Unconstraining Graph-Constrained Group Testing

Bruce Spang
Stanford University, CA, USA
bspang@stanford.edu

Mary Wootters
Stanford University, CA, USA
marykw@stanford.edu

Abstract

In network tomography, one goal is to identify a small set of failed links in a network using as little information as possible. One way of setting up this problem is called *graph-constrained group testing*. Graph-constrained group testing is a variant of the classical combinatorial group testing problem, where the tests that one is allowed are additionally constrained by a graph. In this case, the graph is given by the underlying network topology.

The main contribution of this work is to show that for most graphs, the constraints imposed by the graph are no constraint at all. That is, the number of tests required to identify the failed links in graph-constrained group testing is near-optimal even for the corresponding group testing problem *with no graph constraints*. Our approach is based on a simple randomized construction of tests. To analyze our construction, we prove new results about the size of giant components in randomly sparsified graphs.

Finally, we provide empirical results which suggest that our connected-subgraph tests perform better not just in theory but also in practice, and in particular perform better on a real-world network topology.

2012 ACM Subject Classification Theory of computation → Graph algorithms analysis

Keywords and phrases Group testing, network tomography, random graphs

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2019.46

Category RANDOM

Related Version A full version of the paper is available at `https://arxiv.org/abs/1809.03589`.

Acknowledgements We thank Clément Canonne, Nick McKeown and the anonymous reviewers for helpful comments.

1 Introduction

Suppose you run a network with n switches and m links between the switches. Occasionally links will fail, and it is your goal to find and fix them. One common approach is to have each switch send a test packet on its neighboring links and report the results to a central monitoring system. However, in large networks these monitoring systems – and the volume of data that they produce – can become hard to manage. In light of this, the problem (sometimes called *network tomography* [5]) is: how little information does this central system need to find failing links quickly?

In the version of this problem that we focus on, suppose that some set of at most d links fail. Instead of observing these failures directly, we may send test packets along any connected walks in the network, and we observe whether or not each packet reaches its destination.[1] The goal is to identify any set of up to d failed links while sending as few packets as possible.

[1] We explain a bit more about how this, or tests equivalent to this, might be implemented in Section 2.3.1.

We focus on the *non-adaptive* setting, where the packet walks must be fixed ahead of time. Non-adaptive tests are faster since they allow packets to be sent in parallel, and are easier to implement since these walks can be hard-coded once into the switches.

As observed by [6, 18], this problem is a variant of a well-studied problem called *combinatorial group testing*. Combinatorial group testing, originally motivated by the problem of cheaply testing for disease [10], has been studied since the 1940's and has applications from computational biology to wireless networks. We refer the reader to [11] for a survey. In the combinatorial group testing problem there are m items, at most d of which are "defective." A single test reveals whether or not there are any defectives in a subset T of items. The goal is to identify the defective items, by observing the output of a few tests.

The connection to network tomography is as follows: each link is defective if it fails, and each test T corresponds to a set of links. In the network tomography setting, there is one additional requirement: a test $T \subseteq [m]$ must correspond to a walk that a packet could take through the network. Because of this connection, [6] called this problem "graph-constrained group testing."

Our Question. A natural question is whether graph-constrained group testing is more difficult than classical group testing. That is, whether the additional constraints of graph-constrained group testing lead to significantly more tests. For the unconstrained group testing problem, the state-of-the-art construction uses $O(d^2 \log(m/d))$ tests [25], which nearly matches the lower bound of $\Omega(d^2 \log_d m)$ tests [13, 27]. Thus, our question is as follows:

▶ **Question 1.** *For what graphs G can we solve the graph-constrained group testing problem using $O(d^2 \log(m/d))$ tests?*

Previous work [18] has shown that certain graphs, such as a line, require $\Omega(m)$ tests, far more than the $O(d^2 \log(m/d))$ we would need for the unconstrained problem. However, it is also known that for sufficiently "well-connected" graphs (for example, those with large minimum cuts, many disjoint spanning trees, or constant mixing time), a sublinear number of tests suffice [6, 18]. These works have proposed using large subtrees [18] or random walks [6] as the tests. However, both of these approaches stop short (by polylogarithmic factors or more) of obtaining an $O(d^2 \log(m/d))$ bound for most graphs.

Our contributions. We improve upon the results of [6,18] to show that $O(d^2 \log(m/d))$ tests are sufficient for a wide collection of graphs, including many of the graphs already considered in prior work. Our construction – which is randomized – is quite simple: we sparsify the graph by choosing edges at random, and use the resulting large connected subgraphs. This is similar in flavor to earlier work – for example, [6] considered random walks – but our tests lead to stronger theorems, and also appear to perform better in practice. Moreover, our approach is quite general and works for other variants of the group-testing problem. To illustrate this, we show that it also gives near-optimal results in a model of random link failures. Concretely, our contributions are as follows:

- **$O(d^2 \log(m/d))$ tests suffice for (β, α)-edge expanders.** Our main result, Theorem 8, applies to graphs which are (β, α)-edge expanders, meaning that every set $S \subseteq V$ of size at most βn has at least $\alpha|S|$ edges coming out of it. We show that $O(d^2 \log(m/d))$ tests suffice when β is constant and $\alpha \gtrsim d$. Moreover, if β is sub-constant, then the number of tests required degrades gracefully with β.
  (β, α)-edge expansion is a general notion, and our results imply improved graph-constrained group-testing schemes for several natural classes of graphs like Erdös-Rényi graphs and constant-degree expanders. Moreover, our results are even optimal when applied to certain "counter-example" graphs like the barbell graph which foil earlier work.

Table 1 Summary of general results using connected-subgraph tests to identify any d defective edges for "well-connected" graphs with n vertices and m edges, for various notions of "well-connected." See discussion in Section 3 for slightly more general statements of results in previous work.

Source	Graph	Max. defective edges	Number of Tests
[18]	G has min-cut K	$d \leq \lceil \frac{K-1}{2} \rceil - 1$	$O(d^3 \log(m))$
[6]	G is D-regular, with mixing time τ	$d \leq d_0$ for some $d_0 = \Omega(D/\tau^2)$	$O(\tau^2 d^2 \log(m/d))$
Proposition 2	G has min-cut K	$d \leq \frac{K}{5 \log n}$	$O(d^2 \log(m/d))$
Theorem 8	G is a (β, α)-edge expander	$d \leq 1.99\alpha$	$O\left(\frac{d^2 \log(m/d)}{\beta}\right)$

Our general theorem (Theorem 8) is compared to existing general theorems in Table 1. The results for a few specific families of graphs are shown in Table 2. These results are presented in more detail in Section 4.

- **New results about large connected components of random graphs.** While our construction is quite simple, the analysis requires some delicacy. In order to show that our tests work, we prove new results about giant components in randomly sparsified graphs.

  More precisely, our main technical theorem (Theorem 10) establishes the following. Suppose that $G = (V, E)$ is a (β, α)-edge expander, and let $G(p) = G(V, E')$ be the graph where $E' \subseteq E$ is a random subset where each edge is kept independently with probability p. We show that if $p \geq (1 + \varepsilon)/\alpha$, then for any edge $e \in E$, with probability $\Omega(p\varepsilon)$ then not only does e survive, but also e's connected component in $G(p)$ has size at least βn. This is of a similar flavor to previous work on giant components of randomly sparsified graphs, but with two important differences: first, our result works even if β is small (so the components are "big" but not "giant") and second, we require that any edge e be contained in a large component with decent probability after sparsification. Theorem 10 is stated and proved in Section 5.

- **A general algorithm for graph-constrained group testing problems.** Our approach is to simulate a uniformly random (unconstrained) approach in the constrained setting. Because a uniformly random approach is near-optimal for many variants on the group testing problem, our results extend to variants of the graph-constrained group testing setting. We illustrate this in the full version of this paper, where we show that our algorithm also achieves near-optimal results in the stochastic model (where the link failures are random) [28].

- **Empirical results.** Finally, we present empirical results which suggest that our approach significantly out-performs the random-walk method of [6] and in many cases, nearly matches the performance of unconstrained random tests. On complete graphs and hypercubes it uses less than half the tests of the random-walk method. On a family of graphs often used in datacenter networks (the "fat-tree" topology), our approach is able to find defectives using a nontrivial number of tests while the random-walk method is not.

Table 2 Summary of work on the number of connected-subgraph tests required to identify any $d \leq d_0$ failures, for specific families of graphs. All graphs have n vertices and m edges. Results which meet the near-optimal $O(d^2 \log(m/d))$ tests or which have only the asymptotically optimal restriction $d \leq d_0$ for some $d_0 = \Omega(\text{degree})$ are highlighted in grey.

