Adaptive Graph-Constrained Group Testing

Saurabh Sihag, *Member, IEEE*, Ali Tajer, *Senior Member, IEEE*, and Urbashi Mitra, *Fellow, IEEE*

***Abstract*—This paper considers the problem of adaptive group testing for isolating up to k defective items from a population of size n. There exist restrictions or preferences which determine how the items can be pooled for testing. A graphical model formalizes the pooling restrictions and preferences. Such graph-constrained group testing is investigated in three settings: populations with defectives, populations facing the potential presence of inhibitors, and populations with community structures. Adaptive group testing frameworks are provided for each setting. In populations without inhibitors, existing non adaptive frameworks can isolate the defective items perfectly with $\Theta(k \log n/k)$ number of tests, where is the β-mixing time of a random walk over the underlying graph. This paper provides a two-stage framework that can perfectly isolate up to k defective items for a regular graph using $\Theta(k2 \log nk + k)$ number of tests, thus achieving an approximate gain of a factor of k over the non-adaptive frameworks. This twostage framework's principles are extended to community-structured graphs and graphs with up to r inhibitor items. In particular, when inhibitors are present in the graph, a four-stage group testing framework is proposed. The results show that in the regime $r = O(k)$ for a fully connected graph, $\Theta((k + r) \log n/(k + r) + r \log n)$ tests are sufficient for isolating the defective items. This matches the corresponding necessary condition on tests which scales $(k + r) \log n$. The adaptive graphconstrained group testing framework is also empirically evaluated.**

***Index Terms*—Adaptive algorithms, communities, detection, graphical models, group testing.**

I. Introduction

IN GROUP testing, the central objective is to identify up to k items with a property of interest (such as being defective) in a population of size $h > k$. This is achieved by pooling different subsets of the items and testing the pooled subsets individually. Group testing approaches are effective in practice because they provide significant savings in the number of tests required to isolate the defectives compared with testing each item independently. As a result, group testing algorithms have been studied in a wide range of domains, including healthcare [1], sensor fault detection and diagnosis [2]–[4], diagnosis in digital circuits [5], [6], and more recently for rapid and scalable testing for isolating infected individuals in the COVID-19 pandemic [7]–[9]. Each test on a pool of items indicates the presence of at least one defective item in the pool [10].

Group testing frameworks can be classified based on their structures in multiple dimensions. The nomenclature pertinent to the scope of this paper revolves around *non-adaptive* and *adaptive* frameworks. When the tests are designed independently of each other, the group testing framework is called non-adaptive. In contrast, when the pools of items to be tested are determined sequentially, the framework is called adaptive. In non-adaptive group testing, when the items can be selected to be pooled without any constraints, that is, any item can be pooled arbitrarily with any subset of items, the existing results show that in the regime $k = \Theta(n^\beta)$ for $\beta < 1/3$, $\Theta(k^2 \log n)$ pooled tests are sufficient to isolate the defective items perfectly under a zero-error criterion [10] and $\Theta(k \log n)$ pooled tests are sufficient to isolate the defective items under a vanishing error criterion [11]. In contrast, adaptive group testing approaches facilitate perfect isolation of defective items by using $\Theta(k \log n/k + k)$ tests under different error criteria [12], [13] for $\beta \in (0, 1)$.

The second related taxonomy specifies whether there exists any constraint that limits pooling the items arbitrarily. While in some applications there does not exist a hard constraint on pooling and any subset of items can be randomly accessed simultaneously for pooled testing (such as spectrum hole search for cognitive radios [14]–[16] or anomaly detection [17]), in some domains there exist hard constraints or preferences that specify that each given item can be pooled with only a pre-specified subset of items. Such constraints are governed by underlying relationships among the items depending on the context. Examples of such underlying relationships include proximity, access rights, and correlation among different items. Some specific applications characterized by such constraints include network tomography [18] and testing for infection [19]. For network tomography, a routing mechanism governs the host node's information transfer to the network's receiver node. Therefore, the measurements are characterized by the set of nodes lying on the information transfer path. In infection detection, there is a preference in pooling certain items over others, such as testing of infection in a population, where community structure and the nature of interactions within the population may affect individuals' pooling.

In such settings, the relationships among the items, and subsequently, the constraints on pooling, can be represented by graphical models [20]. In [20], a non-adaptive group testing approach is proposed for a population of n items that form the vertices of an underlying graphical model wherein pooling is restricted by

Manuscript received February 8, 2021; revised August 29, 2021; accepted December 10, 2021. Date of publication December 21, 2021; date of current version January 18, 2022. The work of Saurabh Sihag and Ali Tajer was supported in part by the U. S. National Science Foundation under the CAREER Award ECCS-1554482 and the Grant DMS-1737976, and in part by the RPI-IBM Artificial. Intelligence Research Center. The work of Urbashi Mitra was supported by ARO W911NF1910269, DOE DE-SC0021417, Swedish Research Council 2018-04359, NSF CCF-2008927, and ONR 503400-78050. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yue M. Lu. *(Corresponding author: Ali Tajer.)*

Saurabh Sihag is with the University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: saurabh.sihag@pennmedicine.upenn.edu).

Ali Tajer is with the Department of Electrical, Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180 USA (e-mail: tajer@ecse.rpi.edu).

Urbashi Mitra is with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90007 USA (e-mail: ubli@usc.edu).

Digital Object Identifier 10.1109/TSP.2021.3137026

the graph structure. It is shown that $\Theta(k^2\varpi^2 \log n/k)$ tests can perfectly isolate up to k defectives under the zero-error criterion, where ϖ is the β-mixing time of the random walk over the underlying graph [20]. In this paper, we investigate adaptive, multi-stage group testing approaches under similar graphical constraints on pooling. Our results show that the gain in the number of tests over the non-adaptive approach scales with k, mirroring the gain achieved by the adaptive approaches over non-adaptive approaches under no constraints on pooling [13].

A. Related Literature

There exists an extensive literature on different information-theoretic and computational aspects of group testing. In this subsection, we provide an overview of the literature pertinent to adaptive and constrained group testing as the two key aspects of the framework investigated in this paper. We will also discuss the literature on group testing in the presence of inhibitors. The performance of a group testing framework is measured by the number of tests required to correctly isolate the defective items under different criteria [21] and [22]. In various regimes of the relative number of the defectives with respect to the population size, generally, the number of tests sufficient for recovering the defective items scales sublinearly in the population size [10].

Adaptivity: When a sequence of tests is specified independently of the tests' outcomes, the testing framework is referred to as *non-adaptive* [11], [23], and [24]. In contrast, when the set of tests already performed guide the subsequent tests' design, the framework is referred to as *adaptive* [25]–[27].

The non-adaptive group testing frameworks consist of one stage of multiple tests performed in parallel as the set of items being tested in each test does not depend on the outcomes of the other tests. On the other hand, incorporating limited adaptivity by introducing multiple stages allows recovering the defective items more efficiently and in a broader set of regimes [13], [28]–[31]. For instance, in the two-stage approaches, the first stage consists of a set of non-adaptive tests to reduce the search space for defective items by culling a large fraction of the non-defective items. The second stage depends on the outcome of the first stage and generally consists of individually testing the items retained by the first stage [13], [28], and [29]. The same principle can be used to extend the two-stage frameworks to general multi-stage ones. The tradeoffs between the number of stages and the number of tests have been studied in [30], where different regimes of k and n are characterized for two-, three-, and four-stage group testing to be sufficient to isolate up to k defective items from n total items. A computationally efficient four-stage approach that achieves the optimal number of tests when limited errors in the test outcomes can be tolerated has been studied in [31].

Pooling constraints: Group testing approaches under graphical constraints imposed by the topology have been studied in [4], [20], [32]–[34], and under the community structures have been studied in [35]. The existing studies on graph-constrained group testing are mainly focused on the non-adaptive strategies [4], [20], [32], and [33]. These studies propose different non-adaptive group testing frameworks and characterize sufficient conditions (in terms of the number of tests) for isolating defective items for different classes of graphs. Specifically, the pool of edges, i.e., the connected subgraph to be tested, was determined using deterministic rules in [4] and [32]. The connected subgraphs to be tested were determined randomly in [20] and [33]. Adaptive algorithms for community-aware group testing have been studied in [35] where the group of items to be tested can be divided into multiple communities. The aforementioned graph-constrained group testing frameworks are typically motivated by applications in network tomography, where statistical inference is performed by observing the flow of information over the edges or links in the network [36]. Therefore, to identify any failed links in the network in this context, group testing strategies that employ pooled test on a set of connected links can be leveraged. Pooling constraints are also relevant to applications such as drug discovery and blood screening in bioinformatics. In a broader context, the interactions among different biological modules, ranging from DNAs, genes, and proteins to complex chemical compounds such as drugs have been modeled as networks in this domain [37]. In such networks, a strategic design of pools for testing that accommodates various interactions (such as synergistic or antagonistic) among entities overcomes the drawbacks in random pooling, which ignores these interactions and, therefore, enables more meaningful interpretations into the outcomes of tests [38], [39].

Inhibitors: Testing biological samples for blood screening in practice faces different challenges due to dilution of the testing performance as the pool sizes increase [1] and the presence of blockers or inhibitors in the samples that may inhibit the detection of compounds relevant for the disease [40]. However, in related applications such as drug discovery, it is desirable to pool together compounds that achieve a blocking objective, for instance, a drug blocking a pathogenic protein [41]. Motivated by such scenarios, group testing approaches in the presence of inhibitors have been studied in [13], [41]–[44]. Specifically, lower bounds on the number of tests for non-adaptive group testing with inhibitors have been studied in [42] and a four-stage group testing algorithm whose performance meets the lower bounds in [42] has been proposed in [13]. A probabilistic non-adaptive group testing framework for isolating defectives in the presence of inhibitors was characterized in [43].

B. Contributions

While unconstrained group testing is well-investigated under both adaptive and non-adaptive settings, the existing studies on graph-constrained group testing are mainly limited to the non-adaptive methods [20]. In this paper, *we propose three adaptive graph-constrained group testing frameworks* . The first framework is a two-stage approach that is characterized by random walk based pooling of items for tests in the first stage. The idea of random walk based pooling in this paper is similar to the pooling strategy of [20] with the key distinction that our group testing framework hinges on a generic (and potentially non-uniform) random walk. We show, both theoretically and empirically, that conditions on the number of tests in our two-stage approach are less stringent compared to these for the existing non-adaptive approaches. For a degree regular graph, the results in [20] show that $\Theta(\varpi^2 k^2 \log \frac{n}{k})$ tests are sufficient for isolating up to k defective vertices for one-stage group testing under

a zero-error criterion. Our two-stage group testing framework reduces the number of tests to $\Theta(\varpi^2 k \log \frac{n}{k} + k)$ for isolating up to k defectives, thus showing an improvement of a factor of k over the one-stage group testing frameworks.

Random walks over a graph with loosely connected structures can have high mixing times. This renders pooling based on a random walk in such graphs infeasible. Alternative approaches for pooling, such as pooling random edges and testing the largest connected component formed by them have been adopted in [33]. However, applicability of such approaches is also limited for graphs with loosely connected structures. Motivated by this, we extend the two-stage framework and design a three-stage group testing framework for graph models with community structures. In contrast to [35], our framework in the context of community-structured models is minimally adaptive, and the performance measures are characterized by the zero-error criterion. We also consider the problem of group testing with graph constraints in the presence of inhibitors and extend the principles of the two-stage group testing framework to provide a four-stage group testing framework that requires a number of tests that matches, up to a factor of ϖ^2, to the asymptotically optimal number of tests for the corresponding scenario under no graph constraints established in [13]. Finally, we evaluate the theoretical results and our frameworks' performance in a case study motivated by the spread of infection in a population with underlying communities.

II. Preliminaries

Consider a population of n items labeled by $V \triangleq \{1, \ldots, n\}$. A subset $\mathcal{K} \subseteq V$ of these objects are defective. We define k as the maximum number of defective items.

Definition 1 (Defective vector): Define the binary variable $u_i \in \{0, 1\}$ to signify the state of item $i \in V$, where $u_i = 1$ indicates that it is defective. Accordingly, define the defective vector as $\mathbf{u} \triangleq [u_1, \ldots, u_n]$.

In multi-stage group testing, each stage's design is characterized by the number of pooled tests and the set of items pooled in each test, which we formalize next.

Definition 2: Define $T(A)$ as the number of pooled tests we perform on a subset of items $A \subseteq V$.

Definition 3 (Test matrix): For performing $T(A)$ pooled tests on the items in $A \subseteq V$, we define $\mathbf{X}(T) \in \{0, 1\}^{T(A) \times |A|}$ as a binary matrix whose $(t, i)^{\text{th}}$ element, denoted by x_{ti}, is set according to

$$x_{ti}(T) \triangleq [\mathbf{X}(T)]_{ti} = \begin{cases} 1 & \text{if item } i \in A \text{ is included in test } t \\ 0 & \text{otherwise} \end{cases}. \tag{1}$$

In our designs, we will be controlling the number of items pooled for each test, formalized by bounds on $\|\mathbf{X}(T)\|_\infty$, where $\|\mathbf{X}(T)\|_\infty$ is the ℓ_∞-norm of $\mathbf{X}(T)$, which equals the maximum ℓ_1-norm of any row of $\mathbf{X}(T)$.

Definition 4 (Pooling capacity): We say that the pooling capacity of a test is ℓ if up to ℓ items can be pooled for each individual test. This indicates for a given test matrix $\mathbf{X}(T)$ we have $\|\mathbf{X}(T)\|_\infty \leq \ell$.

We denote the outcome of the test $t \in \{1, \ldots, T(A)\}$ on the items in A by $y_t \in \{0, 1\}$. The outcome $y_t = 1$ indicates that the group of items pooled together in test t includes *at least* one defective item, and the outcome $y_t = 0$ indicates that all items are deemed non-defective. Accordingly, we define $\mathbf{y} \triangleq [y_1, \ldots y_T]$.

