Partial network immunization in Continuous-Time Information Cascades

Argyris Kalogeratos[1], Kevin Scaman[1*,2], Luca Corinzia[1*,3], and Nicolas Vayatis[1]

[1] CMLA – ENS Cachan, CNRS, Université Paris-Saclay, 94230 Cachan, France
[2] MSR-Inria Joint Center, 91120 Palaiseau, France
[3] ETH Zürich, 8092 Zürich, Switzerland
{kalogeratos, scaman, vayatis}@cmla.ens-cachan.fr, lucac@ethz.ch

1 Introduction

Studying the way in which diffusion processes evolve in networks is fundamental for further understanding such complex and dynamic phenomena. In particular, being able to predict the number of nodes that will be reached at the end of the spread when that starts from a known set of initial infection seeds (i.e. seeds' *influence*) is of broad interest. Taking preparatory measures by acting on the network, so as to reduce the reach of a possible future diffusion, is a core administration problem. Traditionally, that has been studied as a way to improve public healthcare (e.g. through vaccination), recently though it has attracted a lot of attention due to the concerns raised by cases of malicious information propagation in social networks (e.g. fake news, rumors).

In the existing literature the role of the spectral radius of the adjacency matrix representing the underlying network has been largely highlighted as a quantity tightly connected with the epidemic threshold over which the reach of the diffusion explodes and becomes comparable with the network size [8,4,1]. Various studies have been presented for virus propagation models and influence maximization [10,6,5].

In this paper we present a brief overview of our recent work on partial node immunization in Continuous-Time Information Cascade Model ($\mathcal{CTIC}$) [7]. $\mathcal{CTIC}$ [2] is a stochastic model allowing propagation rates along edges to vary in time. Relying on previous work, we use the concept of *Hazard radius* introduced in [4], that is highly correlated to the influence and helps us in deriving upper bounds for the influence under the $\mathcal{CTIC}$. We subsequently develop the *NetShape* strategy that enjoys a convex relaxation and, among other influence optimization tasks that we do not go through in this short summary, it can be used for *offline and partial node immunization*. In that scenario, a budget of treatment units is available. Each treatment unit can target a single node, in advance of the diffusion, thereafter reducing node's propagation rates along all of its outgoing edges by a fixed factor.

2 Results

The NetShape method. Formally, let $\mathcal{F}_{ij}(s-\tau_i)$ be the *Hazard function*, an element of the *Hazard matrix* $\mathcal{F}$, representing the propagation rate on edge $i \to j$ at a specific time

* Part of the work has been conducted while author was at CMLA[1].

s after τ_i when node i received the piece of information and got 'infected'. Also, let the *Hazard radius* be the spectral radius of a matrix computed by integrating the Hazard functions over time (i.e. the component-wise integration of the symmetrized $\mathcal{F}$):

$$\rho_H(\mathcal{F}) = \rho\left(\int_0^{+\infty} \frac{\mathcal{F}(t) + \mathcal{F}(t)^\mathsf{T}}{2} dt\right), \tag{1}$$

where $\rho(\cdot) = \max_i |\lambda_i|$, and λ_i are the eigenvalues of the implied input matrix (since we refer to square matrices). By further elaborating results from [4], we have shown that the maximum influence cannot exceed a certain proportion of the network that is non-decreasing with $\rho_H(\mathcal{F})$, and displays a sharp transition between a sub-critical and super-critical regime. Therefore, we solve the following optimization problem over a set of feasible Hazard matrices $\mathbb{F}$ that can be produced by valid actions on the nodes:

$$\mathcal{F}^* = \operatorname{argmin}_{\mathcal{F}\in\mathbb{F}}\ \rho_H(\mathcal{F}). \tag{2}$$

When $\mathbb{F}$ is a convex set, this optimization problem is also convex and the proposed *NetShape* method uses a simple *projected subgradient descent* scheme to solve it. The interested reader is refereed to [7] for more technical details on NetShape algorithm.

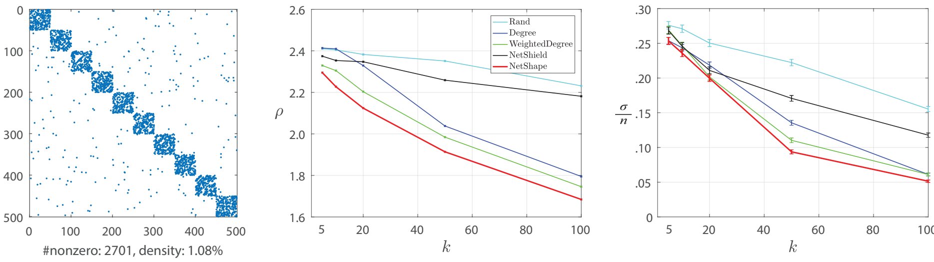
Experimental evaluation. We evaluated the NetShape algorithm for the *offline and partial node immunization* under the $\mathcal{CTIC}$ and compared it with baseline and state-of-the-art approaches for selecting the k nodes to target (k is the provided budget): **i)** random node selection (*Rand*); **ii)** selection of the nodes with the highest out-degree (*Degree*); **iii)** selection of k nodes with highest sum of outgoing edge weight $w_{ij} = \int_0^{+\infty} \mathcal{F}_{ij}(t)dt$ (*WeightedDegree*), actually derived by the optimization of the lower bound LB_1 in [3]; **iv)** the *NetShield* algorithm from [9] (originally designed for total immunization).