Graph	**Source**	**Number of tests**	**Limit d_0 so that recovery of $d \leq d_0$ failures is possible**
Complete Graphs	[6]	$O(d^2 \log(m/d))$	$d_0 = \Omega(n)$
	This work	$O(d^2 \log(m/d))$	$d_0 = \Omega(n)$
D-Regular Expanders ($O(1)$ spectral gap)	[18]	$O(d^3 \log m)$	$d_0 = \Omega(D)$
	[6]	$O(d^2 \log^3(m))$	$d_0 = \Omega(D/\log^2(n))$
	This work	$O(d^2 \log(m/d))$	$d_0 = \Omega(D)$
Erdös-Rényi Graphs $G(n, D/n)$	[6]	$O(d^2 \log^3(m))$	$d_0 = \Omega(D/\log^2(n))$
	This work	$O(d^2 \log(m/d))$	$d_0 = \Omega(D)$
Barbells	[18]	$O(d^3 \log m)$	$d_0 = 1$
	[6]	m (see discussion)	$d_0 = n$
	This work	$O(d^2 \log(m/d))$	$d_0 = \Omega(n)$
Fat-Trees	[18]	$O(d^3 \log m)$	$d_0 = 1$
	This work	$O(d^2 \log(m/d))$	$d_0 = \Omega(D/\log D)$

1.1 Overview of approach

Our approach is quite simple: we just choose random edges and take the large connected components. Intuitively, the reason that this works is because:

1. If we just chose random edges, we would be back in the traditional group testing setting, where random tests are nearly optimal by Proposition 6.
2. We will show that if G is a (β, α)-edge expander, then the graph $G(p)$ formed by choosing random edges has mostly large components, of size at least βn. Intuitively this means that throwing away the few disconnected parts should not matter much.

There are several challenges in making the above intuition rigorous. First, once we throw away the edges that are not in a large connected component, the edges that remain, conditional on remaining, are no longer independent. Thus, the intuition from point 1 above does not quite hold. However, this ends up being reasonably straightforward to deal with.

The second and more interesting challenge is that, while there is a great deal of work on when random sparsifications of graphs have giant components, we need a different result that to the best of our knowledge does not appear in the literature. The first difference between our setting and existing work is that we work with (β, α)-edge expanders; this means that we need to show that there are "decently big" connected components rather than "giant" components if β is small. The second difference is that we must show that each edge e is still contained in a test with high probability. That is, when we pass from G to $G(p) = (V, E')$,

the probability that $e \in E'$ is p; for the analysis in the random case to still work, we need the probability that e is in a large connected component of E' to also be proportional to p.

To address this second challenge, we reduce the question to one about random walks. Taking inspiration from [22] who study random sparsifications of the complete graph, we introduce a process to generate the connected component of a particular edge v, and argue that this process generates a large connected component if and only if an appropriately chosen random walk diverges with decent probability; the we prove it does converge.

The details of the approach are given in Section 5. However, as a warm-up to show why the intuition presented above is believable, we first prove an easy statement of the same flavor where neither of these challenges arise. Proposition 2 below shows that if we replace (β, α)-expansion with the property of having a large min-cut, $O(d^2 \log(m/d))$ tests suffice.

▶ **Proposition 2** (Informal). *There is a constant $C > 0$ so that the following holds. Suppose that $G = (V, E)$ is a graph with $|V| = n$, $|E| = m$. Let $d \geq 1$ be an integer. Suppose that the minimum cut K of G satisfies $K \geq 5(d+1)\log(n)$. Let $\mathcal{T}$ be a set of tests $T \subseteq E$ generated according to the following process:*

- *Initialize $\mathcal{T} = \emptyset$.*
- *For $t = 1, 2, \ldots, Cd^2 \log(m/d)$:*
  - *Let $T \subset E$ be a random set where each edge is included with probability $1/(d+1)$.*
  - *Add T to $\mathcal{T}$.*

Then $\mathcal{T}$ can identify d failed edges, and with high probability each test is connected.

Proof sketch. As we will see in Proposition 6, $\mathcal{T}$ is able to identify d failed edges, and so it suffices to show that a random set T is connected with high probability. Fortunately, this is true:

▶ **Theorem 3** ([21]). *Let K be the minimum cut of G. If $p > \min\left(\frac{5\log n}{K}, 1\right)$, $G(p)$ is connected with probability at least $1 - \frac{1}{n}$.*

By Theorem 3 and a union bound, the probability any test is disconnected is at most $|\mathcal{T}|/n = O\left(\frac{d^2 \log(m/d)}{n}\right)$. ◀

Proposition 2 is already enough to establish order-optimal results for some of the examples shown in Table 2 (for example the complete graph and the fat tree), but for the others we will need Theorem 8 about (α, β)-expanders.

Organization. In Section 2, we formally set up the problem. In Section 3, we survey related work, and we state our theoretical results in Section 4. The proofs of these results follow in Section 5. Finally, we present our empirical results in Section 6.

2 Setup and Preliminaries

We begin with some basic notation and definitions.

2.1 Graph-theoretic preliminaries

Throughout, we will be working with undirected, unweighted graphs $G = (V, E)$ with $|V| = n$, $|E| = m$. For a set of vertices $A \subseteq V$, the *boundary* of A is $\partial A = \{\{u, v\} \in E : u \in A, v \notin A\}$. For a set of edges $B \subseteq E$, we use the notation $N(B)$ to denote the set of vertices v that are endpoints of an edge in B: $N(B) = \{v \in V : \exists u, \{u, v\} \in B\}$.

The minimum cut K of a graph G is defined by $K = \min_{A \subseteq V} |\partial A|$. Our main theorem is about *edge expanders*. We give a slightly more general definition than the usual notion (which would have $\beta = 1/2$ below), so that we can state a more general theorem.

▶ **Definition 4.** *A graph $G = (V, E)$ is a (β, α)-edge expander if for all sets $A \subseteq V$ with $|A| \leq \beta|V|$, $|\partial A| \geq \alpha|A|$.*

We will consider random sparsifications of graphs. For a graph $G = (V, E)$ and $p \in (0, 1)$, $G(p) = (V, E')$ denotes the random graph where $E' \subseteq E$ is generated by including each edge of E in E' independently with probability p. We use $G(n, p) = K_n(p)$ to denote the Erdös-Rényi graph where each edge is included independently with probability p. (Here, K_n denotes the complete graph on n vertices).

2.2 Group testing preliminaries

The combinatorial group testing problem is set up as follows (using slightly non-standard notation in order to be consistent with the graph-constrained set-up below). Let E be a set of size m, and suppose that $B \subseteq E$ is a set of at most d special or "defective" items in E. A *test* $T \subseteq E$ is a collection of items, and we say that the *outcome* of the test T is TRUE if $T \cap B \neq \emptyset$ and FALSE otherwise. We say that a collection of tests $\mathcal{T} \subseteq 2^E$ (here, 2^E denotes the power set of E, consisting of all subsets of E) *can identify up to d defective items in E* if for any $B \subseteq E$ with $|B| \leq d$, B is uniquely determined from the outcomes of the tests $T \in \mathcal{T}$. The goal is to design a collection of tests $\mathcal{T} \subseteq 2^E$ which can identify up to d defective items, so that $|\mathcal{T}|$ is as small as possible. A useful notion in the group testing literature is *disjunctness*, which is a sufficient condition for recovery.

▶ **Definition 5.** *Let E be a set and $\mathcal{T} \subseteq 2^E$. We say that $\mathcal{T}$ is* d-disjunct *if for all $e \in E$, for all $B \subseteq E$ where $|B| \leq d$ and $e \notin B$, there exists a test $T \in \mathcal{T}$ so that $e \in T$ and $B \cap T = \emptyset$.*

If $\mathcal{T}$ is d-disjunct, then $\mathcal{T}$ can identify up to d defective items in E. More precisely, it is not hard to see that the following algorithm will do the job: for each item $e \in E$, declare $e \in B$ if and only if all the tests $T \in \mathcal{T}$ with $e \in T$ had outcome TRUE.

Choosing tests completely at random is a good way to obtain d-disjunct sets.

