Group Testing on a Network: Supplementary Material

Arlei Silva, Ambuj K. Singh
University of California, Santa Barbara {arlei,ambuj}@cs.ucsb.edu

More Details on Algorithms

In the Section Algorithms for Group Testing on Networks, we have introduced two four algorithms for the GTN problem (see Section Problem Definition). Due to space constraints, we will cover some details about the algorithms, including their time complexity analysis, in this section.

The implementation of the proposed algorithms is quite simple. For the case of sampling-based approaches, we compute scoring functions (Δ^g and Δ^{kl}) using union and difference operations over lists of infection events. We leave the optimization of the performance of these algorithms as future work.

Topology-based Algorithms

These algorithms identify groups for testing by maximizing the total weight of within-group edges while guaranteeing that group sizes are at most k.

Greedy-Topology: Initializes groups with a single member and merges the two groups that maximize the total weight of within-group edges if their combined size is at most k. We use a binary heap to store and retrieve candidate group merge operations. The number of iterations (lines 3-4) of Algorithm 2 is $O(|V|)$. In each iteration, updating the heap takes $O(\log(n))$, where n is the current size of the heap, plus the cost of computing scores Δ^g for pairs of groups containing the newly merged group, which takes $O(kd)$. Thus the total complexity is $O(|V|(\log(|V|) + kd))$.

KL-Topology: We initialize groups ($C^{(0)}$) using *Greedy-Topology*. Lines 6-9 of Algorithm 3 are executed stm^2 times. For each iteration, the cost of identifying the optimal vertex swap operation between two groups is $O(k^2d)$ and, so, the total complexity is $O(stm^2k^2d)$.

Sampling-based Algorithms

These algorithms identify groups for testing by minimizing the total expected number of tests over samples from the infection process.

Greedy-Sampling: Similar to *Greedy-Topology*, but performs merge operations that minimize the expected number of tests based on samples. Computing scores Δ^g for new groups takes $O(kqz)$ time and thus the total complexity is $O(|V|(\log(|V|) + kqz))$.

Greedy-Topology: Applies *Greedy-Sampling* for initialization. The cost of identifying optimal vertex swaps between groups is $O(k^2qz)$ and the total complexity is $O(stm^2k^2qm)$.

Code and Data

We have implemented the methods evaluated in this paper in Python (version 3). These implementations and also code for reproducing the experiments are shared as a github repository.[1] We do not own the datasets used in our experiments but they can be obtained from their authors (see details in Experimental Settings in main paper).

SIR Model Parameters

Table 1 shows the parameters used in our SIR simulations for each dataset. We found that these parameters led to a large enough portion of the network to be infected at some point during the epidemic process, allowing us to evaluate testing approaches at multiple prevalence levels. Figure 2 shows typical infection curves—with numbers of susceptible, infected and recovered individuals over time—for each dataset. Figure 1 shows visualizations of the Primary School dataset with subsets of infected vertices for values of prevalence varying from 2% to 32%.

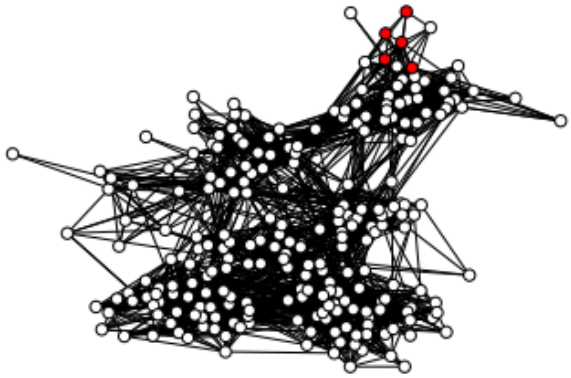
Robustness to Missing Transmission Links

We have evaluated the robustness of group testing approaches to missing transmission links in our experiments. The motivation for such an analysis is that we expect the transmission network to be generated based on contact tracing, which is prone to errors. In particular, we focus on missing links because those are more likely to occur in real settings. From Figure 1, we can notice that most links in the network do not transmit an infection. Thus, we expect our approaches to be robust also to the addition of false edges.

