Graph and Cluster Formation Based Group Testing

Batuhan Arasli Sennur Ulukus
Department of Electrical and Computer Engineering
University of Maryland, College Park, MD 20742
barasli@umd.edu *ulukus@umd.edu*

***Abstract*—We propose a novel infection spread model based on a random connection graph which represents connections between n individuals. Infection spreads via connections between individuals and this results in a probabilistic cluster formation structure as well as a non-i.i.d. (correlated) infection status for individuals. We propose a class of *two-step sampled group testing algorithms* where we exploit the known probabilistic infection spread model. We investigate the metrics associated with two-step sampled group testing algorithms. To demonstrate our results, for analytically tractable *exponentially split cluster formation trees*, we calculate the required number of tests and the expected number of false classifications in terms of the system parameters, and identify the trade-off between them. For such exponentially split cluster formation trees, for zero-error construction, we prove that the required number of tests is $O(\log_2 n)$. Thus, for such cluster formation trees, our algorithm outperforms any zero-error non-adaptive group test, binary splitting algorithm, and Hwang's generalized binary splitting algorithm. Our results imply that, by exploiting probabilistic information on the connections of individuals, group testing can be used to reduce the number of required tests significantly even when infection rate is high, contrasting the prevalent belief that group testing is useful only when infection rate is low.**

I. Introduction

The group testing problem, introduced by Dorfman in [1], is the problem of identifying the infection status of a set of individuals by performing fewer tests than individually testing everyone. The key idea of group testing is to mix test samples of the individuals and test the mixed sample. A negative test result implies that everyone within that group is negative, thereby identifying infection status of an entire group with a single test. A positive test result implies that there is at least one positive in that group, in which case, Dorfman's algorithm goes into a second phase of testing everyone individually.

Since Dorfman's seminal work, various families of algorithms have been studied, such as, adaptive algorithms, where one designs test pools in the $(i+1)$st step by using information from the test results in the first i steps, and non-adaptive algorithms, where every test pool is predetermined and run in parallel. In addition, various forms of infection spread models have been considered, such as, the independent and identically distributed (i.i.d.) model where each person is infected independent of others with probability p, and the combinatorial model where k out of n people are infected uniformly distributed on the sample space of $\binom{n}{k}$ elements. Under these various system models and family of algorithms, group testing problem has been widely studied [2]–[26]. The advantage of group testing is known to diminish when the disease is not rare [27]–[29].

Early works mainly consider combinatorial and i.i.d. probabilistic models. Although there is no general result for arbitrary infection probabilities and arbitrary correlations, [30]–[34] have considered advanced probabilistic models with some structured introduction of correlations or non-identical infection status of individuals. In this paper, we consider a realistic graph-based infection spread model, and use the knowledge of the infection spread model to design group testing algorithms.

First, we propose a novel infection spread model, where individuals are connected via a random connection graph, whose connection probabilities are known, for instance, through location data obtained from individuals' cell phones. The infection starts with a patient zero who is uniformly randomly chosen among n individuals. Then, any individual who is connected to at least one infected individual is also infected. For this system model, we propose a novel family of algorithms which we coin *two-step sampled group testing algorithms*. The algorithm consists of a sampling step, where a set of individuals are chosen to be tested, and a zero-error non-adaptive test step, where selected individuals are tested according to a zero-error non-adaptive group test matrix. In order to select individuals to test in the first step, one of the possible cluster formations that can be formed in the random connection graph, is selected. Then, according to the selected cluster formation, we select exactly one individual from every cluster. After identifying the infection status of the selected individuals with zero-error, we assign the same infection status to the other individuals in the same cluster with identified individuals. The actual cluster formation is not known prior to the test design, and because of that, selected cluster formation can be different from the actual cluster formation. Thus, this process is not necessarily a zero-error group testing procedure.

Our main contributions consist of proposing a novel infection spread model with random connection graph, proposing a two-step sampled group testing algorithm which is based on novel $\mathcal{F}$-separable zero-error non-adaptive test matrices, characterizing the optimal design of two-step sampled group testing algorithms, and presenting explicit results on analytically tractable *exponentially split cluster formation trees*. We identify the optimal sampling function selection, calculate the required number of tests and the expected number of false classifications, and identify the trade-off between them. Our $\mathcal{F}$-separable zero-error non-adaptive test matrix construction is based on taking advantage of the known probability distribution of cluster formations. We consider exponentially split cluster formation trees as a special case, in which we explicitly

calculate the required number of tests and the expected number of false classifications. For zero-error construction, we prove that the required number of tests is less than $4(\log_2 n + 1)/3$ and is of $O(\log_2 n)$, when there are $n\delta$ individuals. We show that, even when we ignore the gain by cluster size δ, our non-adaptive algorithm, in the zero-error setting, outperforms any zero-error non-adaptive group test and Hwang's generalized binary splitting algorithm [35]. Since the number of infections scale as $\frac{n}{\log_2 n}\delta$ in exponentially split cluster formation trees with $n\delta$ individuals, our results show that, we can use group testing to reduce the required number of tests significantly in our system model even when the infection rate is high by using our two-step sampled group testing algorithm.

