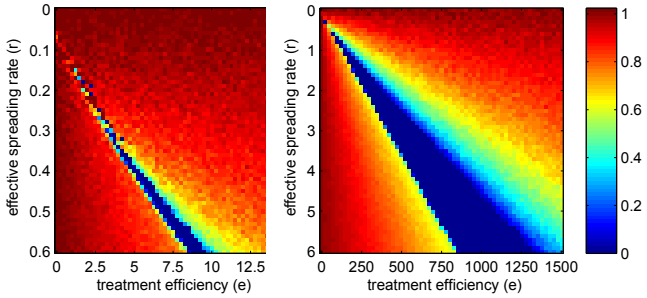
Fig. 3: Efficiency of LRIE compared to LRSR for an Erdös-Rényi network: $N=1000$, $p=0.01$. Small and large ranges of values are used for $r=\beta/\delta$, $e=\rho/\delta$.

1) Erdös-Rényi random networks: Fig. 2 presents simulation results on Erdös-Rényi networks with the same parametrization of the generator, while using different parameter values for r, e, b_{tot}, and two different cases of initial infected population. In all simulations LRIE performs better than the competing strategies. We observe two different behaviors depending on the percentage of initially infected population. If this is low (30% in Fig. 2a), then centrality-based strategies (MN, PRC, LRSR) perform well and eliminate the DP. However, when this percentage is large (100% in Fig. 2b) and the budget b_{tot} is low, only LRIE is able to eliminate the DP. More importantly, MN, PRC, and LRSR seem counter-effective, as they present worse results than the random strategy. The reason, in this case, is that central nodes have many infected neighbors which makes



(a) $e = 4000$ (b) $e = 3000$

Fig. 4: Results for random scale-free networks: $N = 10^4$ nodes, $m = 5$, $r = 2$, and $b_{tot} = 10$ resources.



(a) $b_{tot} = 100$; small r, e (b) $b_{tot} = 10$; large r, e values

Fig. 5: Efficiency of LRIE compared to LRSR for a scale-free network: $N = 1000$, $m = 5$. Small and large ranges of values are used for $r = \beta/\delta$ and $e = \rho/\delta$.

them prone to quick reinfection. Fig. 2c presents a scenario with only moderately effective treatments ($e = 5$). Note that Fig. 2 presents worse case scenarios for centrality-based strategies, whereas they do not provide insights whether those situations are usual or rare extreme cases.

In order to allow for more general observations, we present heat maps that compare LRIE to LRSR (the best among the competitors of LRIE) for a range of parameter values. Note that all the presented heat maps consider totally infected networks at the initial stage, and display three characteristic regions: i) the lower-left region in which both strategies are not able to eliminate the DP, ii) the middle blue region where LRIE eliminates the DP but its competitor fails to do the same, and iii) the upper-right region in which both strategies converge to $N_I(t) = 0$.

Fig. 3a considers a range of realistic parameter values (a low treatment efficiency, and moderate budget of resources). We can see that LRIE is always more efficient than LRSR (i.e. ratio < 1) and, in a large region of the space of parameter values, LRIE increases the relative quality of DP control by 10% or more. Also, there is a thin line where the ratio decreases to 0 indicating that LRIE eliminates the DP while LRSR fails to do so.

Finally, Fig. 3b compares LRIE to LRSR for a larger range of parameter values. In this setting, the efficiency of LRIE increases and the blue region becomes larger than that in Fig. 3a. The fact that such a region is approximately of

Network	DP scenario δ	r	e	b_{tot}	Strategy	AUC$_\downarrow$	$T_{ext\downarrow}$	$N_I(T)_\downarrow$
Twitter subgraph	1	0.2	300	100	RAND	∞	∞	30.6%
					MN	∞	∞	33.4%
					LRSR	246,476	7.70	0%
					MSN	89,671	2.52	0%
					LRIE	**64,425**	**2.07**	**0%**
	1	0.2	200	100	RAND	∞	∞	37.3%
					MN	∞	∞	42.3%
					LRSR	161,195	5.11	43.2%
					LRIE	**87,600**	**3.03**	**0%**
	1	0.2	50	100	RAND	∞	∞	46.4%
					MN	∞	∞	48.5%
					LRSR	∞	∞	48.9%
					MSN	∞	∞	44.4%
					LRIE	∞	∞	**29.2%**
US air traffic	1	2	210	50	RAND	∞	∞	26.1%
					MN	∞	∞	73.8%
					LRSR	3,723	1.81	0%
					MSN	3,235	1.65	0%
					LRIE	**493**	**0.43**	**0%**
	1	2	150	50	RAND	∞	∞	38.9%
					MN	∞	∞	76.6%
					LRSR	∞	∞	76.5%
					MSN	∞	∞	76.4%
					LRIE	**863**	**1.08**	**0%**
	1	2	100	50	RAND	∞	∞	49.7%
					MN	∞	∞	79.0%
					LRSR	∞	∞	79.2%
					MSN	∞	∞	77.4%
					LRIE	∞	∞	**23.1%**

