Model-Based and Graph-Based Priors for Group Testing

Ivan Lau, Jonathan Scarlett, Member, IEEE, and Yang Sun

Abstract—The goal of the group testing problem is to identify a set of defective items within a larger set of items, using suitably-designed tests whose outcomes indicate whether any defective item is present. In this paper, we study how the number of tests can be significantly decreased by leveraging the structural dependencies between the items, i.e., by incorporating prior information. To do so, we pursue two different perspectives: (i) As a generalization of the uniform combinatorial prior, we consider the case that the defective set is uniform over a subset of all possible sets of a given size, and study how this impacts the information-theoretic limits on the number of tests for approximate recovery; (ii) As a generalization of the i.i.d. prior, we introduce a new class of priors based on the Ising model, where the associated graph represents interactions between items. We show that this naturally leads to an Integer Quadratic Program decoder, which can be converted to an Integer Linear Program and/or relaxed to a non-integer variant for improved computational complexity, while maintaining strong empirical recovery performance.

Index Terms—Group testing, model-based sparsity, information-theoretic limits, Ising model, linear programming relaxations.

I. INTRODUCTION

The concept of group testing was first introduced by Robert Dorfman [1] in 1943 to provide a more efficient method of screening for syphilis during conscription for World War II. It has since gained significant attention due to its wide-ranging applications in areas such as biology, communications, and information technology, as well as connections with problems in data science and theoretical computer science (see [2, Section 1.7] and the references therein). Recently, it has also found utility in testing for COVID-19 [3].

The general goal of group testing is to identify a set of “defective” items of size $k$ within a larger set of items of size $n$, based on the outcomes of $t$ suitably-designed tests, where each test outcome indicates whether the test includes at least one defective item. The current state-of-the-art group testing schemes can reliably identify the defective set using an asymptotically optimal number of tests [4]. In this paper, we build on a recent line of works demonstrating that we can further decrease the number of tests by leveraging structural dependencies between the items [5], [6], [7], [8], [9], [10], [11]. We study such dependencies from two different perspectives, adopting a natural generalization of the uniform combinatorial prior for the purpose of theoretical analysis, and introducing a new prior based on Ising models as a more practical alternative that generalizes the i.i.d. prior.

Before outlining the related work and our contributions in more detail, we formally introduce the problem.

A. Problem Setup

1) Observation Model: We have a set of $n$ items, $V = [n] := \{1, \ldots, n\}$, and an unknown defective set $S \subseteq V$ of size $k$. The goal is to recover $S$ via a set of tests, where each test contains a subset of $V$, and the test outcome reveals whether any of the items in the test are defective. We use a binary test matrix $X \in \{0, 1\}^{t \times n}$ to represent the design of the pooled tests, with

$$X_{ij} = \begin{cases} 1 & \text{the j-th item is included in the i-th test;} \\ 0 & \text{otherwise,} \end{cases}$$

for $i \in \{1, \ldots, t\}$ and for $j \in V$. We focus on the non-adaptive setting, in which $X$ must be designed prior to observing any outcomes. The test outcomes, denoted by $Y = (Y_1, \ldots, Y_t) \in \{0, 1\}^t$, may follow one of several observation models. In the noiseless group testing model, which is our main focus, the test outcomes are given by

$$Y_i = \bigvee_{j \in S} X_{ij} = \begin{cases} 1 & \text{any defectives in the i-th test;} \\ 0 & \text{otherwise}. \end{cases}$$

In the noisy group testing model, the test outcomes might be “flipped” by some (random) noise vector $\xi \in \{0, 1\}^t$, in which $\xi_i = 1$ if and only if the outcome of the $i$-th test is flipped. In this model, the corresponding test outcomes are defined as

$$Y_i = \begin{cases} 1 & \text{if } \left(\bigvee_{j \in S} X_{ij}\right) \oplus \xi_i = 1; \\ 0 & \text{otherwise,} \end{cases}$$

where $\oplus$ represents the logical XOR operator. We focus primarily on the case that $\xi$ has i.i.d. Bernoulli($\rho$) entries for some $\rho \in (0, \frac{1}{2})$ (i.e., the symmetric noisy model).
In addition to the above specific models, we will sometimes work with a general probabilistic model, writing
\[(Y_i | X_i) \sim P_{Y_i | X_i, S},\] (4)
where \(X_i, S\) is the sub-vector of the \(i\)-th row of \(X\) containing the entries indexed by \(S\). In all models, we assume conditional independence (given \(X\)) among the tests \(i = 1, \ldots, t\).

2) Probabilistic Priors: We consider settings in which the unknown defective set \(S\) has some (known) prior distribution, i.e., a distribution over all subsets of \(V\). In the literature, the most common priors are as follows (e.g., see [2, Page 205]):
- **Combinatorial prior**: The number of defective items \(k\) is fixed, and each of the \(\binom{n}{k}\) subsets of \(V\) of cardinality \(k\) are equally likely, i.e., \(S\) is uniform over such subsets.
- **i.i.d. prior**: Each item is defective with some probability \(q\), with independence between items. Hence, the average number of defectives is \(k = nq\).

Note that \(k\) and \(q\) may scale with \(n\); most works consider a sub-linear number of defectives (e.g., \(k = \sqrt{n}\) or \(q = \frac{1}{\sqrt{n}}\)).

3) Test Designs: As mentioned in Section I-A1, we consider non-adaptive test designs, where the entire test matrix \(X\) must be designed before observing any test outcomes. This often has significant practical benefits due to allowing parallel testing. For the more practically-oriented parts of this paper, our focus will be on introducing a suitable prior and designing a decoder for it, but that decoder will be applicable to arbitrary test designs. That is, the decoder design and the non-adaptive test design will be decoupled.

In contrast, for our theoretical analysis, we will focus on \(i.i.d.\) Bernoulli testing, in which each item is placed in a given test independently with probability \(\frac{k}{n}\) for some constant \(k > 0\). While there is evidence that slightly more structured random designs can do better [4], [12], [13], such improvements have been absent for approximate recovery under sublinear sparsity, which is our focus. Hence, we believe that the i.i.d. design is as suitable as any for our purposes.

4) Recovery Criteria: Given the test design \(X\) and the test outcomes \(Y\), a decoder forms an estimate \(\hat{S} = \hat{S}(n, k, X, Y)\) of \(S\). A variety of recovery criteria exist for measuring the performance. We focus on approximate recovery criteria, allowing a certain number of false positives (i.e., \(|\hat{S} \setminus S|\) and false negatives (i.e., \(|S \setminus \hat{S}|\)). In our theoretical analysis, we will treat these two equally, allowing up to \(d_{\max} = \lfloor \alpha^* k \rfloor\) of each for some \(\alpha^* \in (0, 1)\). That is, the error probability is
\[
P_{d_{\max}}[\text{err}] := \mathbb{P}\left[d(S, \hat{S}) > d_{\max}\right],\] (5)
where
\[d(S, \hat{S}) = \max \left\{ |S \setminus \hat{S}|, |\hat{S} \setminus S| \right\}.\] (6)

We are interested in the required number of tests to attain asymptotically vanishing error probability, i.e., \(\lim_{n \to \infty} P_{d_{\max}}[\text{err}] = 0\). Permitting a small error probability is commonly given the terminology **probabilistic group testing** or the **small-error criterion**.

The approximate recovery criterion is less stringent than the usual exact recovery criterion (which corresponds to setting \(d_{\max} = 0\) in (5)), and both are of practical interest depending on the application. We focus only on the approximate recovery criterion, but studying exact recovery may be of interest in future work. We also note that our goal is to understand **improved decoding algorithms** for exploiting prior information, rather than improved test designs (though both are of interest).

We briefly note that a distinct line of works for the noiseless setting (or non-stochastic noise settings) requires **zero error probability**. This strict requirement incurs \(k^2\) dependence in the number of tests for exact recovery, though approximate recovery is still possible with \(O(k \log n)\) tests [14].

### B. Related Work

In addition to the related works already mentioned above, we highlight the following aspects of group testing that are particularly relevant to our paper.

1) **Information-Theoretic Limits of Group Testing:** Among many existing theoretical works on group testing, the information-theoretic limits in [15] are particularly relevant (see also [16], [17]). In particular, we highlight the following.

**Lemma I.1 ([15]):** Let \(n\) be the number of items, \(k = \Theta(n^\theta)\) for some \(\theta \in (0, 1)\) be the size of defective set, and suppose that \(S\) is uniform over the set of all \(k\)-sized subsets of \(V\). Fix \(a^* \in (0, 1)\), and let \(d_{\max} = \lfloor a^* k \rfloor\). For the noiseless group testing model, under i.i.d. Bernoulli testing with parameter \(\frac{\ln 2}{k}\), there exists a decoder outputting \(\hat{S}\) such that as \(n \to \infty\), provided that
\[
t \geq \left( k \log_2 \frac{n}{k} \right) (1 + o(1)).\] (7)

Conversely, for any non-adaptive test design\(^1\) and any decoder \(\hat{S}\), in order to achieve \(P_{d_{\max}}[\text{err}] \to 0\), it is necessary that
\[
t \geq (1 - a^*) \left( k \log_2 \frac{n}{k} \right) (1 - o(1)).\] (8)

Observe that the upper and lower bounds match as \(a^* \to 0\).

More generally, recent results in [18], [19] demonstrated that (8) is tight for general designs (i.e., there exists a matching achievability result), whereas Bernoulli designs cannot do better than (7), so they are mildly suboptimal in terms of the dependence on \(a^*\).