▶ **Proposition 6** (See e.g. [11] Theorem 8.1.3)**.** *Let $d \geq 1$. Let E be a set. Consider a random test $T \subseteq E$ such that each $e \in E$ is included in T independently with probability $p = \frac{1}{d+1}$. Let $\mathcal{T} = \{T_1, \ldots, T_\tau\}$, where each $T_i \in \mathcal{T}$ is chosen independently from the above distribution. Then there is a value $\tau = O(d^2 \log(m/d))$ so that $\mathcal{T}$ is d-disjunct with probability at least $1 - 1/m$.*

This is nearly optimal, up to a factor of $O(1/\log d)$:

▶ **Theorem 7** ([13])**.** *Let E be a set of size m and $\mathcal{T} \subseteq 2^E$. If $\mathcal{T}$ is d-disjunct, then $|\mathcal{T}| = \Omega(d^2 \log_d m)$*

2.3 Graph-constrained group testing

Given a graph $G = (V, E)$, the graph-constrained group testing problem on G is the same as the standard group testing problem on a set of items E, with the additional constraint that each test $T \subseteq E$ be a connected subgraph of G. That is, in the network tomography application, a packet must be able to traverse a test T. We say that a *connected-subgraph test* is a set of edges $T \subseteq E$ so that the graph $(N(T), T)$ is connected.

We note that the definition of disjunctness directly applies to the graph-constrained setting, and our goal in this work will be to design d-disjunct collections $\mathcal{T}$ of connected-subgraph tests, such that $\mathcal{T}$ is as small as possible.

2.3.1 Why connected-subgraph tests?

Our work, like existing work on graph-constrained group testing [6, 18, 20], uses connected-subgraph tests. Connected-subgraph tests can be implemented in an actual network in a number of ways:

- The test can be converted into a path on the graph by solving an instance of the Chinese Postman Problem [14]. A packet can then be sent along this path, for instance by using source routing or by adding a rule at each switch on the path matching the packet's source IP, destination IP, and time-to-live.
- We can compute a spanning tree of subgraph. Each switch in the subgraph checks the health of its links, and forwards a bit to its parent in the tree: 1 if its adjacent links are healthy, and 0 otherwise. If the link from a switch to its parent in the tree is down, then it cannot send anything. If the root node receives all 1's, it knows that all the links in the subgraph are healthy; otherwise at least one of the links is broken.

One could also imagine restricting the tests to be, for example, simple paths or trees. Some restrictions on tests do have some advantages in implementation (in particular, tests which are *shortest* paths may be easier to implement than general connected-subgraph tests: for example many networks support equal-cost multipath routing (ECMP) which splits traffic across all the shortest paths between a pair of hosts). However, connected subgraph tests are strictly more powerful than either simple paths or trees for the constrained group-testing problem. We provide some examples which demonstrate this in the full version of this paper [28].

3 Related Work

Boolean network tomography. Most of the work on *boolean network tomography* (that is, the problem of identifying failures in a graph using end-to-end traffic) has a much harsher set-up than the one we consider here, in that both the graph and the tests are taken to be worst-case, or at least very constrained. For example, all the tests may be required to be simple paths starting from a particular vertex. The reason for the harsh set-up is that historically, networks have been quite inflexible, which severely limits both the graph topologies and the sorts of tests that are used. Since identifying failures uniquely is often impossible in these settings, this work has focused on doing as well as possible given the circumstances, for example by finding *any* set of failures that will explain the test outcomes [3, 8] or by finding the most likely set of failures given some underlying distribution [12, 24]. When the input graph and set of allowed tests is worst-case, these problems are hard, and the usual approach is to reduce to some NP-hard problem and use a heuristic or approximation algorithm.

Graph-constrained group testing. More recently, the field of networking has shifted towards flexible datacenter networks, where the tomography problem is still interesting [26, 30]. Modern datacenter networks, however, fundamentally change the constraints of the tomography problem. Datacenters are good expanders [9, 29], so the worst-case assumptions about the graphs can be relaxed. Modern networks are programmable [4]: instead of the network defining what can and cannot be done, operators program networks to do what they want. Thus, the set of allowed tests $\mathcal{T}$ need not be worst-case. This leads to graph-constrained group testing, where we can design the tests, and make assumptions about the connectivity of the underlying network.

We are not the first to investigate group testing with graph constraints (although to the best of our knowledge our work is the first to consider graph-constrained group testing with random failures). Du and Hwang discuss two different group-testing problems on graphs in Chapter 12 of [11], although neither are exactly the same as the setup we consider here. The connection between boolean network tomography and group testing was first observed by [18]. They give results for specific families of graphs including line graphs, grids, and binary trees. Their most general result is that if a graph has d edge-disjoint spanning trees $T_1, \ldots, T_d$ with δ being the maximum diameter of the trees, then $O(d^3 \log m + d \min(\delta + \log^2 n, \delta \log n))$ tests are sufficient to identify at most d failures. (As a corollary, this implies that any graph with minimum cut K can identify $d \leq \lceil \frac{K-1}{2} \rceil - 1$ failed edges, which is what is stated in Table 1). Finally, [20] considers the adaptive version of graph-constrained group testing. They present an adaptive algorithm which, on any graph, uses a number of tests that is within a constant factor of the optimal number.

The work closest to our is that of Cheraghchi et al. [6], who give a randomized construction of connected-subgraph tests via random walks. They show that $O(d^2 \log(m/d))$ tests suffice for *very* well-connected graphs (those with constant mixing time), and $O(d^2 \log^3(m))$ tests suffice for certain expanders and Erdös-Réyni graphs. Their most general result is that for graphs with mixing time τ and where there exists some $c > 0$ such that the degree D_v of each vertex $v \in V$ lies in between $6c^2 d\tau^2 \leq D_v \leq 6c^3 d\tau^2$, at most $O(c^4 \tau^2 d^2 \log(m/d))$ tests are sufficient. We summarize the general results for related work in non-adaptive graph-constrained group testing in Table 1.

Giant components in random graphs. Finally, we mention some related work on the size of giant components of randomly sparsified graphs, since our main technical theorem (Theorem 10) is related to this. This question is well-studied, but we need a slightly different result. One difference is that we work with (β, α)-edge expansion, and in particular our "giant" components need not be so giant if β is small. A second difference is that we require that every edge be contained in a large connected component with constant probability; to the best of our knowledge, existing work does not explicitly give such a guarantee.

The study of giant components in $G(n, p)$ (aka, a randomly sparsified complete graph) was initiated by Erdös and Rényi in [15]. This was extended to sparsifications of the hypercube [1] and sufficiently good expander graphs [16] and [7]. Our approach to Theorem 10 is based on that of Krivelevich and Sudakov [22] who give a simpler argument for existing results on giant components.

Most of these results show that as long as $p \leq \frac{1+\varepsilon}{D}$, where D is the degree of the graph G, then $G(p)$ contains a giant component. We show a similar result: if $p \leq \frac{1+\varepsilon}{\alpha}$, where G is a (β, α)-edge expander, then there exists a connected component of size at least βn. (And moreover, any edge is contained in such a component with decent probability).

4 Results

4.1 Main result

Our group testing scheme is quite simple: the idea is just to choose random edges of the graph, and keep any large-enough connected components. Recall the notation that for a graph $G = (E, V)$, $G(p) = (V, E')$ is the graph where each edge in E is included in E' independently with probability p. Then the randomized algorithm for constructing the tests is given in Algorithm 1.

Algorithm 1 Make-Tests.

input : Graph $G = (V, E)$; number of failed edges d; parameters $\frac{2}{d} \leq \delta \leq 1/3$, $\beta \in (0, \frac{1}{2}]$, and $\tau \in \mathbb{N}$
output : A collection of tests $\mathcal{T} \subseteq 2^E$

```
T ← ∅;
p ← 1/(δd);
for t = 1, ..., τ do
    Draw G' ~ G(p) independently from all the other rounds;
    Find the connected components A_1, A_2, ..., A_r of G(p);
    for each A_i so that |A_i| ≥ βn do
        Add the test A_i to T;
    end
end
return T
```

Our main theorem implies that any (β, α)-edge expander with large enough α admits a group testing scheme with $O(d^2 \log(m/d)/\beta)$ tests.