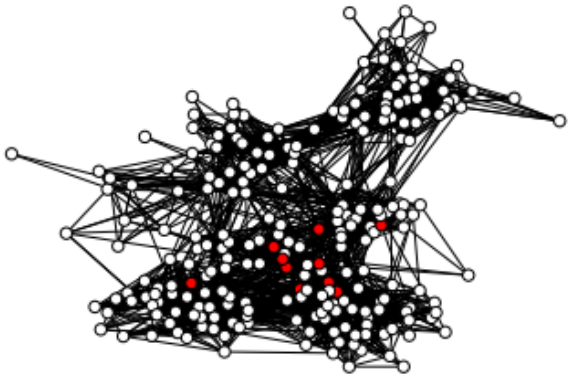
[1] https://github.com/arleilps/group-testing

Dataset	**Transmission rate (τ)**	**Recovery rate (γ)**
Primary School (PS)	40.0	1.5
High School (HS)	40.0	0.2
Company (CP)	40.0	1.0
Conference (CF)	40.0	0.2
Erdos-Renyi (ER)	1.0	1.0
Gaussian Rand. Part. (GRP)	0.1	0.1
Gowalla (GW)	10.0	1.0

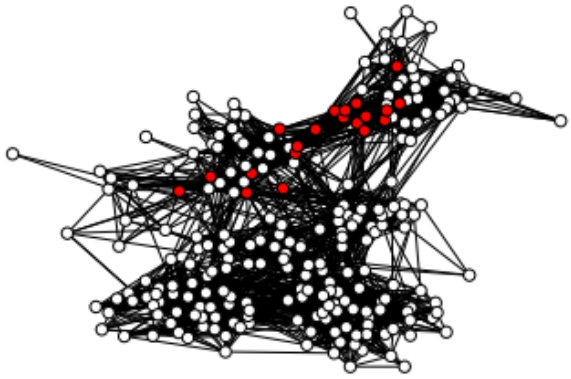
Table 1: SIR parameter settings for each dataset. The number of seeds N was set to 1 for all datasets. These parameters lead to the infection curves shown in Figure 2, where a large enough part of the network is infected at some point in the process.