Definition 5 (Noiseless group testing): Fix the set of items $A \subseteq V$ and the number of pooled tests $T(A)$. For a given defective vector $\mathbf{u}$ and a given test matrix $\mathbf{X}(T)$, the outcome of a noiseless group test is specified by

$$y_t = \bigvee_{i \in A} u_i \cdot x_{ti}(T), \tag{2}$$

where $\vee$ denotes the Boolean inclusive *or*.

In this paper, we focus on noiseless group testing, and we are interested in the zero-error criterion for the exact recovery of defective items. In unconstrained (conventional) group testing, there is no restriction on the identities of items to be pooled in each test. In *graph-constrained* group testing, in contrast, there are two distinctions:

- **Pooling restrictions:** Restrictions or preferences on pooling may be imposed by a pre-specified set of conditions that characterize the environment. These restrictions specify the set of items that a given item $i \in V$ can be pooled with. We denote the set of items that item i can be pooled with by $N_i \subseteq V$. For instance, for the identification of congested links in network tomography, end-to-end measurements that consist of information flow between two different points in the network are commonly leveraged. When the points are not directly connected, the measurements must follow a feasible path between them in the network.
- **Pooling preferences:** Furthermore, while item i can be pooled with all items in V, we assume that there is a pre-specified bias profile, according to which i will be pooled with different items with possibly distinct likelihoods. For instance, in network tomography, it may be logistically easier to transmit information over certain paths in the network, thus inducing preferences in pooling.
  Another application in which pooled tests may be characterized by preferences is drug discovery in bioinformatics. Drug discovery typically involves identifying "lead compounds" which are more active than other biological compounds in biological assays [41]. Broadly, the identities of compounds pooled together may determine their activity (for instance, synergistic or antagonistic). Specifically, a pool of individually inactive compounds may give an active result when pooled together in a test. This phenomenon is called *synergism*, which is crucial to combination therapies in the pharmaceutical industry [41]. Therefore, strategic pooling determined by the topology of molecular interaction networks is a better approach for statistical analysis in bioinformatics than random pooling [38] and [39]. These aspects imply the alignment of the pooled testing approaches applied in different applications in drug discovery with the characteristics of pooling preferences for graph-constrained group testing discussed in this paper. However, not all applications may benefit from such preferences in pooling. For instance, in the problem of isolating infection spread in a population, pooling of samples collected from members known to have more frequent

inter-personal interactions (such as family members, co-workers, etc.) may be more likely over an arbitrary pooling of samples, thus inducing a bias in group testing. If the infection is more likely to be contagious than sporadic, this bias in pooling may limit the advantages offered by group testing.

We note that the pooling restrictions or preferences implicit in a graph-constrained setting motivate our choice of the zero-error criterion in this paper. Note that when the pooling of items is constrained by the graph topology, the symmetry in identifying the defective items is lost, i.e., some defective items may be harder to identify than others. This is in contrast to unconstrained group testing, in which different items are generally defective with equal probabilities. Such a symmetry can be leveraged to decrease the complexity of a non-adaptive group testing procedure to $O(k \log n/k)$ by considering a probabilistic error criterion as opposed to a zero-error criterion, which has complexity $O(k^2 \log n/k)$. However, such a probabilistic recovery relaxation in error criterion is not as straightforward in the context of graph-constrained group testing and the choice of a zero-error criterion is comparatively more robust.

We formalize pooling restrictions and preferences by an *undirected* and *connected* graph $G \triangleq (V, E)$, in which the vertices represent the items, edge connections signify the restrictions, and edge weights capture preferences. When vertices $i, j \in V$ are connected, we denote the edge connecting them by $(i, j) \in E$. Hence, given the definition of N_i, N_i specifies the set of immediate neighbors of vertex i. The weight of the edge (i, j) is denoted by w_{ij}. We assume that the edge weights are normalized and fall in the range $[w, 1]$ for some $w \in (0, 1]$. When $(i, j) \notin E$, we set $w_{ij} = 0$. Accordingly, for each vertex $i \in V$ we define the weighted degree as

$$\mathsf{deg}_w(i) \triangleq \sum_{j \in V} w_{ij}. \tag{3}$$

The following definition formalizes the notion that pooling designs that enable the construction of the test matrix $\mathbf{X}(T)$ should conform to the structure of graph G.

Definition 6 (Graph-constrained pooling): For a given set of items V, and an associated undirected and weighted graph $G = (V, E)$, we say that pooling is graph-constrained when the items pooled in each test form an induced path or an induced cycle in G.

In this paper, our focus will be on weight-bounded graphs.

Definition 7 ((ν, d)-regular graph): A graph is called (ν, d)-regular if for all $i \in V$ we have $\mathsf{deg}_w(i) \in [\nu, d]$ for given constants $d \geq \nu > 0$.

We remark that a fully-connected, unweighted graph is equivalent to an (n, n)-regular graph with $w = 1$ and an unweighted, degree regular graph with degree ν is equivalent to a (ν, ν)-regular graph with $w = 1$. While we primarily focus on general (ν, d)-regular graphical models, we will also provide results for the special cases in which graph G includes non-overlapping community structures, which are formalized next.

Definition 8 (Community-structured graphs): We say that graph G consists of M communities if G is partitioned to M disjoint subgraphs $\{G_m = (V_m, E_m) : m \in \{1, \ldots, M\}\}$, such that, $n_m \triangleq |V_m|$. The intra-community edges in G_m have weight s_m, and the inter-community edges have weight q, i.e.,

$$w_{ij} = \begin{cases} s_m & \text{if } \ (i, j) \in E_m \qquad \forall m \in \{1, \ldots, M\} \\ q & \text{if } i \text{ and } j \text{ belong to different communities} \end{cases}. \tag{4}$$

We also remark that the parameters s_m and q can be selected independently. The neighbors of vertex $i \in V_m$ within the community G_m are denoted by $N_i^m \triangleq V_m \cap N_i$. We denote the minimum weight degree of any vertex in community $\mathcal{G}_m$ by ν_m, where clearly, $\nu_m \leq \nu$. Hence, subgraph $\mathcal{G}_m$ is (ν_m, d)-regular. An example of a community-structured graph with $k = 1$ defective is illustrated in Fig. 1. In biological applications, the study of biological compounds that "inhibit" or block the activity of other entities is of particular interest. For instance, in drug discovery, an example of the activity of the lead compounds is their binding to pathogenic proteins, that eventually inhibit their activity. On the one hand, such inhibitors block the detection of pathogenic proteins in a population of biomolecules if their identity is unknown. On the other hand, inhibitor compounds are crucial in designing drugs [44]. However, not all lead compounds may have a uniform activity profile. For instance, a lead compound (inhibitor) may inhibit only a subset of pathogenic proteins or a pathogenic protein may be inhibited only by a specific pool of lead compounds. Such interactions have been modeled by "immune-defectives" graphs in [44]. Therefore, motivated by such biological applications, in which the presence of inhibitory compounds may inhibit the detection of items of interest (in this case, defective items), we also consider the setting in which the items may consist of inhibitors, where the activity of an inhibitor is defined next.

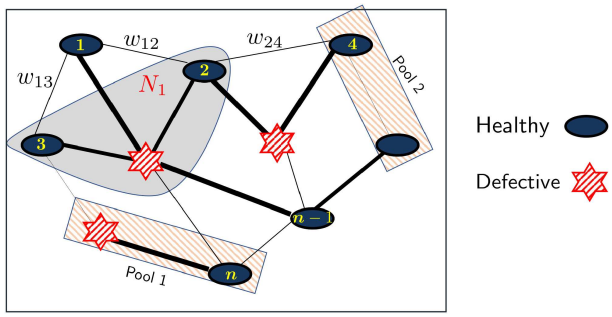

Fig. 1. A population with $(n-3)$ healthy and $k = 3$ defective items. It is a $(2w, 5)$-regular graph, since $2 \leq N_i \leq 5$ and for the weights we have $w \leq w_{ij} \leq 1$. The pooling capacity is $\ell = 2$. The connectivity structure specifies the pooling constraints. The set of items that item 1 can be pooled with, i.e., N_1, is shown in the solid shaded region. Two tests are depicted in hashed shaded regions (Pool 1 and Pool 2). The outcome of test on Pool 1 is $y_t = 1$ and the outcome of the test on Pool 2 is $y_t = 0$.

Definition 9 (Inhibitor): Distinct from the defective and non-defective items, an object is called an inhibitor if its presence in a pool renders the output of testing the pool *meaningless*.

Hence, under the potential presence of inhibitors among the objects, we have ternary outputs for the tests. Specifically, the outcome of a test t is $y_t \in \{0, 1, \mathrm{e}\}$, where $y_t = \mathrm{e}$ indicates that the objects pooled in test t contain at least one inhibitor. We assume that there can be at most r inhibitor items among the n items, when inhibitors are present. We denote the set of inhibitors by $\mathcal{J} \subset V$. In multi-stage group testing, in principle, all the test matrices and test outputs are leveraged to form an estimate for

$\mathcal{K}$. In the group testing frameworks discussed in this paper, the preliminary stages will be used to winnow out the non-defective vertices of the graph, and the final decisions will be formed based on individual tests in the last stage.

III. Graph-Constrained Pooling

This section provides the key ingredients of the multi-stage graph-constrained group testing algorithms presented in Section IV for different settings. Selecting and grouping the items such that the selection conforms to the underlying graph is facilitated via a random walk discussed in Section III-A. The random walk model is then leveraged to construct the test matrices in Section III-B.

A. Random Walk Over the Graph

To construct the test matrices that conform to the pooling restrictions imposed by a general (ν, d)-regular graph G, we adopt a random walk over G. Specifically, each test is designed independently of the rest through a random walk. The vertices visited by the random walk represent the items to be pooled. Each walk's origin is selected randomly, and its length is bounded by the pooling capacity ℓ. Each vertex may be visited more than once, and the random walks in different tests are allowed to cross paths. In this subsection, we formalize a generic random walk process over a given graph G. We also provide the definitions and notations that we will leverage in describing the algorithms and their attendant analyses. The definitions of the random walk described in the context of G readily extend to its connected subgraphs as well, which we will leverage for community-aware group testing and group testing with inhibitors.

We define Π as an $n \times n$ transition matrix that models a sequential random walk on G. The element of Π at coordinate (i, j) signifies the probability that a random walk moves from vertex i to vertex j, and it is defined as the weight of edge w_{ij} relative to the aggregate weight of the edges emitting from vertex i, i.e.,

$$[\Pi]_{ij} \triangleq \frac{w_{ij}}{\sum_{j \in N_i} w_{ij}}, \qquad \forall i, j \in V. \tag{5}$$

Denote the stationary probability distribution of the random walk by $\boldsymbol{\pi} \triangleq [\pi_1, \ldots, \pi_n]$, representing the distribution of the position in the graph if we run a random walk for an infinite number of steps. Accordingly, define

$$\pi_{\mathsf{max}} \triangleq \max_{i \in V} \pi_i, \quad \text{and} \quad \pi_{\mathsf{ratio}} \triangleq \max_{i,j \in V} \frac{\pi_i}{\pi_j}\ . \tag{6}$$

Corresponding to a graph G and its stationary probability distribution $\boldsymbol{\pi}$, the *mixing time* quantifies the length of time after which the distribution of the vertices visited by any random walk on G becomes point-wise close to the stationary distribution $\boldsymbol{\pi}$ [45], and it is formally defined next.

Definition 10 (β-mixing time): For a random walk of length τ over $G = (V, E)$ with the stationary probability distribution $\boldsymbol{\pi}$, define $\boldsymbol{\pi}_v^\tau$ as the position distribution when the random walk starts at vertex $v \in V$. The β-mixing time with respect to the ℓ_∞ norm is the smallest integer τ such that

$$\|\boldsymbol{\pi}_v^t - \boldsymbol{\pi}\|_\infty \leq \beta\ , \qquad \forall t \geq \tau, \tag{7}$$

for any $v \in V$. We denote the β-mixing time by ϖ.

The definitions provided in this section readily extend to random walks over the connected subgraphs of G too. To represent these metrics for a subgraph $F \subseteq G$, we use the notations $\Pi(F)$, $\pi_{\mathsf{max}}(F)$, $\pi_{\mathsf{ratio}}(F)$, and $\boldsymbol{\pi}(F)$.

B. Test Matrix Construction

The *efficiency* of the group testing framework is quantified by the number of tests required for successfully isolating the defective items. The following two definitions play pivotal roles in distinguishing the efficiency of different test matrices.

Definition 11 ((σ, ψ, n)-selector Matrix): A Boolean matrix $\mathbf{X} \in \{0,1\}^{T \times n}$ is (σ, ψ, n)-selector for integers $1 \leq \psi \leq \sigma < n$ if any submatrix of A constructed by choosing σ columns of A, contains at least ψ distinct rows of the identity matrix I_σ.

As established in [13], in a two-stage group testing framework, if the test matrix of the first stage is a $(2k, k+1, n)$-selector, then it is possible to perfectly identify all the defective items in the absence of inhibitors.

Definition 12 (k-disjunct Matrix): A matrix is said to be k-disjunct if no set of k columns has a Boolean sum that is a superset of any other single column.

In unconstrained testing, a test matrix $\mathbf{X}(T)$ that is k-disjunct and its respective test outcomes can be decoded using existing algorithms for the perfect isolation of up to k defective items in the noiseless setting [46]. This property is leveraged in [20] to design a one-stage group testing framework with graph constraints. In this context, we add the following remark [47].

Remark 1: A $(k+1, k+1, n)$-selector matrix is equivalent to a k-disjunct matrix with n columns.

