Discrete Mathematics and Theoretical Computer Science DMTCS vol. **20**:1, 2018, #9

Non-adaptive Group Testing on Graphs

Hamid Kameli*

Department of Mathematical Sciences, Shahid Beheshti University, G.C. Iran

received 27th Jan. 2016, revised 3rd May 2017, 6th Jan. 2018, accepted 25th Jan. 2018.

Grebinski and Kucherov (1998) and Alon et al. (2004-2005) studied the problem of learning a hidden graph for some especial cases, such as hamiltonian cycle, cliques, stars, and matchings, which was motivated by some problems in chemical reactions, molecular biology and genome sequencing. The present study aims to present a generalization of this problem. Graphs G and H were considered, by assuming that G includes exactly one defective subgraph isomorphic to H. The purpose is to find the defective subgraph by performing the minimum non-adaptive tests, where each test is an induced subgraph of the graph G and the test is positive in the case of involving at least one edge of the defective subgraph H. We present the first upper bound for the number of non-adaptive tests to find the defective subgraph by using the symmetric and high probability variation of Lovász Local Lemma. Finally, we present the first non-adaptive randomized algorithm that finds the defective subgraph by at most $\frac{3}{2}$ times of this upper bound with high probability.

Keywords: Group testing on graphs, Non-adaptive algorithm, Combinatorial search, Learning a hidden subgraph

1 Introduction

In the classic *group testing* problem which was first introduced by Dorfman [11], there is a set of n items including at most d defective items. The purpose of this problem is to find the defective items with the minimum number of tests. Every test consists of some items and each test is positive if it includes at least one defective item. Otherwise, the test is negative. There are two types of algorithms for the group testing problem, *adaptive* and *non-adaptive*. In adaptive algorithm, the outcome of previous tests can be used in the future tests and in non-adaptive algorithm all tests perform simultaneously and the defective items are obtained by considering results of all tests.

Regarding some extensions of classical group testing, we can refer to *group testing on graphs*, *complex group testing*, *additive model*, *inhibitor model*, etc. (see [12, 13, 17] for more information). Aigner [1] proposed the problem of group testing on graphs, in which we look for one defective edge of the given graph G by performing the minimum adaptive tests, where each test is an induced subgraph of the graph G and the test is positive in the case of involving the defective edge.

In the present paper, the problem of *non-adaptive group testing on graphs* was considered by assuming that there is one defective subgraph (not necessarily induced subgraph) of G isomorphic to a graph H

*Partially supported by Shahid Beheshti University, G.C.

ISSN 1365–8050

and our purpose is to find the defective subgraph with minimum number of non-adaptive tests. Each test F is an induced subgraph of G and the test result is positive if and only if F includes at least one edge of the defective subgraph. In this study we provide the first non-adaptive algorithm for this problem. Our problem is a generalization of the problem of *non-adaptive learning a hidden subgraph* studied in [2, 4, 14]. In the problem of learning hidden graph, the graph G is a complete graph. In other words, let $\mathcal{H}$ be a family of labeled graphs on the set $V = \{1, 2, ..., n\}$. In this problem the goal is to reconstruct a hidden graph $H \in \mathcal{H}$ by minimum number of tests, where a test $\mathcal{F} \subset V$ is positive if the subgraph of H induced by $\mathcal{F}$, contains at least one edge. Otherwise the test is negative. Alon and Asodi [2] for the problem of non-adaptive learning a general graph presented a lower bound based on the size of the maximum independent set. Their bound is almost tight for the random graph $G(n, \frac{1}{2})$. Chang et al. [9] provided the best adaptive algorithm that learns the general hidden graph of n vertices, with at most $m \log n + 10m + 3n$ tests when the hidden graph has m edges. Also they proved $\left\lceil \log \sum_{i=0}^{m} \binom{\binom{n}{2}}{i} \right\rceil$ adaptive tests are required to identify the hidden graph H(with m edges) drawn from the family of all graphs with n vertices.

The problem of learning a hidden graph was emphasized in some models as follows:

K-vertex model: In this model, each test has at most k vertices.

Additive model: Based on this model, the result of each test F is the number of edges of H induced by F. This model is mainly utilized in bioinformatics and was studied in [7, 15].