We refer the reader to [2], [4], [15], [20] and the references therein for other works on the information-theoretic limits of group testing. These works typically adopt an information-theoretic decoder, roughly amounting to a brute force search over the space of defective sets, which is computationally intractable.

A distinct line of works has sought schemes that require not only order-optimal number of tests \(t\), but also computationally feasible decoder, running in time polynomial in \(n\) (e.g., see [2, Chapters 2 and 3] and the references therein), or even polynomial in \(k \log n\) (e.g., see [21, Table 1] and the references therein). Our theoretical focus, however, is on information-theoretic limits.

\(^1\)Note that the dependence on \(a^*\) here only enters via the \(o(1)\) term.

\(^2\)In [15] this was stated specifically for the i.i.d. design, but the general case is given in [17, Thm. 1].
2) Decoding Techniques: In the theoretical part of our paper, we will consider an information-theoretic decoder (similar to [15]) whose definition is quite technical, so its details are deferred to later (see Appendix A-II). In the practically-oriented part of our paper, we will use techniques based on Linear Programming and Quadratic Programming; here we briefly outline some existing works that adopted similar approaches.

We first discuss the noiseless setting. Given the test matrix $X$ and the test outcome vector $Y$, a natural decoding rule is to choose $\hat{S}$ to be the smallest set of items consistent with the test results; this was termed the smallest satisfying set (SSS) algorithm in [22]. Since this poses a potentially hard combinatorial optimization problem, it was proposed in [23] to relax it to a linear program. The resulting optimization problems can be written in a unified manner as follows:

$$\text{minimize}_{u} \sum_{j=1}^{n} u_j \quad \text{subject to} \quad u \in C_{\text{noiseless}}$$

(9)

where

$$C_{\text{noiseless}} = \left\{ u \in \{0, 1\}^n : \sum_{j=1}^{n} X_{ij} u_j = 0 \right\}$$

(10)

and $C_{\text{noiseless}}^{[0,1]}$ is defined analogously with $[0, 1]^n$ in place of $\{0, 1\}^n$. Denoting the resulting estimate by $\hat{u}$ (with suitable rounding to $[0, 1]^n$ in the relaxed case), the defective set is estimated according to $\hat{S} = \{ j \in [n] : \hat{u}_j = 1 \}$.

In the noisy setting, it was proposed in [23] to weigh the sparsity-based objective in (9) with an additional term indicating how many test results are flipped compared to their nominal outcome, leading to the following:

$$\text{minimize}_{u, \xi} \sum_{j=1}^{n} u_j + \eta \sum_{i=1}^{t} \xi_i \quad \text{subject to} \quad (u, \xi) \in C_{\text{noisy}}$$

(11)

where

$$C_{\text{noisy}} = \left\{ (u, \xi) \in \{0, 1\}^n \times \{0, 1\}^t : \sum_{j=1}^{n} X_{ij} u_j = \xi_i \right\}$$

and $C_{\text{noisy}}^{[0,1]}$ is defined analogously with $[0, 1]^n \times [0, 1]^t$ in place of $\{0, 1\}^n \times \{0, 1\}^t$. Here the parameter $\eta$ balances the trade-off between having a sparse solution and having few flipped tests.

It is shown in [24, Sec. 5.3] that the non-relaxed variants of (9) and (11) can be interpreted as the maximum a posteriori decoding rules under an i.i.d. prior, and i.i.d. symmetric noise in the noisy case. Specifically, for this interpretation in the noisy case, one should set $\eta = \log \frac{q}{p}$, where $q$ is the defectivity probability and $p$ is the noise parameter.

3) Group Testing With Prior Information: Here we briefly outline some existing group testing literature where non-uniform models on the defective set have been considered.

A common choice of non-uniform prior is the i.-non-i.d. prior, in which each item is randomly defective with independence across items, but the associated probabilities may differ [26], [27]. This is an important special case of prior knowledge in group testing, but it is limited due to the absence of interactions/correlations between items.

Most related to our theoretical part is the concurrent work of [6], who considered essentially the same model that we will consider: The defective set is known to lie in some set $S$, with each $S \in S$ being a size-$k$ (or size at most $k$) subset of $\{1, \ldots, n\}$. However, apart from the model, their work is very different from ours. In particular, they focus on zero-error exact recovery, which is a much more stringent goal than small-error approximate recovery, and accordingly, considerably more tests are required for non-adaptive strategies. In addition, the focus in [6] is on scaling laws on the number of tests, whereas we are interested in how the underlying constants change even when the scaling laws match.

There are several additional works that are related to the more practical part of our work, and to our paper more generally. We briefly outline them here, and discuss the main differences and advantages/disadvantages in Section III-B.

A recent line of works has studied community-aware group testing [9], [10], [11], in which the items are arranged into known clusters, and those in the a common cluster are (relatively) highly correlated. Various algorithms were proposed based on first identifying highly infected communities, and then using high-prevalence testing strategies (possibly even one-by-one testing) on the individuals therein. This is another important special case of prior information, and is practically well-motivated, e.g., in testing contagious diseases where certain individuals share a house, office, etc., but in this paper we are also interested in broader kinds of structure beyond communities/clusters alone.

Two algorithms in [8] are proposed for exploiting information from contact tracing with two different test models. In particular, in the case of a binary-valued testing model (as we consider), they propose a Generalized Approximate Message Passing (GAMP) algorithm based on a dynamic infection model with

3Throughout the paper, by “Appendix” we mean “Supplementary Material”, attached as a separate document.

4Specifically, the upper bound for non-adaptive testing in [6] is $O(k \log |S|)$, whereas our less stringent criterion will give $O(\log |S|)$ scaling.
known infection times and physical proximity levels. Further graph-based models are given in [5] and [7]. In [5] the graph is modeled as being random, and after its realization is produced, anyone with a path to a “patient zero” becomes infected. Along similar lines, in [7], given a known graph, each edge is dropped with some fixed probability, and then each connected component is independently (fully) defective with some fixed probability.

We note that various theoretical guarantees appear in [5], [7], [9], [10], [11] (as well as [6] discussed above), but they are largely incomparable to ours due to some combination of the following: (i) focusing on more specific defectivity models; (ii) giving an analysis that does not attempt to optimize constants (or sometimes even logarithmic factors), at least in the non-adaptive case that we focus on; and (iii) providing bounds (e.g., on the average number of errors) without attempting to characterize their asymptotic behavior, which is our focus.

C. Contributions

Our main contributions are as follows:

- **Theoretical bounds for model-based priors:** Focusing primarily on the noiseless setting, we characterize the information-theoretic limits for approximate recovery when the space of defective sets, \( S \), is a strict subset of all size-\( k \) sets, and \( |S| \ll \binom{n}{k} \). We specialize our general result to several specific examples, and illustrate the gains offered by this prior information, both in terms of \(|S|\) and other useful combinatorial properties.

- **Using model priors and relaxation-based decoders:** We introduce the ising model as a natural and flexible model for capturing dependencies in group testing, and show that it naturally leads to a Quadratic Programming (QP) based decoder, which can also be further relaxed to a more computationally efficient Linear Program (LP). This is in contrast with previous approaches exploiting prior information, which often lead to Belief Propagation (BP) based decoders [8], [9]. We provide synthetic numerical experiments demonstrating the improvement of our decoders over standard baselines that do not exploit prior information.

II. THEORETICAL BOUNDS FOR MODEL-BASED PRIORS

In this section, we study the information-theoretic limit of group testing with prior information, focusing on the approximate recovery criterion given in (5). Specifically, we generalize the standard combinatorial prior as follows: The number of defective items \( k \) is still fixed, but the defective set \( S \) is now uniform over some subset \( S' \), where each element of \( S \) is a size-\( k \) subset of \([n]\). In the case that \(|S| \ll \binom{n}{k}\), this amounts to having significant prior knowledge compared to the standard combinatorial prior. Accordingly, we focus primarily on the following scaling regime:

\[
|S| = 2^{\beta k \log_2 \frac{\tau}{2} (1+o(1))} \quad \text{for some } \beta \in (0, 1),
\]

(12)

We observe that \( \beta \) close to one corresponds to having little prior information, whereas \( \beta \) close to zero amounts to having substantial prior information. Varying \( \beta \) between 0 and 1 covers a wide range of distinct scaling regimes, analogous to the consideration of \( k = O(n^\theta) = n^{\theta(1+o(1))} \) with \( \theta \in (0, 1) \).

Since the preceding setup is a direct counterpart to that of model-based compressive sensing [28], we adopt the terminology model-based priors.

A. General Achievability and Converse Bounds

In this subsection, we provide general achievability and converse bounds for model-based priors. These bounds are typically not easy to evaluate directly, but they serve as the starting point for applying to specific cases and simplifying.

Throughout this section, we focus on the noiseless case for concreteness and relative ease of analysis, but sometimes also discuss noisy settings.