TABLE II: Results of the simulations on two real networks. An infinite value for AUC and extinction time T_{ext} means that the number of infected nodes reached a non-zero stable infection state $N_I(T)$ in our simulations. $T = 16$ for the Twitter subgraph, and $T = 2$ for the US air traffic dataset.

the form $a\beta \leq \rho \leq b\beta$, with $a, b \in \mathbb{R}$, is characteristic of the DRA problem and indicates that, similarly to the epidemic threshold in δ/β in the absence of control [6], the diffusion process also displays a sudden switch in its behavior when ρ/β reaches a certain threshold, depending on the strategy.

2) Scale-free random networks: Scale-free networks are extremely prone to epidemics due to the existence of highly connected nodes. The behavior of the compared strategies are similar to the Erdös-Rényi case (see Fig. 4 and Fig. 5), except that the DP is more aggressive. In Fig. 4a, some of the strategies do not manage to converge, despite initiated with a low percentage of infected nodes. As expected, MN is more efficient in this case compared to PRC and LRSR since node degree is more significant attribute in a scale-free network than in a network with uniform random connections. Heat maps in Fig. 5 present similar characteristics to those of Fig. 3 for Erdös-Rényi networks, except there is a slightly improved performance for LRIE relatively to LRSR.

D. Simulations on real-world networks: Specifically:

- *The US air traffic* for the year 2010, containing $2,939$ nodes and $30,501$ edges. The nodes correspond to the US airports that serviced domestic and international flights, and those non-US airports that serviced flights

to US, during that year[4].

- *A Twitter subgraph* extracted from $1,000$ ego-networks of the social network [24]. The resulting network contains $81,306$ nodes and $1,342,303$ edges.

Tab. II summarizes the simulation results on these networks, where three scenarios were considered:

- *High treatment efficiency:* For the most efficient strategies, the DP reaches zero in reasonable time. In this case, AUC and extinction time are good quality metrics for the comparison of strategies.
- *Moderate treatment efficiency:* Only LRIE is able to eliminate the DP, thus showing the substantial improvement of the method over its competitors.
- *Low treatment efficiency:* In this case, the considered strategies suppress the epidemic but none of them eliminates it. However, LRIE still achieves a far lower stable state infection than its competitors.

In all three regions, LRIE seems robust and substantially outperforms its competitors. Note that for low treatment efficiency, centrality-based strategies become counter-effective with even higher stable infection level than that of RAND. Intuitively, this result is due to the fact that at the beginning of the DP the whole network is infected. Although central nodes have many infected neighbors and are prone to fast reinfections, these strategies will keep their focus on these highly connected nodes and will hence fail to clear the central part of the network, which results in a high stable infection level. Contrary, LRIE indirectly tends to contain the infected nodes in *clusters* and reduces infection's scattering. Even with low treatment efficiency, LRIE will first focus on the periphery and gradually contain the DP to the central part of the network, achieving a significantly lower stable infection percentage.

VI. Conclusion and future work

In this paper, we investigated the use of real-time information for allocating treatment resources in a network so as to suppress a diffusion process. First, we presented a model formulation as a *dynamic resource allocation problem* (DRA). Then, we proposed the novel *Largest Reduction in Infectious Edges* (LRIE) control strategy which minimizes the second-order approximation of the cost associated with a diffusion process. We analyzed that this approach minimizes the number of *infectious edges* that can transmit the infection to healthy nodes, and thus reduces the scattering of the infection across the network. Simulations on various randomly generated and two real-world networks showed that the proposed LRIE strategy is the most effective and robust among the compared strategies. Finally, we released a version of our simulation software for research use.

[4]Source of data: OpenFlights, `http://Openflights.org`. Available at T. Opsahl's web post: `http://wp.me/poFcY-Vw`.

Our plans for future work include a deeper theoretical study of the DRA problem as well as the development of better strategies. Notably, the combination of the short-term criterion used by LRIE and some kind of longer-term planning would certainly improve the control performance. This long-term planning could benefit from the particular characteristics of social networks, such as clusters.

References

[1] Nicholas A. A. Christakis and James H. H. Fowler. The Spread of Obesity in a Large Social Network over 32 Years. *NEJM*, 357(4):370–379, July 2007. ISSN 0028-4793. doi: 10.1056/nejmsa066082.