We note that in unconstrained testing, the two-stage group testing framework that leverages a $(2k, k+1, n)$-selector test matrix in the first stage to isolate a superset of defectives followed by individual testing of the isolated items is known to require a factor k fewer tests compared with one-stage group testing. Specifically, in the one-stage framework in [10], it was shown that $\Theta(k^2 \log n/k)$ pooled tests are sufficient to form a k-disjunct matrix, and consequently, to perfectly isolate up to k defective items. On the other hand, in [13] it was shown that $\Theta(k \log n/k)$ pooled tests are sufficient to form a $(2k, k+1, n)$-selector matrix, and subsequently, $\Theta(k \log n/k + k)$ tests are sufficient to perfectly isolate up to k defective items. Parallel to these results for unconstrained group testing, we will show that the same level of gain in the number of tests (i.e., factor k) can also be achieved in graph-constrained group testing when we use multi-stage group testing versus non-adaptive one-stage group testing [20].[1]

IV. Algorithms and Main Results

In this subsection, we provide multi-stage graph-constrained group testing algorithms for (ν, d)-regular graphs. The common

[1]Noisy settings with up to e incorrect test outcomes can be accommodated in the one-stage group testing framework by considering a more restrictive (k, e)-disjunct matrix design [20, Definition 2]. We remark that noisy settings with up to e incorrect test outcomes can similarly be accommodated in the two-stage group testing framework by modifying Definition 11 to a more restrictive selector matrix design, $(2k, k+1, n, e)$-selector matrix. A $(2k, k+1, n, e)$-selector matrix is characterized by the property that any submatrix constructed by a random selection of $2k$ columns contains at least e copies of $k+1$ distinct rows of the identity matrix I_{2k}.

Algorithm 1: Forming a Test Matrix $\mathbf{X}(T)$ on Subgraph F.

input subgraph F with n_F number of vertices, pooling capacity ℓ, number of tests T
set matrix $\mathbf{X}(T) = \mathbf{0}_{T\times n}$
for test $t \in \{1, \ldots, T\}$ **do**
  initialize $v_{\mathsf{prev}} = 0$, $v = 0$
  sample vertex v from V according to $\boldsymbol{\pi}(F)$ and set $[\mathbf{X}(T)]_{tv} = 1$
  for $j \in \{2, \ldots, \ell\}$ **do**
    set $v_{\mathsf{prev}} = v$
    sample vertex v from the neighborhood of v_{prev} according to $\Pi(F)$
    set $[\mathbf{X}(T)]_{tv} = 1$
  end for
end for
return test matrix $\mathbf{X}(T)$

Algorithm 2: Decoding $\mathbf{X}(T)$ and $\mathbf{y}$.

input test matrix $\mathbf{X}(T)$, test outcomes $\mathbf{y}$
initialize empty set $\mathcal{H} = \varnothing$
for column $j \in \{1, \ldots, n_F\}$ **do**
  If $\mathbf{y}$ covers j-th column of $\mathbf{X}(T)$ **then**
    add j to $\mathcal{H}$
  end if
end for
return set $\mathcal{H}$

theme of the algorithms in different settings is that the preliminary stages are intended to remove the non-defective vertices of the graph, deferring the final decisions about the defective items to the last stage. Forming the test matrices constitutes the core of these algorithms. We start by providing two subroutines for any subgraph F spanning n_F vertices in G that are instrumental to filtering out the non-defective vertices in F using pooled testing. The first subroutine (Algorithm 1) provides the steps for constructing a test matrix using graph-constrained pooling over F. The second subroutine (Algorithm 2) provides the steps for decoding the test outcomes and the test matrix jointly.

The decision on the item represented by j-th column in $\mathbf{X}(T)$ being discarded for further decision making hinges on Step 4 in Algorithm 2, which tests if the respective column of $\mathbf{X}(T)$ is *covered* by $\mathbf{y}$, i.e., the positions of 1's in j-th column of $\mathbf{X}(T)$ form a subset of the positions of $1's$ in $\mathbf{y}$.

A. (ν, d)-Regular Graphs with no Inhibitors

We first discuss the two-stage group testing framework for recovering $\mathcal{K}$ perfectly in the absence of inhibitors. The two stages of the framework for successful recovery of $\mathcal{K}$ over (ν, d)-regular graph $\mathcal{G}$ are described below.

1. **Filtration:** We conduct T_{F} parallel pooled tests to form a test matrix $\mathbf{X}(T_{\mathsf{F}})$ according to Algorithm 1 and determine the test outcome vector $\mathbf{y}$. Test matrix $\mathbf{X}(T_{\mathsf{F}})$ and decision vector $\mathbf{y}$ are subsequently used by Algorithm 2 to distill the set of candidates to the set $\mathcal{H}$ (generated by Algorithm 2). We are interested in identifying a superset of $\mathcal{K}$ with $2k-1$ elements. It is shown in [13, Theorem 3] that if $\mathbf{X}(T_{\mathsf{F}})$ is a $(2k, k+1, n)$-selector matrix, then the output of Algorithm 2 consists of $2k-1$ vertices that form a superset of $\mathcal{K}$. We reprise the proof of this claim from [13] in Appendix A for completeness.
2. **Isolation:** Each of the vertices in Stage 1 is tested individually. This leads to identifying all defective items.

Therefore, the total number of tests performed is

$$T_{\mathsf{total}} \triangleq T_{\mathsf{F}} + |\mathcal{H}| \,, \tag{8}$$

and when $\mathbf{X}(T_{\mathsf{F}})$ is a $(2k, k+1, n)$-selector matrix, we have $T_{\mathsf{total}} = T_{\mathsf{F}} + 2k - 1$. Since the construction of $\mathbf{X}(T_{\mathsf{F}})$ is stochastic due to random walks involved, $\mathbf{X}(T_{\mathsf{F}})$ being a $(2k, k+1, n)$-selector matrix is ensured only stochastically. For this purpose, we define a Bernoulli random variable

$$b \triangleq \begin{cases} 1 & \mathbf{X}(T_{\mathsf{F}}) \text{ is a } (2k, k+1, n)\text{-selector matrix} \\ 0 & \text{otherwise} \end{cases} . \tag{9}$$

Therefore, the objective is to appropriately select T_{F}, such that, we have $\mathbb{P}(b = 1) \geq 1 - \epsilon$ for some $\epsilon \in [0, 1/2)$. We next formalize sufficient conditions on T_{F} for selector matrix construction and the total number of tests T_{total} for perfect isolation of $\mathcal{K}$.

Theorem 1 (Two-stage group testing): When graph G is (ν, d)-regular and the β-mixing time constant is set to $\beta = \frac{\pi_{\mathsf{max}}}{n}$, if ℓ and T_{F} satisfy

$$\ell = O\left(\frac{1}{k\pi_{\mathsf{max}}\pi_{\mathsf{ratio}}\varpi}\right) \,, \tag{10}$$

$$T_{\mathsf{F}} = \Theta\left(\varpi^2\pi_{\mathsf{ratio}}\left(k\log\frac{n}{k} + \log\frac{1}{\epsilon}\right)\right) \,, \tag{11}$$

then when $k = O(\frac{\nu w}{\varpi^2\pi_{\mathsf{ratio}}})$ we have the following guarantees:

1) The filtration process identifies $2k-1$ elements forming a superset of $\mathcal{K}$ with a probability at least $1-\epsilon$, i.e.,
$$\mathbb{P}(\mathcal{K} \subset \mathcal{H} \text{ and } |\mathcal{H}| = 2k-1) \geq 1 - \epsilon \,. \tag{12}$$
2) T_{total} number of tests is sufficient to perfectly recover $\mathcal{K}$ with probability at least $1-\epsilon$, where
$$T_{\mathsf{total}} = \Theta\left(\varpi^2\pi_{\mathsf{ratio}}\left(k\log\frac{n}{k} + \log\frac{1}{\epsilon}\right) + k\right) \,. \tag{13}$$

Proof: See Appendix B. □

From Theorem 1, we make the following observations.

Remark 2 (Optimal scaling behavior): The total number of tests T_{total} sufficient for isolating $\mathcal{K}$ with high probability scales as $\Theta(\varpi^2\pi_{\mathsf{ratio}}k\log\frac{n}{k} + k)$, achieving a linear scaling behavior in k. This matches the scaling behavior of k in the information-theoretic lower bound of $k\log\frac{n}{k}$ on the number of tests for isolating k defectives from n items without constraints on pooling [10].

Remark 3 (Tradeoff between graph sparsity and defectiveness): There exists an inherent tradeoff between the size of the defective set (k) and the sparsity of the graph, captured by ν. Specifically, a larger value for ν implies that there exist more paths that do not contain a defective node for a given k. In adaptive pooling, in order to effectively winnow out the non-defective items, it is desirable to have as many defective-free pools as possible. This is, however, hindered as k increases for a given ν, which leads to an increase in number of tests T_{total}. It can be even observed that in certain settings, if the value of k exceeds a certain level, perfect isolation of the defective items

may become infeasible. For instance, consider a 10-vertex graph in which the vertices $V = \{1, \ldots 10\}$ are connected in a circle ($\nu = 2$). Assume that nodes $\{1, 3, 5, 7, 9\}$ are defective. In this example, we will never be able to isolate all defectives when $\ell > 1$ because the pooled test will always return an outcome of one. Such an interplay between k and the pooling capacity ℓ and graph sparsity ν is also observed in Theorem 1, where it is shown that recovery is guaranteed only in the regime $k = O(\frac{\nu w}{\pi_{\mathsf{ratio}} \varpi^2})$ which reflects a tradeoff between the sparsity of the graph (captured by ν) and the feasible maximum number of defectives k that can be successfully isolated, with an increase in ν implying a broader set of feasible values for k.

Remark 4 (Connection to non-adaptive group testing): Comparison with the results for the non-adaptive graph-constrained group testing indicates that the total number of tests in our framework scales by a factor k slower than $\Theta(\varpi^2 \pi_{\mathsf{ratio}} k^2 \log \frac{n}{k})$ tests sufficient in the non-adaptive framework, indicating a significant reduction in complexity. Such a comparison can be carried out by leveraging the tradeoff mentioned above. Specifically, Theorem 1 considers a given graph with an arbitrary value of ν, and imposes a condition (upper bound) on k in terms of ν to provide sufficient conditions for recovering the defective items. On the other hand, the analysis in [20] takes the dual approach of letting k change freely and imposing a condition (lower bound) on ν as a function of k. This condition of [20] is equivalent to $\nu = \Omega(k \varpi^2 \pi_{\mathsf{ratio}} / w)$ in our framework. We can re-interpret the results in [20] from the perspective of feasibility of k for a given (ν, d)-regular graph, where both non-adaptive and adaptive graph-constrained group testing frameworks work for the same regimes of k.

Remark 5 (Matrix design): The conditions on the pooling capacity (length of random walk) ℓ and the number of pooled tests T_{F} are the key factors influencing the design of $\mathbf{X}(T_{\mathsf{F}})$. The factors ϖ, π_{max}, and π_{ratio} are indicators of the bias in the random walk over the graph, and increasing the bias implies an increase in all these metrics. Based on Theorem 1, we note that T_{F} increases with an increase in the β-mixing time ϖ and the metrics π_{max} and π_{ratio} of the random walk. On the other hand, for a fixed k, an increase in ϖ implies that the upper bound on ℓ in (10) shrinks. Clearly, a stronger bias necessitates a larger number of pooled tests T and a smaller pooling capacity ℓ.

Remark 6 (Mixing time): The β-mixing time ϖ significantly influences the feasibility of group testing and is impacted by the graph's connectivity. Specifically, if the graph is 'loosely' connected, a higher β-mixing time implies that T_{F} random walks are not guaranteed to cover all vertices in the graph. Therefore, a desirable property for the graph-constrained random walk is the ability to mix rapidly (i.e., for instance, ϖ being poly-logarithmic in n) to avoid bottlenecks in the coverage of vertices [48]. We also remark that if the mixing time of a graph is not known, it can be estimated by existing algorithms in the literature [49].

Remark 7 (Restrictions on network access): In certain applications such as network tomography, only some parts of the network may be accessible and only a subset of vertices can be used as starting or ending points of the random walk. In these applications, we can modify our analysis along similar lines as in [20] to show that the number of tests T_{F} increases by a factor of ϖ^2, while preserving the other dependencies on k and n.

The following corollary summarizes the results for two-stage group testing based on an unbiased random walk for a degree-regular graph with $w = 1$ and $d = \nu$. In this setting, the β-mixing time can be approximated by $\varpi = \Theta(\log \beta \log n/(1-\eta))$ for a degree-regular graph, where η is the eigenvalue of Π with the second highest absolute value and $(1-\eta)$ is the eigengap of the random walk over G [45].

Corollary 1 (Degree-regular graph): When graph G is (ν, ν)-regular with $w = 1$ and the β-mixing time constant is set to $\beta = n^{-2}$, if the parameters satisfy

$$\ell = O\left(\frac{n(1-\eta)}{k \log^2 n}\right), \; T_{\mathsf{F}} = \Theta\left(\frac{\log^4 n}{1-\eta}\left(k \log \frac{n}{k} + \log \frac{1}{\epsilon}\right)\right), \tag{14}$$

then when $k = O(\frac{\nu(1-\eta)^2}{\log^4 n})$ we have the following guarantees:

1) The filtration process identifies $2k-1$ elements forming a superset of $\mathcal{K}$ with a probability at least $1-\epsilon$, i.e.,
$$\mathbb{P}(\mathcal{K} \subset \mathcal{H} \;\text{ and }\; |\mathcal{H}| = 2k-1) \geq 1-\epsilon\,. \tag{15}$$
2) T_{total} number of tests is sufficient to perfectly recover $\mathcal{K}$ with probability at least $1-\epsilon$, where
$$T_{\mathsf{total}} = \Theta\left(\left(\frac{k}{1-\eta} \log^4 n \log \frac{n}{k} + \log \frac{1}{\epsilon}\right) + k\right)\,. \tag{16}$$

From Corollary 1, we note that the number of tests T_{total} scales as $\Theta(k \log^4 n \log \frac{n}{k})$ for a degree-regular graph. We next formalize the result for a fully-connected graph, in which case we have $\pi_{\mathsf{ratio}} = 1$.

Corollary 2 (Fully connected graph): For a fully connected graph, the β-mixing time is $\varpi = 1$ and the total number of tests T_{total} scales as $\Theta(k \log n/k + k)$.

Note that the information-theoretic lower bound on the number of tests for recovering k defective items out of n total number of items scales as $k \log n/k$ [10]. Therefore, in the asymptote of large n, when $k = O(n)$, the number of tests for the two-stage group testing framework has an optimal scaling behavior.