1) Achievability: The analysis in [15], which takes \( S \) to be the entire space of \( k \)-sized subsets of \([n]\), contains the combinatorial terms

\[
N_\tau^\text{default} = \binom{k}{\tau} \binom{n-k}{\tau} \quad \text{for } \tau = 0, 1, \ldots, k,
\]

(13)
counting the number of ways that the correct defective set \( S \) can have \( \tau \) of its items removed, and a different \( \tau \) items added (from \([n]\) \(\setminus\) \(S\)), to produce some incorrect defective set \( S' \) of cardinality \( k \). Note that the value \( N_\tau^\text{default} \) is the same for each realization of the defective set \( S \) by symmetry. Hence, for \( \tau = 0, 1, \ldots, k \), we can rewrite (13) as

\[
N_\tau^\text{default} = \max_{S : |S|=k} |S' : |S'| = k \text{ and } d(S, S') = \tau|,
\]

(14)

where \( d(S, S') = \max\{|S \setminus \hat{S}, |\hat{S} \setminus S|\} \) is the distance measure associated with our approximate recovery criterion.

In our analysis, we generalize (14) for the setting of model-based priors, defining

\[
N_\tau = \max_{S \in S} |S' \in S : d(S, S') = \tau|.
\]

(15)

Coupling this with a suitably-modified information-theoretic decoder, we analyze how the final bound on the number of tests \( t \) changes, leading to the following analog of [15, Equation (3.47)].

**Theorem II.1:** Let \( n \) be the number of items, \( k = \Theta(n^\theta) \) for some \( \theta \in (0, 1) \) be the defective set size, \( S \) an arbitrary subset of the entire space of \( k \)-sized subsets of \([n]\), and \( S \) drawn uniformly from \( S \). Fix \( \alpha \in (0, 1) \), and let \( d_{\text{max}} = [\alpha^* k] \). Let \( X \in \{0, 1\}^{\alpha n} \) be a random binary matrix with i.i.d Bernoulli \((\frac{\tau}{2})\) entries for some \( \nu \in [0, 1] \). Under the noiseless group testing model, the error probability \( P_{d_{\text{max}}}(\text{err}) \) of the information-theoretic threshold decoder in Appendix A-II vanishes asymptotically as \( n \to \infty \), provided that

\[
t \geq \max_{\alpha \in [0, 1]} \frac{\log_2 N_{\lfloor \alpha k \rfloor}}{e^{-1-\alpha} H_2(e^{-\alpha}) (1 + o(1))},
\]

(16)

where \( H_2(\cdot) \) is measured in bits, and \( N_\tau \) is given in (15).

**Proof:** See Appendix A.

This expression can be simplified by substituting the standard choice \( \nu = \ln 2 \), which makes the denominator in (16) equal to \( 2^{-(1-\alpha)H_2(2^{-\alpha})} \), in particular simplifying to 1 when \( \alpha = 1 \).
However, as we will see shortly, this choice of \( \nu \) is not always optimal.

While we have focused on the noiseless setting in Theorem II.1, symmetric noise is also straightforward to handle in the same way as [15]. Non-symmetric noise can also be considered (e.g., see [29, App. A]), but in such cases, the final expressions tend to be more complicated (e.g., not closed form, more free parameters, etc.).

2) **Converse**: Next, we provide an information-theoretic converse, i.e., a hardness result that holds for any non-adaptive test design and decoder. To seek an ideal balance between the convenience of analysis vs. generality, we consider two separate cases depending on whether the decoder is constrained to output an element of \( S \), and accordingly define

\[
S_{\text{dec}} = \begin{cases} S & \text{decoder always outputs} \hat{S} \in S \\ \{S : |S| = k\} & \text{unconstrained decoder output}. \end{cases}
\]  

(17)

In the second case, the reason we may still constrain \(|S| = k\) is that outputting a size-\( k \) set is without loss of optimality for the approximate recovery criterion that we consider [16], [30].

We then introduce the following quantity:

\[
\tilde{N}_{d_{\text{max}}} = \max_{S \in S_{\text{dec}}} \max_{|S'| = \nu} |S' : d(S,S') \leq d_{\text{max}}|,
\]

(18)

which is closely related to \( \sum_{|S| = k} N_{\tau} \) (see (15)), but is slightly different due to the order of sum/max and the possibility that \( S_{\text{dec}} \) differs from \( S \).

**Theorem II.2**: Let \( n \) be the number of items, \( k = \Theta(n^{\theta}) \) for some \( \theta \in (0,1) \) be the defective set size, \( S \) an arbitrary subset of the entire space of \( k \)-sized subsets of \([n]\), and \( S \) drawn uniformly from \( S \). Fix \( \alpha^* \in (0,1) \), and let \( d_{\text{max}} = \lfloor \alpha^*k \rfloor \). Let \( X \in \{0,1\}^{\times n} \) be an arbitrary non-adaptive test design. Then, under the noiseless group testing model with approximate recovery, for any decoder to attain \( P_{d_{\text{max}}}[\text{err}] \rightarrow 1 \) as \( n \rightarrow \infty \), it must hold that

\[
t \geq \left( \log_2 |S| - \log_2 \tilde{N}_{d_{\text{max}}} \right) \cdot (1 + o(1)),
\]

(19)

where \( \tilde{N}_{d_{\text{max}}} \) is given in (18).

**Proof**: See Appendix B. \( \square \)

To simplify the analysis, we will sometimes apply this result to cases where the decoder must output an element of \( S \), i.e., the first case in (17). However, by a basic triangle inequality argument, e.g., as used in [31, Sec. 4.2.2], a converse for such decoder translates to a converse for general decoders with \( d_{\text{max}} \) replaced by \( 2d_{\text{max}} \). In particular, the resulting converse bounds will always coincide in the limit \( \alpha^* \rightarrow 0 \), which we believe to be the regime of primary interest.

For noisy group testing models in which the noiseless outcome is passed through a binary channel, we can obtain a similar result to Theorem II.2 with \( P_{d_{\text{max}}}[\text{err}] \rightarrow 0 \) instead of \( P_{d_{\text{max}}}[\text{err}] \rightarrow 1 \), and with the right-hand side of (19) being divided by the channel capacity. Specifically, the analysis based on Fano’s inequality in [31, Sec. 4.1.2] readily applies, and \( \tilde{N}_{d_{\text{max}}} \) directly appears therein in the same way as (19). We also expect that the stronger statement \( P_{d_{\text{max}}}[\text{err}] \rightarrow 1 \) can be obtained following the ideas of [32, Thm. 1], but we omit this direction since it is not our main focus.

**B. Cardinality-Bounded Structure**

In this case, we specialize our general bounds to the case that only \(|S|\) is known, without more fine-grained bounds on each value of \( N_{\tau} \). This is motivated by the fact that given knowledge of \( S \), one would expect that \(|S|\) is typically easy to bound, whereas the individual \( N_{\tau} \) may be much more complex.

For the achievability part, we characterize (16) using the following trivial bound:

\[
N_{\tau} \leq \min \{ N_{\text{default}}, |S| \},
\]

(20)

and show that this leads to the following corollary.

**Corollary II.3**: Under the setup of Theorem II.1, if \(|S| = 2(\beta k \log_2 \frac{n}{k})^{1+o(1)} \) for some \( \beta \in (0,1) \), then the error probability \( P_{d_{\text{max}}}[\text{err}] \) of the information-theoretic threshold decoder (Appendix A-II) vanishes asymptotically provided that

\[
t \geq \left( \beta \exp(\nu(1 - \max\{\alpha^*, \beta\})) \right) \left( \frac{k \log_2 n}{k} \right) (1 + o(1)).
\]

(21)

**Proof**: See Appendix C. \( \square \)

Based on [15, Thm. 3], we expect the minimizing \( \nu \) to approach \( \ln 2 \) when \( \max\{\alpha^*, \beta\} \rightarrow 1 \). For fixed \( \max\{\alpha^*, \beta\} \), we can show by differentiation that the minimizing \( \nu \) satisfies

\[
\text{H}_2(\exp(-\nu \max\{\alpha^*, \beta\})) = -\max\{\alpha^*, \beta\} \cdot \log_2(1 - \exp(-\nu \max\{\alpha^*, \beta\}))
\]

(22)

which appears to have no closed-form solution for general \( \max\{\alpha^*, \beta\} \). However, this can be solved numerically; from Fig. 1, we see that the optimal \( \nu \) starts higher and gradually decreases to \( \ln 2 \).
In the group testing literature, it is common to measure the performance using the following limiting ratio, or “rate”:

$$\lim_{n \to \infty} \frac{\log_2 \binom{n}{k}}{t},$$

(23)

where \(k\) and \(t\) implicitly depend on \(n\). This is motivated by the fact that the prior uncertainty without prior information is \(\log_2 \binom{n}{k}\). Accordingly, in our setting, a natural counterpart is the following:

$$\lim_{n \to \infty} \frac{\log_2 |S|}{t}.$$  

(24)

We will consider both of these limiting ratios, since (23) is also useful for the purpose of comparing to what one would attain in the absence of prior information (beyond knowledge of \(k\)). When \(|S| = 2^{(\beta k \log_2 \frac{n}{k})(1+o(1))}\) and \(k = o(n)\) (as we consider), the two limiting ratios are simply related via \(\lim_{n \to \infty} \frac{\log_2 \binom{n}{k}}{t} = \frac{1}{\beta} \lim_{n \to \infty} \frac{\log_2 |S|}{t}\).