II. System Model

We consider a group of n individuals. The *random infection vector* $U = (U_1, U_2, \ldots, U_n)$ represents the infection status of the individuals. Here U_i is a Bernoulli random variable with parameter p_i. If individual i is infected then $U_i = 1$, otherwise $U_i = 0$. Note U_i's need not be independent. A *patient zero random variable* Z is uniformly distributed over the set of individuals. Patient zero is the first person to be infected. So far, the infection model is identical to the traditional combinatorial model with $k = 1$ infected among n individuals.

Random connection graph $\mathscr{C}$ is a random graph where vertices represent the individuals and edges represent the connections between the individuals. $p_{\mathscr{C}}$ denotes the probability distribution of the random graph $\mathscr{C}$ over the support set of all possible edge realizations. $p_{\mathscr{C}}$ is a known probability distribution, prior to the design of the group testing algorithm. Random connection graph $\mathscr{C}$ is an undirected random graph with vertex set $V_{\mathscr{C}} = [n]$, with each vertex representing a unique individual, and a random edge set $E_{\mathscr{C}} = \{e_{ij}\}$ which represents connections between the individuals.

In our model, if there is a path in $\mathscr{C}$ between two individuals, then their infection status are equal, i.e., the infection spreads from patient zero Z to everyone that is connected to patient zero. A realization of the random graph $\mathscr{C}$ consists of clusters of individuals, where a cluster is a subset of vertices in $\mathscr{C}$ such that all elements in a cluster are connected with each other and none of them is connected to any vertex that is not in the cluster. The set of all clusters in a realization of the random graph $\mathscr{C}$ is a partition of $[n]$. In a random connection graph, formation of clusters in $\mathscr{C}$ along with patient zero Z determine the status of the infection vector. Thus, instead of the specific structure of the graph $\mathscr{C}$, we focus on the cluster formations in $\mathscr{C}$. Working on the cluster formation structures, rather than the random connection graph itself, is equally informative for the sake of designing group tests, and cluster formations are sufficient statistics. For a given $p_{\mathscr{C}}$, we can calculate the probabilities of possible cluster formations in $\mathscr{C}$.

The *random cluster formation* variable F is distributed over $\mathcal{F}$ as $\mathbb{P}(F = F_i) = p_F(F_i)$, for all $F_i \in \mathcal{F}$, where $\mathcal{F}$ is a subset of the set of all partitions of the set $\{1, 2, \ldots, n\}$. In our model, we know the set $\mathcal{F}$ (i.e., the set of cluster formations that can occur) and the probability distribution p_F, since we know $p_{\mathscr{C}}$. Let us denote $|\mathcal{F}|$ by f and, let $|F_i| = \sigma_i$. Without loss of generality, for $i < j$, we assume $\sigma_i \leq \sigma_j$. Let S_j^i be the jth subset of the partition F_i where $i \in [f]$ and $j \in [\sigma_i]$. For tractability, we investigate a specific class of $\mathcal{F}$ which satisfies the following: For all i, F_i can only be obtained by partitioning some elements of F_{i-1}. This assumption results in a tree like structure for cluster formations and we call $\mathcal{F}$ sets that satisfy this condition *cluster formation trees*. Cluster formation trees may arise in real-life clustering scenarios, e.g., if individuals belong to a hierarchical structure. For instance, an individual may belong to a professor's lab, then to a department, then to a building, then to a campus.

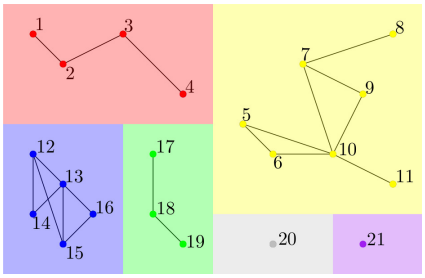


Fig. 1: Realization of a random connection graph $\mathscr{C}$ and its cluster formation. We show each cluster with a different color.