We next discuss the extension of the two-stage group testing framework in Theorem 1 to graphs with community structures. In practical scenarios, vertices within the same community likely have higher affinity than the vertices in different communities, i.e., the edge weights satisfy $s_m \gg q$ and, therefore, the graph G may consist of loosely-connected subgraphs. Since the β-mixing time ϖ depends on the eigengap of Π, it increases rapidly for loosely-connected graphs [45]. Thus, the two-stage group testing framework in Theorem 1 may be practically infeasible in practice for graphs with communities. Using this motivation, we next provide a framework graphs with community structures.

B. (ν, d)-Regular Graphs with Communities

For community-aware group testing in the absence of inhibitors, we include an additional *scanning* stage prior to the filtration step to identify the communities with at least one defective item. In this stage, each community in the graph is treated as an independent entity, and the test determines whether the community consists of at least one defective vertex. We denote the set of communities deemed to have at least one defective vertex by the preliminary stage by $\mathcal{M} \subseteq M$. The number of tests performed in this step is denoted by T_{s}. The subsequent stages

for each community in $\mathcal{M}$ follow similar steps as discussed in Section IV-A for two-stage group testing.

1. **Scanning:** Each community $m \in M$ is treated as an individual entity. We conduct T_{s} number of tests, where in each test, a pool of a subset of communities in M are tested for the presence of a defective item. In this stage, we use an existing optimal group testing method for non-adaptive group testing based on disjunct matrix construction to determine the set $\mathcal{M}$ [10].
2. **Filtration:** For each community $m \in \mathcal{M}$, we conduct $T_{\mathsf{F},m}$ parallel pooled tests, with the pooling capacity ℓ_m on G_m. The test matrix $\mathbf{X}(T_{\mathsf{F},m})$ is formed using similar steps as described in Algorithm 1. The outcomes of tests and $\mathbf{X}(T_{\mathsf{F},m})$ are jointly decoded using similar steps as in Algorithm 2 to a set of vertices denoted by $\mathcal{H}_m$. If $\mathbf{X}(T_{\mathsf{F},m})$ is a $(2k, k+1, n_m)$-selector matrix, then $\mathcal{H}_m$ consists of $2k-1$ vertices that form a superset of the defective items in community m.
3. **Isolation:** Finally, each of the $|\mathcal{H}_m|$ selected vertices from community $m \in \mathcal{M}$ is tested individually, and the outcomes from all communities in M are combined to isolate $\mathcal{K}$.

Therefore, the total number of tests to isolate the defective set $\mathcal{K}$ for graphs with communities is given by

$$T_{\mathsf{total}} \triangleq T_{\mathsf{s}} + \sum_{m \in \mathcal{M}} \left(T_{\mathsf{F},m} + |\mathcal{H}_m|\right) . \tag{17}$$

Successfully isolating $\mathcal{K}$ is predicated upon perfectly isolating the communities $\mathcal{M}$ in the scanning stage, and appropriately constructing matrices $\mathbf{X}(T_{\mathsf{s}})$ and $\mathbf{X}(T_{\mathsf{F},m})$. Determining $\mathcal{M}$ is a standard group testing problem, and we use an existing optimal strategy in the scanning stage. Subsequently, the efficiency of community-aware group is characterized by $\{T_{\mathsf{F},m} : m \in \mathcal{M}\}$. We formalize this with the help of a Bernoulli random variable b_m for community m defined as

$$b_m \triangleq \begin{cases} 1 & \mathbf{X}(T_{\mathsf{F},m}) \text{ is a } (2k, k+1, n_m)\text{-selector matrix} \\ 0 & \text{otherwise} \end{cases} .$$

Therefore, the objective is to determine an appropriate number of pooled tests based on random walks for each community $m \in \mathcal{M}$ such that we have $\mathbb{P}(b_m = 1) \geq 1 - \epsilon$ for some $\epsilon \in [0, 1/2)$. The following lemma characterizes the number of tests sufficient for identifying communities with defective vertices in the scanning stage.

Lemma 1 (Community identification): T_{s} tests is sufficient for perfectly identifying the communities that have at least one defective vertex, where T_{s} is defined as

$$T_{\mathsf{s}} \triangleq \begin{cases} \Theta\left(k^2 \log \frac{M}{k}\right) & \text{if} \quad k \leq \sqrt{M} \\ M & \text{otherwise} \end{cases} . \tag{18}$$

The proof of Lemma 1 follows directly from the classical group testing results on disjunct matrix construction driven non-adaptive group testing [10]. We remark that the dependence of T_{s} on k in (18) can be scaled from k^2 down to k by adopting a vanishing error criterion [10]. However, to maintain consistency with the subsequent stages, we have provided the results corresponding to the zero error criterion in Lemma 1.

We next provide a sufficient condition for a community-aware graph-constrained group testing framework to isolate the defective items from the communities in $\mathcal{M}$. The following theorem characterizes the number of tests for the test matrix $\mathbf{X}(T_{\mathsf{F},m})$ being a $(2k, k+1, n_m)$-selector matrix and the total number of tests required for isolating up to k defective vertices in the graph. For this purpose, we denote the β_m-mixing time for graph G_m by ϖ_m.

Theorem 2 (Community-aware group testing): In community $m \in \mathcal{M}$, when the β_m-mixing time constant for G_m is set to $\beta_m = n_m^{-2}$, if the parameters satisfy

$$\ell_m = O\left(\frac{1}{k \pi_{\mathsf{max}}(G_m) \pi_{\mathsf{ratio}}(G_m) \varpi_m}\right) , \tag{19}$$

$$T_{\mathsf{F},m} = \Theta\left(\varpi_m^2 \pi_{\mathsf{ratio}}(G_m) \left(k \log \frac{n_m}{k} + \log \frac{1}{\epsilon_m}\right)\right) , \tag{20}$$

then when $k = O(\frac{\nu_m w}{\pi_{\mathsf{ratio}}(G_m) \varpi_m^2})$ we have the following guarantees:

1) The filtration stage identifies $2k-1$ elements forming a superset of $\mathcal{K}_m$ with probability at least $1 - \epsilon$, i.e.,
$$\mathbb{P}(\mathcal{K}_m \subset \mathcal{H}_m \text{ and } |\mathcal{H}_m| = 2k-1) \geq 1 - \epsilon_m . \tag{21}$$
2) T_{total} number of tests is sufficient to perfectly recover $\mathcal{K}$ with probability at least $\prod_{m \in \mathcal{M}} (1 - \epsilon_m)$, where
$$T_{\mathsf{total}} = T_{\mathsf{s}} + \sum_{m \in \mathcal{M}} \left(T_{\mathsf{F},m} + 2k - 1\right) . \tag{22}$$

The proof of Theorem 2 follows directly from the proof of Theorem 1. From Theorem 2, we observe that the number of tests for isolating up to k defectives from community m, $T_{\mathsf{F},m}$ has linear dependence on k. We argue that the number of tests for each community can be further improved by incorporating more side information when available, such as the number of defectives per community [35]. The conditions on $T_{\mathsf{F},m}$ and ℓ_m collectively determine the design of test matrix $\mathbf{X}(T_{\mathsf{F},m})$ for community m.

C. (ν, d)-Regular Graphs with Inhibitors

A four-stage framework for identifying defective items in the presence of inhibitors is proposed in [13]. The first three stages of the strategy in [13] employ selector matrices for filtering out potential inhibitors, defectives, and non-defectives, respectively. The final and fourth stage individually tests the retained items. We extend this approach to the graph-constrained testing setting by leveraging random walks for constructing the selector matrices in the first three stages. When there are up to r inhibitors, the different stages of a four-stage group testing framework are formalized as follows.

1. **Benchmark specification:** In the first stage, our goal is to identify a certain pool of items $A_{\mathsf{bs}} \subseteq V$ that tests positive and does not contain any inhibitor, i.e., $A_{\mathsf{bs}} \cap \mathcal{J} = \varnothing$ and $A_{\mathsf{bs}} \cap \mathcal{K} \neq \varnothing$. In this stage, we conduct T_{bs} pooled tests on graph G with pooling capacity ℓ_{bs} to form a test matrix according to steps similar to Algorithm 1. We adopt pooled testing to identify A_{bs} such that it isolate as many items as possible with a substantially smaller number of tests compared with testing each item individually. Note that in this stage, a pooled test has an outcome 1 if and only if the pool of items tested has at least one defective item and no inhibitors. If $\mathbf{X}(T_{\mathsf{bs}})$ is a $(k+r, r+1, n)$-selector matrix, there exists a pool A_{bs}, such that, $|A_{\mathsf{bs}}| = n/(k+r)$ [13].

The set A_{bs} enables us to isolate a superset of all inhibitors from the items $V \backslash A_{\mathsf{bs}}$ in the next stage.

2. **Inhibitor detection:** In this stage, our goal is to identify a superset of the set of inhibitors $\mathcal{J}$. For this purpose, we first exclude the vertices in the set A_{bs} and their associated edges from the graph to focus on the subgraph with the vertices $V_{\mathsf{id}} = V \backslash A_{\mathsf{bs}}$. The number of vertices in the remaining graph is denoted by $n_{\mathsf{id}} \triangleq n - |A_{\mathsf{bs}}|$. We conduct T_{id} number of pooled tests with pooling capacity ℓ_{id} on the subgraph $G_{\mathsf{id}} \triangleq (V_{\mathsf{id}}, E_{\mathsf{id}})$ spanned by vertices V_{id} and a set of edges $E_{\mathsf{id}} \subseteq E$ that consists of all edges (u, v) such that $u, v \in V_{\mathsf{id}}$. Using steps similar to those in Algorithm 1, the pooled tests enable us to form a test matrix $\mathbf{X}(T_{\mathsf{id}})$ of dimension $T_{\mathsf{id}} \times n_{\mathsf{id}}$. In this stage, the outcome y_t of a pooled test is determined by coupling the t^{th} pool of items in $\mathbf{X}(T_{\mathsf{id}})$ with the pool A_{bs}. Since the pool A_{bs} consists of at least one defective vertex and no inhibitors, the outcome of a pooled test in this stage is e if and only if the t^{th} pool consists of at least one inhibitor and it is 1, otherwise. By using the principles from Section IV-A, we note that a superset of inhibitors $\mathcal{J}$ can be identified by using steps similar to those in Algorithm 2 with trivial transformation to the outcome vector [13]. Therefore, by leveraging a test matrix $\mathbf{X}(T_{\mathsf{id}})$ that is a $(2r, r+1, n_{\mathsf{id}})$-selector matrix, we can identify a set of vertices $A_{\mathsf{id}} \subset V$ such that $|A_{\mathsf{id}}| = 2r - 1$ and A_{id} is a superset of $\mathcal{J}$.
3. **Filtration:** In this stage, our goal is to cull the non-defective vertices from the subgraph in G obtained by removal of set of vertices A_{id}. For this purpose, we exclude the vertices in the set A_{id} and adopt a random walk for collecting data from on the subgraph $G_{\mathsf{f}} \triangleq (V_{\mathsf{f}}, E_{\mathsf{f}})$ with vertices $V_{\mathsf{f}} = V \backslash A_{\mathsf{id}}$, such that the number of vertices in G_{f} is given by $n_{\mathsf{f}} \triangleq n - |A_{\mathsf{id}}|$ and the set of edges $E_{\mathsf{f}} \subseteq E$ that consists of all edges (u, v) such that $u, v \in V_{\mathsf{f}}$. We conduct T_{f} number of pooled tests with pooling capacity ℓ_{f}. The pooled tests enable us to form a test matrix $\mathbf{X}(T_{\mathsf{f}})$ of dimension $T_{\mathsf{f}} \times n_{\mathsf{f}}$ using steps similar to those in Algorithm 1. We next decode $\mathbf{X}(T_{\mathsf{f}})$ and its respective outcomes using steps similar to those in Algorithm 2. Therefore, if $\mathbf{X}(T_{\mathsf{f}})$ is a $(2k, k+1, n_{\mathsf{f}})$-selector matrix, Algorithm 2 yields an outcome A_{f} such that $|A_{\mathsf{f}}| = 2k - 1$ and A_{f} is a superset of $\mathcal{K}$.
4. **Isolation:** Finally, we individually test the vertices in $A_{\mathsf{id}} \cup A_{\mathsf{f}}$ isolated in stages 2 and 3.

In this context, the total number of tests sufficient for successful isolation of defective vertices $\mathcal{K}$ is given by

$$T_{\mathsf{total}} \triangleq T_{\mathsf{bs}} + T_{\mathsf{id}} + T_{\mathsf{f}} + |A_{\mathsf{id}}| + |A_{\mathsf{f}}| \,, \tag{23}$$

where the total number of pooled tests $T_{\mathsf{bs}} + T_{\mathsf{id}} + T_{\mathsf{f}}$ determines the efficiency of the four-stage group testing framework. Therefore, the numbers of pooled tests in the first three stages must be selected such that they facilitate constructing appropriate selector test matrices in their respective stages with high likelihood. This can be compactly conveyed by using Bernoulli random variable b_m for stage m defined as

$$b_m \triangleq \begin{cases} 1 & \mathbf{X}(T_m) \text{ is an appropriate selector matrix for stage } m \\ 0 & \text{otherwise} \end{cases},$$

where $T_1 = T_{\mathsf{bs}}$, $T_2 = T_{\mathsf{id}}$, and $T_3 = T_{\mathsf{f}}$. Hence, in stages 1–3, we construct test matrices $\{T_{\mathsf{bs}}, T_{\mathsf{id}}, T_{\mathsf{f}}\}$ via random walks by appropriately selecting the number of tests such that $\mathbb{P}(\sum_{m=1}^{3} b_m = 3) \geq 1 - \epsilon$ for $\epsilon \in [0, 1/2)$. Next, we provide sufficient conditions for perfectly recovering $\mathcal{K}$ in the presence of inhibitors $\mathcal{J}$. For this purpose, we denote mixing time constants in the first three stages by ϖ_{bs}, ϖ_{id}, and ϖ_{f}, respectively. The following lemmas capture the design specifications for the test matrices in different stages.