Substituting (21) into (24), we obtain

$$\lim_{n \to \infty} \frac{\log_2 |S|}{t} = \lim_{n \to \infty} \frac{\beta k \log_2 \frac{n}{k}}{t} = \frac{H_2(\exp(-\nu \max\{\alpha^*, \beta\}))}{\exp(\nu(1-\max\{\alpha^*, \beta\}))},$$

(25)

where \(\nu\) takes the value that minimizes (21). It follows that for a fixed value of \(\max\{\alpha^*, \beta\}\), decreasing \(\min\{\alpha^*, \beta\}\) does not affect the limit. The limit is shown in Fig. 2 (Top). As discussed above, the limit of \(\lim_{n \to \infty} \frac{\log_2 \binom{n}{k}}{t}\) is simply a \(\frac{1}{\beta}\) factor higher than (25). For fixed \(\max\{\alpha^*, \beta\}\), if we have \(\alpha^* \leq \beta\), then decreasing \(\alpha^*\) does not affect the limit; however, if we have \(\beta \leq \alpha^*\), then decreasing \(\beta\) still increases the limit. In particular, as \(\beta \to 0\) with fixed \(\alpha^*\), the resulting value becomes unbounded, indicating an arbitrarily large reduction in the number of tests compared to no prior information. The limits for both cases are shown in Fig. 2 (Bottom).

As seen in Fig. 2 (Bottom), for all \(\beta < 1\), we need strictly fewer tests than \(\log_2 \binom{n}{k}\) (corresponding to 1 on the y-axis), i.e., the achievability bound in the absence of prior information (Lemma I.1). Furthermore, smaller \(\beta\) naturally leads to smaller \(t\). On the other hand, we observe in Fig. 2 (Bottom) that a smaller value \(\max\{\alpha^*, \beta\}\) actually leads to smaller \(\lim_{n \to \infty} \frac{\log_2 |S|}{t}\). This may be in part due to the crude upper bound in (20) becoming looser for small \(\beta\) (it is tight for \(\beta = 1\)).

For the converse, if all that we know about \(S\) is a cardinality bound, then we have the following generalization of (8).

**Corollary II.4:** Consider the setup of Theorem II.2, and suppose that

$$|S| = 2^{(\beta k \log_2 \frac{n}{k})(1+o(1))}$$

(26)

for some \(\beta \in (0, 1)\). Then, for any decoder to attain \(P_{\max\{\text{err}\}} \not\to 1\) as \(n \to \infty\), it must hold that

$$t \geq \max\{0, \beta - \alpha^*\} \left(k \log_2 \frac{n}{k}\right) \cdot (1+o(1)).$$

(27)

**Proof:** See Appendix C.

A regime of particular interest is \(\alpha^*\) becoming arbitrarily small, meaning that the recovery is required to be increasingly close to exact. In this case, the lower bound approaches \((\beta k \log_2 \frac{n}{k})(1+o(1))\), which is simply \((\log_2 |S|)(1+o(1))\) as one would expect.

We observe that Corollary II.4 is non-vacuous only if \(\beta > \alpha^*\). While this may appear limiting, this condition is unavoidable: It is straightforward to show that a feasible choice of \(S\) (subject to the cardinality bound (26)) is the set of all \(k\)-sparse binary vectors within distance \(\beta k (1+o(1))\) of \(\{0, \ldots, k\}\). In this case, if \(\alpha^* > \beta\), then simply returning \(\{0, \ldots, k\}\) will be considered non-erroneous with probability one, even when there are no tests.

We will shortly see that both Corollary II.3 and Corollary II.4 can be significantly improved when more specific knowledge of the structure of \(S\) (beyond \(|S|\) alone) is available.

### C. Block Structure

We now look at an extreme scenario where we have exact knowledge of all \(\{N_r\}_{r=d_{\max}+1}^k\) in (15), and illustrate that the achievable can be further improved as compared to using only the cardinality bound (20). Specifically, we consider the scenario
where the $n$ items have the following structure: The $n$ items are arranged in $n/q$ groups of size $q$, and each group is either all defective or all nondefective, with $k/q$ defective groups.

This model can be viewed as capturing (albeit in a highly simplified manner) the fact that infections or defects may occur in groups in applications. It is also a special case of community-aware models studied in [9], [10]. Unlike those works, we do not seek to identify new test designs to handle community structure, but instead explore the benefits of improved decoding alone (while maintaining i.i.d. designs).

Then, $N_{\tau} = 0$ if $\tau$ is not a multiple of $q$, while if $\tau$ is a multiple of $q$, then

\[ N_{\tau} = \frac{k/q}{\tau/q} \left( \frac{n-k}{\tau/q} \right). \]  

(28)

The exact knowledge of $N_{\tau}$ gives the following improved achievability result. We note that while this result in itself is not "surprising", it serves as a useful illustration of the benefit of characterizing each $N_{\tau}$ individually in Theorem II.1, instead of only $|S|$.

**Corollary II.5:** Let $1 \leq q \leq k \leq n$ be integers such that $n$ and $k$ are multiples of $q$, $k = \Theta(n^q)$ for some $\theta \in (0, 1)$, and $d_{\text{max}} = \Theta(k)$. Consider $S$ defined according to the above block-structured model. Let $X \in \{0, 1\}^{k \times n}$ be a random binary matrix with i.i.d. Bernoulli$(\frac{q}{k})$ entries for some $\nu \in [0, 1]$. Under the non-adaptive, noiseless, and approximate recovery setting group testing model, as the number of items $n \to \infty$, the error probability $P_{d_{\text{max}}}^{\text{err}}$ of the information-theoretic threshold decoder (see Appendix A-II) vanishes asymptotically provided that

\[ t \geq \frac{1}{q} H_2(e^{-\nu}) (k \log_2 \frac{n}{k})(1 + o(1)). \]  

(29)

**Proof:** See Appendix D. □

The minimizing $\nu$ in (29) is $ln 2$, which gives the condition

\[ t \geq \frac{1}{q} (k \log_2 \frac{n}{k})(1 + o(1)) \]

\[ = \left( \log_2 \left( \frac{n}{q} \right) \right)(1 + o(1)) = (\log_2 |S|)(1 + o(1)). \]  

(30)

Observe that $\alpha^*$ is absent, as it only affects the higher-order asymptotics.

For fixed $q$, we simply have $\lim_{n \to \infty} \frac{\log_2 |S|}{k} = 1$ and $\lim_{n \to \infty} \frac{\log_2 (\nu^q)}{k} = \nu$. Hence, the required number of tests is a $1/q$ fraction of that without prior information (Lemma I.1).

For comparison, we now consider the achievability result obtain by only considering the naive cardinality bound (20) on $N_{\tau}$, instead of (28). Note that the value of $\beta$ in (12) evaluates to $\beta = \frac{\log_2 |S|}{k \log_2 \left( \frac{n}{k} \right)} = \frac{1}{q}(1 + o(1))$. For fixed $\alpha^* = d_{\text{max}}/k \in (0, 1)$, it follows from the discussion in Section II-B that the coefficient of $(k \log_2 \frac{n}{k})(1 + o(1))$ is given by

\[ \frac{1}{q} \left( \inf_{\nu > 0} \frac{H_2(e^{-\nu} \max(\alpha^*, 1/q))}{H_2(e^{-\nu} \max(\alpha^*, 1/q))} \right) > \frac{1}{q}, \]  

(31)

and is no longer independent of $\alpha^*$. From Fig. 2, we see that if $\max\{\alpha^*, 1/q \} \leq \alpha^*$, then the values are lower than 1 (i.e., more tests are needed), and significantly so when $\max\{\alpha^*, 1/q \} = \alpha^*$ (i.e., $d_{\text{max}}$ is small and $q$ is large). This illustrates that having a good control over $\{N_{\tau} \}_{\tau = d_{\text{max}} + 1}$ can give a significantly tighter achievability bound.

The converse is also simple in this case, and could likely be obtained by slightly modifying the analyses in [9], [10], but we provide a self-contained proof for completeness.

**Corollary II.6:** Consider the setup of Corollary II.5, but with an arbitrary non-adaptive test matrix $X$, and suppose that $d_{\text{max}} = \lceil \alpha^* k \rceil$ for some $\alpha^* \in (0, 1)$. Then, for any decoder that outputs $\hat{S} \subseteq S$ to attain $P_{d_{\text{max}}}^{\text{err}} \not\to 1$ as $n \to \infty$, it must hold that

\[ t \geq (1 - \alpha^*) \left( \frac{k}{q} \log_2 \frac{n}{k} \right)(1 + o(1)) \]

\[ = (1 - \alpha^*)(\log_2 |S|)(1 + o(1)). \]  

(32)

**Proof:** See Appendix D. □

Similarly to Lemma I.1, the achievability and converse match to within a factor of $1 - \alpha^*$. The converse is non-vacuous for all $\alpha^* \in (0, 1)$, demonstrating that we can overcome the limitation of Corollary II.4 discussed above given more precise knowledge about $S$ rather than only the cardinality.

### D. Imbalanced Structure

We now turn to a more complex example that further demonstrates the possible gaps between the various bounds of interest. We consider the following setup: The $n$ items consist of two disjoint blocks of size $n_1 = o(n)$ and $n_2 = n - n_1 = n(1 - o(1))$ items, and there are $k_1$ defectives in the first block and $k_2 = k - k_1$ in the second. We let $S$ be the set of all $k$-size subsets satisfying these conditions. This model captures (in a simplified manner) scenarios where some items are known a priori to be more likely to be defective, analogous to previous works on exact recovery such as [26], [27].