▶ **Theorem 8.** *There are constants $c, C > 0$ so that the following holds. Suppose that $G = (V, E)$ is a graph with $|V| = n, |E| = m$. Let $d \geq 1$ be an integer, and $\frac{2}{d} \leq \delta \leq \frac{1}{3}$. Suppose that G is a (β, α)-edge-expander with $\beta \in (0, 1/2]$ and such that $\alpha \geq 1$ satisfies*

$$\alpha \geq d\left(\frac{1}{2} + \delta\right). \tag{1}$$

Let $\mathcal{T}$ be the set of tests returned by Algorithm 1 run with parameters δ, β, and

$$\tau = Cd^2 \log(m/d) e^{(1+\delta)/\delta}.$$

Then with probability at least $1 - m^{-cd}$, $\mathcal{T}$ is d-disjunct. Further,

$$|\mathcal{T}| \leq \frac{Cd^2 \log(m/d) e^{(1+\delta)/\delta}}{\beta}.$$

If β, δ are constant, then the number of tests required is $O(d^2 \log(m/d))$, which is nearly optimal even for the unconstrained group testing problem (that is, it nearly matches Theorem 7). Moreover, the requirement on the expansion factor α is nearly tight. That is, in (1), we may take $\alpha = d\left(\frac{1}{2} + \delta\right) d$ for any constant $\delta > 0$, while still maintaining the near-optimal $O(d^2 \log(m/d))$ test complexity. On the other hand, such a statement could not hold for α much smaller than $d/2$. More precisely, we show below in Proposition 9 that the degree D of the graph G must be at least $d/2$ to obtain any nontrivial bound on $|\mathcal{T}|$. Since for any $\gamma > 0$, there exist D-regular (β, α)-edge expanders with $\alpha = (1 - \gamma)D$ for small $\beta \sim \gamma$, this implies that the cut-off for α of $d/2$ in Theorem 8 cannot be improved.

▶ **Proposition 9.** *Let $G = (V, E)$ be a D-regular graph with $|E| = m$. If $\mathcal{T} \subseteq 2^E$ is a collection of d-disjunct connected-subgraph tests, for $d \geq 2D - 2$, then $|\mathcal{T}| \geq m$.*

Proof. Suppose that $\mathcal{T} \subseteq 2^E$ with $|\mathcal{T}| < m$. Then there is some edge $e = \{u, v\}$ so $\{e\} \notin \mathcal{T}$. Let $B = \partial(\{u, v\})$ be the set of edges adjacent to e, so $|B| = 2D - 2 \leq d$. Then the only connected-subgraph test $T \subset E$ so that $e \in T$ but $T \cap B = \emptyset$ is $\{e\}$, which by assumption is not in $\mathcal{T}$. Thus, $\mathcal{T}$ is not d-disjunct. (See Figure 1). ◀

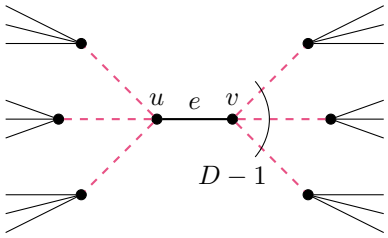


Figure 1 Proof of Proposition 9. Suppose that the number of edges that may fail is $d \geq 2D-2$ where D is the degree of the graph. If the set of tests $\mathcal{T}$ does not contain the singleton $\{e\}$ for an edge $e = \{u, v\}$, then the set B of dashed edges – all of the neighbors of e – provides an counter-example to d-disjunctness.

We instantiate Theorem 8 for several differnt families of graphs in Appendix A; the results are summarized in Table 2. We note that our results also extend to a model with random link failures; we defer this discussion to the full version of this paper [28].

5 Proofs

In this section, we prove Theorem 8. Our proof is based on the following theorem, which implies that any edge e is reasonably likely to be contained in a large connected component of $G(p)$.

▶ **Theorem 10.** *Let $\beta \in (0, 1/2)$ and $\alpha \geq 1$, and let $G = (V, E)$ be a graph with $|V| = n, |E| = m$ so that, for any set $A \subset V$ of size $2 \leq |A| \leq \beta n$, we have $|\partial A| \geq \alpha |A|$. For an edge $e \in E$, let C_e denote the connected component of $G(p)$ containing e, or $\emptyset$ if there is no such connected component (that is, if $e \notin G(p)$ was deleted), and let $|C_e|$ denote the number of vertices in C_e.*

Choose any $\varepsilon \in (0, 1/3)$, and suppose that $p \geq \frac{1+\varepsilon}{\alpha}$. Then for all edges $e \in E$,

$$\mathbb{P}\left(|C_e| \geq \beta n\right) \geq \frac{p\varepsilon}{8}. \tag{2}$$

Before we prove Theorem 10, we discuss how it can be used to prove Theorem 8.

Let $G = (V, E)$, and let $G(p) = (V, E')$ be the random sparsification. Fix $B \subseteq E$ with $|B| = d$ and $e \in E$. The following lemma shows that with high probability, at least one of the tests in $\mathcal{T}$ will separate e from B. Then one can union bound over all choices for e and B to conclude that $\mathcal{T}$ is d-disjunct. We defer the details to the full version of this paper [28].

▶ **Lemma 11.** *Consider one draw of $G(p)$ in Make-Tests, and let $T_1, T_2, \ldots$ be the connected components of $G(p)$ which have size at least βn. Fix $B \subseteq E$ with $|B| = d$ and $e \in E$. Then*

$$\mathbb{P}\left(\exists i \text{ s.t. } e \in T_i \text{ and } B \cap T_i = \emptyset\right) \geq \frac{\delta p}{8} \cdot (1-p)^d.$$

Proof. We have

$$\begin{aligned}
\mathbb{P}\left(\exists i \text{ s.t. } e \in T_i \text{ and } B \cap T_i = \emptyset\right) &= \mathbb{P}\left(|C_e| \geq \beta n \text{ and } B \cap C_e = \emptyset\right) \\
&\leq \mathbb{P}\left(|C_e| \geq \beta n \text{ and } B \cap E' = \emptyset\right) \\
&= \mathbb{P}\left(|C_e| \geq \beta n \mid B \cap E' = \emptyset\right) \cdot \mathbb{P}\left(B \cap E' = \emptyset\right).
\end{aligned}$$

We have $\mathbb{P}\left(B \cap E' = \emptyset\right) = (1-p)^d$, since this is just the probability that all the edges in B survive in $G(p)$.

For our fixed e, B, let $\bar{G} = (V, E \setminus B)$ be the graph with all the edges in B removed. Consider the distribution of $G(p)$ conditioned on the event that $B \cap E' = \emptyset$. This is the same as the distribution of $\bar{G}(p)$. To see this, notice that for $A \subseteq E \setminus B$, the random sets of $A \cap E'$ and $B \cap E'$ are independent. Let $\overline{C_e}$ be the connected component containing e in $\bar{G}(p)$. Then

$$\mathbb{P}\left(|C_e| \geq \beta n \mid B \cap E' = \emptyset\right) = \mathbb{P}\left(|\overline{C_e}| \geq \beta n\right).$$

Notice that since G is a (β, α)-edge expander with $\alpha \geq d\left(\frac{1}{2} + \delta\right)$, then for any set $A \subset V$ of size $2 \leq |A| \leq \beta n$, we have $|\partial A| \geq \alpha|A| - d \geq \left(\alpha - \frac{d}{2}\right)|A|$. Thus, we may apply Theorem 10 to $\bar{G}$. Choose $p = \frac{1+\delta}{d\delta}$, and apply Theorem 10. Our assumption (1) that $\alpha \geq d\left(\frac{1}{2} + \delta\right)$ and the choice of p implies that $p \geq \frac{1+\delta}{\alpha - d/2}$, we conclude that $\mathbb{P}\left(|\overline{C_e}| \geq \beta n\right) \geq \frac{\delta p}{8}$. Putting things together proves the lemma. ◀

Proof of Theorem 10. As in the theorem statement, suppose $p \geq (1+\varepsilon)/\alpha$ for some $\varepsilon \in (0, 1/3)$, and let $G = (V, E)$ be a (β, α)-edge expander. Write $G(p) = (V, E')$, so that $E' \subseteq E$. Choose $p \geq (1+\varepsilon)/\alpha$.

Consider the probability (u, v) is in a large component,

$$\mathbb{P}\left(C_{(u,v)} \geq \beta n\right) = \mathbb{P}\left((u,v) \in E'\right) \cdot \mathbb{P}\left(|C_{(u,v)}| \geq \beta n \mid (u,v) \in E'\right).$$

Condition on the event that $(u, v) \in E'$, and imagine building $C_{(u,v)}$ by starting with $\{(u, v)\}$ and building the set outwards, one at a time. More precisely, consider the following randomized process:

- $S_0 \leftarrow \{(u, v)\}$, $B_0 \leftarrow \emptyset$
- For $t = 0, 1, 2, \ldots$:
  1. Let $N(S_t)$ be the set of vertices in V adjacent to S_t.
  2. Let $U_t = \partial(N(S_t)) \setminus B_t$ be the set of unvisited edges that lie on the boundary of $N(S_t)$.
  3. If $|U_t| = 0$, **break**.
  4. $S_{t+1} \leftarrow S_t$, $B_{t+1} \leftarrow B_t$.
  5. Choose an edge $e \in U_t$ arbitrarily.
  6. With probability p, declare that e has survived and add it to S_{t+1}.
  7. Otherwise (with probability $1 - p$) add e to B_{t+1}.