Shortest path test: In this model, each test u, v indicates the length of the shortest path between u and v in the hidden graph and if no path exists, it returns ∞. More information about this model and the result is given in [21]. Further, this model is regarded as a canonical model in the evolutionary tree literature [16, 18, 22].

There are various families of hidden graphs to study. However, a large number of recent studies have focused on hamiltonian cycles and matchings [4, 6, 14], stars and cliques [2], graph of bounded degree [7, 15], general graphs [5, 7] . Here, we present a short survey of known results on these problems by using adaptive and non-adaptive algorithms.

Grebinski and Kucherov [14] suggested an adaptive algorithm to learn a hidden Hamiltonian cycle by $2n \log n$ tests, which achieves the information lower bound for the number of tests needed. Further, Chang et al.[8] improved their results to $(1 + o(1))n \log n$.

Alon et al. [4] proposed an upper bound $(\frac{1}{2} + o(1))\binom{n}{2}$ on learning a hidden matching using non-adaptive tests. Bouvel et al. [7] developed an adaptive algorithm to learn a hidden matching with at most $(1 + o(1))n \log n$ tests. In addition, Chang et al. [8] improved their result to $(1 + o(1))\frac{n \log n}{2}$.

Alon and Asodi [2] developed an upper bound $O(n \log^2 n)$ on learning a hidden clique using non-adaptive tests. Also they proved an upper bound $k^3 \log n$ on learning a hidden $K_{1,k}$ using non-adaptive tests. Bouvel et al. [7] presented two adaptive algorithms to learn hidden star and hidden clique with at most $2n$ tests. Chang et al. [8] improved their results on learning hidden star and hidden clique to $(1 + o(1))n$ and $n + \log n$, respectively.

Grebinski and Kucherov [15] gave tight bound of $\theta(dn)$ and $\theta(\frac{n^2}{\log n})$ non-adaptive tests on learning a hidden d-degree-bounded and general graphs in additive model, respectively. Angluin and Chen [5] proved that a hidden general graph can be identified with $12m \log n$ adaptive tests where m (unknown) is the number of edges in the hidden graph.

Group testing can be implemented in finding pattern in data, DNA library screening, and so on (see

[12, 13, 19, 20] for an overview of results and more applications). Learning hidden graph, especially hamiltonian cycle and matchings, is mostly applied in genome sequencing, DNA physical mapping, chemical reactions and molecular biology (see [5, 8, 14, 23] for more information about these applications). Regarding the present study, the main motive behind investigating the problem of non-adaptive group testing on graphs is the application of this problem in chemical reactions. In chemical reactions, we are dealing with a set of chemicals, some pairs of which may involve a reaction. Moreover, before testing, we know some pairs have no reaction. When some chemicals are combined in one test, a reaction takes place if and only if at least one pair of the chemicals reacts in the test. The present study aimed to identify which pairs are reacted using as few tests as possible. Therefore, we can reformulate this problem as follows. Suppose that there are n vertices and two vertices u and v are adjacent if and only if two chemicals u and v may involve a reaction. The reaction of each pair of the chemicals indicates a defective edge and finding all there types of pairs is equal to find the defective subgraph. As we know some pairs have no reaction, the graph G is not necessarily a complete graph.

2 Notation

Throughout this paper, we suppose that H is a subgraph of G with k edges. Moreover, we assume that G contains exactly one defective subgraph isomorphic to H.

We denote the maximum degree of H by $\Delta = \Delta(H)$. Also, $G[X]$ denotes the subgraph of G induced by $X \cap V(G)$ and for any vertex $v \in G$, $N_H(v)$ stands for the set of neighbours of the vertex v in the graph H. Hereafter, we assume that the subgraph H has no isolated vertex, because in the problem of group testing on graphs, just edges are defective.