Although we do not provide its details, the case $k_1 = k(1 - o(1))$ and $k_2 = k - k_1 = o(k)$ (i.e., almost all of the $k$ defective items are concentrated in the smaller block) turns out to be straightforward: Similarly to the block example, the achievability part gives $\lim_{n \to \infty} \frac{\log_2 |S|}{k} = 1$, and the converse matches this up to multiplication by $1 - \alpha^*$. We instead focus on the more challenging case in which $k_1 \gamma k_1 = k$ for some $\gamma \in (0, 1)$, i.e., both $k_1$ and $k_2$ are linear in $k$, while taking $n_1 = \Theta(n^\gamma)$ for some $c \in (\theta, 1)$ (with $k = \Theta(n^\gamma)$).

**Corollary II.7:** Let $n$ be the number of items, and $k = \Theta(n^\gamma)$ for some $\theta \in (0, 1)$ be the size of an unknown defective set $S \subseteq [n]$. Fix $\alpha^* \in (0, 1)$, and let $d_{\text{max}} = \lceil \alpha^* k \rceil$. Let $S$ be defined as above with parameters $n_1, n_2, k_1, k_2$ satisfying $n_1 = \Theta(n^\gamma)$ for some $c \in (\theta, 1)$ and $n_2 = n - n_1$, as well as $k_1 = k \gamma k_1$ and $k_2 = (1 - \gamma) k_1$ for some $\gamma \in (0, 1)$. Let $X \in \{0, 1\}^{k \times n}$ be a random binary matrix with i.i.d. Bernoulli$(\frac{q}{k})$ entries for some $\nu \in [0, 1]$. Under the noiseless group testing model, the error probability $P_{d_{\text{max}}}^{\text{err}}$ of the information-theoretic threshold decoder (see Appendix A-II) vanishes asymptotically as $n \to \infty$, provided...

---

6By interpreting the problem as recovering $k/q$ defective groups out of the $n/q$ groups, Lemma I.1 gives the same achievability bound.

7Because the blocks are large, the maximum number of items that can be defective or nondefective is small and $q$ is large. This illustrates that having a good control over $N_{\tau}$ in the block example can give a significantly tighter achievability bound.

---
that
\[
t \geq \max_{\alpha \in [\max\{\alpha^*,1-\gamma\},1]} \frac{(1-\theta) - \gamma(1-c) - (1-\alpha)(c-\theta)}{e^{-(1-\alpha)\nu}H_2(e^{-\alpha\nu})} \\
\times (k \log_2 n)(1 + o(1)) \tag{33}
\]

**Proof:** See Appendix E. \qed

Based on various numerical calculations, we observed that the maximizing \(\alpha\) is often at one of the endpoints, i.e., \(\alpha = \max\{\alpha^*,1-\gamma\}\) or \(\alpha = 1\). This is supported by the following.

**Corollary II.8:** Under the setup of Corollary II.7, if \(\alpha^* \leq 1-\gamma\) and
\[
(\frac{\nu}{e-\theta})(1-\gamma)(1-c) > 0.029, \tag{34}
\]

then the achievability bound (33) can be simplified to
\[
t \geq \max \left\{ (1-\theta) \frac{(1-\gamma)e^\gamma}{H_2(e^{-(1-\gamma)nu})}k \log_2 n, \right. \\
\left. \times \frac{(1-\theta) - \gamma(1-c)}{H_2(e^{-\nu})}k \log_2 n \right\}(1 + o(1)). \tag{35}
\]

**Proof:** See Appendix E. \qed

We note that the condition \(\alpha^* \leq 1-\gamma\) is mild since we typically want small \(\alpha^*\), and the condition (34) is also mild since it is only violated when \(\gamma\) and/or \(c\) are very close to one (we have \(\frac{1}{\ln 2} > 1\), and we will discuss below how the optimal \(\nu\) is at least \(\ln 2\)).

We now consider minimizing (35) with respect to \(\nu\). We show in (104) of Appendix E that \(\log_2|S| = (1-\theta) - \gamma(1-c)(k \log_2 n)(1 + o(1))\), and combining this with (35) gives the following for the limit introduced in (24) upon optimizing \(\nu\):
\[
\lim_{n \to \infty} \frac{\log_2 |S|}{n} = \max_{\nu > 0} \min \left\{ H_2(e^{-\nu}), \right. \\
\left. \frac{H_2(e^{-(1-\gamma)\nu})}{e^{\gamma\nu}} \left(1 + \frac{\gamma(c-\theta)}{(1-\gamma)(1-\theta)} \right) \right\}. \tag{36}
\]

We denote the two minimands by
\[
f_1(\nu) := H_2(e^{-\nu}), \tag{37}
\]
\[
f_2(\nu) := \frac{H_2(e^{-(1-\gamma)\nu})}{e^{\gamma\nu}} \left(1 + \frac{\gamma(c-\theta)}{(1-\gamma)(1-\theta)} \right), \tag{38}
\]
and analyze their maximizing \(\nu\) separately:

- The first minimand, \(f_1\), is maximized at \(\nu = \ln 2\). It follows that if \(1 = f_1(\ln(2)) \leq f_2(\ln(2))\), or equivalently,
\[
\frac{c-\theta}{1-\theta} \geq 1 - \gamma \left(\frac{2^\gamma}{H_2(2^{(1-\gamma)\nu})} - 1\right), \tag{39}
\]
then (36) can be simplified to 1.

- For the second minimand \(f_2\), we can show by differentiation that, similar to (22), the maximizing \(\nu\) satisfies
\[
H_2 \left( e^{-(1-\gamma)\nu} \right) = (\gamma-1) \log_2 \left( 1 - e^{-(1-\gamma)\nu} \right), \tag{40}
\]
which appears to have no closed-form solution in general. The numerically optimized \(\nu\) gives similar behavior to that shown in Fig. 1 (with \(1-\gamma\) in place of \(\max\{\alpha^*,\beta\}\)). Defining \(\nu^* := \arg\max_{\nu > 0} f_2(\nu)\) accordingly, it follows that if \(f_2(\nu^*) \leq f_1(\nu^*)\), or equivalently,
\[
\frac{c-\theta}{1-\theta} \leq 1 - \gamma \left(\frac{H_2(e^{-\nu})}{e^{-\nu(1-\gamma)}} - 1\right), \tag{41}
\]
then (36) can be simplified to \(f_2(\nu^*)\).

While these cases on \(\frac{c-\theta}{1-\theta}\) are not exhaustive, we found numerically that one of the two usually holds; see Fig. 3, where the horizontal parts correspond to \(\nu = \ln 2\), the diagonal parts to \(\nu = \nu^*\), and a seemingly imperceptible (but non-zero) curved region in between corresponds to other \(\nu\) values.

To assess the tightness of our achievability results, we now turn to the converse, for which we have the following.

**Corollary II.9:** Consider the setup of Corollary II.7, but with an arbitrary non-adaptive test matrix \(X\). If \(\alpha^* > 1 - \gamma\), then for any decoder that outputs \(S \in S\) to attain \(P_{d_{\text{max}}[\text{err}]} \neq 1\) as \(n \to \infty\), it must hold that
\[
t \geq (1-\alpha^*) \left(\frac{c-\theta}{1-\theta} \right) \left(k \log_2 \frac{n}{k}\right)(1 + o(1)). \tag{42}
\]
If \(\alpha^* \leq 1 - \gamma\), then for any decoder to attain \(P_{d_{\text{max}}[\text{err}]} \neq 1\) as \(n \to \infty\), it must hold that
\[
t \geq \left(1 - \gamma(1-c) - \alpha^* \right) \left(k \log_2 \frac{n}{k}\right)(1 + o(1)) \tag{43}
\]
\[
\left(\beta - \alpha^* \right) \left(k \log_2 \frac{n}{k}\right)(1 + o(1)). \tag{44}
\]

**Proof:** See Appendix D-I. \qed

The limit \(\alpha^* \to 0\) is of particular interest, and gives a converse approaching \(\beta k \log_2 \frac{n}{k}\), which corresponds to a horizontal line with value 1 in Fig. 3. Hence, we see that the achievable and converse match for many values of \(\gamma\) and \(\frac{c-\theta}{1-\theta}\), but do not always do so. In fact, if we move beyond i.i.d. designs, then we can adopt the block-structured design outlined in [2, Sec. 5.6] (essentially running group testing separately on the two groups of items);
then, a minor modification of the analysis therein reveals that the resulting bound matches the converse for all values of $\gamma$ and $e^{-\theta \rho}$. Hence, this example illustrates that sometimes adjusting the decoder alone is enough to maintain optimality in the presence of prior information, but sometimes it is necessary to also modify the test design.

III. ISING MODEL PRIORS AND PRACTICAL DECODING RULES

The model-based setup considered in Section II is convenient for theoretical analysis, but has notable limitations that may prevent its use in practice. Firstly, it assumes that $k$ is pre-specified, which is typically unrealistic when infections are random (e.g., i.i.d. infections would lead to $k$ following a binomial distribution). Secondly, in general, $S$ may be a complicated set of subsets, and accordingly may be inconvenient to specify explicitly for use in the group testing algorithm, let alone clear how to effectively exploit once it is specified.

Motivated by these considerations, in this section, we introduce a natural generalization of the i.i.d. prior that is versatile, admits a compact representation, and naturally leads to practical decoding algorithms for exploiting the availability of prior information.

A. Ising Model

We incorporate prior knowledge in the form of an Ising model. Since the $\{-1, 1\}$-valued Ising model is more standard than the $\{0, 1\}$-valued counterpart, we work with $\bar{u} \in \{-1, 1\}^n$ with $+1$ for defectives and $-1$ for non-defectives. Hence, $u \in \{0, 1\}^n$ and $\bar{u} \in \{-1, 1\}^n$ are related via $\bar{u} = 2u - 1$, where 1 is a vector of 1s.