Lemma 2 (Benchmark creation): When graph G is (ν, d)-regular and its β_{bs}-mixing time constant is set to $\beta_{\mathsf{bs}} = \frac{\pi_{\mathsf{max}}}{n}$, if ℓ_{bs} and T_{bs} satisfy

$$\ell_{\mathsf{bs}} = O\left(\frac{1}{r \pi_{\mathsf{max}} \pi_{\mathsf{ratio}} \varpi_{\mathsf{bs}}}\right), \tag{24}$$

$$T_{\mathsf{bs}} = \Theta\left(\pi_{\mathsf{ratio}} \varpi_{\mathsf{bs}}^2 \frac{(r+1)}{k}\left((k+r)\log\frac{n}{k+r} + \log\frac{1}{\epsilon}\right)\right), \tag{25}$$

then when the maximum number of inhibitors satisfies $r = O(\frac{\nu w}{\varpi_{\mathsf{bs}}^2 \pi_{\mathsf{ratio}}(G)})$, the test matrix $\mathbf{X}(T_{\mathsf{bs}})$ is a $(k+r, r+1, n)$-selector matrix with a probability at least $1 - \epsilon > 0$.

Proof: The proof follows directly from the construction of a $(\alpha, k+1, n)$-selector matrix in Appendix B. □

Lemma 3 (Inhibitor detection): When graph G_{id} is (ν_{id}, d)-regular and its mixing time constant is set to $\beta_{\mathsf{id}} = \frac{\pi_{\mathsf{max}}(G_{\mathsf{id}})}{n_{\mathsf{id}}}$, if ℓ_{id} and T_{id} satisfy

$$\ell_{\mathsf{id}} = O\left(\frac{1}{r \varpi_{\mathsf{id}} \pi_{\mathsf{max}}(G_{\mathsf{id}}) \pi_{\mathsf{ratio}}(G_{\mathsf{id}})}\right), \tag{26}$$

$$T_{\mathsf{id}} = \Theta\left(\pi_{\mathsf{ratio}}(G_{\mathsf{id}}) \varpi_{\mathsf{id}}^2 \left(r \log n_{\mathsf{id}} + \log\frac{1}{\epsilon}\right)\right), \tag{27}$$

then when the maximum number of inhibitors satisfies $r = O(\frac{\nu_{\mathsf{id}} w}{\varpi_{\mathsf{id}}^2 \pi_{\mathsf{ratio}}(G_{\mathsf{id}})})$, the test matrix $\mathbf{X}(T_{\mathsf{id}})$ is a $(2r, r+1, n_{\mathsf{id}})$-selector matrix with a probability at least $1 - \epsilon$.

Proof: The proof follows directly from Appendix B where we replace k with r as a parameter in the construction of selector matrix. □

Note that Lemma 3 is equivalent to the results for the construction of a $(2k, k+1, n)$-selector matrix in Theorem 1 with parameters in terms of r. Furthermore, lemmas 2 and 3 provide distinct conditions on the number of inhibitors r for successful construction of the respective selective matrices in stages 1 and 2. Since $\nu_{\mathsf{id}} \leq \nu$, we argue that the condition in Lemma 3 is more stringent and therefore, captures the overall condition on the number of inhibitors.

Lemma 4 (Filtration): When graph G_{f} is (ν_{f}, w)-regular and ts mixing time constant is set to $\beta_{\mathsf{f}} = \frac{\pi_{\mathsf{max}}(G_{\mathsf{f}})}{n_{\mathsf{f}}}$, if ℓ_{f} and T_{f} satisfy

$$\ell_{\mathsf{f}} = O\left(\frac{1}{\nu \varpi_{\mathsf{f}} \pi_{\mathsf{max}}(G_{\mathsf{f}}) \pi_{\mathsf{ratio}}(G_{\mathsf{f}})}\right), \tag{28}$$

$$T_{\mathsf{f}} = \Theta\left(\pi_{\mathsf{ratio}}(G_{\mathsf{f}}) \varpi_{\mathsf{f}}^2 \left(k \log\frac{n_{\mathsf{f}}}{2k} + \log\frac{1}{\epsilon}\right)\right), \tag{29}$$

then when the maximum number of defective items satisfies $k = O(\frac{\nu_{\mathsf{f}} w}{\varpi_{\mathsf{f}}^2 \pi_{\mathsf{ratio}}(G_{\mathsf{f}})})$, the test matrix $\mathbf{X}(T_{\mathsf{f}})$ is a $(2k, k+1, n_{\mathsf{f}})$-selector matrix with a probability at least $1 - \epsilon$.

The proof directly follows the proof of Theorem 3. Lemmas 2–4 collectively establish the following overall performance for the

four-stage group testing framework. From lemmas 3–4, we note that the number of pooled tests T_{id} and T_{f} in stages 2 and 3 depend on the properties of their respective random walks on reduced graphs G_{id} and G_{f}, respectively.

Theorem 3 (Group testing with inhibitors): If the subgraphs G_{id} and G_{f} have connected edge structures and their mixing time constants are set according to lemmas 2–4, the four-stage group testing framework isolates the defective items with probability at least $(1-\epsilon)^3$, when the pooling capacities and the number of tests follow those specified by lemmas 2–4.

In general, it is known that the number of tests for recovering up to k defective items and r inhibitors from a pool of n items scales at least as fast as $(k+r)\log n$ [40]. When we have a fully connected graph G (i.e., an (n,n)-regular graph) with an unbiased random walk, the mixing time constants for all stages are set as $\varpi_{\mathsf{bs}} = \varpi_{\mathsf{id}} = \varpi_{\mathsf{f}} = 1$. In this setting, the following theorem characterizes the scaling behavior of the number of tests for four-stage group testing.

Theorem 4 (Fully connected graph): When graph G is fully connected, by performing an unbiased random walk, the number of tests sufficient for recovering up to k defective vertices in the presence of up to r inhibitor vertices is specified as follows.

- In the regime $r = O(k)$, T_{total} grows as

$$\Theta\left((k+r)\log\frac{n}{k+r} + r\log n\right).$$

- In the regime $k = o(r)$, T_{total} grows as $\Theta(r^2\log n)$.

Remark 8: Based on the scaling behaviors of T_{total} in different regimes in Corollary 4, we remark that the asymptotic scaling behaviors of T_{total} matches with those for non-adaptive group testing in [43]. In the regime $r = O(k)$, the number of tests for the four-stage group testing framework scales almost linearly in r and k. On the other hand, in the regime $k = o(r)$, the number of tests scales as r^2.

V. Numerical Results

In Section V-A, we empirically evaluate the constructional aspects of graph-constrained group testing algorithms in this paper and compare them with those in the existing literature. In Section V-B, we implement the algorithms presented for a case study where various aspects are motivated by the realistic settings of infectious diseases.

A. Selector Matrix Construction

Selector matrix construction is instrumental to the filtration stage in group testing algorithms in Section IV-A and Section IV-B as well as benchmark creation, inhibitor detection, and filtration stages in the four-stage group testing algorithm in Section IV-C. In our numerical evaluations, we assess the construction of a $(2k, k+1, n)$-selector matrix and compare the complexity of graph-constrained construction of a $(2k, k+1, n)$-selector matrix against that of a $(k+1, k+1, n)$-selector matrix (or a k-disjunct matrix with n columns) which is relevant to the existing one-stage frameworks.

We consider a (40, 40)-regular graph G with number of vertices $n = 100$, minimum edge weight $w = 1$, and the maximum number of defective items $k = 2$. In this setting, each vertex has 40 vertices in its immediate neighborhood. We first evaluate

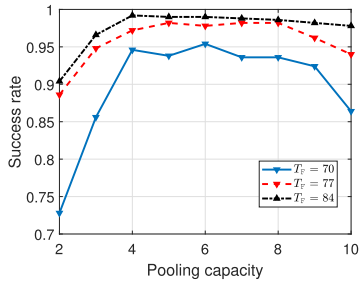

Fig. 2. Likelihood of forming a $(k+1, k+1, n)$-selector matrix versus pooling capacity ℓ.

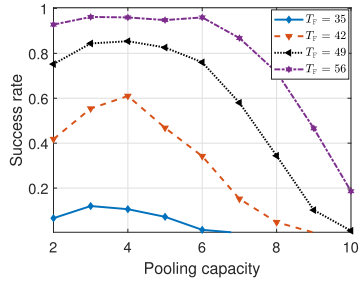

Fig. 3. Likelihood of forming a $(2k, k+1, n)$-selector matrix versus pooling capacity ℓ.

the likelihoods of successful construction of different selector matrices versus the length of a random walk over G. Figure 2 depicts the success rate of forming a $(k+1, k+1, n)$-selector matrix using random walk construction characterized by different number of tests T_{F} and pooling capacity ℓ over 1000 random instances. Similarly, Fig. 3 depicts the success rate of forming a $(2k, k+1, n)$-selector matrix using random walk construction over 1000 random instances. These two figures indicate that for a smaller number of tests, the likelihood of forming the test matrix with desired properties increases up to a certain pooling capacity (4 in this case), followed by a sharp decline as the length of the random walk is further increased beyond 5 for $(2k, k+1, n)$-selector matrix and a gradual decline in the case of $(k+1, k+1, n)$-selector matrix. These observations indicate that there is an upper limit on the pooling capacity beyond which the likelihood of successful selector matrix construction diminishes. This is consistent with the implications of Theorem 1 and the discussion in Remark 3. These experiments also indicate that the construction of a $(k+1, k+1, n)$-selector matrix is more robust to the variations in pooling capacity than that of a $(2k, k+1, n)$-selector matrix. Furthermore, we observe that the number of tests required for constructing $(2k, k+1, n)$-selector matrix with a given likelihood is less and becomes closer to half of that sufficient for the construction of a $(k+1, k+1, n)$-selector matrix with a similar likelihood as the likelihood becomes closer to 1. This observation confirms what is expected analytically, that is constructing a $(2k, k+1, n)$-selector matrix can be achieved by a factor of k smaller number of tests than that for the construction of a $(k+1, k+1, n)$-selector matrix.

We also evaluate the likelihood of constructing a $(2k, k+1, n)$-selector matrix for graphs with different degrees ν in Fig. 4. For the results in this figure, we set the pooling capacity fixed as

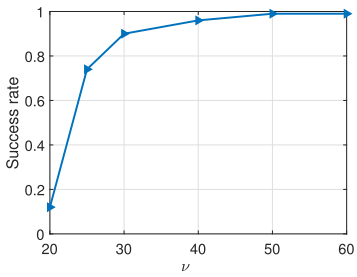

Fig. 4. Empirical likelihood of forming a $(2k, k+1, n)$-selector matrix versus degree ν for graphs of size $n = 100$, number of tests, $T_{\mathsf{F}} = 60$ and pooling capacity, $\ell = 4$.

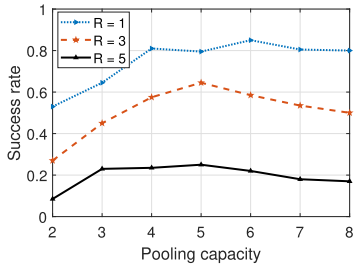

Fig. 5. Empirical likelihood of forming a $(2k, k+1, n)$-selector matrix versus R for graphs of size $n = 100$ and number of tests, $T_{\mathsf{F}} = 60$.

$\ell = 4$ and number of tests $T_{\mathsf{F}} = 60$ and vary ν. As expected from the results in Theorem 1, we observe that the success rate of forming the selector matrix improves with an increase in connectivity and, therefore, a decrease in the mixing time of the graph, as implied by increasing ν.

We next evaluate the likelihood of constructing a $(2k, k+1, n)$-selector matrix for graphs with different characteristics of random walk in Fig. 5. Specifically, we consider a graph with $M = 4$ communities and set $s_m = s, \forall m \in \{1, 2, 3, 4\}$. To reflect the bias of the random walk in the graph, we define $R \triangleq \frac{s}{q}$. Figure 5 illustrates the success rate for forming a selector matrix for $T_{\mathsf{F}} = 60$ samples over 1000 random instances for different R. Clearly, as the bias increases, the success rate drops for the same number of samples as indicated by Theorem 1.

B. Infection Detection

In this section, we evaluate our framework in the context of infection spread in a human population whose interactions are modeled by a community-structured graph. We use the community-aware group testing framework described in Section IV-B for this setup. Note that this framework was designed for noiseless testing, which is not a realistic setting in practice. Therefore, we characterize the performance of the pooled test adopted for group testing

1) Infection Test: We motivate the test performance by existing epidemiology studies. We note that the relative costs of false-positive errors and false-negative errors of a test can be different in practice. For instance, a significant number of false-negative errors in testing a population leads to infected individuals labeled as healthy. Therefore, tolerating a high rate of false-negatives diminishes the test's efficacy since it does not isolate the infected individuals. Thus, tests are designed with a focus on minimizing the false-negative error. However, in pooled testing strategies, the sensitivity of the test may depend on different factors such as the number of infected individuals in the pool [1], viral loading in a sample collected from an individual, and the number of tests that can be performed using one sample [50]. Motivated by these considerations, we assume that the false-positive error rate is 0 and use the following approximation for the success rate for when ℓ individuals are pooled together to be tested, and the pool consists of i number of infected individuals [1]:

$$\mathsf{P}(i, \ell) \triangleq 1 - \alpha^{\frac{i}{\ell\gamma}} , \tag{30}$$

where α and γ are context-specific parameters. For instance, for HIV testing we have $\alpha = 0.00033$ and $\gamma = 0.179$ [1]. Clearly, a test's error rate increases with an increase in ℓ and decreases in the number of infected individuals i. Since collecting samples is straightforward and less resource-intensive for mass testing in various contexts [7], the dilution effect is compensated by collecting enough samples that allow repeated testing of the pooled individuals. Specifically, if p samples are available per individual, the pooled test can be performed p times, increasing the overall success rate to

$$\mathsf{P}_p(i, \ell) \triangleq 1 - \alpha^{\frac{pi}{\ell\gamma}} \geq \mathsf{P}_p(\ell) \triangleq 1 - \alpha^{\frac{p}{\ell\gamma}} . \tag{31}$$

Clearly, p controls the likelihood of noiseless setting being observed in practice, and more samples per individual increases this likelihood. In our experiments, we evaluate the testing performance for different p.