3 Main result

Throughout this paper, let $H_1, H_2, \ldots, H_m$ be all the subgraphs of G isomorphic to H. For $1 \leq l \leq t$, let $\mathcal{F}_l$ be a random set obtained by choosing each vertex of $V(G)$ randomly and independently with probability p. For simplicity of notation we write F_l as an induced subgraph of G on the vertices of $\mathcal{F}_l$. For the random subset $\mathcal{F}$ of $V(G)$, we say the random test $F = G[\mathcal{F}]$, distinguishes between two distinct subgraphs H_i and H_j if and only if F contains an edge of one of the subgraphs H_i and H_j and contains no edge of the other. In other words exactly one of the following events happens

$$E(F \cap H_i) \neq \varnothing \; and \; E(F \cap H_j) = \varnothing$$

or

$$E(F \cap H_i) = \varnothing \; and \; E(F \cap H_j) \neq \varnothing$$

For any i, j, l, where $1 \leq i \neq j \leq m$ and $1 \leq l \leq t$, we define $A^l_{i,j}$ to be the event that the test F_l cannot distinguish between H_i and H_j. Also, let $A_{i,j}$ denote the event where there is no test $F \in \{F_1, F_2, \ldots, F_t\}$ that distinguishes between H_i and H_j. So we would like to bound the probability that none of the bad events $A_{i,j}$ occur. In such cases, when there is some relatively small amount of dependence between events, one can use a powerful generalization of the union bound, known as the Lovász Local Lemma. The main device in establishing the Lovász Local Lemma is a graph called the dependency graph. Let $A_1, A_2, \ldots, A_n$ be events in an arbitrary probability space. A graph $D = (V, E)$ on the set of vertices $V = \{1, 2, \ldots, n\}$ is a dependency graph for events $A_1, A_2, \ldots, A_n$ if for each

$1 \leq i \leq n$ the event A_i is mutually independent of all the events $\{A_j : \{i, j\} \notin E\}$. We state the Lovász Local Lemma as follows.

Lemma A. [3] (*Lovász Local Lemma, Symmetric Case*). *Suppose that $A_1, A_2, \ldots, A_n$ are events in a probability space with $Pr(A_i) \leq p$ for all i. If the maximum degree in the dependency graph of these events is d, and if $ep(d+1) \leq 1$, then*

$$Pr\Big(\bigcap_{i=1}^{n} \overline{A_i}\Big) > 0,$$

where e is the basis of the natural logarithm.

To find the maximum degree in the dependency graph of the events $A_{i,j}$, we define the parameter $r_G(H)$ as follows. Set $r_G(H, H_i)$ to be the number of subgraphs of G isomorphic to H have common vertex with H_i, i.e., $r_G(H, H_i) = |\{H_j : 1 \leq j \leq m, j \neq i, V(H_i) \cap V(H_j) \neq \varnothing\}|$. Also, define

$$r_G(H) = \max_{1 \leq i \leq m} r_G(H, H_i).$$

In Theorem 1, we show that there are t tests, $F_1, F_2, \ldots, F_t$, such that for every i and j, there is a test $F \in \{F_1, F_2, \ldots, F_t\}$ that distinguishes between H_i and H_j. So if H_i is the defective subgraph, then for every non-defective subgraph H_j, there exists a test $F \in \{F_1, F_2, \ldots, F_t\}$ that distinguishes between the defective subgraph H_i and non-defective subgraph H_j. Therefore, by these tests we can find the defective subgraph.

Theorem 1. *Let H be the defective subgraph of G and $H_1, H_2, \ldots, H_m$ be all the subgraphs of G isomorphic to H. There are t induced subgraphs $F_1, \ldots, F_t$ of G such that for each pair of H_i and H_j, at least one of $F_1, \ldots, F_t$ can distinguish between H_i and H_j, where $k = |E(H)|$, $\Delta = \Delta(H)$,*

$$t = 1 + \left\lceil \frac{\ln(4er_G(H)) + \ln m}{\ln \frac{1}{1-P_{k,\Delta}}} \right\rceil,$$

$P_{k,\Delta} = \frac{1}{2k\Delta}\left(1 - \frac{1}{2\Delta}\right)^{2\Delta-1}\left(1 - \sqrt{\frac{1}{2k\Delta}}\left(1 - \frac{1}{2\Delta}\right)^{\Delta-1}\right)^{2\Delta-2}$*, and e is the basis of the natural logarithm.*

In order to prove Theorem 1, first we should find the probability that tests $F_1, F_2, \ldots, F_t$, distinguish between each pair of subgraphs H_i and H_j. Thus, finding the upper bound for the probability of occurring the bad event $A_{i,j}$ is essential. Accordingly, we should find the lower bound of probability that the random test F_l can distinguish between two subgraphs H_i and H_j.