Considering the items $V = [n]$ as vertices, the Ising model is defined as

$$P(\bar{u}) = \frac{1}{Z} \exp \left( \sum_{(j,j') \in E} \lambda_{jj'} \bar{u}_j \bar{u}_{j'} - \sum_{j \in V} \phi_j \bar{u}_j \right),$$

(44)

where

- $G = (V, E)$ is a graph with vertices $V = [n]$ and edges $E \subseteq [n] \times [n]$, where edges represent interactions among vertices.
- For each edge $(j, j') \in E$, $\lambda_{jj'}$ is the edge strength. We focus mainly on the case of a common non-negative value $\lambda_{jj'} = \lambda$ (higher means that connected vertices are more likely to have the same status), but differing values and/or negative values are also straightforward to handle.
- For each vertex $j \in V$, $\phi_j$ is a parameter controlling the sparsity; again, we focus primarily on the case of a common non-negative value $\phi_j = \phi$ ($\phi = 0$ means around half the items are defective (+1), whereas $\phi \gg 0$ means that very few are);
- $Z$ is a normalizing constant.

We believe that this model serves as a natural way to incorporate prior information, with edges corresponding to interactions (e.g., making infections more likely).

For the more standard case of a 0-1 valued defectivity vector, we can substitute $u = 2\bar{u} - 1$ into (44) to obtain

$$P(u) = \frac{1}{Z} \exp \left( \lambda \sum_{(j,j') \in E} (2u_j - 1)(2u_{j'} - 1) - \phi \sum_{j \in V} (2u_j - 1) \right).$$

(45)

In the case that $E = \emptyset$, this model reduces to the standard i.i.d. prior, with a one-to-one correspondence between our parameter $\phi$ and the defectivity probability $q$. Recall that in this case, previous works have considered the (Integer) Linear Programming decoding methods stated in (9) (noisy) and (11) (noisy), and the integer versions can be interpreted as maximizing $P(u|X, Y)$ [24]. We will derive the analogs for the Ising model, re-using the definitions of $C_n$ and $C_n$ in (10) and (12) (and their relaxations $C_n$ and $C_n$).

B. Advantages, Limitations, and Comparisons

An advantage of adopting the Ising model is that it is a versatile and well-studied model, with existing applications including statistical physics [33], image processing [34], computational biology [35], natural language processing [36], and social network analysis [37]. As a result, an extensive range of graph learning algorithms are already available (e.g., see [38] and the references therein) and could potentially be used provided that suitable historical data is available. Further advantages are highlighted at the end of this subsection when comparing with previous approaches.

On the other hand, it is also important to note the limitations of the model. Notably, accurate learning of the graph (and/or the associated edge strengths $\{\lambda_{jj'}\}_{(j,j') \in E}$) can require significant amounts of data that may not always be available. Inaccurate graph learning may degrade the performance, though we provide some experimental evidence of robustness in Section IV. Finally, as we discuss further below, our model is static, and thus does not model dynamic behavior that may arise in applications such as medical testing (e.g., via the well-known SIR model [39]).

Finally, we give some discussion on this model compared to previous models (see Section I-B3 for their summary):

- In [9], [10], [11], the items are arranged into known clusters with high correlations within clusters. The Ising model can also capture such structure, e.g., by including a large number of edges (or even all edges) within each cluster, but also has the added versatility of allowing arbitrary cross-community edges. On the other hand, unlike the preceding works, we do not attempt to give theoretical guarantees on the number of tests for our graph-based model.

- The works [5], [7], [8] also allow general graphs, and can thus model structure beyond clustering alone. In all of these works, adding an edge to the graph amounts to increasing the chances of the two nodes simultaneously being defective. However, the graphs in these works and ours are all used in a different manner, making a direct comparison difficult. A potential limitation of [5], [7] is the “all or nothing” nature of defectivity in each connected
component (after possible edge removals), though this may be reasonable in many cases of interest. The model in [8] appears to be especially powerful when precise knowledge infection times, dynamics, and proximity are known. In contrast, ours is a static model suited to scenarios where interactions are known to have occurred, but not necessarily the finer details (though such details could potentially be incorporated into varying $\{\lambda_{jj'}\}_{(j,j')\in E}$). Overall, our model is not intended to “replace” any other, but we believe that it is a useful addition with a good balance between modeling power and simplicity. The following two aspects appear to be unique to our model compared to the existing ones:

- By considering negative values of $\lambda_{jj'}$, we can easily model negative associations between items (i.e., if one

- Our model lends itself naturally to relaxation-based decoding methods based on Quadratic Programming and Linear Programming, in contrast to existing techniques that instead lead to message passing algorithms [8], [9]. Overall, in view of the preceding advantages and limitations, we believe that the Ising model is primarily suited to settings that (i) are static in nature, (ii) have historical data or prior knowledge available for (at least approximately) inferring a graph and an edge strength $\lambda$ (or varying edge strengths $\{\lambda_{jj'}\}_{(j,j')\in E}$), and (iii) benefit from the use of simple and efficient decoding algorithms.

C. Quadratic Programming Decoder

In the following, we provide a simple theorem showing that maximum a posteriori (MAP) decoding rule (i.e., choosing $u$ to maximize $P(u|X, Y)$) naturally leads to an Integer Quadratic Program.

**Theorem III.1:** In the noiseless setting, under an Ising model prior with parameter $\lambda, \phi$, the MAP decoder has the same solution as the following minimization problem:

$$\min_{u} - \left( \lambda \sum_{(j,j')\in E} (2u_{j} - 1)(2u_{j'} - 1) - \phi \sum_{j\in V} (2u_{j} - 1) \right)$$

subject to $u \in C_{\text{noiseless}}$. \(\blacksquare\)

Moreover, in the case of i.i.d. symmetric noise with parameter $\rho \in (0, \frac{1}{2})$, the MAP decoder has the same solution as the following minimization problem (upon keeping $u$ and discarding $\xi$):

$$\min_{u, \xi} - \left( \lambda \sum_{(j,j')\in E} (2u_{j} - 1)(2u_{j'} - 1) - \phi \sum_{j\in V} (2u_{j} - 1) \right)$$

$$- \eta \sum_{t=1}^{T} \xi_{t}$$

subject to $(u, \xi) \in C_{\text{noisy}}$, \(\blacksquare\)

where $\eta = \log \frac{1}{1-\rho}$.

**Proof:** See Appendix F.

As with the LP formulations in (9) and (11), we can relax the constraints in (46) and (47) by replacing $C_{\text{noiseless}}$ and $C_{\text{noisy}}$ by $c_{[0,1]}^{\text{noiseless}}$ and $c_{[0,1]}^{\text{noisy}}$. This removes the seemingly difficult integer-valued nature of the problems, though as we discuss further below, it does not immediately imply that they can be solved efficiently. We also note that the objective function (47) can be rewritten in a matrix-vector form that is more typical of QP problems (and similarly for (46)):

$$-(\lambda(2u - 1)^T Q(2u - 1) - \phi 1^T (2u - 1)) - \eta 1^T \xi,$$  \(\blacksquare\)

where $Q$ is the “upper” adjacency matrix for the graph $G$, i.e., it is upper triangular since we only count each edge between $u_j$ and $u_{j'}$ once, and all entries on the main diagonal are zeros because there are no self-edges.

D. Linearized Quadratic Programming Decoder

Since the adjacency matrix $Q$ in (48) is not positive semidefinite, the optimization problem is non-convex. This is somewhat unfortunate, because even with real-valued linear constraints, non-convex QPs are NP hard to solve in general [40], [41].

To overcome this potential difficulty, we apply an idea introduced in [42] and surveyed in [43], which “linearizes” the QP at the expense of adding further constraints. In general there can be up to $O(n^2)$ such additional constraints, but in our setting there is only $O(|E|)$, which is much smaller than $O(n^2)$ for typical (sparse) graphs. We call this approach Linearized QP.

In both the noiseless and noisy cases, the only non-linear terms in the objective are the $u_j u_{j'}$ terms. We convert the Integer QP into an Integer LP by replacing each such term $u_j u_{j'}$ by a new binary variable $u_{jj'} \in \{0, 1\}$, and these variables $\{u_{jj'}\}_{(j,j')\in E}$ are included as binary optimization variables. We need to ensure that the property $u_{jj'} = u_j u_{j'}$ is maintained, and to do so, it suffices to constrain $u_j \geq u_{jj'}$, $u_{j'} \geq u_{jj'}$, and $u_j + u_{j'} - u_{jj'} \leq 1$ [43]. In the noiseless case, this leads to the following: $u$ is equal to

$$\text{argmin}_{u \in C_{\text{noiseless}}} \left( \lambda \sum_{(j,j')\in E} (2u_{j} - 1)(2u_{j'} - 1) - \phi \sum_{j\in V} (2u_{j} - 1) \right)$$

$$= \text{argmin}_{u \in C_{\text{noiseless}}} \left( \lambda \sum_{(j,j')\in E} (4u_{j} u_{j'} - 2u_{j} - 2u_{j'}) - \phi \sum_{j\in V} 2u_{j} \right)$$

$$= \text{argmin}_{u \in C_{\text{noiseless}}} \left( 2\lambda \sum_{(j,j')\in E} u_{jj'} - \lambda \sum_{(j,j')\in E} (u_{j} + u_{j'}) - \phi \sum_{j\in V} u_{j} \right)$$