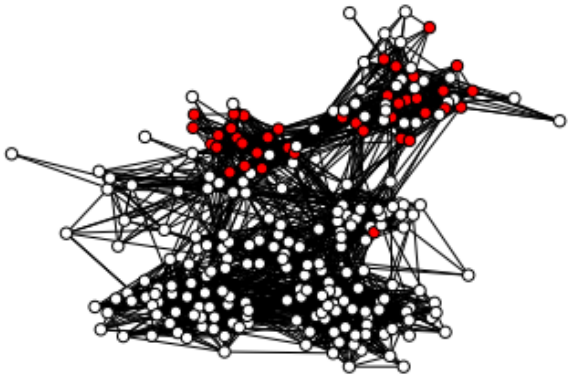
(a) Prevalence=2%



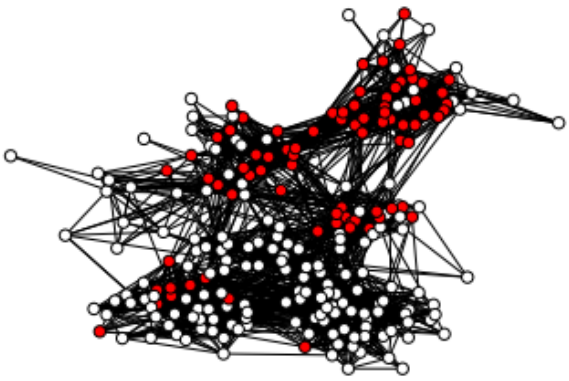
(b) Prevalence=4%



(c) Prevalence=8%

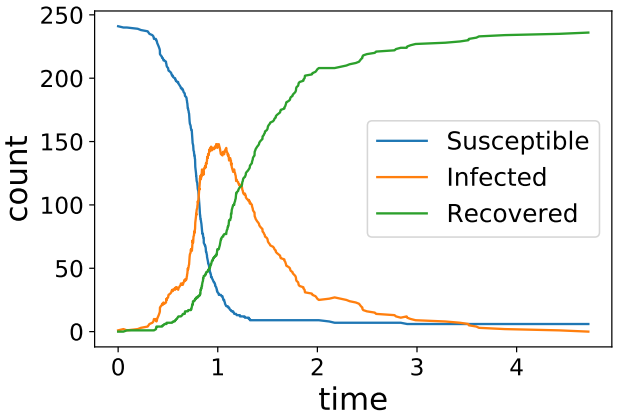


(d) Prevalence=16%

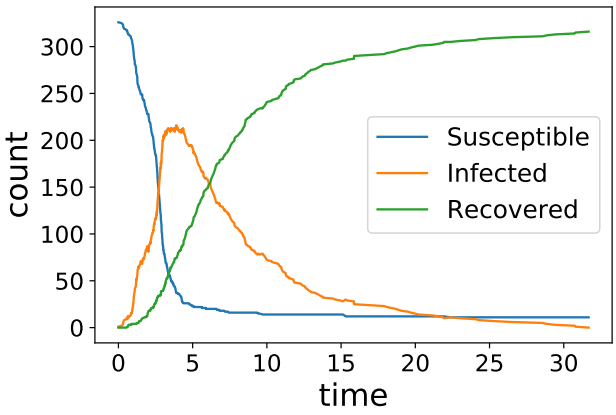


(e) Prevalence=32%

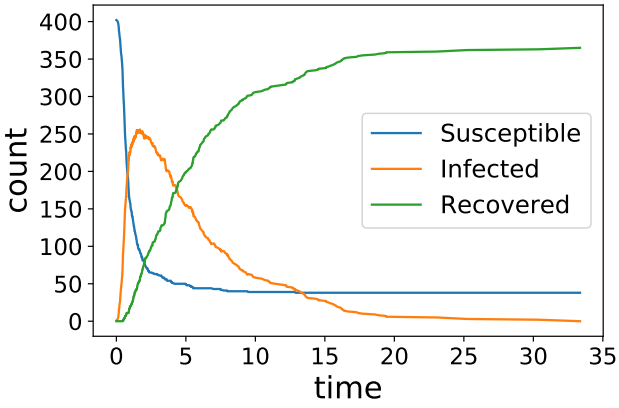
Figure 1: Examples of infections on the Primary School dataset for varying values of prevalence from 2-32%. Red nodes are infected and white nodes are not infected. Better seen in color.



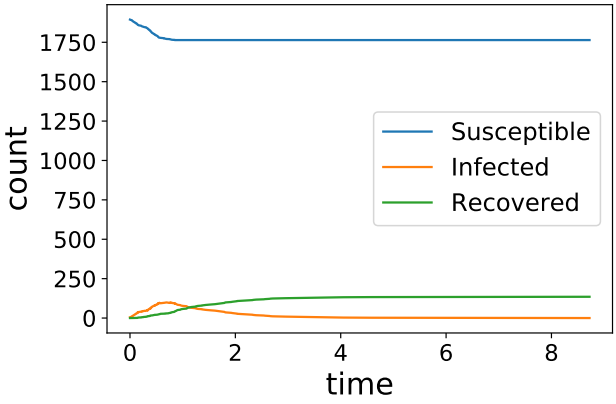
(a) Primary School (PS)



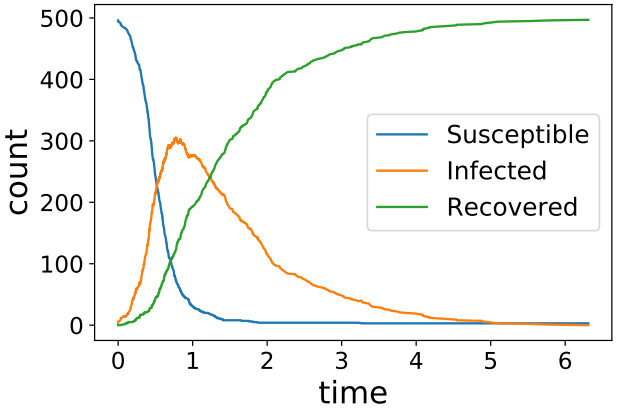
(b) High School (HS)



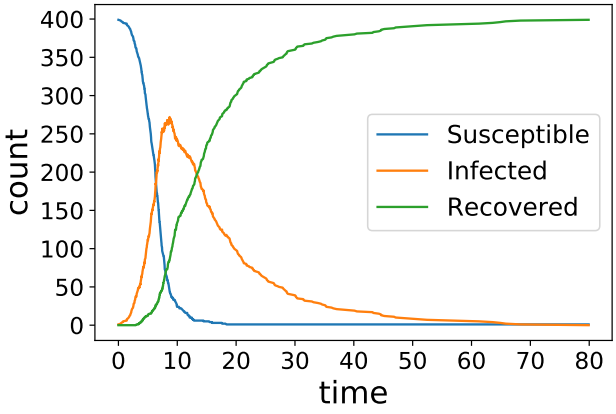
(c) Conference (CF)



(d) Gowalla (GW)



(e) Erdos-Renyi (ER)



(f) Gaussian Rand. Part (GRP)

Figure 2: Representative infection curves for SIR model using our parameter settings for each dataset. Because Gowalla is directed, with mostly small connected components, the infection only reaches a small fraction of the network.