We introduce *two-step sampled group testing algorithms* in this paper. Two-step sampled group testing algorithms consist of two steps in both testing phase and decoding phase. To design a two-step sampled group testing algorithm, we first pick one of the cluster formations in $\mathcal{F}$ to be the *sampling cluster formation*. The selection of F_m is a design choice. We define the *sampling function*, M, to be a function of F_m. The sampling function selects which individuals to be tested by selecting exactly one individual from every subset that forms the partition F_m. Let the infected set among the sampled individuals be denoted by K_M. In the second step, a zero-error non-adaptive group test is performed on the sampled individuals and the infection status of the selected $\sigma_m = |F_m|$ individuals is identified with zero-error.

The *test matrix* $\boldsymbol{X}$ is a non-adaptive test matrix of size $T \times \sigma_m$, where T is the required number of tests. Let $U^{(M)}$ denote the infection status vector of the sampled individuals. Then, we have the following test result vector y,

$$y_i = \bigvee_{j \in [\sigma_m]} X_{ij} U_j^{(M)}, \quad i \in [T] \tag{1}$$

By using the test result vector y, in the first decoding step, infection status of the sampled individuals are identified with zero-error. In the second stage of decoding, depending on F_m and the infection status of the sampled individuals, other non-tested individuals are estimated by assigning the same infection status to all of the individuals that share the same cluster in the cluster formation F_m.

Finally, we have two metrics to measure the performance of a group testing algorithm: The required number of tests T, which is the number of rows of $\boldsymbol{X}$, and the expected number of false classifications which we denote by $E_f = \mathbb{E}[d_H(U \oplus \hat{U})]$,

where $d_H(\cdot)$ is the Hamming weight of a binary vector and $\hat{U}$ is the estimated infection status vector. False classifications occur only in the second step of our algorithm.

III. Motivating Example

In this example, there are $n = 10$ individuals, and a cluster formation tree with $f = 3$ levels:

$$F = \begin{cases} F_1 = \{\{1,2,3\},\{4,5\},\{6,7,8,9,10\}\}, & \textit{w.p. } 0.4 \\ F_2 = \{\{1,2\},\{3\},\{4,5\},\{6,7,8,9,10\}\}, & \textit{w.p. } 0.2 \\ F_3 = \{\{1,2\},\{3\},\{4,5\},\{6,7\},\{8,9,10\}\}, & \textit{w.p. } 0.4 \end{cases}$$

First, we find the optimal sampling functions, M, for all selections of F_m. M selects exactly one individual from each subset that forms F_m by definition and a false classification occurs only when one of the sampled individuals has a different infection status than one of the individuals in its cluster in F_m. This scenario occurs only when F_m is at a higher level than the realized F in the cluster formation tree $\mathcal{F}$, where we refer to F_1 as the top level of the cluster formation tree and F_f as the bottom level.

To find the optimal M, one must consider the possible false classifications and minimize E_f. For $F_m = F_1$, we have: If M samples individual 1 or 2 from the cluster $S_1^1 = \{1,2,3\}$, a false classification occurs if $F = F_2$ and the cluster $\{1,2\}$ is infected, in that case, individual 3 is falsely classified as infected. Similar false classification occurs when $F = F_3$ and the cluster $\{1,2\}$ is infected. Similarly, in these cases, if individual 3 is infected, again, individual 3 is falsely classified as non-infected. Thus, for cluster $\{1,2,3\}$, when either individual 1 or 2 is sampled, the expected number of false classifications is:

$$(p_F(F_2) + p_F(F_3))(p_Z(1) + p_Z(2) + p_Z(3)) = 0.18 \quad (2)$$

where p_Z is the pmf of patient zero Z. Similarly, when individual 3 is sampled from the cluster $\{1,2,3\}$, individuals 1 and 2 are falsely classified when $F = F_2$ or $F = F_3$ and either the cluster $\{1,2\}$ or individual 3 is infected. Thus, in that case, the expected number of false classifications is:

$$2(p_F(F_2) + p_F(F_3))(p_Z(1) + p_Z(2) + p_Z(3)) = 0.36 \quad (3)$$

Thus, (2) and (3) imply that, for cluster $S_1^1 = \{1,2,3\}$, the optimal M should select either individual 1 or 2 for testing. For cluster $S_2^1 = \{4,5\}$, the selection of sampled individual is indifferent and results in 0 expected false classification. For $S_3^1 = \{6,7,8,9,10\}$, similar calculations yield that the optimal M should select one of $\{8,9,10\}$. Similar analysis follows for other two choices of F_m, $F_m = F_2$ and $F_m = F_3$; see [36].