In the next theorem, based on some following lemmas, we show that the probability of distinguishing between H_i and H_j has the minimum value whenever $V(H_i) = V(H_j)$ and $|E(H_i) \setminus E(H_j)| = 1$.

Theorem 2. *Let $k = |E(H)|$ and $\Delta = \Delta(H)$. For every $1 \leq i \neq j \leq m$ and $1 \leq l \leq t$, we have*

$$Pr\left(\overline{A_{i,j}^{l}}\right) \geq 2p^2(1-p)^{2\Delta}(1-\epsilon), \tag{1}$$

where $p = \sqrt{\frac{\epsilon}{k}}\,(1-\epsilon)^{\Delta-1}$.

Lemma 1. *Let T be a graph with n vertices, k edges, and maximum degree Δ. Pick, randomly and independently, each vertex of T with probability p, where $p = \sqrt{\frac{\epsilon}{k}}(1 - \frac{\epsilon}{k})^{(\Delta-1)}$. If F is the set of all chosen vertices, then $T[F]$ has no edges, with probability at least $1 - \epsilon$.*

To prove this lemma, we need high probability variation of Lovász Local Lemma.

Lemma B. [10] *Let $B_1, B_2, \ldots, B_k$ be events in a probability space. Suppose that each event B_i is independent of all the events B_j but at most d. For $1 \leq i \leq k$ and $0 < \epsilon < 1$, if $Pr(B_i) \leq \dfrac{\epsilon}{k}(1 - \dfrac{\epsilon}{k})^d$, then $Pr\Big(\bigcap_{i=1}^{k} \overline{B_i}\Big) > 1 - \epsilon$.*

Proof of Lemma 1: Let $E(T) = \{e_1, e_2, \ldots, e_k\}$. For $1 \leq i \leq k$, we define B_i to be the event that $e_i \in E(T[F])$, so $Pr(B_i) = p^2$. Since vertices are chosen randomly and independently, the event B_i is independent of the event B_j if and only if edges e_i and e_j have no common vertex. So the maximum degree of the dependency graph is at most $2(\Delta - 1)$. Since $p^2 \leq \dfrac{\epsilon}{k}\Big(1 - \dfrac{\epsilon}{k}\Big)^{2(\Delta-1)}$, by Lemma B, $Pr\Big(\bigcap_{i=1}^{k} \overline{B_i}\Big) > 1 - \epsilon$. Hence, $T[F]$ has no edges, with probability at least $1 - \epsilon$. □

To find the probability of distinguishing between H_i and H_j and then prove Theorem 2, we consider the following three cases.

Case 1: $V(H_i) = V(H_j)$, $|E(H_i) \setminus E(H_j)| = 1$.

Case 2: $|V(H_i) \setminus V(H_j)| \geq 1$.

Case 3: The induced subgraph on $V(H_i) - V(H_j)$ has at least one edge.

Lemma 2. *If $V(H_i) = V(H_j)$ and $|E(H_i) \setminus E(H_j)| = 1$, then*

$$Pr\Big(E(F_l \cap H_i) \neq \varnothing, E(F_l \cap H_j) = \varnothing\Big) \geq p^2(1 - p)^{2\Delta}(1 - \epsilon),$$

where $p = \sqrt{\frac{\epsilon}{k}}\,(1 - \epsilon)^{\Delta - 1}$.

Proof: Let $e = \{u, v\} \in E(H_i) \setminus E(H_j)$. Consider the induced subgraph H' of G, where $V(H') = V(H_j) \setminus \Big(u \cup v \cup N_{H_j}(u) \cup N_{H_j}(v)\Big)$. Note that if $u, v \in \mathcal{F}_l$ and $H_j \cap F_l$ has no edges of H_j, then $E(F_l \cap H_i) \neq \varnothing$ and $E(F_l \cap H_j) = \varnothing$. Also, one can see that $u, v \in \mathcal{F}_l$ and $H_j[F_l]$ has no edges if the following events hold

1. $u, v \in \mathcal{F}_l$,
2. $N_{H_j}(u) \cap \mathcal{F}_l = \varnothing$ and $N_{H_j}(v) \cap \mathcal{F}_l = \varnothing$,
3. $H'[\mathcal{F}_l]$ has no edges.