\(\blacksquare\)

$^8$Even if we symmetrize $Q$, the fact that the diagonals are zero (but the off-diagonals contain non-zeros) prevents the matrix from being PSD. For instance, if there are two nodes connected by a single edge, the symmetrized matrix becomes $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ which has eigenvalues $-1$ and 1.
Hence, we are left with the following optimization problem:

\[
\min_{\{u_{jj'}\}_{(j,j')}\in E} \left( - 2\lambda \sum_{(j,j')\in E} u_{jj'} - \lambda \sum_{(j,j')\in E} (u_j + u_j') - \phi \sum_{j\in V} u_j \right)
\]

subject to \(u \in C_{\text{noiseless}}\) (or \(C_{\text{noiseless}}^{[0,1]}\) if relaxed)

\[
\begin{align*}
&u_{jj'} \in \{0, 1\}, \text{ (or } [0, 1]\text{ if relaxed)} \\
&u_j \geq u_{jj'} \\
&u_{jj'} \geq u_{jj''} \\
&u_j + u_{jj'} - u_{jj''} \leq 1.
\end{align*}
\]

(52)

Similarly for the noisy case, the objective function is given as

\[
\min_{\{u_{jj'}, \xi\}_{(j,j')}\in E} \left( - 2\lambda \sum_{(j,j')\in E} u_{jj'} - \lambda \sum_{(j,j')\in E} (u_j + u_j') - \phi \sum_{j\in V} u_j \right) - \eta \sum_{i=1}^{t} \xi_i
\]

subject to \((u, \xi) \in C_{\text{noisy}}\) (or \(C_{\text{noisy}}^{[0,1]}\) if relaxed)

\[
\begin{align*}
&u_{jj'} \in \{0, 1\}, \text{ (or } [0, 1]\text{ if relaxed)} \\
&u_j \geq u_{jj'} \\
&u_{jj'} \geq u_{jj''} \\
&u_j + u_{jj'} - u_{jj''} \leq 1,
\end{align*}
\]

(53)

where \(\eta = \log\left(\frac{p}{1-p}\right)\). We observe that the relaxed versions are indeed Linear Programs, for which numerous efficient solvers exist both in theory and in practice. While the integer versions remain equivalent to those given in Theorem III.1, this equivalence may be lost upon relaxing the problems. Despite this, we will see in the following section that the relaxed versions still give strong empirical performance, analogous to the findings for standard group testing in [23].

IV. NUMERICAL EXPERIMENTS

In this section, we provide some simple proof-of-concept experiments for the QP and LP based decoding rules introduced in the previous section, without attempting to be comprehensive.\(^9\)

A. Ising Model Details

We consider the Ising model, focusing initially on two simple choices of the graph. The first is a square grid graph where each vertex is connected with its (up to 4) neighboring vertices (up, down, left and right). The second graph, with the same size as the first one, is further divided into several blocks of equal size.

Within each block, all vertices are still connected in the form of grid, but there are no connections among different blocks. Two examples of these types of graphs are shown in the top row of Fig. 4. Roughly, the grid graph can be viewed as capturing a simple form of proximity-based correlation, and the block graph can be viewed as capturing a simple form of grouped/clumped structure.

We set the graph size as \(28 \times 28\) and the block size as \(7 \times 7\) (16 blocks in total). We manually set the parameters \(\lambda\) and \(\phi\) to be 0.5 and 0.006 for the grid graph, and 0.6 and 0.035 for the block graph, which we found to be suitable for producing defectivity patterns that are sparse (as desired for group testing) while also exhibiting visible correlations (as desired for understanding the gains of prior information). We apply Gibbs sampling [34] to (approximately) generate samples from these models, using 1000 iterations. The generated samples are shown in the bottom row of Fig. 4.

B. Implementation

For each generated sample with size \(28 \times 28\), we flatten it as a vector \(\mathbf{u}' \in \{-1, 1\}^n\) with \(n = 784\), and form \(\mathbf{u} \in \{0, 1\}^n\) by replacing \(-1s\) by 0s. The test matrix is randomly drawn according to a Bernoulli design, where each item \(u_j\) is included in each test independently with probability \(\frac{\ln 2}{k}\).

For both graphs, we keep \(\mathbf{u}\) fixed to the samples indicated in Fig. 4, and for each trial of our experiment, we only change the test matrix \(\mathbf{X}\) applied on them. We generate 50 trials to get the averaged \(\text{FP}/k\), \(\text{FN}/k\), and time for each algorithm. For each number of tests, the same 50 test matrices are used for three different algorithms and for two different noise levels, where \(\rho \in \{0, 0.01\}\).

We use Gurobi [44] as the mathematical optimization solver to implement LP, QP, and Linearized QP, and compare these three decoding algorithms’ results using three metrics: the normalized number of false positives (FP/k), the normalized number of false negatives (FN/k), and the computation time. Note that by “LP” we mean the decoder in (9) (noiseless) or (11) (noisy), which

\(^9\)For reproducibility, the code for our experiments is available at https://github.com/ethangela/priors_group_testing.
is only based on sparsity and does not use the graph information. For all continuous relaxations, we form the \{0, 1\}-valued estimate by rounding to the nearer value.

### C. Results

We present the results in Figs. 5 (grid) and 6 (block). We show the error rates FP\(/k\) (left column) and FN\(/k\) (middle column), and the computation time (right column), both for the non-relaxed and relaxed (“Re”) variants, and in both noiseless and noisy settings.

We find that for both graphs and for both noise levels, the recovery performance of QP improve on that of LP, and often significantly so (note that we use a logarithmic y-scale). The only exception is that when \(\rho = 0.01\), LP tends to generate fewer FPs, possibly due to there being too few tests to reliably mark many items as positive (hence the very high FN). These results support the idea that by incorporating the Ising model as prior
information, we are able to attain more accurate recovery with fewer tests. Moreover, although QP is naturally slower than LP, we still manage to reduce its computation time to a large extent by utilizing Linearized QP, which shows almost identical recovery performance as that of QP. Therefore, we believe that Linearized QP provides a good solution for practical use.

Perhaps surprisingly, in Figs. 5 and 6, the computation times are similar for the non-relaxed and relaxed variants, except for LP in the final sub-figure of Fig. 6, and for certain other cases with few tests (e.g., 100). However, as we exemplify using larger $72 \times 72$ graphs in Fig. 11 in the supplementary material, the gap can be significant for Linearized QP as well. Fig. 11 also highlights the fact that both relaxed and non-relaxed QPs can be slow to solve (as discussed previously), whereas the relaxed Linearized QP can be much faster than the non-relaxed variant.

Overall, it is difficult to pinpoint the precise factors that make these methods faster or slower, particularly when the time may not even be monotonically increasing with respect to the number of tests (this was also observed in [24], with the intuition being that more tests can help to rule out suboptimal parts of the search space faster). Nevertheless, our results suggest that both the non-relaxed and relaxed variants can be practically feasible, with the latter typically being preferable when computation is a bottleneck, and with (relaxed) Linearized QP being preferable to QP for large problem sizes.

D. Sensitivity to Model Mismatch

In most practical scenarios, one cannot know the graph $G$ and edge strength $\lambda$ perfectly (e.g., they may only be approximately learned from historical data). Here we consider the effect of mismatch with respect to each of these, albeit without seeking to be comprehensive.

**Graph mismatch:** In this experiment, we consider varying degrees of mismatch in the graph estimate. To do so, starting with the true graph, we remove a certain fraction of its edges uniformly at random, and simultaneously add the same number of edges uniformly chosen from the original non-edges. Clearly, the higher the fraction, the ‘more mismatched’ the graph estimate is.

We consider two separate scenarios: (i) noiseless tests, the grid structure, and 300 tests, and (ii) noisy tests, the block structure, and 500 tests. In both cases, we vary the fraction of removed (and re-added) edges from 0 to $\frac{1}{2}$. All other settings are the same as before.

The results are presented in Fig. 7. Naturally, for all algorithms that use the prior information, the performance worsens as the degree of mismatch increases. Nevertheless, we observe graceful degradation rather than a sharp drop in performance, with the FN curves lying entirely below those of the LP approach, and the FP curves doing the same until the degree mismatch becomes considerable.

**Parameter mismatch:** Here we consider the effect of varying the edge strength $\lambda$ that the algorithm assumes, which may now differ from the true value. We consider two scenarios: (i) noiseless tests, the grid structure, and 300 tests, and (ii) noisy tests, the block structure, and 500 tests. In both cases, we vary the fraction of removed (and re-added) edges from 0 to $\frac{1}{2}$, with the intuition being that more tests can help to rule out suboptimal parts of the search space faster). Nevertheless, our results suggest that both the non-relaxed and relaxed variants can be practically feasible, with the latter typically being preferable when computation is a bottleneck, and with (relaxed) Linearized QP being preferable to QP for large problem sizes.

![Fig. 7. Results under graph mismatch: (Top) noiseless, grid graph, 300 tests, (Bottom) noisy, block graph, 500 tests.](image)
E. Example on Real-World Data

While the grid and block graphs exhibit certain properties of interest (e.g., clustering structure), there are naturally limitations to what they can capture. To address this limitation, here we consider a graph created from real-world data. Specifically, we consider an ‘ego-Facebook’ dataset \cite{45} collected from 4039 survey Facebook users (nodes) consisting of 88234 ‘social interactions’ (edges), and we reduce its size by randomly selecting 500 nodes and extracting the corresponding sub-network.