This process is illustrated in Figure 2.

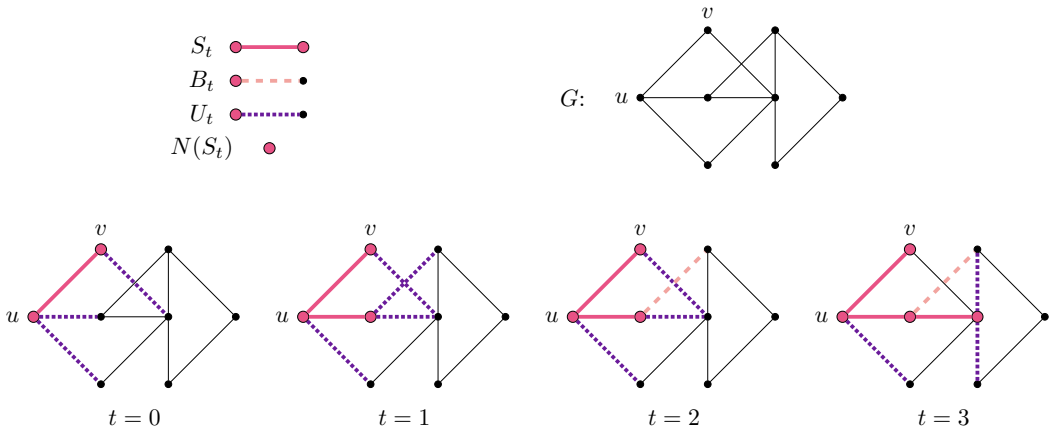


Figure 2 First few steps of the process to build $C_{(u,v)}$ in the proof of Theorem 10. The edge from U_t which we chose ended up being in E' in steps $t = 1$ and $t = 3$, but not $t = 2$.

It is not hard to see that this process terminates at the first time $t_{\max}$ so that $|U_{t_{\max}}| = 0$, and when it does, $N(S_{t_{\max}})$ is distributed identically to the set of vertices in $C_{(u,v)}$. Thus, to bound $|C_{(u,v)}|$ with high probability, we can bound the set $|N(S_t)|$ with high probability. Notice that S_t is a tree; thus, $|S_t| = |N(S_t)|-1$, and so it suffices to show that $|S_{t_{\max}}| > \beta n-1$ with high probability. To that end, we will show that, as long as $|S_t| \leq \beta n-1$, the probability that $|U_t| = 0$ is very small.

At each step t, we either add an edge to B_t or to S_t, so $|S_t| + |B_t| = t$. Let X_t be the random variable which is 1 if we added an edge to S_t in step t; thus $X_t \sim \text{Ber}(p)$ and $|S_t| = 1 + \sum_{i=1}^{t} X_i$.

Suppose that S_t is nonempty and $|S_t| \leq \beta n - 1$, so $2 \leq |N(S_t)| \leq \beta n$. By our expansion assumption, we have $|\partial N(S_t)| \geq \alpha |S_t|$, and so

$$\begin{aligned}|U_t| &= |\partial(N(S_t)) \setminus B_t| \geq \alpha|S_t| - |B_t| = \alpha|S_t| - (t - |S_t|) = (1+\alpha)|S_t| - t \\ &= (1+\alpha)\left(1 + \sum_{i=1}^{t} X_i\right) - t \geq \sum_{i=1}^{t}((1+\alpha)X_i - 1) + \alpha.\end{aligned}$$

Thus,

$$\begin{aligned}\mathbb{P}\left(\exists t > 0 \quad \text{s.t.} \quad |S_t| \leq \beta n - 1 \text{ and } |U_t| = 0\right) &= \mathbb{P}\left(\exists t > 0 \quad \text{s.t.} \quad |S_t| \leq \beta n - 1 \text{ and } |U_t| \leq 0\right) \\ &\leq \mathbb{P}\left(\exists t > 0 \quad \text{s.t.} \quad \sum_{i=1}^{t}((1+\alpha)X_i - 1) \leq -\alpha\right).\end{aligned}$$

Let $Y_i = (\alpha+1)X_i - 1$, so that $Y_i = \alpha$ with probability p, and $Y_i = -1$ with probability $1-p$. Let $Z_t = \sum_{i=1}^{t} Y_i$ be a random walk. The above shows that

$$\mathbb{P}\left(\exists t > 0, |S_t| \leq \beta n - 1 \text{ and } |U_t| = 0\right) \leq \mathbb{P}\left(\exists t > 0, Z_t \leq -\alpha\right). \tag{3}$$

Let $\tau_k(Y_i)$ denote the smallest t so that $\left|\sum_{i=1}^{t} Y_i\right| \geq |k|$ and $\text{sign}(\sum_{i=1}^{t} Y_i) = \text{sign}(k)$. We claim that $\mathbb{P}\left(\exists t > 0, Z_t \leq -\alpha\right) \leq 1 - \inf_{M>0} \mathbb{P}\left(\tau_M(Y_i) \leq \tau_{-\alpha}(Y_i)\right)$. Indeed, suppose that there is some $t > 0$ so that $Z_t \leq -\alpha$. Then there is some sufficiently large $M > \alpha t$ so that Z_t could not have reached M in t steps, and so the random walk does not arrive at M before arriving at $-\alpha$. Thus

$$\begin{aligned}\mathbb{P}\left(\exists t > 0, Z_t \leq -\alpha\right) &\leq \mathbb{P}\left(\exists M, \tau_M(Y_i) > \tau_{-\alpha}(Y_i)\right) \\ &= \sup_M \mathbb{P}\left(\tau_M(Y_i) > \tau_{-\alpha}(Y_i)\right) = 1 - \inf_{M>0} \mathbb{P}\left(\tau_M(Y_i) \leq \tau_{-\alpha}(Y_i)\right).\end{aligned}$$

Thus our goal is to show that $\mathbb{P}\left(\tau_M(Y_i) \leq \tau_{-\alpha}(Y_i)\right)$ is bounded away from zero for all M; then there will be some constant probability that the edge-exploration process will make progress as long as $|S_t| \leq \beta n - 1$.

▷ Claim 12. Using the notation above, for all $M > 0$,

$$\mathbb{P}\left(\tau_M(Y_i) \leq \tau_{-\alpha}(Y_i)\right) \geq \frac{\varepsilon}{8}.$$

The proof of the claim can be found in the full version of this paper [28]. The main idea is as follows: suppose for simplicity that $\alpha \in \mathbb{Z}$. Then we replace the walk defined with the Y_i's by one whose steps W_i are ± 1, where the probability of 1 is chosen so that the the probability that W_i reaches α before -1 is at most p. Analyzing this walk is then an instance of the Asymmetric Gambler's Ruin problem (see, for example, [17]).

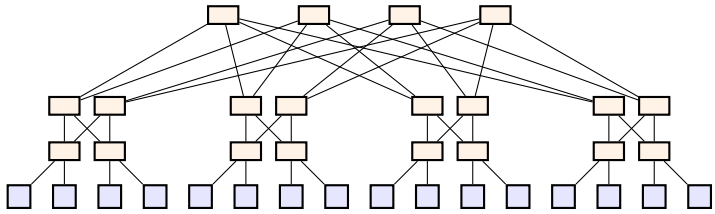


Figure 3 A small example of the fat tree topology with $n = 36, m = 48$. In our experiments, we consider larger versions ($n = 80, m = 256$ and $n = 45, m = 108$).

Using the claim and the discussion following (3), we have that

$$\mathbb{P}\left(\exists t > 0 \text{ s.t. } |S_t| \leq \beta n - 1, |U_t| = 0\right) \leq 1 - \varepsilon/8.$$

Finally,

$$\begin{aligned}\mathbb{P}\left(|C_{(u,v)}| \geq \beta n\right) &= \mathbb{P}\left(|C_{(u,v)}| \geq \beta n \mid (u,v) \in E'\right) \cdot \mathbb{P}\left((u,v) \in E'\right) \\ &= p \cdot \left(1 - \mathbb{P}\left(\exists t \text{ s.t. } |S_t| \leq \beta n - 1, |U_t| = 0\right)\right) \geq \frac{p\varepsilon}{8}.\end{aligned}$$

This completes the proof. ◀

6 Empirical Results

In this section, we numerically compare Algorithm 1 to existing work. We compare to the random walk based approach of [6] and to the randomized group testing *without* graph constraints of Proposition 6. We find that the random subgraph tests perform nearly as well as the unconstrained versions in most settings, and often perform significantly better than the random walk approach.