2) Distribution of Defectives: We considered the scenarios of *independent* defectives and *graph-constrained* defectives. We randomly selected a set of k vertices, denoted by $\mathcal{H}$, from which a set $\mathcal{J} \subseteq \mathcal{H}$ is infected (or turn defective). In the case of independent defectives, we assumed that the vertices in $\mathcal{H}$ are selected independently at random from the set V. For graph-constrained defectives, we assumed that the vertices in $\mathcal{H}$ form a connected subgraph. The probability associated with a set of vertices $\mathcal{H}$ being exposed to the infection is given by

$$\mathbb{P}(\mathcal{H}) = \frac{1}{Z_k} \sum_{(u,v)\in E, u\in\mathcal{H}, v\in\mathcal{H}} w_{uv} , \tag{32}$$

where Z_k is the normalizing factor such that $\mathbb{P}(\mathcal{H})$ is a valid probability measure. Each vertex in $\mathcal{H}$ can be defective with probability 0.9, which determines $\mathcal{J}$. In the context of infectious disease, we are more likely to encounter graph-constrained defectives. For comparison, we evaluated the community-aware group testing framework over both cases of distributions of defectives in the graph.

3) Community-Aware Group Testing: To evaluate the efficiency in a population with community structures, we consider a population of $n = 10000$ individuals scatterred in $M = 5$ communitys and set $\nu = d = 400$. We assume that there were at most $k = 10$ infected individuals. The intra-community edge weight s_m in community m is selected randomly from $[0.67, 1]$ and we set $q = 1/150$. We evaluate the performance for graph-consistent infections as well as independent infections and depict the respective performances of the group testing framework in Figs. 6 and 7. For our experiments, we set the pooling capacity to $\ell = 37$, which is verified to achieve maximum success rate in the construction of selector matrices by graph-constrained pooling on individual communities. The number of pooled tests is set to be 1200 for each community, beyond which we did not

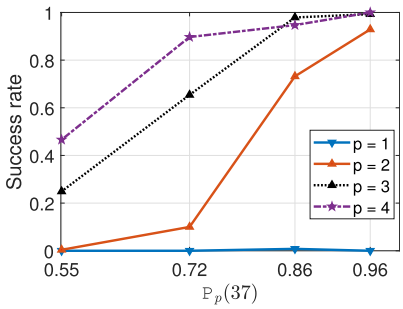

Fig. 6. Likelihood of successfully isolating of infected vertices versus sensitivity of the test for different number of samples per vertex. The set of infected vertices are selected randomly according to (32).

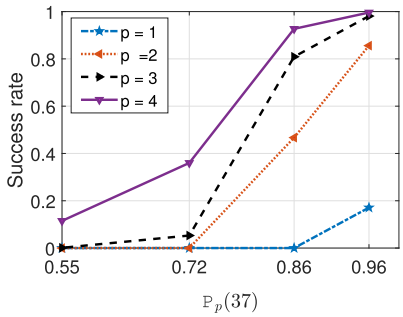

Fig. 7. Likelihood of successfully isolating of infected vertices versus sensitivity of the test for different number of samples per vertex. The infected vertices are selected randomly and independently.

observe any noticeable improvement in the rate of forming a selector matrix. In our experiments, we observe that collecting up to 3 samples per individual resulted in a very high success rate when the test had high sensitivity (greater than 0.85) for both scenarios of the distribution of defective vertices in the graph. Furthermore, we observe that the group testing framework has a higher success rate for scenarios with graph-consistent defectives as compared to that with independent defectives except for the case when $p = 1$.

VI. Conclusions

In this paper, we have considered the problem of adaptive group testing where pooling and testing the groups are constrained with pre-specified restrictions. A graphical model has been adopted to represent the constraints. We have proposed three group testing frameworks. First, a two-stage group testing framework has been proposed for the perfect recovery of the defective items. This approach has been subsequently extended to a three-stage group-testing framework for graphical models that are community-structured. This setting is motivated by settings where the graphs may be loosely connected. Finally, when the inhibitors can exist in the graph, the two-stage group testing framework has been extended to a four-stage group testing framework, where the additional stages are responsible for identifying and filtering out the potential inhibitors. We have characterized sufficient conditions for all three frameworks that specify the number of tests sufficient for ensuring a perfect recovery of the defectives. The graph-constrained group testing framework has also been evaluated in a case study on infection spread in a population with underlying communities.

Appendix A
Proof of Validity of Algorithm 2

In this section, we provide the proof that Algorithm 2 outputs a set of $2k-1$ items that form a superset of $\mathcal{K}$. The proof follows the same arguments as those in the proof of [13, Theorem 3]. Let there be $k' \leq k$ defectives in n items, such that $|\mathcal{K}| = k'$. In the noiseless setting, the outcome vector $\mathbf{y}$ is a Boolean sum of the k' columns associated with the defective set $\mathcal{K}$. Therefore, we claim that besides the columns associated with items in $\mathcal{K}$, there are at most $2k-k'-1$ columns completely covered by $\mathbf{y}$. This can be proved by contradiction. To begin, let there be more than $2k-k'-1$ columns besides those associated with $\mathcal{K}$ that are completely covered by $\mathbf{y}$. Next, consider a submatrix of $\mathbf{X}(T)$ that consists of k' columns in $\mathcal{K}$ and any random set of $2k-k'$ additional columns that are covered by $\mathbf{y}$. We denote this set of $2k-k'$ columns by $\mathcal{Z} \triangleq \{z_1, \ldots, z_{2k-k'}\}$, where $z_i \in V$ and $\mathcal{Z} \cap \mathcal{K} = \varphi$. By definition of a $(2k, k+1, n)$-selector matrix, this submatrix consists of at least $k+1$ rows of identity matrix I_{2k}. Therefore, at least one of such $k+1$ rows has an entry 1 in a column in the set $\mathcal{Z}$. We denote the index of this row by x. Since the entries in columns associated with $\mathcal{K}$ all have entry 0 in row x, this implies that the corresponding entry in $\mathbf{y}$ is 0. However, this contradicts the assumption that $\mathbf{y}$ covers all columns in the set $\mathcal{Z}$ for $\mathbf{X}(T)$. Thus, we conclude that by discarding all columns not covered by $\mathbf{y}$, Algorithm 2 isolates $2k-1$ columns, k' of which correspond to the k' defectives.

Appendix B
Proof of Theorem 1

The number of tests T_{total} hinges on characterizing the number of pooled tests in the filtration stage, i.e. T_{F}, for the construction of a $(2k, k+1, n)$-selector matrix. Therefore, in this proof, our focus is on the design of $(2k, k+1, n)$-selector matrix. From the definition of selector matrix in Definition 11, note that a selector matrix is characterized by the presence of rows of an identity matrix in any submatrix formed by selecting a certain number of columns. Therefore, we start the proof by characterizing the probability that a random walk visits one vertex but not any other in a given set of vertices in Lemma 5. Subsequently, we will leverage Lemma 5 to characterize the total probability of the events that violate the construction of a $(2k, k+1, n)$-selector matrix.

To begin the proof, we consider a random walk of length ℓ such that the sequence of vertices visited by the random walk is given by $\mathbf{W} = [v_0, \ldots, v_\ell]$ where $v_0 \in V$ is a randomly initialized state of the random walk and v_i is the vertex visited by the random walk in i^{th} time step for $v_i \in V, \forall i \in \{1, \ldots, \ell\}$. We define χ_u as the probability that the random walk visits vertex u, such that

$$\chi_u \triangleq \mathbb{P}(u \in \{v_0, \ldots, v_\ell\} \text{ where } \mathbf{W} = [v_0, \ldots, v_\ell]) \,. \quad (33)$$

We also define χ_u^Q as the probability that the random walk visits a vertex u but not any vertex in a given set $Q \subset V$, such that,

$|Q| \leq k$. Therefore,

$$\chi_u^Q \triangleq \mathbb{P}(u \in \{v_0, \ldots, v_\ell\}, \{v_0, \ldots, v_\ell\} \cap Q = \varnothing, \text{ where } \mathbf{W} = [v_0, \ldots, v_\ell]) \ . \tag{34}$$

For brevity, we denote the event $\{\{v_0, \ldots, v_\ell\} \cap Q = \varnothing\}$ by $\mathcal{D}$ and its complement by $\bar{\mathcal{D}}$. Then, based on the analysis of a random walk with random initialization, we formalize the property of χ_u^Q for any choice of u and Q in Lemma 5.

Lemma 5: For a node $u \in V$ and a set of vertices $Q \subset V$ such that $u \notin Q$, we have

$$\chi_u^Q = \Omega\left(\frac{\ell \min_v \pi_v}{\varpi}\right) \ , \tag{35}$$

for the choice of pooling capacity ℓ and the regime for k specified in Theorem 1.

Proof: See Appendix D. □

Next, to find the number of tests T_{F} in Theorem 1, we characterize the events when the matrix $\mathbf{X}(T_{\mathsf{F}})$ is not $(\alpha k, k+1, n)$-selector. Specifically, note that an $(\alpha k, k+1, n)$-selector matrix is equivalent to the following definition from [29, Theorem 5] which states that: For any two disjoint set of vertices $F_1 \subset V$ and $F_2 \subset V$, such that, $|F_1| = k$ and $|F_2| = (\alpha - 1)k$, there always exists at least one row for which $[\mathbf{X}(T_{\mathsf{F}})]_{iu} = 1$ for some column $u \in F_2$ in the test matrix $\mathbf{X}(T_{\mathsf{F}})$ and $[\mathbf{X}(T_{\mathsf{F}})]_{ij} = 0, \forall j \in F_1 \cup F_2 \backslash \{u\}$. In this context, using a union bound and Lemma 5, the probability that $\mathbf{X}(T_{\mathsf{F}})$ is not $(\alpha k, k+1, n)$-selector can be upper bounded by

$$\begin{aligned}
&\mathbb{P}(\mathbf{X}(T_{\mathsf{F}}) \text{ is not } (\alpha k, k+1, n)\text{-selector}) \\
&\leq \binom{n-1}{\alpha k - 1}\binom{\alpha k}{k+1}\left(1 - \frac{((\alpha-1)k)\ell\pi_{\mathsf{max}}}{\varpi}\right)^{T_{\mathsf{F}}} \qquad (36) \\
&\leq \exp\Bigg((\alpha k - 1)\log\frac{n-1}{(\alpha k - 1)} + (k+1)\log\alpha \\
&\quad + T_{\mathsf{F}}\log\left(1 - \frac{(\alpha-1)k\ell\pi_{\mathsf{max}}}{\varpi}\right) + (\alpha+1)k\Bigg) \ , \qquad (37)
\end{aligned}$$

where (37) follows from (36) by using the upper bound $\log\binom{n_1}{n_2} \leq n_2 \log\frac{n_1}{n_2} + n_2$ for any pair of positive integers $n_1 \geq n_2$. Then, the upper bound in (36) is strictly less than $\epsilon > 0$ if we have

$$T_{\mathsf{F}} \geq \frac{(\alpha k - 1)\log\frac{n-1}{\alpha k - 1} + (k+1)\log\alpha + (\alpha+1)k - \log\epsilon}{-\log\left(1 - \frac{((\alpha-1)k)\ell\pi_{\mathsf{max}}}{\varpi}\right)} \tag{38}$$

which is further simplified by using Taylor series expansion of $\log(1-x)$ to the condition

$$T_{\mathsf{F}} \geq \frac{\varpi}{((\alpha-1)k)\ell\pi_{\mathsf{max}}}\left((\alpha k - 1)\log\frac{n-1}{\alpha k - 1} + (k+1)\log\alpha + (\alpha+1)k - \log\epsilon\right). \tag{39}$$

Since we have $\ell = O(\frac{1}{k\pi_{\mathsf{max}}\pi_{\mathsf{ratio}}\varpi})$, the condition on the number of tests T can be further simplified to

$$T_{\mathsf{F}} = \Theta\left(\frac{\varpi^2\pi_{\mathsf{ratio}}}{\alpha - 1}\left((\alpha k - 1)\log\frac{n-1}{\alpha k - 1} + (k+1)\log\alpha + (\alpha+1)k - \log\epsilon\right)\right), \tag{40}$$

which completes the proof for $\alpha = 2$.

APPENDIX C
AUXILIARY RESULTS

In this section, we provide two auxiliary results which are instrumental to the proof of Lemma 5 in Appendix B. We first establish the following lemma which characterizes the probability that a random walk of length ℓ visits a given node more than z times.

Lemma 6: For any instance of random walk of length ℓ, there exists $z = \Theta(\pi_{\mathsf{ratio}}\varpi)$, such that, the probability that the random walk visits a vertex u more than z times is at most $\chi_u/4$.

Proof: Let A_i be a Bernoulli random variable that takes the value 1 if the random walk visits vertex u at step $i \geq 0$ and 0, otherwise. Then, for any step $i \geq \varpi$, we have

$$\mathbb{E}[A_i] \leq \max_{v \in V}\pi_v + \beta \ , \tag{41}$$

where (41) follows from the definition of β-mixing time in Definition 10. By leveraging $\beta = \pi_{\mathsf{max}}/n$, we simplify (41) to $\mathbb{E}[A_i] \leq 2\max_{v\in V}\pi_v$. The number of times the random walk visits vertex u between times t_1 and t_2 is given by $A_{t_1}^{t_2} \triangleq \sum_{i=t_1}^{t_2} A_i$ and we have

$$\mathbb{E}[A_\varpi^\ell] \leq 2\ell \max_{v\in V}\pi_v \ . \tag{42}$$

Therefore, by Markov's inequality, we have

$$\mathbb{P}(A_\varpi^\ell \geq \zeta) < \frac{2\ell\max_{v\in V}\pi_v}{\zeta} \ . \tag{43}$$

By setting $\zeta = \Omega(\pi_{\mathsf{ratio}}\varpi)$, and by leveraging (50), we get $\mathbb{P}(A_\varpi^\ell \geq \zeta) < \frac{\chi_u}{4}$. Next, by setting $z = \varpi + \zeta$, we get

$$\mathbb{P}(A_0^\ell \geq z) \leq \mathbb{P}(A_\varpi^\ell \geq z - \varpi) < \frac{\chi_u}{4} \ . \tag{44}$$

□

Next, we provide Lemma 7 which characterizes the distribution of an event at current step of the random walk in a probability space conditioned on any event that is more than ϖ steps away from the current state of the random walk.