In the second step, for the sampled individuals, we design zero-error non-adaptive test matrices. We give the test matrix design for $m = 3$; for the other two choices of F_m, see [36].

For $m = 3$, when smallest indexed individuals are selected when a choice set is indifferent, we have $M = \{1,3,4,6,8\}$, and the set of all possible infected sets is $\mathcal{P}(K_M) = \{\{1\},\{3\},\{1,3\},\{4\},\{6\},\{8\},\{6,8\}\}$. Each element of $\mathcal{P}(K_M)$ needs to be assigned to a distinct result vector,

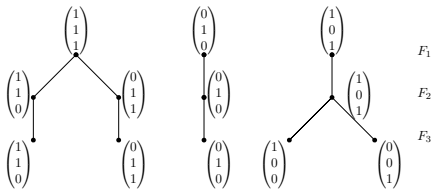


Fig. 2: $\mathcal{F}$ with assigned result vectors for each node.

thus, there need to be 7 distinct result vectors. Moreover, result vectors that are assigned to the sets that are union of some other sets need to be OR of the result vectors that are assigned to the sets that are the subsets of them. We give a tree structure representation with assigned result vectors of length 3, which is shown in Fig. 2 where each unique node is assigned a unique vector except for the nodes that do not get partitioned. Note that every unique node in the tree corresponds to a unique element of $\mathcal{P}(K_M)$. Corresponding test matrix $\boldsymbol{X}$ is,

	1	3	4	6	8
Test 1	1	0	0	1	0
Test 2	1	1	1	0	0
Test 3	0	1	0	0	1

Resulting expected numbers of false classifications for $F_m = (F_1, F_2, F_3)$ are $0.58, 0.4$ and 0, respectively. F_m is a design choice and one can use time sharing between different choices of m, depending on the system specifications and achieve intermediate values of expected false classifications and required number of tests. Generally, there is a trade-off between the number of tests and the number of false classifications, and we can formulate optimization problems for system requirements, such as finding a time sharing distribution for F_m that minimizes the number of tests for a desired level of false classifications, or vice versa.

IV. Proposed Algorithm and Analysis

In our *$\mathcal{F}$-separable matrix* construction, we aim to construct binary matrices that have n columns, and for each possible infected subset of the selected individuals, there must be a corresponding distinct result vector. Let $\boldsymbol{X}_S$ be the sub-matrix of $\boldsymbol{X}$ that is obtained by columns of $\boldsymbol{X}$ with indices in S, where S is a subset of $[n]$ and $\boldsymbol{X}$ has n columns. A binary matrix $\boldsymbol{X}$ is $\mathcal{F}$-separable if

$$\bigvee_{i \in [S_1]} \boldsymbol{X}_{S_1}^{(i)} \neq \bigvee_{i \in [S_2]} \boldsymbol{X}_{S_2}^{(i)} \quad (4)$$

is satisfied for all distinct subsets S_1 and S_2 in $\mathcal{P}(K_M)$, where $\boldsymbol{X}_S^{(i)}$ denotes the ith column of $\boldsymbol{X}_S$ and $\mathcal{P}(K_M)$ denotes the set of possible infected subsets of the selected individuals. The decoding process is a mapping from the result vectors to the infected sets and thus, we require the distinct result vector property to guarantee zero-error decoding.

Theorem 1 *In a two-step sampled group testing algorithm with the given sampling cluster formation F_m and the sampling function M over a cluster formation tree structure defined by $\mathcal{F}$ and p_F, with uniform patient zero distribution*

p_Z over $[n]$, the expected number of false classifications given $F = F_\alpha$, denoted by $E_{f,\alpha}$, is

$$\sum_{i\in[\sigma_m]} \left(\frac{|S^\alpha(M_i)|}{n} \cdot |S_i^m \backslash S^\alpha(M_i)| + \sum_{S_j^\alpha \subseteq S_i^m \backslash S^\alpha(M_i)} \frac{|S_j^\alpha|^2}{n} \right)$$

and the expected number of false classifications is

$$E_f = \sum_{\alpha > m} p_F(F_\alpha) E_{f,\alpha}$$

where $S^\alpha(M_i)$ is the subset in the partition F_α which contains the ith selected individual.

To characterize the optimal choice of the sampling function M, first, for $i \in [f]$ and $k \in [n]$, we define $\beta_i(k)$ functions as

$$\sum_{j>i} p_F(F_j) \left(|S^j(k)| \cdot |S_i^m \backslash S^j(k)| + \sum_{S_l^j \subseteq S_i^m \backslash S^j(k)} |S_l^j|^2 \right)$$

where $S^j(k)$ is the subset in partition F_j that contains k.