Below, we test the following randomized constructions of tests:

- Our approach, Algorithm 1 (called "Subgraph" in the figures). We use $p = 1/(d+1)$ and include only the largest connected component of $G(p)$.
- The random walk approach of [6] (called "Random walks" in the figures). We empirically estimate the mixing time τ by picking a node at random and finding the first time that the total variation distance to the equilibrium distribution is less than $1/(2cn)^2$, as per the definition in [6], where c is defined so that the graph $G = (V, E)$ has $D \leq \deg(v) \leq cD$ for each $v \in V$. We fix a constant $\ell > 0$ and run each random walk for $\lceil \frac{\ell n D}{c^3 d \tau} \rceil$ steps. We tried a few of values of ℓ and present the best results for each graph: for the complete graph we chose $\ell = 1$ and for all other graphs we chose $\ell = 4$.
- Unconstrained random approach (called "Random" in the figures). We include each edge in a test with probability $p = 1/(d+1)$, and ignore the graph constraints.

We consider four types of graphs. The first three – random regular graphs, complete graphs, and hypercubes – are idealized graphs that may or may not capture real networks. For our last graph, we choose the "Fat-Tree" graph [23], originally designed for use in supercomputers and which is now widely used in datacenter networks [2]. As the name suggests, this is a "fattened" tree, where the fatness (number of links) near the top of the tree is greater than the fatness near the leaves. (See Figure 3).

We perform two types of experiments:

1. In the first type of experiment, we compare the probability of obtaining a d-disjunct matrix from any of these three randomized approaches. Unfortunately, it is computationally intense to determine whether or not a given collection $\mathcal{T}$ of tests is d-disjunct, and so we are only able to do this for small d ($d = 1$ and $d = 2$).
2. In the second type of experiment, we are trying to understand the performance of our method for larger d. Since determining d-disjunctness is computationally infeasible for large d, instead we choose d random defectives and estimate the probability of success under each of the three methods. We note that, as shown in the full version of this paper [28], our algorithm is order-optimal for random defectives.

6.1 Number of tests required for d-disjunctness

First, we estimate the number of tests required for d-disjunctness for Algorithm 1 and compare it to [6] and randomized group testing without graph constraints. We find that for many graphs, our approach requires roughly the same number of tests as group testing without constraints.

Figure 4 (resp. Figure 5) shows the probability that a randomly generated test matrix is 1-disjunct (resp. 2-disjunct) for various graphs, algorithms, and numbers of tests. Each point is the empirical mean of 200 independent trials, and we plot error bars of width $1/\sqrt{200} \approx 0.07$. (Notice that by Hoeffding's inequality, the probability that the true average lies outside the error bars is at most $1/e^2$).

Algorithm 1 performs similarly to the nearly optimal randomized group testing procedure of Proposition 6. Notably, for the Fat-Tree graph, Algorithm 1 significantly outperforms the approach of [6].

6.2 Number of tests required for d random failures

As mentioned above, determining d-disjunctness is computationally infeasible for larger d, and so to assess larger d we consider performance on random failures, which we address in the full version [28].

For our experiments with random failures, we focus on the fat-tree topology. The main reasons for this are (a) that the "Fat-Tree with random failures" set-up is perhaps the most relevant for real-life applications, and (b) the other topologies yield graphs that look similar, but the differences between the three approaches are less pronounced.

We find that Algorithm 1 significantly out-performs the random walk approach of [6], but performs less well as d grows. In this graph once d becomes much larger than 5, even the random group testing construction without graph constraints requires at least m tests.

Figure 6 shows the probability that a set of tests correctly identifies d random failures, where the probability is taken over both the tests and the failures. Each point is the empirical mean of 200 independent trials, and we plot error bars of width $1/\sqrt{200} \approx 0.07$. As above, by a Hoeffding bound the probability that the true mean lies outside the error bars is at most e^{-2}.

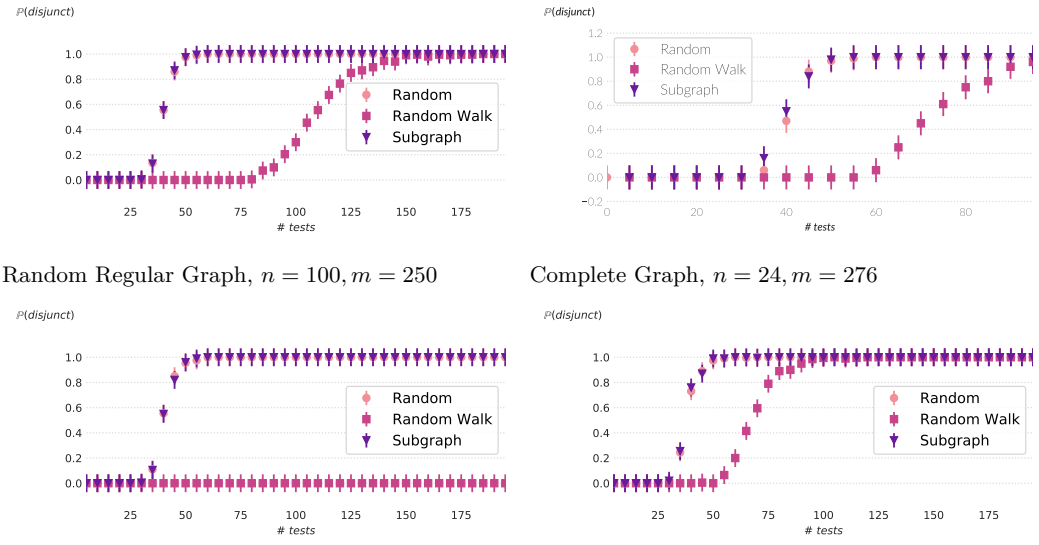
7 Conclusion

We have given a simple randomized construction which shows that for many graphs, graph-constrained group testing is possible with a near-optimal number of tests. Our results – which are proved by analyzing a particular random walk – improve over previous work, and also apply to a wider range of graphs. However, many open questions remain, and we conclude with a few of these here.

1. Both our approach and the approach of [6] give randomized constructions. Derandomizing these constructions remains a fascinating open question. Such a derandomization would be especially useful if it allowed a node to extremely efficiently determine which of its neighbors a test packet should be sent to next, using only minimal information stored in the packet.

2. While (β, α)-expansion is reasonably general, it is not completely general. For example, hypercubes are not very good (β, α)-expanders, but the result of [18] implies that (since they have many disjoint spanning trees) hypercubes are reasonably good for the graph-constrained group-testing problem: $O(d^3 \log m)$ tests suffice to identify d defectives for $d \lesssim \log(n)$. It seems possible that one could modify our analysis using the approach of [1] – which shows that random sparsifications of hypercubes have large connected components with high probability – to obtain a good result for hypercubes as well. Thus, it is an open question to see how well our approach works for hypercubes, but more generally if there is some quantity (more general than (β, α)-edge expansion) which precisely captures when our approach works and when it does not.

3. In the full version [28], we show that simple paths are not as powerful as connected-subgraph tests for graph-constrained group testing. However, in practice it is often the case that simple paths (and especially shortest paths) are easier to implement. It would be interesting to characterize the limitations of graph-constrained group testing when the tests are restricted to (shortest) simple paths.



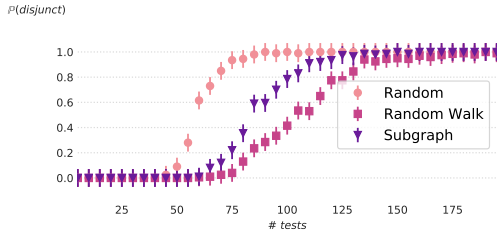
Random Regular Graph, $n = 100, m = 250$

Complete Graph, $n = 24, m = 276$

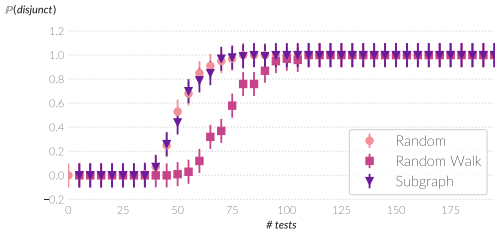
Fat Tree, $n = 80, m = 256$

Hypercube, $n = 64, m = 192$

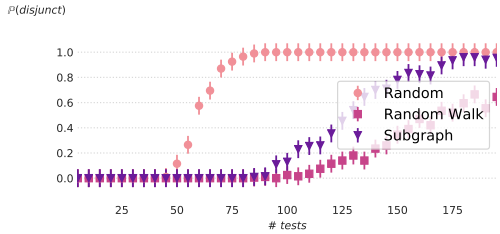
Figure 4 Probability that a randomly generated test matrix with a certain number of tests is 1-disjunct for various graphs. Each point is the mean of 200 trials, with error bars of $1/\sqrt{200}$. "Subgraph" is our approach, "Random Walk" the approach of [6], "Random" the nearly-optimal randomized construction for unconstrained group testing.



