Semi-metric topology characterizes epidemic spreading on complex networks

David Soriano-Paños[1,2], Felipe Xavier Costa[3,1,4], and Luis M. Rocha[3,1]
[1] Instituto Gulbenkian de Ciência, 2780-156 Oeiras (Portugal).
[2] GOTHAM lab, Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza, 50018 Zaragoza (Spain).
[3] Systems Science and Industrial Engineering Department, Binghamton University, Binghamton, NY 13902, USA.
[4] Department of Physics, State University of New York at Albany, Albany, NY 12222, USA.

The interplay between network sparsification and spreading phenomena constitutes an important challenge in the Big Data Era. Specifically, the interest of this problem is twofold: pruning the large set of redundant connections usually included in high resolution databases reduces the computational cost of the simulations but also unveils the primary subgraphs sustaining the spread of a pathogen. Mounting evidence in the literature suggest that network sparsification relying on global information allows for a better retrieval of spreading dynamics than just removing the weakest connections. For instance, Mercier et al. [1] found that the effective resistance, which accounts for the relevance of a given edge in the ensemble of paths connecting their two nodes, outperforms weight thresholding in preserving SIR dynamics. Recently, Correia et al. [2] also showed that the metric backbone, containing all the connections with non-zero edge betweenness, provides more solid foundations for network sparsification than relying on local information and constitutes a primary subgraph for disease transmission.

Despite the relevance of the metric backbone, determining the network features limiting the reconstruction of epidemic outbreaks from the union of shortest paths remains an open problem. In this talk, we will introduce a new method to construct synthetic metric backbones with tunable properties, allowing us to explore the relevance of different network features for the latter problem. Our results suggest that both the relative size of the backbone, i.e. the fraction of edges with non-zero edges betweenness in the network, and the semi-metric distortion values, quantifying the redundancy of the removed paths, are crucial for the reconstruction of outbreaks generated by a SI model. Therefore, our ability to reconstruct the outbreaks from sparsified networks comes from an interplay between the quantity and quality of paths removed while pruning the network.

Building on the latter result, we will propose a new sparsification process relying on the semi-metric distortion distribution. Figure 1 shows that this method allows for a better retrieval of the dynamics on the original network while not disrupting its functionality, in comparison to other methods studied in the literature such as those relying on effective resistance or weights to remove connections. Finally, we will generalize our analysis to other epidemic dynamics such as the SIS or SIR models, showing that metric backbone fails in capturing the global prevalence of diseases but provides an accurate microscopic ranking of the different individual exposures to the circulating pathogen.

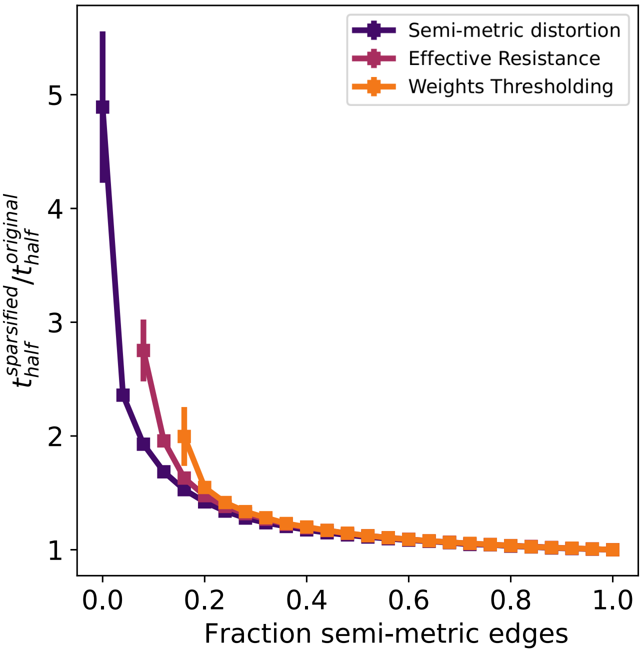

Fig. 1. Ratio between the time for a SI dynamics to reach half of the network in the sparsified configuration $t_{half}^{sparsified}$ and the original network $t_{half}^{original}$ as a function of the number of edges removed during sparsification. Note that this number of edges is indicated in terms of fraction of semi-metric edges, which are those not belonging to any shortest path in the network. Each color correspond to a different sparsification method based on a given edge feature: semi-metric distortion (blue), effective resistance (purple) and weights (orange). Curves are interrupted when sparsification breaks the giant component of the network. The results are obtained by averaging 200 realizations for 50 different infectious seed for an infectivity $\beta = 0.5$ in a network constructed from face-to-face interactions in a school [3].

[1] A. Mercier, S. Scarpino and C. Moore. PLOS Computational Biology **18**(11), e1010650 (2022).

[2] R.B. Correia, A. Barrat and L.M. Rocha. PLOS Computational Biology **19**, e1010854 (2023).

[3] DJA Toth et al. Journal of The Royal Society Interface **12**(108), 20150279 (2015).