It is straightforward to check that the aforementioned events are independent. Also, one can see that the event $u, v \in \mathcal{F}_l$ occurs with probability p^2. Since $|N_{H_j}(u) \cup N_{H_j}(v)| \leq 2\Delta$, we have

$$\begin{aligned} Pr\Big(N_{H_j}(u) \cap \mathcal{F}_l = \varnothing, N_{H_j}(v) \cap \mathcal{F}_l = \varnothing\Big) &= \\ Pr\Big(\mathcal{F}_l \cap \big(N_{H_j}(u) \cup N_{H_j}(v) \setminus \{u, v\}\big) = \varnothing\Big) &\geq (1-p)^{2\Delta}. \end{aligned}$$

Set $E(H') = k'$. If $k' = 0$, then $F_l \cap H'$ has no edges and $Pr\big(E(F_l \cap H') = \varnothing\big) = 1$. Suppose that $k' \geq 1$. Since $k \geq k'$, we have $p^2 = \dfrac{\epsilon}{k}(1-\epsilon)^{2\Delta-2} \leq \dfrac{\epsilon}{k'}(1-\dfrac{\epsilon}{k'})^{2\Delta-2}$. Each vertex of the induced subgraph H' is chosen with probability p. So by Lemma 1, the induced subgraph on $\mathcal{F}_l \cap V(H')$ has no edges, with probability at least $1-\epsilon$. In other words, $Pr\big(E(F_l \cap H') = \varnothing\big) \geq 1-\epsilon$. Since the events are independent, we have

$$Pr\Big(E(F_l \cap H_i) \neq \varnothing, E(F_l \cap H_j) = \varnothing\Big) \geq p^2(1-p)^{2\Delta}(1-\epsilon),$$

as desired. □

Lemma 3. *If $|V(H_i) \setminus V(H_j)| \geq 1$, then*

$$Pr\Big(E(F_l \cap H_i) \neq \varnothing, E(F_l \cap H_j) = \varnothing\Big) \geq p^2(1-p)^{\Delta}(1-\epsilon),$$

where $p = \sqrt{\frac{\epsilon}{k}}\,(1-\epsilon)^{\Delta-1}$.

Proof: Since H has no isolated vertex, there exists at least one edge $e = \{u, v\} \in E(H_i) \setminus E(H_j)$. Let $v \in V(H_i) \cap V(H_j)$ and $u \in V(H_i) \setminus V(H_j)$. Suppose that H' is an induced subgraph of H_j, where $V(H') = V(H_j) \setminus (v \cup N(v))$. Set $|E(H')| = k'$. Similar to the proof of Lemma 2, $E(F_l \cap H_i) \neq \varnothing$ and $E(F_l \cap H_j) = \varnothing$ if the following independent events hold

1. $u, v \in \mathcal{F}_l$,
2. $N_{H_j}(v) \cap \mathcal{F}_l = \varnothing$,
3. $H'[\mathcal{F}_l]$ has no edges.

Since $|N_{H_j}(v)| \leq \Delta$, the probability that $N_{H_j}(v) \cap \mathcal{F}_l = \varnothing$ is at least $(1-p)^{\Delta}$. The rest of proof is similar to Lemma 2, so

$$Pr\Big(E(F_l \cap H_i) \neq \varnothing, E(F_l \cap H_j) = \varnothing\Big) \geq p^2(1-p)^{\Delta}(1-\epsilon),$$

as desired. □

Lemma 4. *If the induced subgraph on $V(H_i) \setminus V(H_j)$ has at least one edge, then*

$$Pr\Big(E(F_l \cap H_i) \neq \varnothing, E(F_l \cap H_j) = \varnothing\Big) \geq p^2(1-\epsilon),$$

where $p = \sqrt{\frac{\epsilon}{k}}\,(1-\epsilon)^{\Delta-1}$.