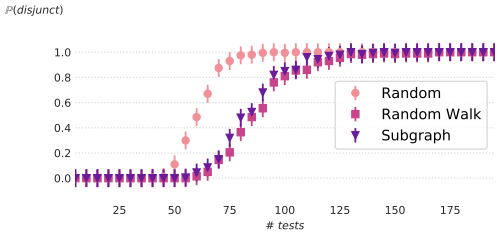
Random Regular Graph, $n = 16, m = 32$



Complete Graph, $n = 9, m = 36$

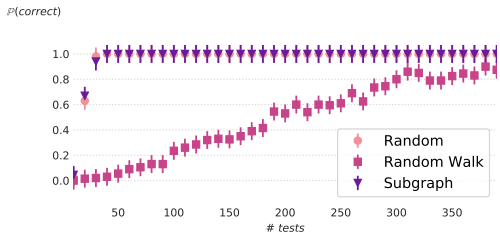


Fat Tree, $n = 20, m = 32$

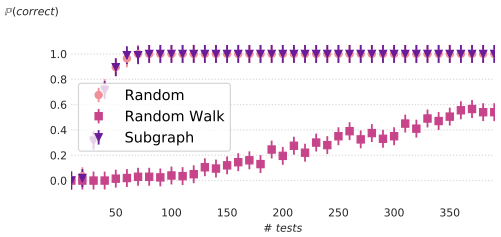


Hypercube, $n = 16, m = 32$

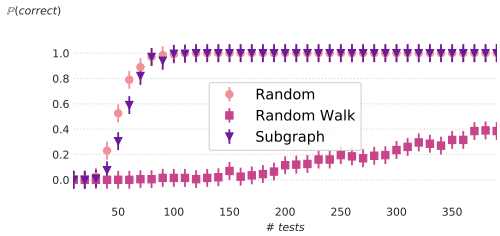
Figure 5 Probability that a randomly generated test matrix with a certain number of tests is 2-disjunct for various graphs. Each point is the mean of 200 trials, and with error bars of $1/\sqrt{200}$. "Subgraph" is our approach, "Random Walk" the approach of [6], "Random" the nearly-optimal randomized construction for unconstrained group-testing.



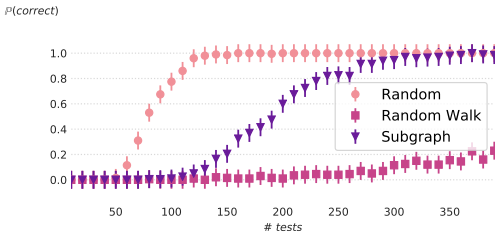
Fat Tree, $n = 80, m = 256, d = 1$



Fat Tree, $n = 80, m = 256, d = 2$



Fat Tree, $n = 80, m = 256, d = 3$



Fat Tree, $n = 80, m = 256, d = 5$

Figure 6 The probability that tests generated from various schemes with a certain number of tests correctly identifies d random failed edges. "Subgraph" is our approach, "Random Walk" the approach of [6], "Random" the nearly-optimal randomized construction for unconstrained group testing.

References

1 Miklós Ajtai, János Komlós, and Endre Szemerédi. Largest random component of a k-cube. *Combinatorica*, 2(1):1–7, 1982. doi:10.1007/BF02579276.

2 Mohammad Al-Fares, Alexander Loukissas, and Amin Vahdat. A scalable, commodity data center network architecture. *SIGCOMM*, 2008. URL: http://dblp.org/rec/conf/sigcomm/Al-FaresLV08.

3 Yigal Bejerano and Rajeev Rastogi. Robust Monitoring of Link Delays and Faults in IP Networks. *INFOCOM*, 1:134–144, 2003. doi:10.1109/INFCOM.2003.1208666.

4 Pat Bosshart, Dan Daly, Glen Gibb, Martin Izzard, Nick McKeown, Jennifer Rexford, Cole Schlesinger, Dan Talayco, Amin Vahdat, George Varghese, and David Walker. P4 - programming protocol-independent packet processors. *Computer Communication Review*, 44(3):87–95, 2014. doi:10.1145/2656877.2656890.

5 Rui Castro, Mark Coates, Gang Liang, Robert Nowak, and Bin Yu. Network Tomography: Recent Developments. *Statistical Science*, 19(3):499–517, August 2004. doi:10.1214/088342304000000422.

6 Mahdi Cheraghchi, Amin Karbasi, Soheil Mohajer, and Venkatesh Saligrama. Graph-Constrained Group Testing. *CoRR*, 2010. URL: http://dblp.org/rec/journals/corr/abs-1001-1445.

7 Fan Chung and Linyuan Lu. The Volume of the Giant Component of a Random Graph with Given Expected Degrees. *SIAM J. Discrete Math.*, 20(2):395–411, January 2006. doi:10.1137/050630106.

8 Amogh Dhamdhere, Renata Teixeira, Constantine Dovrolis, and Christophe Diot. NetDiagnoser - troubleshooting network unreachabilities using end-to-end probes and routing data. *CoNEXT*, page 1, 2007. doi:10.1145/1364654.1364677.

9 Michael Dinitz, Michael Schapira, and Gal Shahaf. Large low-diameter graphs are good expanders. In *26th Annual European Symposium on Algorithms, ESA 2018, August 20-22, 2018, Helsinki, Finland*, pages 71:1–71:15, 2018. doi:10.4230/LIPIcs.ESA.2018.71.

10 Robert Dorfman. The detection of defective members of large populations. *The Annals of Mathematical Statistics*, 14(4):436–440, 1943.

11 Ding-Zhu Du and Frank K Hwang. *Combinatorial Group Testing and Its Applications*, volume 12 of *Series on Applied Mathematics*. World Scientific Publishing Co. Pte. Ltd., 2 edition, 1999. doi:10.1142/9789812798107.

12 Nick G Duffield. Network Tomography of Binary Network Performance Characteristics. *IEEE Trans. Information Theory*, 2006. URL: https://dblp.org/rec/journals/tit/Duffield06.

13 Arkadii Georgievich D'yachkov and Vladimir Vasil'evich Rykov. Bounds on the length of disjunctive codes. *Problemy Peredachi Informatsii*, 18(3):7–13, 1982.

14 Jack Edmonds and Ellis L Johnson. Matching, Euler tours and the Chinese postman. *Mathematical Programming*, 5(1):88–124, 1973. doi:10.1007/BF01580113.

15 Paul Erdös and Alfréd Rényi. On random graphs I. *Publicationes Mathematicae Debrecen*, 6:290–298, 1959.

16 Alan M Frieze, Michael Krivelevich, and Ryan R Martin. The emergence of a giant component in random subgraphs of pseudo-random graphs. *Random Struct. Algorithms*, 24(1):42–50, 2004. doi:10.1002/rsa.10100.

17 Charles Miller Grinstead and James Laurie Snell. *Introduction to probability*. American Mathematical Soc., 2012.

18 Nicholas J A Harvey, Mihai Patrascu, Yonggang Wen, Sergey Yekhanin, and Vincent W S Chan. Non-Adaptive Fault Diagnosis for All-Optical Networks via Combinatorial Group Testing on Graphs. *INFOCOM*, pages 697–705, 2007. doi:10.1109/INFCOM.2007.87.

19 Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(04):439–562, October 2006. doi:10.1090/S0273-0979-06-01126-8.

20 Amin Karbasi and Morteza Zadimoghaddam. Sequential group testing with graph constraints. In *2012 IEEE Information Theory Workshop*, pages 292–296. IEEE, 2012. doi:10.1109/itw.2012.6404678.

21 David R Karger. Using Randomized Sparsification to Approximate Minimum Cuts. *SODA*, 1994. URL: https://dblp.org/rec/conf/soda/Karger94.

22 Michael Krivelevich and Benny Sudakov. The phase transition in random graphs - A simple proof. *Random Struct. Algorithms*, 2013. URL: https://dblp.org/rec/journals/rsa/KrivelevichS13.

23 Charles E Leiserson. Fat-trees: universal networks for hardware-efficient supercomputing. *IEEE transactions on Computers*, 100(10):892–901, 1985.