Theorem 2 *For sampling cluster formation F_m, the optimal choice of M that minimizes the expected number of false classifications is*

$$M_i = \underset{k \in S_i^m}{\arg\min}\ \beta_m(k) \tag{5}$$

where M_i is the ith selected individual. The number of tests is constant and independent of the choice of M.

We call matrices that satisfy (4) as *$\mathcal{F}$-separable matrices* and algorithms that use them as test matrices as *$\mathcal{F}$-separable non-adaptive group tests*. The key idea of designing an $\mathcal{F}$-separable matrix is to determine the set $\mathcal{P}(K_M)$ for a given M and $\mathcal{F}$ so that we can find binary column vectors for each selected individual where all of the corresponding possible result vectors are distinct. For a cluster formation tree with assigned result vectors to each node, a sufficient condition for achievability of $\mathcal{F}$-separable matrices is as follows:

> Let u be a node with Hamming weight $d_H(u)$. The number of descendants of u with constant Hamming weights i must be less than $\binom{d_H(u)}{i}$ for all i which must hold for all u. Number of nodes with constant Hamming weight i must be less than $\binom{T}{i}$ for all i. Hamming weights of the nodes must decrease from ancestors to descendants.

We define $\lambda_{S_i^j}$ as the number of unique ancestor nodes of the set S_i^j and λ_j as the number of unique sets S_a^b in $\mathcal{F}$ at and above the level F_j. Then, we have the following lower bound.

Theorem 3 *For given $\mathcal{F}$ and F_m for $m < f$, the number of required tests for an $\mathcal{F}$-separable non-adaptive group test, i.e., the number of rows of the test matrix $\boldsymbol{X}$, must satisfy*

$$T \geq \max \left\{ \max_{j\in[\sigma_m]} (\lambda_{S_j^m} + 1),\ \lceil \log_2(\lambda_m + 1) \rceil \right\} \tag{6}$$

with addition of 1's removed in (6) when $m = f$.

Theorem 3 is a converse argument, and in fact, the given lower bound is not always achievable. In the next section, we introduce a family of cluster formation trees which we call *exponentially split cluster formation trees*. For this analytically tractable family of cluster formation trees, we achieve the lower bound in Theorem 3 *order-wise*.

V. Exponentially Split Cluster Formation Trees

A cluster formation tree $\mathcal{F}$ is an *f level exponentially split cluster formation tree* if it has 2^{i-1} nodes at level F_i for each $i \in [f]$, and at level F_i, every node has $2^{f-i}\delta$ individuals where δ is a constant, and each node has exactly 2 descendants in one level down, and F is uniformly distributed on $\mathcal{F}$.

Due to the symmetry, for any choice F_m, each element of S_i^m has the same $\beta_m(i)$ value for all $i \in \sigma_m$. Thus, M selects individuals from each set arbitrarily and we can pick any sampling function that selects one element from each S_i^m. By Theorem 1, the expected number of false classifications is

$$E_f = \frac{\delta}{3f} \left(2^{f-m+2} + 2^{m-f+1} - 6 \right) \tag{7}$$

which takes its maximum value when $F_m = F_1$, and minimum value when $F_m = F_f$ as $E_f = 0$. Since the choice of F_m is a design parameter, one can use time sharing between the possible selections of F_m to achieve intermediate values.

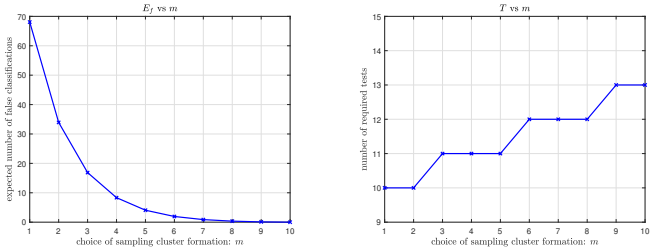
Theorem 4 *For an f level exponentially split cluster formation tree, at level f, there exists an $\mathcal{F}$-separable test matrix, $\boldsymbol{X}$, with not more than $\frac{4}{3}f$ rows, i.e., an upper (achievable) bound for the number of tests is $\frac{4}{3}(\log_2 n + 1)$ for n individuals. Conversely, this is also the capacity order-wise, since the number of tests must be greater than f.*

In an exponentially split cluster formation tree structure with f levels, expected total number of infections is,

$$\sum_{i=1}^{f} \frac{1}{f} 2^{f-i}\delta = \frac{\delta}{f}(2^f - 1) = O\left(\frac{n}{\log_2 n} \right) \tag{8}$$

To compare our results, for fairness, we focus on the zero-error case in our system model, i.e., $F_m = F_f$. Resulting M selects 2^{f-1} individuals and the resulting number of tests is between f and $\frac{4}{3}f$, i.e., $O(\log_2 n)$. This results in a gain scaled with δ. However, to fairly compare our results, we ignore this gain and compare the performance of the second step of our algorithm. To avoid confusion, let $\delta = 1$, i.e., $n = 2^{f-1}$.