In Fig. 9, we visualize the resulting adjacency matrix (Left) and degree distribution (Right). While there is still some block structure, we observe that this graph is significantly more complex than those above, with several “clusters” of varying sizes, and a number of “heavily connected” nodes connected to multiple clusters. For the Ising model, we choose $\lambda$ and $\phi$ as 1.5 and 1.0, respectively. A resulting defective set from Gibbs...
sampling is also depicted in Fig. 9 (Left) using vertical and horizontal lines.

We apply group testing on this graph in the same way as the synthetic graphs, and present the results in Fig. 10. Overall, we observe generally similar performance patterns to those of the simpler grid/block graphs (including computation time), particularly as the number of tests increases.

REFERENCES

[1] R. Dorfman, “The detection of defective members of large populations,” Ann. Math. Statist., vol. 14, no. 4, pp. 436–440, 1943.
[2] 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, 2019.
[3] M. Aldridge and D. Ellis, “Pooled testing and its applications in the COVID-19 pandemic,” in Pandemics: Insurance and Social Protection, M. d. C. Bosdo-Penas, J. Eisenberg, and S. Sahin, Eds. Berlin, Germany: Springer, 2022, pp. 217–249.
[4] A. Coja-Oghlan, O. Gebhard, M. Hahn-Klimroth, and P. Loick, “Optimal group testing,” in Proc. Conf. Learn. Theory, 2020, pp. 1374–1388.
[5] B. Arasli and S. Ulukus, “Group testing with a graph infection spread model,” 2021. [Online]. Available: https://arxiv.org/abs/2101.05792
[6] M. Gonen, M. Langberg, and A. Sprintson, “Group testing on general set-systems,” in Proc. IEEE Int. Symp. Inf. Theory, 2022, pp. 874–879.
[7] H. Nikpey, J. Kim, X. Chen, S. Sarkar, and S. S. Bidokhti, “Group testing with correlation under edge-faulty graphs,” in Proc. IEEE Int. Symp. Inf. Theory, 2022, pp. 880–885.
[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. IEEE Int. Conf. Acoust., Speech Signal Process., 2021, pp. 8168–8172.
[9] P. Nikolopoulos, S. R. Srinivasasaranadh, T. Guo, C. Fragouli, and S. Diggavi, “Group testing for connected communities,” in Proc. Int. Conf. Artif. Intell. Statist., 2021, pp. 2341–2349.
[10] S. Ahn, W.-N. Chen, and A. Özgür, “Adaptive group testing on networks with community structure,” in Proc. IEEE Int. Symp. Inf. Theory, 2021, pp. 1242–1247.
[11] P. Nikolopoulos, S. R. Srinivasasaranadh, T. Guo, C. Fragouli, and S. Diggavi, “Group testing for overlapping communities,” in Proc. IEEE Int. Conf. Commun., 2021, pp. 1–7.
[12] O. Johnson, M. Aldridge, and J. Scarlett, “Performance of group testing algorithms with near-constant tests per item,” IEEE Trans. Inf. Theory, vol. 65, no. 2, pp. 707–723, Feb. 2019.
[13] N. Tan, W. Tan, and J. Scarlett, “Performance bounds for group testing with doubly-regular designs,” IEEE Trans. Inf. Theory, to be published, doi: 10.1109/TIT.2022.3210593.
[14] M. Cheraghchi, “Noise-resilient group testing: Limitations and constructions,” in Proc. Int. Symp. Found. Comput. Theory, 2009, pp. 62–73.
[15] J. Scarlett and V. Cevher, “Phase transitions in group testing,” in Proc. 27th Annu. ACM-SIAM Symp. Discrete Algorithms, 2016, pp. 40–53.
[16] J. Scarlett and V. Cevher, “Limits on support recovery with probabilistic models: An information-theoretic framework,” IEEE Trans. Inf. Theory, vol. 63, no. 1, pp. 593–620, Jan. 2017.
[17] J. Scarlett and V. Cevher, “How little does non-exact recovery help in group testing?,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., 2017, pp. 6090–6094.
[18] L. V. Truong, M. Aldridge, and J. Scarlett, “On the all-or-nothing behavior of Bernoulli group testing,” IEEE J. Sel. Areas Inf. Theory, vol. 1, no. 3, pp. 669–680, Nov. 2020.
[19] J. Niles-Weed and I. Zadik, “It was "all" for "nothing": Sharp phase transitions for noiseless discrete channels,” in Proc. Conf. Learn. Theory, 2021, pp. 3546–3547.
[20] M. B. Malyutov, “The separating property of random matrices,” Math. Notes Acad. Sci. USSR, vol. 23, no. 1, pp. 84–91, 1978.
[21] E. Price and J. Scarlett, “A fast binary splitting approach to non-adaptive group testing,” in Proc. Int. Conf. Randomization Comput. Comput., 2020, pp. 13:1–13:20.
[22] M. Aldridge, L. Baldassini, and O. Johnson, “Group testing algorithms: Bounds and simulations,” IEEE Trans. Inf. Theory, vol. 60, no. 6, pp. 3671–3687, Jun. 2014.
[23] D. Malioutov and M. Malyutov, “Boolean compressed sensing: LP relaxation for group testing,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., 2012, pp. 3305–3308.
[24] L. Ciampiconi, B. Ghosh, J. Scarlett, and K. S. Meel, “A MaxSAT-based framework for group testing,” in Proc. AAAI Conf. Artif. Intell., 2020, pp. 10144–10152.
[25] H. Zabeti, N. Dexter, I. Lau, L. Unruh, B. Adcock, and L. Chindelevitch, “Group testing large populations for SARS-CoV-2,” 2021. [Online]. Available: https://www.medrxiv.org/content/10.1101/2021.06.03.21258258
[26] T. Li, C. L. Chan, W. Huang, T. Kaced, and S. Jaggi, “Group testing with prior statistics,” in Proc. IEEE Int. Symp. Inf. Theory, 2014, pp. 2346–2350.
[27] T. Kealy, O. Johnson, and R. Piechocki, “The capacity of non-identical adaptive group testing,” in Proc. IEEE 52nd Allerton Conf. Commun. Control Comput., 2014, pp. 101–108.

[28] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, “Model-based compressive sensing,” IEEE Trans. Inf. Theory, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[29] J. Scarlett, “Noisy adaptive group testing: Bounds and algorithms,” IEEE Trans. Inf. Theory, vol. 65, no. 6, pp. 3646–3661, Jun. 2019.

[30] G. Reeves and M. C. Gastpar, “Approximate sparsity pattern recovery: Information-theoretic lower bounds,” IEEE Trans. Inf. Theory, vol. 59, no. 6, pp. 3451–3465, Jun. 2013.

[31] J. Scarlett and V. Cevher, “An introductory guide to Fano’s inequality with applications in statistical estimation,” in Information-Theoretic Methods in Data Science, M. R.D. Rodrigues and Y. C. Eldar, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2021, pp. 487–528.

[32] J. Scarlett and V. Cevher, “Converse bounds for noisy group testing with arbitrary measurement matrices,” in Proc. IEEE Int. Symp. Inf. Theory, 2016, pp. 2868–2872.

[33] R. J. Glauber, “Time-dependent statistics of the Ising model,” J. Math. Phys., vol. 4, no. 2, pp. 294–307, 1963.

[34] S. Geman and D. Geman, “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,” IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.

[35] R. Durbin, S. R. Eddy, A. Krogh, and G. Mitchison, Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[36] C. D. Manning and H. Schütze, Foundations of Statistical Natural Language Processing. Cambridge, MA, USA: MIT Press, 1999.

[37] S. Wasserman and K. Faust, Social Network Analysis: Methods and Applications, vol. 8. Cambridge, U.K.: Cambridge Univ. Press, 1994.

[38] A. Y. Lokhov, M. Vuffray, S. Misra, and M. Chertkov, “Optimal structure and parameter learning of Ising models,” Sci. Adv., vol. 4, no. 3, 2018, Art. no. e1700791.

[39] W. O. Kermack and A. G. McKendrick, “A contribution to the mathematical theory of epidemics,” Proc. Roy. Soc. London (Series A), vol. 115, no. 772, pp. 700–721, 1927.

[40] R. M. Karp, “Reducibility among combinatorial problems,” in Complexity of Computer Computations, R. E. Miller, J. W. Thatcher, and J. D. Bohlinger, Eds. New York, NY, USA:Plenum Press, 1972, pp. 85–103.

[41] S. A. Cook, “The complexity of theorem-proving procedures,” in Proc. 3rd ACM Symp. Theory Comput., 1971, pp. 151–158.

[42] F. Glover and E. Woolsey, “Technical note–Converting the 0-1 polynomial programming problem to a 0-1 linear program,” Oper. Res., vol. 22, no. 1, pp. 180–182, 1974.

[43] S. Burer and A. N. Letchford, “Non-convex mixed-integer nonlinear programming: A survey,” Surv. Operations Res. Manage. Sci., vol. 17, no. 2, pp. 97–106, 2012.

[44] Gurobi Optimization, LLC. Gurobi Optimizer Reference Manual. 2022. Accessed: Mar. 10, 2022. [Online]. Available: https://www.gurobi.com

[45] J. Leskovec and J. Mcauley, “Learning to discover social circles in ego networks,” in Proc. Adv. Neural Inf. Process. Syst., 2012, pp. 539–547.