Lemma 7: For any i and j, such that, $j \geq i + \varpi$, we have

$$\mathbb{P}(v_i = u \,|\, v_j = v, \mathcal{R}_i) - \mathbb{P}(v_i = u \mid v_j = v)| \leq \frac{2\beta}{\min_v \pi_v - \beta} \ ,$$

where $\mathcal{R}_i$ is any event that depends only on the states of the random walk up to time i given by $[v_0, \ldots, v_i]$.

The proof of follows from (45) and [20, Proposition 23].

APPENDIX D
PROOF OF LEMMA 5

We start by characterizing the probability χ_u for any vertex $u \in V$. Recall that the position distribution, i.e., the distribution of the most recent vertex visited by the random walk at step ℓ is given by $\boldsymbol{\pi}_{v_0}^\ell$. We denote the probability that the random walk visited u at step ℓ by $\pi_{v_0}^\ell(u)$. Then, when we have $\ell \geq \varpi$, it follows from the definition of β-mixing time in Definition 10 that

$$\|\boldsymbol{\pi}_{v_0}^\ell - \boldsymbol{\pi}\|_\infty \leq \beta \ . \tag{45}$$

Next, for a random walk of length ℓ, we consider the vertices visited by the random walk at the time steps in the set $\mathbf{W}_{\mathsf{r}} \triangleq \{0, \varpi, 2\varpi, \ldots, \ell\}$, such that, $|\mathbf{W}_{\mathsf{r}}| = \lceil \ell/\varpi \rceil$. Let the sequence of vertices visited by the random walk at the time steps in $\mathbf{W}_{\mathsf{r}}$ be given by $V_{\mathbf{W}_{\mathsf{r}}}$. Therefore, probability that the vertex u is

visited by the random walk of length ℓ is lower bounded by the probability that the random walk visited u at time steps in $\mathbf{W}_{\mathsf{r}}$, i.e., $\chi_u \geq \mathbb{P}(u \in V_{\mathbf{W}_{\mathsf{r}}})$. Note that the distribution of the vertices visited by the random walk after ϖ steps is β-close to the stationary distribution $\boldsymbol{\pi}$. Therefore, using (45) and union bound, we have

$$\mathbb{P}(u \notin V_{\mathbf{W}_{\mathsf{r}}}) \leq \left(1 - \min_{v \in V} \pi_v + \beta\right)^{\ell/\varpi} , \tag{46}$$

For any $\beta \leq \frac{1}{2} \min_{v \in V} \pi_v$, we have

$$\mathbb{P}(u \notin V_{\mathbf{W}_{\mathsf{r}}}) \leq \left(1 - \frac{1}{2} \min_{v \in V} \pi_v\right)^{\ell/\varpi} , \tag{47}$$

$$\leq \exp\left(-\frac{\ell \min_{v \in V} \pi_v}{2\varpi}\right) , \tag{48}$$

$$\leq 1 - \Omega\left(\frac{\ell \min_{v \in V} \pi_v}{\varpi}\right) . \tag{49}$$

From (49), we directly obtain

$$\chi_u = \Omega\left(\frac{\ell \min_{v \in V} \pi_v}{\varpi}\right) . \tag{50}$$

The rest of the proof focuses on characterizing χ_u^Q, for which we adopt the high level ideas similar to that in [20] and leverage auxiliary results on different events associated with the random walk in Appendix C. We note that

$$\chi_u^Q \geq \mathbb{P}(\mathcal{D}, u \in \mathbf{W}, \mathcal{Y}_u^z) , \tag{51}$$

where $\mathcal{Y}_u^z$ represents the event that u is not visited by the random walk in first 2ϖ steps and not more than z times overall in ℓ steps. Furthermore, from (51), we have

$$\chi_u^Q \geq \mathbb{P}(\mathcal{D}, u \in \mathbf{W}, \mathcal{Y}_u^z) , \tag{52}$$

$$= \mathbb{P}(u \in \mathbf{W}, \mathcal{Y}_u^z)(1 - \mathbb{P}(\bar{\mathcal{D}} | u \in \mathbf{W}, \mathcal{Y}_u^z)) . \tag{53}$$

Note that

$$\mathbb{P}(u \in \mathbf{W}, \mathcal{Y}_u^z) \geq \mathbb{P}(u \in \mathbf{W}) + \mathbb{P}(\mathcal{Y}_u^z) - \mathbb{P}(u \in \mathbf{W} \cup \mathcal{Y}_u^z) , \tag{54}$$

$$= \frac{\chi_u}{2} . \tag{55}$$

where (55) follows from (54) using (33) and Lemma 6. Therefore, by showing that $\mathbb{P}(\bar{\mathcal{D}} | u \in \mathbf{W}, \mathcal{Y}_u^z)$ is upper bounded by a constant, the proof of claim (35) in Lemma 5 is complete. In this context, when conditioned on $\mathcal{Y}_u^z$, the event $u \in \mathbf{W}$ is a union of at most k disjoint events of the form $\{v_i = v\}$ for $i > 2\varpi$. Therefore, using [20, Proposition 18], we have

$$\mathbb{P}(\bar{\mathcal{D}} | u \in \mathbf{W}, \mathcal{Y}_u^z) \leq k \max_{i \in [2\varpi+1, \ell]} \mathbb{P}(\bar{\mathcal{D}} |\, v_i = u, \mathcal{Y}_u^z) .$$

To upper bound $\mathbb{P}(\bar{\mathcal{D}} |\, v_i = u, \mathcal{Y}_u^z)$ in (56), we will leverage the auxiliary results in Appendix C regarding the behavior of random walk and provide a general upper bound for any $i \in \{2\varpi + 1, \ldots, \ell\}$. For this purpose, we consider a random walk which is initialized as $v_0 = u_0$ such that $u_0 \notin Q$ and where the vertex $v_i = u$ for some $i > 2\varpi$. Then, we select the time points $t_1 \triangleq \varpi$, $t_2 \triangleq i - \varpi$ and $t_3 \triangleq i + \varpi$, such that, $t_1 < t_2 < t_3 < \ell$. Next, we divide sequence of vertices $\mathbf{W}$ visited by the random walk into 4 parts: $\mathbf{W}_0^{t_1}, \mathbf{W}_{t_1+1}^{t_2}, \mathbf{W}_{t_2+1}^{t_3}$, and $\mathbf{W}_{t_3+1}^{\ell}$, where $\mathbf{W}_j^i$ denotes the set of vertices visited by the random walk between steps j and i, for $j < i$. Using union bound, we have

$$\mathbb{P}(\bar{\mathcal{D}} | v_i = u, v_0 = u_0) \leq \omega_1 + \omega_2 + \omega_3 + \omega_4 , \tag{56}$$

where

$$\omega_1 \triangleq \mathbb{P}(\mathbf{W}_0^{t_1} \cap Q \neq \varnothing | v_i = u, v_0 = u_0) , \tag{57}$$

$$\omega_2 \triangleq \mathbb{P}(\mathbf{W}_{t_1+1}^{t_2} \cap Q \neq \varnothing | v_i = u, v_0 = u_0) , \tag{58}$$

$$\omega_3 \triangleq \mathbb{P}(\mathbf{W}_{t_2+1}^{t_3} \cap Q \neq \varnothing | v_i = u, v_0 = u_0) , \tag{59}$$

$$\omega_4 \triangleq \mathbb{P}(\mathbf{W}_{t_3+1}^{\ell} \cap Q \neq \varnothing | v_i = u, v_0 = u_0) . \tag{60}$$

Note that the section of the random walk that spans $\mathbf{W}_{t_3+1}^{\ell}$ is sufficiently far, i.e., greater than ϖ steps from time steps i and 0. Therefore, by using (45) and union bound for the number of steps in the random walk and the number of possible vertices that can lie in Q, we have

$$\omega_4 \leq \ell k (\max_{v \in V} \pi_v + \beta) . \tag{61}$$

Next, note that without the conditioning on the event $v_i = u$, the probability of the event $W_{t_2+1} \cap Q \neq \varnothing$ conforms to the statistics of the distribution which is β-close to the stationary distribution of random walk over G. By using Lemma 7, we note that conditioning on the event $v_i = u$ shifts the distribution of $W_{t_2+1} \cap Q \neq \varnothing$ by a factor of at most $\ell k \beta'$. Therefore, by using (45) and (45) in Lemma 7, we get

$$\omega_2 \leq \ell k (\max_{v} \pi_v + \beta + \beta') , \tag{62}$$

where $\beta' \triangleq \frac{2\beta}{\max_v \pi_v - \beta}$ is the upper bound in (45). To analyze ω_3, we note that

$$\mathbb{P}(v_{i+1} \in Q | v_i = u, v_0 = u_0) \leq \phi_{\mathsf{max}}^k , \tag{63}$$

where ϕ_{max}^k is the sum of the largest k transition probabilities at any vertex in G. Equation (63) follows by noting that there are at most k vertices that lie in the set Q in the neighborhoods of u. Furthermore, $\mathbb{P}(v_{i+1} \in Q | v_i = u, v_0 = u_0) = \mathbb{P}(v_{i+1} \in Q | v_i = u)$ because of the Markov property of random walk. Similarly, for any time $j \in \{i+1, \ldots, \ell\}$, we have

$$\mathbb{P}(v_j \in Q | v_i = u) \leq \phi_{\mathsf{max}}^k . \tag{64}$$

Note that for the (ν, d)-regular graph with weights in $[w, 1]$, a vertex can have a maximum degree of wd, and therefore, a maximum of wd vertices in its immediate neighborhood. Similarly, if the vertex has the minimum degree of ν, it can have at least ν vertices and at most ν/w vertices in its immediate neighborhood. Therefore, by using these observations and leveraging (5), we note that

$$\phi_{\mathsf{max}}^k \leq \frac{k}{w\nu} . \tag{65}$$

For any time $j \in \{t_2 + 1, \ldots, t_3\}$, using (45) in Lemma 7, we have

$$\mathbb{P}(v_j \in Q | v_i = u, v_0 = u_0) \leq \phi_{\mathsf{max}}^k + \beta' . \tag{66}$$

Therefore, by union bound over the number of time steps and using (65), we get

$$\omega_3 \leq 2\frac{\varpi k}{w\nu} + 2\varpi\beta' . \tag{67}$$

The probability ω_1 is analyzed in a similar fashion as ω_3. Note that the Markov chain over the vertices is time reversible

and conditioning on v_0 changes the probability by at most β'. Therefore, we have

$$\omega_1 \leq \frac{\varpi k}{w\nu} + \varpi\beta' \tag{68}$$

By combining (61), (62), (67), and (68), we get

$$\mathbb{P}(\bar{\mathcal{D}}|v_i = u, v_0 = u_0) \leq 3\frac{\varpi k}{w\nu} + 3\beta'(\varpi + \ell k) + 2\ell k \max_v \pi_v. \tag{69}$$

Next, to relax the condition on the initialization of the random walk, we note that

$$\mathbb{P}(\bar{\mathcal{D}}|v_i = u) \leq \mathbb{P}(v_0 \in Q|v_i = u) + \mathbb{P}(\bar{\mathcal{D}}|v_i = u, v_0 = u_0)\ . \tag{70}$$

Note that the probability $\mathbb{P}(v_0 \in Q|v_i = u) \leq \mathbb{P}(v_0 \in Q) + \beta'$ from Lemma 7 and $\mathbb{P}(v_0 \in Q) \leq k \max_v \pi_v$. Therefore, from (69) and (70), we have

$$\mathbb{P}(\bar{\mathcal{D}}|v_i = u) \leq 3\frac{\varpi k}{w\nu} + 3\beta'(\varpi + \ell k) + 3\ell k \max_v \pi_v\ . \tag{71}$$

Next, we note that to evaluate the probability $\mathbb{P}(\bar{\mathcal{D}}|v_i = u, \mathcal{Y}_u^z)$, we shrink the probability space to that conditioned on $\mathcal{Y}_u^z$. Therefore, $\mathbb{P}(\bar{\mathcal{D}}|v_i = u, \mathcal{Y}_u^z)$ is not significantly larger (say by no more than a factor of $1 + \epsilon$ with $\epsilon \ll 1$) than $\mathbb{P}(\bar{\mathcal{D}}|v_i = u)$ if the event that the random walk does not visit a vertex u more than z times is close to 1. By taking a union bound on the number of times a random walk visits an vertex u, we get

$$\mathbb{P}(\bar{\mathcal{D}}|v_i = u, \mathcal{Y}_z^u) \leq (1+\epsilon)z \times \left(3\frac{\varpi k}{w\nu} + 3\beta'(\varpi + \ell k) + 3\ell k \pi_{\mathsf{max}}\right)\ , \tag{72}$$

where by noting that $z = \Theta(\pi_{\mathsf{ratio}}\varpi)$ from Lemma 6, we can bound (72) by a constant if we have

$$\nu w = \Omega(\varpi^2 k \pi_{\mathsf{ratio}})\ , \qquad \ell = O\left(\frac{1}{k\pi_{\mathsf{max}}\pi_{\mathsf{ratio}}\varpi}\right)\ , \tag{73}$$

and β in Theorem 1 for sufficiently large n. By leveraging (72) under the conditions in (73) and (53), the proof of Lemma 5 is complete.

REFERENCES

[1] N. T. Nguyen, H. Aprahamian, E. K. Bish, and D. R. Bish, "A methodology for deriving the sensitivity of pooled testing, based on viral load progression and pooling dilution," *J. Transl. Med.*, vol. 17, no. 1, pp. 1–10, Aug. 2019, Art. no. 252.

[2] M. T. Goodrich and D. S. Hirschberg, "Efficient parallel algorithms for dead sensor diagnosis and multiple access channels," in *Proc. ACM Symp. Parallelism Algorithms Architectures*, 2006, pp. 118–127.