From (8), the expected number of infections is $\frac{2^f - 1}{f} = O(\frac{n}{\log_2 n})$. When the infections scale faster than $\sqrt{n}$, non-adaptive tests with zero-error criterion cannot perform better than individual testing [27]. Since our algorithm results in $O(\log_2 n)$ tests, it outperforms all non-adaptive algorithms in the literature. We also compare our results with Hwang's generalized binary splitting algorithm [35], even though it is an adaptive algorithm and it assumes the prior knowledge of exact number of infections. Hwang's algorithm results in $k \log_2(n/k) + O(k)$ tests and attains the capacity of adaptive group testing [29], [37]. The resulting mean value of the



(a) (b)

Fig. 3: (a) Expected number of false classifications vs the choice of F_m. (b) Number of tests vs the choice of F_m.

required number of tests with Hwang's algorithm is $O\left(\frac{n}{\log_2 n}\right)$ which scales much faster than our result of $O(\log_2 n)$.

VI. Numerical Results

A. Exponentially Split Cluster Formation Tree Based System

In the first simulation, we have an exponentially split cluster formation tree with $f = 10$ levels and $\delta = 1$. For this system of $n = 2^{f-1}\delta = 512$ individuals, for each sampling cluster formation choice F_m, from $m = 1$ to $m = 10$, we calculate the expected number of false classifications and the minimum required number of tests. In Fig. 3(a), we plot the expected number of false classifications which meets the analytical expressions we found in Section V. While calculating the minimum number of required tests, for each choice of F_m, our program finds the minimum T that satisfies the sufficient achievability condition that we presented in Section IV. We plot the minimum required number of tests in Fig. 3(b).

We compare our zero-error construction results with a variation of Hwang's binary splitting algorithm [35], [37], presented in [38]. The required number of tests in our algorithm scales with $O(\log_2 n)$, resulting in 13 tests at level $m = f = 10$, as seen in Fig. 3(b), while the average number of required tests for Hwang's algorithm scales as $O\left(\frac{n}{\log_2 n}\right)$, and is approximately 172 in this case. Further, when we remove the assumption of known number of infections, we have to use the binary splitting algorithm presented originally in [39], which results in a number of tests that is not lower than individual testing, i.e., $n = 512$ tests.

B. Arbitrary Random Connection Graph Based System

In our second simulation, we present an arbitrary $\mathscr{C}$ with 20 individuals, shown in Fig. 4(c), where the edges realize independently with probabilities shown on them (zero probability edges are not shown). There are $2^9 = 512$ possible cluster formations which do not yield a cluster formation tree, yet we still apply our ideas designed for cluster formation trees. For each one of the 512 possible selections of m, we plot the expected number of false classifications in Fig. 4(a) and the required number of tests in Fig. 4(b). Since the system is arbitrary unlike the first simulation, the resulting expected number of false classifications is not monotonically decreasing when we sort the resulting required number of tests in the increasing order for the choices of F_m. In Fig. 4(a), we