24 Hung Xuan Nguyen and Patrick Thiran. The Boolean Solution to the Congested IP Link Location Problem - Theory and Practice. *INFOCOM*, pages 2117–2125, 2007. doi:10.1109/INFCOM.2007.245.

25 Ely Porat and Amir Rothschild. Explicit Non-Adaptive Combinatorial Group Testing Schemes. *WINE*, cs.DS, 2007. arXiv:0712.3876v5.

26 Arjun Roy, Hongyi Zeng, Jasmeet Bagga, and Alex C Snoeren. Passive Realtime Datacenter Fault Detection and Localization. *NSDI*, 2017. URL: https://dblp.org/rec/conf/nsdi/RoyZBS17.

27 Miklós Ruszinkó. On the Upper Bound of the Size of the r-Cover-Free Families. *Journal of Combinatorial Theory*, pages 302–310, 1994. URL: https://dblp.org/rec/journals/jct/Ruszinko94.

28 Bruce Spang and Mary Wootters. Unconstraining graph-constrained group testing. *arXiv preprint*, 2018. arXiv:1809.03589.

29 Asaf Valadarsky, Gal Shahaf, Michael Dinitz, and Michael Schapira. Xpander - Towards Optimal-Performance Datacenters. *CoNEXT*, 2016. URL: http://dblp.org/rec/conf/conext/ValadarskySDS16.

30 Hongyi Zeng, Peyman Kazemian, George Varghese, and Nick McKeown. Automatic Test Packet Generation. *IEEE/ACM Transactions on Networking*, 22(2):554–566, April 2013. doi:10.1109/TNET.2013.2253121.

A Instantiations of Theorem 8

In this appendix, we instantiate Theorem 8 for several families of graphs and compare them to existing results. We remark that some of these families (D-regular expanders, or $G(n,p)$) are natural candidates, while others (like a barbell graph) are concocted to show the difference between our theorem and that of previous work. A summary is shown in Table 2, and we go into the details below.

A.1 Complete Graphs

The mixing time for a complete graph is constant, so [6] gives an optimal construction requiring $O(d^2 \log(m/d))$ tests. Theorem 8 gives the same result.

A.2 D-Regular Expander Graphs with Constant Spectral Gap

D-regular expander graphs are D-regular graphs which are very "well-connected." One way of measuring this is the *spectral gap*, that is the difference between the largest eigenvalue D of the adjacency matrix A_G and the second-largest eigenvalue, λ. (We refer the reader to the excellent survey [19] for more background on expander graphs.) We consider families of D-regular graphs G whose second largest eigenvalue λ is bounded away from D by a constant: $\lambda \leq D(1-c)$ for some constant $c \in (0,1)$ independent of n.

For larger $D = \Omega(\log^2 n)$, [6] show that $O(d^2 \log^3(m))$ tests are sufficient, which is optimal up to logarithmic factors. For smaller D, in particular when D is a constant, the best previously known result guarantees $O(d^3 \log(m/d))$ tests [18].

As we will see below, Theorem 8 guarantees $O(d^2 \log(m/d))$ tests are sufficient for all D. In order to apply Theorem 8, we relate the second largest eigenvalue to edge-expansion as follows:

▶ **Theorem 13** (See [19] Theorem 4.11). *Let $G = (V, E)$ be a finite, connected, D-regular graph and let λ be its second eigenvalue. Then G is $(1/2, \alpha)$-edge expander with*

$$\frac{D-\lambda}{2} \leq \alpha \leq \sqrt{2D(D-\lambda)}.$$

By plugging in Theorem 13 to Theorem 8 (with constant and sufficiently small $\delta > 0$) we obtain the following corollary:

▶ **Corollary 14.** *Let G be a D-regular expander with second largest eigenvalue $\lambda \leq D(1-c)$ for some constant $c > 0$. Then for any $d < cD/2$, there is a collection $\mathcal{T}$ of connected subgraph tests so that $\mathcal{T}$ is d-disjunct and $|\mathcal{T}| = O(d^2 \log(m/d))$.*

Notice that by Proposition 9, the restriction that $d \leq d_0$ for some $d_0 = O(D)$ is necessary.

A.3 Erdös-Rényi Graphs

Like the D-regular expanders above, an Erdös-Rényi random graph $G(n, p)$ on n nodes with parameter $0 \leq p \leq 1$ is well-connected, and has good spectral properties with high probability. However, these graphs are not D-regular so we consider them separately.

For larger $p = \Omega(\log^2 n/n)$, [6] show that $O(d^2 \log^3(m))$ tests are sufficient to guarantee d-disjunctness. Theorem 8 can improve this to $O(d^2 \log(m/d))$, with only the restriction $p \geq p_0$ for some $p_0 = \Omega(1/n)$.

In order to apply Theorem 8, we use the following lemma from [6], which implies that $G(n, D/n)$ for $D = \Omega(d \log n)$ has $(1/2, \alpha)$ edge expansion for $\alpha \geq (2+\varepsilon)d$. Theorem 8 immediately implies that $O(d^2 \log(m/d))$ tests are sufficient.

▶ **Lemma 15** ([6] Lemma 32). *For every $\phi < 1/2$ there is an $\alpha > 0$ such that a random graph $G = G(n, p)$ with $p \geq \alpha \ln n/n$ has edge expansion $\alpha \geq \phi D$ with probability $1 - o(1)$.*

A.4 Barbells

One of the advantages of our result over previous work is that the notion of (β, α)-edge expansion captures a more general notion of "well-connected" than is captured by minimum cuts or mixing times. As an extreme example of this, consider a barbell graph G, which we define as two copies of the complete graph $K_{n/2}$ on $n/2$ vertices, connected by one edge e (Figure 7).

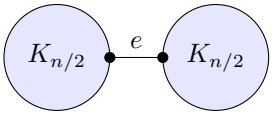


Figure 7 A barbell graph.

This graph is great for graph-constrained group testing: we test the connecting edge e on its own, then then use $O(d^2 \log(m/d))$ tests to identify up to d failures in each of the two copies of $K_{n/2}$. Thus, $O(d^2 \log(m/d))$ tests are sufficient. However, this is a worst-case

graph for existing work. It has a minimum cut of one edge, so [18] only allows $d = 1$. The mixing time of this graph is quite large: the probability of reaching the center edge is $O(1/n)$, so certainly $\tau = \Omega(n)$. The degree condition of [6] is not satisfied as $D \leq n/2$ which is not $O(1/n^2)$. However, even if we could ignore this condition, [6] would use at least $\tilde{O}(n^2 d^2 \log(m/d))$ tests (or, therefore $O(m)$ tests since this is the naive solution).

On the other hand, our results using (β, α)-expansion match the intuition that this example should be easy. Setting $\beta = \frac{1}{4}$ and $\alpha \geq \frac{n}{2} - \frac{n}{4} = \frac{n}{4}$, Theorem 8 gives an $O(d^2 \log(m/d))$ bound.

This is example is meant to highlight the difference between our work and existing work. While the barbell graph is unlikely to be used in practice, it illustrates the intuition that (β, α)-connectivity does better capture somewhat "clustery" graphs – that is, graphs with higher connectivity in some areas than in others – than either minimum cuts or mixing time. This notion may be useful for real-life networks; for example, networks may have higher connectivity within a rack than between racks.

A.5 Fat-Trees

The "Fat-Tree" graph [23] was originally designed for use in supercomputers, and is now widely used in datacenter networks [2]. As the name suggests, this is a "fattened" tree, where the fatness (number of links) near the top of the tree is greater than the fatness near the leaves.

We illustrate the fat-tree topology in Figure 3. For some parameter $D > 0$, each node in the fat-tree has degree D or $D/2$. The Fat-Tree consists of a "core" of $\left(\frac{D}{2}\right)^2$ nodes and a set of D pods. Each pod consists of a complete bipartite graph where each layer has $D/2$ nodes. Each node in the core has one edge to one node in the top layer of each pod.

The minimum cut of the fat tree is $D/2$ and it has $n = 5D^2/4$ nodes. If $D/2 \geq 5(d+1)\log(5D^2/4)$, or equivalently $d \leq \frac{D}{20\log(5D/4)} - 1 = O(D/\log D)$, Proposition 2 gives an upper bound of $O(d^2 \log(m/d))$ on the number of tests required.

On the other hand, the guarantee of [18] gives a graph-constrained group testing scheme for the fat-tree topology with $O(d^3 \log(m/d))$ tests. It is not clear what the mixing time of the fat-tree is, so we do not compare to the bound from [6] which is in terms of the mixing time. (However, as shown in Section 6, it seems that our approach requires fewer tests than that of [6] on this graph).