For our empirical evaluation we used an artificial random network with $n = 500$ nodes generated as follows: 10 equally-sized Erdős Rényi clusters were first created with edge creation probability $p = 0.1$, then their adjacency matrices were synthesized in a block-diagonal structure with a uniform inter-cluster rewiring probability $p' = 0.001$. Fig. 1a shows the structure of the adjacency matrix. Finally, the weights of the created edges (i.e. the transmission probabilities) were generated using a *trivalency model* that picks values uniformly at random from the set $\{low\colon 0.1, medium\colon 0.2, high\colon 0.5\}$.

Each treatment budget can be assigned to a single node and, here, we assume that it can cause a fixed decrease of 70% in that node's propagation rate along all of its edges. Fig. 1b, c plot the curves (average values and stds) of two evaluation measures for our simulations over a set of budget sizes. For each k value, we run 1000 simulations and each simulation starts from nodes of high influence. The measure reported in Fig. 1 c is the influence of the selected seeds, i.e. the expected proportion $\frac{\sigma}{n}$ of infected nodes at the end of the process. Also, the measure plotted in Fig. 1 c is the spectral radius $\rho_H(\mathcal{F})$ of the Hazard matrix that NetShape minimizes as a proxy for influence reduction.

Note that our purpose was to test in a meaningful parametrization scenario where the spectral radius of the original network would have been close to 1 and, thus, its decrease could cause a non-negligible reduction to the influence.

Performance results. The brief reported results show that: **i)** NetShape optimizes the spectral radius $\rho_H(\mathcal{F})$ (an empirical proof of correctness for our optimization scheme), **ii)** effectively minimizes the influence (verifying the relevance of our optimization to



(a) (b) (c)

Fig. 1. Comparison of NetShape's performance against competitors on an artificially generated random network. Tested k values: $\{5, 10, 20, 50, 100\}$. (a) The structure of the generated non-symmetric, block-diagonal adjacency matrix (here plotted as binary matrix); (b) spectral radius $\rho_H(\mathcal{F})$ vs. budget k; (c) influence: the expected proportion of infected nodes $\frac{\sigma}{n}$ vs. budget k.

the influence), **iii)** outperforms the competitors in both previous points; the largest difference is observed -as one could expect- when a moderate amount of treatments are available. In conclusion, the presented approach seems promising and we plan to investigate its potential generalization to other influence optimization problems.

References

1. Chen, C., Tong, H., Prakash, B.A., Tsourakakis, C.E., Eliassi-Rad, T., Faloutsos, C., Chau, D.H.: Node immunization on large graphs: Theory and algorithms. IEEE Transactions on Knowledge and Data Engineering 28(1), 113–126 (2016)
2. Chen, W., Lakshmanan, L.V., Castillo, C.: Information and influence propagation in social networks. Synthesis Lectures on Data Management 5(4), 1–177 (2013)
3. Khim, J.T., Jog, V., Loh, P.L.: Computing and maximizing influence in linear threshold and triggering models. In: Proceedings of the Advances in Neural Information Processing Systems 29. pp. 4538–4546 (2016)
4. Lemonnier, R., Scaman, K., Vayatis, N.: Tight bounds for influence in diffusion networks and application to bond percolation and epidemiology. In: Advances in Neural Information Processing Systems. pp. 846–854 (2014)
5. Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., Glance, N.: Cost-effective outbreak detection in networks. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 420–429 (2007)
6. Ohsaka, N., Akiba, T., Yoshida, Y., Kawarabayashi, K.i.: Fast and accurate influence maximization on large networks with pruned Monte-Carlo simulations. In: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 138–144 (2014)
7. Scaman, K., Kalogeratos, A., Corinzia, L., Vayatis, N.: A spectral method for activity shaping in continuous-time information cascades. ArXiv e-prints (Sep 2017)
8. Scaman, K., Lemonnier, R., Vayatis, N.: Anytime influence bounds and the explosive behavior of continuous-time diffusion networks. In: Advances in Neural Information Processing Systems. pp. 2017–2025 (2015)
9. Tong, H., Prakash, B.A., Tsourakakis, C., Eliassi-Rad, T., Faloutsos, C., Chau, D.H.: On the vulnerability of large graphs. In: Proceedings of the IEEE International Conference on Data Mining. pp. 1091–1096 (2010)
10. Wang, Y., Chakrabarti, D., Wang, C., Faloutsos, C.: Epidemic spreading in real networks: An eigenvalue viewpoint. In: Proceedings of the IEEE International Symposium on Reliable Distributed Systems. pp. 25–34 (2003)