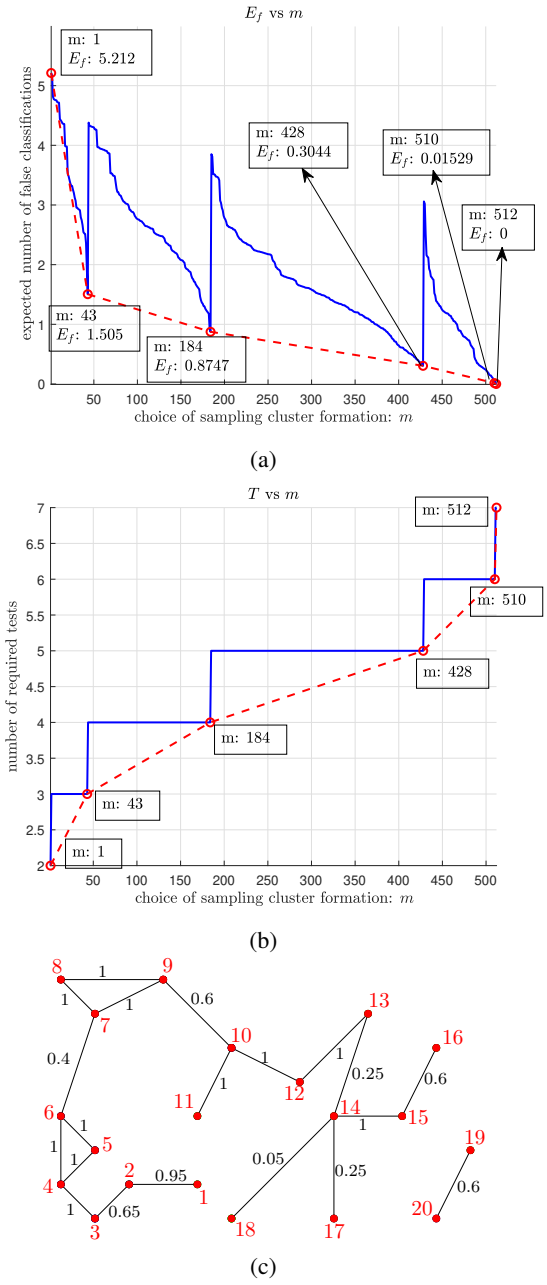


Fig. 4: (a) Expected number of false classifications vs the choice of F_m. (b) Number of tests vs the choice of F_m. (c) Random connection graph.

mark the choices of m that result in the minimum number of expected false classifications within each required number of test range. By using time sharing between these choices, dotted red lines between them can be achieved.

We calculate the average number of required tests for Hwang's generalized binary splitting algorithm by using the results of [35], [37], [38] as in the first simulation and find that the average number of required tests is 16.4. Similar to the first simulation, binary splitting algorithm presented originally in [39] which does not require the exact number of infections, cannot perform better than individual testing.

REFERENCES

[1] R. Dorfman. The detection of defective members of large populations. *Annals of Mathematical Statistics*, 14(4):436–440, December 1943.

[2] D. Z. Zhu and F. K. Hwang. *Combinatorial Group Testing and Its Applications*. 2nd edition, 1999.

[3] J. Wolf. Born again group testing: Multiaccess communications. *IEEE Trans. on Info. Theory*, 31(2):185–191, 1985.

[4] G. K. Atia and V. Saligrama. Boolean compressed sensing and noisy group testing. *IEEE Trans. on Info. Theory*, 58(3):1880–1901, 2012.

[5] T. Wadayama. Nonadaptive group testing based on sparse pooling graphs. *IEEE Trans. on Info. Theory*, 63(3):1525–1534, 2017.

[6] C. Wang, Q. Zhao, and C. Chuah. Optimal nested test plan for combinatorial quantitative group testing. *IEEE Trans. on Signal Processing*, 66(4):992–1006, 2018.

[7] S. Wu, S. Wei, Y. Wang, R. Vaidyanathan, and J. Yuan. Partition information and its transmission over boolean multi-access channels. *IEEE Trans. on Info. Theory*, 61(2):1010–1027, 2015.

[8] C. Shangguan and G. Ge. New bounds on the number of tests for disjunct matrices. *IEEE Trans. on Info. Theory*, 62(12):7518–7521, 2016.

[9] J. Scarlett and O. Johnson. Noisy non-adaptive group testing: A (near-)definite defectives approach. *IEEE Trans. on Info. Theory*, 66(6):3775–3797, 2020.

[10] J. Scarlett and V. Cevher. Near-optimal noisy group testing via separate decoding of items. In *IEEE ISIT*, pages 2311–2315, 2018.

[11] J. Scarlett. Noisy adaptive group testing: Bounds and algorithms. *IEEE Trans. on Info. Theory*, 65(6):3646–3661, 2019.

[12] A. Mazumdar. Nonadaptive group testing with random set of defectives. *IEEE Trans. on Info. Theory*, 62(12):7522–7531, 2016.

[13] T. Kealy, O. Johnson, and R. Piechocki. The capacity of non-identical adaptive group testing. In *Allerton Conference*, pages 101–108, 2014.

[14] O. Johnson, M. Aldridge, and J. Scarlett. Performance of group testing algorithms with near-constant tests per item. *IEEE Trans. on Info. Theory*, 65(2):707–723, 2019.

[15] H. A. Inan, P. Kairouz, M. Wootters, and A. Ozgur. On the optimality of the kautz-singleton construction in probabilistic group testing. In *Allerton Conference*, pages 188–195, 2018.