[3] A. Cohen, A. Cohen, and O. Gurewitz, "Secure group testing," in *Proc. IEEE Int. Symp. Inf. Theory*, 2016, pp. 1391–1395.

[4] N. J. Harvey, M. Patrascu, Y. Wen, S. Yekhanin, and V. W. Chan, "Non-adaptive fault diagnosis for all-optical networks via combinatorial group testing on graphs," in *Proc. IEEE Int. Conf. Comput. Commun.*, Barcelona, Spain, 2007, pp. 697–705.

[5] A. B. Kahng and S. Reda, "New and improved BIST diagnosis methods from combinatorial group testing theory," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 25, no. 3, pp. 533–543, Mar. 2006.

[6] S. P. Prabhu, "Techniques for enhancing test and diagnosis of digital circuits," Ph.D. dissertation, Dept. Elect. Comput. Eng., Virginia Tech, Blackburg, VA, USA, 2015.

[7] C. Gollier and O. Gossner, "Group testing against COVID-19," *Covid Econ.*, vol. 1, no. 2, pp. 32–42, Apr. 2020.

[8] R. Goenka, S.-J. Cao, C.-W. Wong, A. Rajwade, and D. Baron, "Contact tracing enhances the efficiency of COVID-19 group testing," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, 2021, pp. 8168–8172.

[9] A. Heidarzadeh and K. Narayanan, "Two-stage adaptive pooling with RT-qPCR for COVID-19 screening," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, 2021, pp. 8148–8152.

[10] M. Aldridge, O. Johnson, and J. Scarlett, "Group testing: An information theory perspective," *Found. Trends Commun. Inf. Theory*, vol. 15, no. 3–4, pp. 196–392, Feb. 2019.

[11] G. K. Atia and V. Saligrama, "Boolean compressed sensing and noisy group testing," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1880–1901, Mar. 2012.

[12] D. Du and F. Hwang, *Pooling Designs and Nonadaptive Group Testing: Important Tools for DNA Sequencing*. Singapore: World Scientific, 2006.

[13] A. De Bonis, L. Gasieniec, and U. Vaccaro, "Optimal two-stage algorithms for group testing problems," *SIAM J. Comput.*, vol. 34, no. 5, pp. 1253–1270, 2005.

[14] A. Tajer, R. Castro, and X. Wang, "Adaptive sensing of congested spectrum bands," *IEEE Trans. Inf. Theory*, vol. 58, no. 9, pp. 6110–6125, Sep. 2012.

[15] A. Sharma and C. R. Murthy, "Group testing-based spectrum hole search for cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 63, no. 8, pp. 3794–3805, Feb. 2014.

[16] N. Michelusi and U. Mitra, "Cross-layer estimation and control for cognitive radio: Exploiting sparse network dynamics," *IEEE Trans. Cogn. Commun. Netw.*, vol. 1, no. 1, pp. 128–145, Mar. 2015.

[17] A. Tajer, V. V. Veeravalli, and H. V. Poor, "Outlying sequence detection in large datasets: A data-driven approach," *IEEE Signal Process. Mag.*, vol. 31, no. 5, pp. 44–56, Sep. 2014.

[18] W. Xu, E. Mallada, and A. Tang, "Compressive sensing over graphs," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2011, pp. 2087–2095.

[19] A. Deckert, T. Bärnighausena, and N. N. Kyeia, "Simulation of pooled-sample analysis strategies for COVID-19 mass testing," *Bull. World Health Org.*, vol. 98, pp. 590–598, Jul. 2020.

[20] M. Cheraghchi, A. Karbasi, S. Mohajer, and V. Saligrama, "Graph-constrained group testing," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 248–262, Jan. 2012.

[21] M. Aldridge, L. Baldassini, and O. Johnson, "Group testing algorithms: Bounds and simulations," *IEEE Trans. Inf. Theory*, vol. 60, no. 6, pp. 3671–3687, Mar. 2014.

[22] D. Du, F. K. Hwang, and F. Hwang, *Combinatorial Group Testing and Its Applications*, 2nd ed. Singapore: World Scientific, 2000.

[23] E. Porat and A. Rothschild, "Explicit nonadaptive combinatorial group testing schemes," *IEEE Trans. Inf. Theory*, vol. 57, no. 12, pp. 7982–7989, Aug. 2011.

[24] C. L. Chan, S. Jaggi, V. Saligrama, and S. Agnihotri, "Non-adaptive group testing: Explicit bounds and novel algorithms," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 3019–3035, May 2014.

[25] J. M. Hughes-Oliver and W. H. Swallow, "A two-stage adaptive group-testing procedure for estimating small proportions," *J. Amer. Statist. Assoc.*, vol. 89, no. 427, pp. 982–993, Sep. 1994.

[26] A. Cohen, A. Cohen, S. Jaggi, and O. Gurewitz, "Secure adaptive group testing," in *Proc. IEEE Int. Symp. Inf. Theory*, 2018, pp. 2589–2593.

[27] P. Damaschke, "Adaptive group testing with a constrained number of positive responses improved," *Discrete Appl. Math.*, vol. 205, pp. 208–212, May 2016.

[28] M. Mézard and C. Toninelli, "Group testing with random pools: Optimal two-stage algorithms," *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1736–1745, Mar. 2011.

[29] H.-L. Fu, "Group testing with multiple mutually-obscuring positives," *Lecture Notes Comput. Sci.*, vol. 7777, pp. 557–568, Jan. 2013.

[30] P. Damaschke and A. S. Muhammad, "Randomized group testing both query-optimal and minimal adaptive," in *Proc. Int. Conf. Curr. Trends Theory Pract. Comput. Sci.*, Spindleruv Mlyn, Czech Republic, 2012, pp. 214–225.

[31] J. Scarlett, "An efficient algorithm for capacity-approaching noisy adaptive group testing," in *Proc. IEEE Int. Symp. Inf. Theory*, Paris, France, 2019, pp. 2679–2683.

[32] S. Luo, Y. Matsuura, Y. Miao, and M. Shigeno, "Non-adaptive group testing on graphs with connectivity," *J. Combinatorial Optim.*, vol. 38, no. 1, pp. 278–291, Jan. 2019.

[33] B. Spang and M. Wootters, "Unconstraining graph-constrained group testing," in *Proc. Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*, Cambridge, MA, USA, vol. 145, Sep. 2019, pp. 46:1–46:20.

[34] A. Karbasi and M. Zadimoghaddam, "Sequential group testing with graph constraints," in *Proc. IEEE Inf. Theory Workshop*, Lausanne, Switzerland, 2012, pp. 292–296.
[35] P. Nikolopoulos, T. Guo, C. Fragouli, and S. Diggavi, "Community aware group testing," 2020, *arXiv:2007.08111*.
[36] R. Castro, M. Coates, G. Liang, R. Nowak, and B. Yu, "Network tomography: Recent developments," *Stat. Sci.*, vol. 19, no. 3, pp. 499–517, 2004.
[37] M. Lotfi Shahreza, N. Ghadiri, S. R. Mousavi, J. Varshosaz, and J. R. Green, "A review of network-based approaches to drug repositioning," *Brief. Bioinf.*, vol. 19, no. 5, pp. 878–892, 2018.
[38] K. Li, D. Precup, and T. J. Perkins, "Pooled screening for synergistic interactions subject to blocking and noise," *PLoS One*, vol. 9, no. 1, 2014, Art. no. e85864.
[39] M. Hann, B. Hudson, X. Lewell, R. Lifely, L. Miller, and N. Ramsden, "Strategic pooling of compounds for high-throughput screening," *J. Chem. Inf. Comput. Sci.*, vol. 39, no. 5, pp. 897–902, 1999.
[40] M. Farach, S. Kannan, E. Knill, and S. Muthukrishnan, "Group testing problems with sequences in experimental molecular biology," in *Proc. Compression Complexity Sequences*, Salerno, Italy, 1997, pp. 357–367.
[41] J. M. Hughes-Oliver, "Pooling experiments for blood screening and drug discovery," in *Screening*. Berlin, Germany: Springer, 2006, pp. 48–68.
[42] A. De Bonis, "New combinatorial structures with applications to efficient group testing with inhibitors," *J. Combinatorial Optim.*, vol. 15, no. 1, pp. 77–94, Jan. 2008.
[43] A. Ganesan, S. Jaggi, and V. Saligrama, "Non-adaptive group testing with inhibitors," in *Proc. IEEE Inf. Theory Workshop*, Jeju Island, Korea, 2015, pp. 1–5.
[44] A. Ganesan, S. Jaggi, and V. Saligrama, "Learning immune-defectives graph through group tests," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 3010–3028, May 2017.
[45] D. A. Levin and Y. Peres, *Markov Chains and Mixing Times*, 2nd ed. Providence, RI, USA: American Mathematical Society, 2017, vol. 107.
[46] A. J. Macula, "Error-correcting nonadaptive group testing with de-disjunct matrices," *Discrete Appl. Math.*, vol. 80, no. 2–3, pp. 217–222, Dec. 1997.
[47] B. S. Chlebus and D. R. Kowalski, "Almost optimal explicit selectors," in *Proc. Int. Symp. Fundamentals Comput. Theory. Lubeck*, Germany: Springer, 2005, pp. 270–280.
[48] V. Guruswami, "Rapidly mixing Markov chains: A comparison of techniques," 2016, *arXiv:1603.01512*.
[49] A. R. Molla and G. Pandurangan, "Distributed computation of mixing time," in *Proc. Int. Conf. Dist. Comp. Netw.*, Hyderabad, India, 2017, pp. 1–4.
[50] C. M. Verdun *et al.*, "Group testing for SARS-CoV-2 allows for up to 10-fold efficiency increase across realistic scenarios and testing strategies," *Frontiers in Public Health*, vol. 9, Aug. 2021, Art. no. 1205, doi: 10.3389/fpubh.2021.583377.

Saurabh Sihag (Member, IEEE) received the B.Tech. and M.Tech. degrees in electrical engineering from the Indian Institute of Technology Kharagpur, Kharagpur, India, in 2016, and the Ph.D. degree in electrical engineering from Rensselaer Polytechnic Institute, Troy, NY, USA, in 2020. He was the recipient of the 2021 Charles M. Close '62 Doctoral Prize by the Department of Electrical, Computer and Systems Engineering at Rensselaer Polytechnic Institute. He is currently the Clinical Research in ALS and related disorders for Therapeutic Develoment (CReATe) Consortium Postdoctoral Research Fellow with the University of Pennsylvania, Philadelphia, PA, USA. His research interests include statistical signal processing, information theory, high-dimensional statistics, and network neuroscience.

Ali Tajer (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2002 and 2004, respectively. During 2007 and 2010, he was with Columbia University, New York, NY, USA, where he received the M.A. degree in statistics and the Ph.D. degree in electrical engineering, and during 2010 and 2012, he was a Postdoctoral Research Associate with Princeton University, Princeton, NJ, USA. He is currently an Associate Professor of Electrical, Computer, and Systems Engineering with Rensselaer Polytechnic Institute, Troy, NY, USA. His recent publications include an edited book *Advanced Data Analytics for Power Systems* (Cambridge University Press, 2020). His research interests include mathematical statistics, statistical signal processing, and network information theory, with applications in wireless communications, and power grids. He was the recipient of an NSF CAREER Award in 2016 and AFRL Faculty Fellowship in 2019. He is currently an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He was also the Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS, Guest Editor of the *IEEE Signal Processing Magazine*, Editor of the IEEE TRANSACTIONS ON SMART GRID, Editor of the *IET Transactions on Smart Grid*, and Guest Editor-in-Chief of the IEEE TRANSACTIONS ON SMART GRID – special issue on Theory of Complex Systems with Applications to Smart Grid Operations.

Urbashi Mitra (Fellow, IEEE) received the B.S. and the M.S. degrees from the University of California, Berkeley, CA, USA, and the Ph.D. degree from Princeton University, Princeton, NJ, USA. She is currently the Gordon S. Marshall Professor in Engineering with the University of Southern California, Los Angeles, CA. USA, with appointments in Electrical and Computer Engineering and Computer Science. She was the Inaugural Editor-in-Chief of the IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL AND MULTI-SCALE COMMUNICATIONS. She was a Member of the IEEE Information Theory Society's Board of Governors (2002-2007, 2012-2017), the IEEE Signal Processing Society's Technical Committee on Signal Processing for Communications and Networks (2012-2016), the IEEE Signal Processing Society's awards Board (2017-2018), and the Chair/Vice-Chair of the IEEE Communication Theory Technical Committee (2017-2020). She was the recipient of the 2021 USC Viterbi School of Engineering Senior Research Award, the 2017 IEEE Women in Communications Engineering Technical Achievement Award, a 2015 U.K. Royal Academy of Engineering Distinguished Visiting Professorship, a 2015 U.S. Fulbright Scholar Award, a 2015-2016 U.K. Leverhulme Trust Visiting Professorship, IEEE Communications Society Distinguished Lecturer, 2012 Globecom Signal Processing for Communications Symposium Best Paper Award, 2012 U.S. National Academy of Engineering Lillian Gilbreth Lectureship, the 2009 DCOSS Applications & Systems Best Paper Award, 2001 Okawa Foundation Award, 2000 Ohio State University's College of Engineering Lumley Award for Research, and a 1996 National Science Foundation CAREER Award. She has been an Associate Editor for the following IEEE publications: TRANSACTIONS ON SIGNAL PROCESSING, TRANSACTIONS ON INFORMATION THEORY, JOURNAL OF OCEANIC ENGINEERING, and TRANSACTIONS ON COMMUNICATIONS. Dr. Mitra has held visiting appointments with King's College, London, U.K., Imperial College, London, U.K., the Delft University of Technology, Delft, the Netherlands, Stanford University, Stanford, CA, USA, Rice University, Houston, TX, USA, and the Eurecom Institute, Biot, France. Her research interests include wireless communications, structured statistical methods, communication and sensor networks, biological communication systems, detection and estimation and the interface of communication, sensing and control.