Evaluating Graph Neural Networks for Epidemic Source Detection: A Benchmark Study

M. Sterchi[1,2] and L. Hilfiker[3]

1. University of Applied Sciences and Arts Northwestern Switzerland, Switzerland
2. University of Zurich, Switzerland
3. University of Bern, Switzerland

Detecting the source of an epidemic is crucial for effective containment. Graph Neural Networks (GNNs) have recently been proposed as a solution to this task for outbreaks on contact networks [1–4]. While these studies offer limited comparisons with other methods, a benchmark study evaluating GNNs against a broad range of traditional source detection techniques developed over the past 14 years is lacking. The question remains: what is the true potential of GNNs in the context of epidemic source detection? With the present study we take a first step towards a comprehensive evaluation of GNNs for source detection on empirical networks. We address the single-source detection problem on static, undirected contact networks, modeled using a continuous-time *susceptible-infectious-recovered* (SIR) process with known parameters. Given the network and a snapshot of epidemic states at a known duration after the outbreak began, the goal is to estimate the true source node. For this, we adopt a standard (convolutional) GNN architecture [5]. We benchmark the GNN's performance against several methods representative of traditional approaches: these include the Jordan centrality method (CENTRALITY) [6], the soft margin estimator (SME) [7], and a factorized (mean-field) likelihood method utilizing Monte Carlo estimates of node state probabilities (MCMF) [8]. Additionally, we include two baselines: random selection of the source (RANDOM) and an untrained GNN.

Detection performance is evaluated on a variety of networks, revealing two key findings: (i) GNNs offer no substantial advantage over meaningful benchmarks and perform consistently worse than SME and MCMF, and (ii) GNNs exhibit particularly poor performance for small- to medium-sized outbreaks, which is relevant as an epidemic may still be contained at this stage. Fig. 1 shows the results for one of the networks we analyzed. Overall, our findings suggest that the related work may overestimate the potential of GNNs for the source detection problem. Nonetheless, we plan to further explore the GNN design space and to investigate the impact of feature augmentation in future work.

[1] Dong, M. et al., *CIKM*, (2019). DOI
[2] Shah, C. et al., *arXiv*, (2020). arXiv Link
[3] Sha, H. et al., *IEEE DSAA*, (2021). DOI
[4] Ru, X. et al., *AAAI*, (2023). DOI
[5] Morris, Ch. et al., *AAAI*, (2019). DOI
[6] Zhu, K. et al., *IEEE TNET* 24, 1 (2016). DOI
[7] Antulov-Fantulin, N. et al., *Phys. Rev. Lett.* 114, 248701 (2015). DOI
[8] Sterchi, M. et al., *Scientific Reports* 13, 1 (2023). DOI
[9] Haraldsdottir, S. et al., *J Acquir Immune Defic Syndr* 5, 4 (1992).

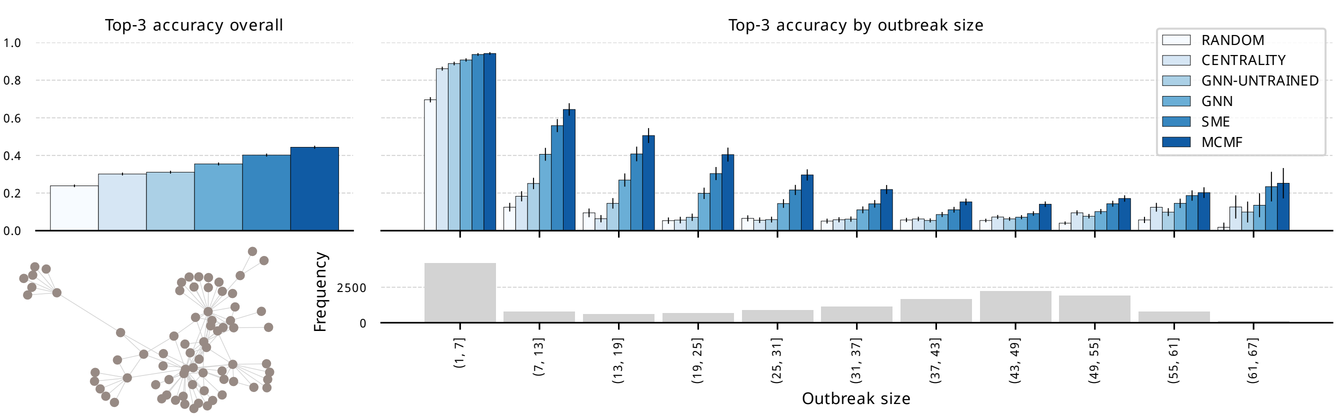

Figure 1: Results for a network of 75 individuals in Iceland (114 edges) [9]. The test outbreaks have a basic reproduction number of $R_0 \approx 2$ and the snapshots are taken when 40% of the nodes are infected, on average. **Top left:** Top-3 accuracy for all simulated test outbreaks. **Top right:** Top-3 accuracy for test outbreaks categorized according to size. **Bottom left:** Iceland network. **Bottom right:** Absolute frequencies of categorized test outbreaks.