[16] E. Karimi, F. Kazemi, A. Heidarzadeh, K. R. Narayanan, and A. Sprintson. Non-adaptive quantitative group testing using irregular sparse graph codes. In *Allerton Conference*, pages 608–614, 2019.

[17] O. Gebhard, M. Hahn-Klimroth, D. Kaaser, and P. Loick. Quantitative group testing in the sublinear regime. Available at arXiv: 1905.01458.

[18] M. Falahatgar, A. Jafarpour, A. Orlitsky, V. Pichapati, and A. T. Suresh. Estimating the number of defectives with group testing. In *IEEE ISIT*, pages 1376–1380, 2016.

[19] A. Coja-Oghlan, O. Gebhard, M. Hahn-Klimroth, and P. Loick. Information-theoretic and algorithmic thresholds for group testing. *IEEE Trans. on Info. Theory*, pages 1–1, 2020.

[20] C. L. Chan, S. Jaggi, V. Saligrama, and S. Agnihotri. Non-adaptive group testing: Explicit bounds and novel algorithms. *IEEE Trans. on Info. Theory*, 60(5):3019–3035, 2014.

[21] S. Cai, M. Jahangoshahi, M. Bakshi, and S. Jaggi. Efficient algorithms for noisy group testing. *IEEE Trans. on Info. Theory*, 63(4):2113–2136, 2017.

[22] S. Bondorf, B. Chen, J. Scarlett, H. Yu, and Y. Zhao. Sublinear-time non-adaptive group testing with $o(k \log n)$ tests via bit-mixing coding. Available at arXiv: 1904.10102.

[23] M. Aldridge. Individual testing is optimal for nonadaptive group testing in the linear regime. *IEEE Trans. on Info. Theory*, 65(4):2058–2061, 2019.

[24] A. Agarwal, S. Jaggi, and A. Mazumdar. Novel impossibility results for group-testing. In *IEEE ISIT*, pages 2579–2583, 2018.

[25] A. Heidarzadeh and K. Narayanan. Two-stage adaptive pooling with rt-qpcr for covid-19 screening. Available at arXiv: 2007.02695.

[26] H. B. Chen and F. K. Hwang. Exploring the missing link among d-separable, $\bar{\text{d}}$-separable and d-disjunct matrices. *Discrete Applied Mathematics*, 155(5):662 – 664, 2007.

[27] M. Ruszinko. On the upper bound of the size of the r-cover-free families. *Journal of Combinatorial Theory, Series A*, 66(2):302 – 310, 1994.

[28] L. Riccio and C. J. Colbourn. Sharper bounds in adaptive group testing. *Taiwanese Journal of Mathematics*, 4(4):669–673, 2000.

[29] M. Aldridge, O. Johnson, and J. Scarlett. *Group Testing: An Information Theory Perspective*. 2019.

[30] T. Li, C. L. Chan, W. Huang, T. Kaced, and S. Jaggi. Group testing with prior statistics. In *IEEE ISIT*, pages 2346–2350, 2014.

[31] S. D. Lendle, M. G. Hudgens, and B. F. Qaqish. Group testing for case identification with correlated responses. *Biometrics*, 68(2):532–540, 2012.

[32] P. Nikolopoulos, T. Guo, C. Fragouli, and S. Diggavi. Community aware group testing. Available at arXiv: 2007.08111.

[33] Y-J. Lin, C-H. Yu, T-H. Liu, C-S. Chang, and W-T. Chen. Positively correlated samples save pooled testing costs. Available at arXiv:2011.09794.

[34] S. Ahn, W-N. Chen, and A. Ozgur. Adaptive group testing on networks with community structure. Available at arXiv:2101.02405.

[35] F. K. Hwang. A method for detecting all defective members in a population by group testing. *Journal of the American Statistical Association*, 67(339):605–608, 1972.

[36] B. Arasli and S. Ulukus. Group testing with a graph infection spread model. Available at arXiv:2101.05792.

[37] L. Baldassini, O. Johnson, and M. Aldridge. The capacity of adaptive group testing. In *2013 IEEE International Symposium on Information Theory*, pages 2676–2680, 2013.

[38] A. Allemann. An efficient algorithm for combinatorial group testing. In *Information Theory, Combinatorics, and Search Theory: In Memory of Rudolf Ahlswede*, pages 569–596, 2013.

[39] M. Sobel and P. A. Groll. Group testing to eliminate efficiently all defectives in a binomial sample. *Bell System Technical Journal*, 38(5):1179–1252, 1959.