[2] Jure Leskovec, Lars Backstrom, and Jon Kleinberg. Meme-tracking and the dynamics of the news cycle. In *Proc. of the ACM SIGKDD*, pages 497–506, 2009.

[3] Wojciech Galuba, Karl Aberer, Dipanjan Chakraborty, Zoran Despotovic, and Wolfgang Kellerer. Outtweeting the twitterers - predicting information cascades in microblogs. In *3rd Workshop on Online Social Networks (WOSN)*, 2010.

[4] Reuven Cohen, Shlomo Havlin, and Daniel Ben-Avraham. Efficient immunization strategies for computer networks and populations. *PRL*, 91(24):247901, 2003.

[5] Hanghang Tong, B Aditya Prakash, Tina Eliassi-Rad, Michalis Faloutsos, and Christos Faloutsos. Gelling, and melting, large graphs by edge manipulation. In *Proc. of the ACM CIKM*, pages 245–254, 2012.

[6] Yang Wang, Deepayan Chakrabarti, Chenxi Wang, and Christos Faloutsos. Epidemic spreading in real networks: An eigenvalue viewpoint. In *Proc. of the IEEE SRDS*, pages 25–34, 2003.

[7] Christian M Schneider, Tamara Mihaljev, Shlomo Havlin, and Hans J Herrmann. Suppressing epidemics with a limited amount of immunization units. *Physical Review E*, 84(6):061911, 2011.

[8] B Aditya Prakash, Hanghang Tong, Nicholas Valler, Michalis Faloutsos, and Christos Faloutsos. Virus propagation on time-varying networks: Theory and immunization algorithms. In *MLKDD*, pages 99–114. Springer, 2010.

[9] Victor M Preciado, Michael Zargham, Chinwendu Enyioha, Ali Jadbabaie, and George Pappas. Optimal vaccine allocation to control epidemic outbreaks in arbitrary networks. In *Proc. of the IEEE CDC*, pages 7486–7491, 2013.

[10] Victor M. Preciado, Michael Zargham, Chinwendu Enyioha, Ali Jadbabaie, and George J. Pappas. Optimal resource allocation for network protection: A geometric programming approach. *CoRR*, abs/1309.6270, 2013.

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

[12] Ken TD Eames and Matt J Keeling. Contact tracing and disease control. In *Proc. of the Royal Society of London. Series B: Biological Sciences*, pages 2565–2571. The Royal Society, 2003.

[13] Christian Borgs, Jennifer Chayes, Ayalvadi Ganesh, and Amin Saberi. How to distribute antidote to control epidemics. *Random Structures and Algorithms*, 37(2):204–222, 2010.

[14] Reza Yaesoubi and Ted Cohen. Dynamic health policies for controlling the spread of emerging infections: Influenza as an example. *PLoS ONE*, 6(9): e24043, 09 2011. doi: 10.1371/journal.pone.0024043.

[15] Petra Klepac, Ottar N. Bjørnstad, C. Jessica E. Metcalf, and Bryan T. Grenfell. Optimizing reactive responses to outbreaks of immunizing infections: Balancing case management and vaccination. *PLoS ONE*, 7(8):e41428, 2012. doi: 10.1371/journal.pone.0041428.

[16] Graeme A Forster and Christopher A Gilligan. Optimizing the control of disease infestations at the landscape scale. In *Proc. of the NAS*, pages 4984–4989, 2007.

[17] MHR Khouzani, Saswati Sarkar, and Eitan Altman. Optimal control of epidemic evolution. In *Proc. of the IEEE INFOCOM*, pages 1683–1691, 2011.

[18] Ali Khanafer and Tamer Başar. Information spread in networks: Control, games, and equilibria. In *Information Theory and Applications Workshop (ITA)*, 2014.

[19] Piet Van Mieghem, Jasmina Omic, and Robert Kooij. Virus spread in networks. *IEEE/ACM TON*, 17(1):1–14, 2009.

[20] Ayalvadi Ganesh, Laurent Massoulié, and Don Towsley. The effect of network topology on the spread of epidemics. In *Proc. of the IEEE INFOCOM*, pages 1455–1466, 2005.

[21] Mark Newman. *Networks: An Introduction*. Oxford University Press, New York, NY, USA, 2010.

[22] P Erdős and A Rényi. On the evolution of random graphs. *Selected Papers of A. Rényi*, 2:482–525, 1976.

[23] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[24] Julian J. McAuley and Jure Leskovec. Learning to discover social circles in ego networks. In *Proc. of the NIPS*, pages 548–556, 2012.