Proof: Let $e=(u,v)\in E(H_i)\setminus E(H_j)$. If the following independent events hold

1. $u,v\in\mathcal{F}_l$,
2. $H_j[\mathcal{F}_l]$ has no edges,

then $E(F_l\cap H_i)\neq\varnothing$ and $E(F_l\cap H_j)=\varnothing$. Since $p^2=\frac{\epsilon}{k}(1-\epsilon)^{2\Delta-2}\leq\frac{\epsilon}{k}\left(1-\frac{\epsilon}{k}\right)^{2\Delta-2}$, by Lemma 1, $Pr\left(E(F_l\cap H_j)=\varnothing\right)\geq 1-\epsilon$. Also one can see that

$$Pr\Big(E(F_l\cap H_i)\neq\varnothing\Big)\geq Pr\Big(e\in E(F_l)\Big)=Pr\big(u,v\in\mathcal{F}_l\big)=p^2.$$

Consequently, $Pr\big(E(F_l\cap H_i)\neq\varnothing, E(F_l\cap H_j)=\varnothing\big)\geq p^2(1-\epsilon)$. □

Proof of Theorem 2: Let $E(H_i)\cap E(H_j)=\{f_1,f_2,\ldots,f_r\}$ and $E(H_i)\setminus E(H_j)=\{e_1,e_2,\ldots,e_{k-r}\}$. As previously mentioned, the event $\overline{A_{i,j}^l}$ occurs if and only if the test F_l distinguish between H_i and H_j. In other words,

$$E(F_l\cap H_i)\neq\varnothing\ and\ E(F_l\cap H_j)=\varnothing$$

or

$$E(F_l\cap H_j)\neq\varnothing\ and\ E(F_l\cap H_i)=\varnothing.$$

It is easy to check that

$$Pr\left(\overline{A_{i,j}^l}\right)=Pr\Big(E(F_l\cap H_i)\neq\varnothing, E(F_l\cap H_j)=\varnothing\Big)+Pr\Big(E(F_l\cap H_j)\neq\varnothing, E(F_l\cap H_i)=\varnothing\Big).$$

In the following we prove $Pr\big(E(F_l\cap H_i)\neq\varnothing, E(F_l\cap H_j)=\varnothing\big)\geq p^2(1-p)^{2\Delta}(1-\epsilon)$ and with the completely similar proof we can prove $Pr\big(E(F_l\cap H_j)\neq\varnothing, E(F_l\cap H_i)=\varnothing\big)\geq p^2(1-p)^{2\Delta}(1-\epsilon)$.

It is easy to check, for every $1\leq q\leq k-r$,

$$Pr\Big(E(F_l\cap H_i)\neq\varnothing, E(F_l\cap H_j)=\varnothing\Big)\geq Pr\Big(e_q\in E(F_l\cap H_i), E(F_l\cap H_j)=\varnothing\Big).$$

So to find the lower bound for this probability, we need to consider the following three cases.

Case 1: $V(H_i)=V(H_j)$, $|E(H_i)\setminus E(H_j)|=1$.

By Lemma 2, it is clear

$$Pr\big(E(F_l\cap H_i)\neq\varnothing, E(F_l\cap H_j)=\varnothing\big)\geq p^2(1-p)^{2\Delta}(1-\epsilon).$$

Case 2: $|V(H_i)\setminus V(H_j)|\geq 1$.

By Lemma 3, we have

$$Pr\big(E(F_l\cap H_i)\neq\varnothing, E(F_l\cap H_j)=\varnothing\big)\geq p^2(1-p)^{\Delta}(1-\epsilon)\geq p^2(1-p)^{2\Delta}(1-\epsilon).$$

Case 3: The induced subgraph on $V(H_i)-V(H_j)$ has at least one edge.

By Lemma 4,

$$Pr\big(E(F_l\cap H_i)\neq\varnothing, E(F_l\cap H_j)=\varnothing\big)\geq p^2(1-\epsilon)\geq p^2(1-p)^{2\Delta}(1-\epsilon).$$

So for every $1 \le i \ne j \le m$ and $1 \le l \le t$, $Pr\left(\overline{A^l_{i,j}}\right) \ge 2p^2(1-p)^{2\Delta}(1-\epsilon)$. $\square$

In order to prove Theorem 1, we present an upper bound for the probability of occurring the bad events $A_{i,j}$ for every $1 \le i \ne j \le m$.

Theorem 3. *Let $k = |E(H)|$ and $\Delta = \Delta(H)$. For every $1 \le i \ne j \le m$, we have*

$$Pr(A_{i,j}) \le (1 - P_{k,\Delta})^t, \tag{2}$$

where $P_{k,\Delta} = \frac{1}{2k\Delta}\left(1 - \frac{1}{2\Delta}\right)^{2\Delta-1}\left(1 - \sqrt{\frac{1}{2k\Delta}}\left(1-\frac{1}{2\Delta}\right)^{\Delta-1}\right)^{2\Delta-2}$.

