Graph Partition Based Adaptive Group Testing Strategy

Jingyi Zhang Lenwood S. Heath

Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061

Abstract

Within an epidemic, mass testing is an essential method to bring healthy individuals back to regular social activities and to promptly identify asymptomatic infected individuals. However, the **testing capacity**, especially for new emerging epidemics, is generally insufficient to meet global health needs. Testing strategies are needed to quickly identify infected individuals within a community to control disease spread. Since a (physical) contact network provides information about potential disease transmission, we propose a group testing strategy based on graph partitioning to address these challenges. Our testing strategy will automatically partition the contact network into a certain number of near-balanced subgraphs and determine the testing priority based on the known testing results and estimated prevalence. A testing order will be assigned to each subgraph according to its risk level and the available testing capacity. Besides, an improved infectious disease transmission model will be introduced to simulate the virus's dynamic spread.

Strategy Overview

We describe the overall adaptive testing strategy as follows:

Figure 1. The overview of the graph partition based adaptive testing strategy, given as a processing cycle over time

We first partition the social contact network into subgraphs to form testing groups and apply Dorfman's group testing strategy without exceeding the upper limit of the test. Then, after round t, we estimate the current prevalence and determine the number of group tests for round $t+1$. Meanwhile, we raise the testing priority of the neighboring group of known infected individuals. Next, perform group testing again and keep iterating through each step of the strategy until estimated prevalence reaches zero.

Contact Network Generation Models

We generated random graphs as the illustration of the social contact network using the following generation models in a wide range of parameter settings, and used the obtained networks as inputs for our strategy.

- Barabási–Albert (BA) model
- Chung-Lu model
- Erdős–Rényi (ER) model
- Small World Network
- Stochastic Block Model (SBM)
- Waxman's Model

Graph Partition

Given an undirected and unweighted social contact network $G = (V, E)$, where individuals are considered as the vertices V, and the edges E denote the social contact between individuals, we partition G into m subgraphs $G_1, G_2, ..., G_m$. The population size is $|V| = N$. k is the maximum number of individuals allowed to pool together for testing using the given testing method. Thus, we suggested k as the upper size bound of each subgraph and M_G as the partition score function in order to partition it into m almost equal size subgraphs without exceeding size k.

$$M_G(G_i, G_j, k) = \begin{cases} \sum_{u \in G_i, v \in G_j} | E_{u,v} |, & | G_i \cup G_j | \leq k \\ -1, & \text{otherwise} \end{cases}$$

Individuals within each subgraph $G_i = (V_i, E_i)$ share correlated infection probabilities induced by the social contact network.

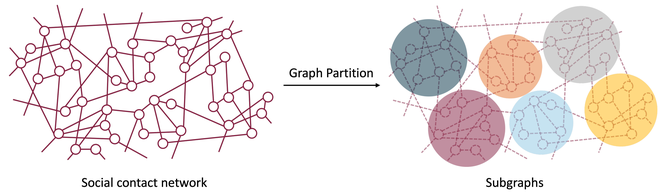

Figure 2. The graph partition algorithm that maximise the intra-subgraph edges without exceeding the upper group testing size bound.

Extended SEIR Model

At any discrete time t between 0 and T, each individual $V_i \in V$ can be in one of the following states: *S* (Susceptible), *E* (Exposed), *U* (Superspreaders), *I* (Infectious), *R* (Removed).

New infections only occur during social contact between infected and susceptible individuals if they have an edge in G. An individual i moves from *S* to *E* with probability β_j if one of i's neighbors $j \in I$. The infected individual's infection probability is a random number drawn from an exponential distribution, shown in Fig. 3B.

After a set incubation period, the *Exposed* individual becomes infectious, and will move to *Infected* state. Each infected individual draws a random recovery period from a normal distribution (mean μ_r, std σ_r). After the recovery period, the infected agent moves to the *Remove* state. Individuals that already moved into *R* will not be reinfected.

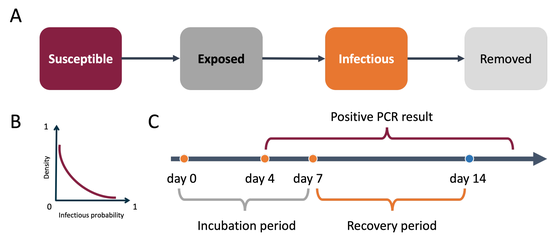

Figure 3. A. The extended SEIR model. B. The exponential distribution used for modeling the infectious heterogeneity. C. Distinct periods of the extended SEIR model.

Group Testing Strategy

For the testing scenario, we consider individuals in each subgraph G_i as a testing group C_i. The group test X pools the samples from each individuals of a testing group and outputs a testing result. If any one of the testing group is infected, the testing result is positive and the $X = 1$. Otherwise, $X = 0$. Specifically, $X_{i,t}$ represents the group test result on testing group C_i at time t.

As we adopt the two-stage testing design from Dorfman's method, once $X_i = 1$, a follow-up second round of testing will be performed to test every individuals in C_i. And we will update X_i with the number of infected individuals detected in the second round.

Due to limitation of the testing capacity, we set the maximum number of tests per day to B. Let $Q_{i,t}$ denotes the number of tests performed on testing group C_i at time t, where

$$Q_{i,t} = \begin{cases} 1, & X_{i,t} = 0 \\ |C_i| + 1, & X_{i,t} \geqslant 1 \end{cases}$$

Thus, at any discrete time t, $\sum_{i=0}^{m} Q_{i,t} \leqslant B$. To ensure that the total number of tests does not exceed the testing capacity at any time, we propose an adaptive group testing strategy. Let

$$p_t = \frac{\eta \sum_{i=0}^{m} X_{i,t-1}}{\sum_{i=0}^{m} |C_i|},$$

where η is a weight factor. We will estimate the current prevalence p_t based on the previous day's test results.

Next, we adjust the number of group tests $z_{t+1} = \frac{\mu I}{k p_t}$ in the following day, where μ is a weight factor to scale the number of tests. We assume that the test capacity is insufficient to test all groups at once. Hence, we will assign a testing order to each testing group based on the previous day's test results. A higher priority will be given to groups with positive test results from neighbors. We will only test the top z_{t+1} group.

Results

We simulate the proposed strategy and baselines using the extended SEIR model on six random graph generation models. We compare testing strategies in terms of total number of tests, the maximum outbreak size, the maximum number of secondary transmission, and the number of uninfected individuals. We also analyze the running time of the methods.

Here we take Barabási–Albert model as an example.

Testing strategy	Total number of tests	Maximum outbreak size
Individual testing	9020	23.58
Random group testing	1628	13.83
Graph-partition based group testing	1668	21.32
Our group testing strategy	1430	10.10

Table 1. Simulation results on the Barabási–Albert model with N = 500, group size is 10, and testing capacity is 70.

We found that our method significantly reduced the number of tests needed for screening the entire population. The peak of the outbreak was also lowered and postponed, with no significant spike throughout the simulation. Furthermore, 68.54% of people were protected from infection using to our approach, compared to 36% with individual testing strategy. We observed similar performance on different network generation models. Since the network structure is heterogeneous, our method can save about 60-85% of tests comparing to the individual testing method.

References

[1] R. Dorfman. The detection of defective members of large populations. *Annals of Mathematical Statistics*, 14:436–440, 1943.

mailto: jingyi19@vt.edu, heath@vt.edu