Proof: Since $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_t \subset V(G)$ are chosen randomly and independently, the events $A^1_{i,j}, \ldots, A^t_{i,j}$ are mutually independent. So

$$Pr(A_{i,j}) = Pr(A^l_{i,j})^t.$$

By Theorem 2, we have $Pr\left(\overline{A^l_{i,j}}\right) \ge 2p^2(1-p)^{2\Delta}(1-\epsilon)$. According to $p = \sqrt{\frac{\epsilon}{k}}\,(1-\epsilon)^{\Delta-1}$, we set $\epsilon = \frac{3}{\Delta}$ to almostly maximize the lower bound of good events $\overline{A^l_{i,j}}$. So $Pr\left(\overline{A^l_{i,j}}\right) \ge P_{k,\Delta}$, where

$$P_{k,\Delta} = \frac{6}{k\Delta}\left(1 - \frac{3}{\Delta}\right)^{2\Delta-1}\left(1 - \sqrt{\frac{3}{k\Delta}}\left(1 - \frac{3}{\Delta}\right)^{\Delta-1}\right)^{2\Delta}.$$

Therefore, $Pr(A_{i,j}) = Pr(A^l_{i,j})^t \le (1 - P_{k,\Delta})^t$. $\square$

Now, we can prove Theorem 1.

Proof of Theorem 1: By Theorem 3, for every $1 \le i \ne j \le m$, $Pr(A_{i,j}) \le (1-P_{k,\Delta})^t$. Now we prove that if $t > \dfrac{\ln(4er_G(H)) + \ln m}{\ln \frac{1}{1-P_{k,\Delta}}}$, then by Lovász Local Lemma, with positive probability no event $A_{i,j}$ occurs.

We construct the dependency graph whose vertices are the events $A_{i,j}$, where $1 \le i, j \le m$. Two events $A_{i,j}$ and $A_{i',j'}$ are adjacent if and only if $\Big(V(H_i) \cup V(H_j)\Big) \cap \Big(V(H_{i'}) \cup V(H_{j'})\Big) \ne \varnothing$. Remember that $r_G(H) = \max_i r_G(H, H_i)$, where $r_G(H, H_i)$ is the number of subgraphs of G isomorphic to H including common vertex with H_i. For the fixed $A_{i,j}$, there are at most $r_G(H)$ subgraph $H_{i'}$ isomorphic to H such that $V(H_i) \cap V(H_{i'}) \ne \varnothing$. We can choose $H_{j'}$ with $m-1$ ways. So it is easy to check that the maximum degree in the dependency graph is at most $4r_G(H)(m-1)$. Accordingly, if

$$t > \frac{\ln(4er_G(H)) + \ln m}{\ln \frac{1}{1-P_{k,\Delta}}},$$

then $e\,(1-P_{k,\Delta})^t\,\big(4r_G(H)(m-1)+1\big) < 1$, and by Lovász Local Lemma

$$Pr\Big(\bigcap_{i,j} \overline{A_{i,j}}\Big) > 0.$$

Therefore, if $t = 1 + \lceil \dfrac{\ln(4er_G(H)) + \ln m}{\ln \frac{1}{1-P_{k,\Delta}}} \rceil$, then with positive probability no event $A_{i,j}$ occurs. Thus, for each pair of H_i and H_j there is a test $F \in \{F_1, F_2, \ldots, F_t\}$ that distinguishes between H_i and H_j. □

We can obtain $t = 1 + \lceil \frac{2\ln m}{\ln \frac{1}{1-P_{k,\Delta}}} \rceil$ if we use union bound. In fact the Lovász Local Lemma is better when the dependencies between events are rare.

Based on this theorem there are t tests which distinguish between each pair of H_i and H_j with positive probability. However, an algorithm is essential to find these tests with high probability if we are interested in finding these tests.

Theorem 4. *Let H be the defective subgraph of G with k edges. If $t = \dfrac{\ln \frac{m^2}{\delta}}{\ln \frac{1}{1-P_{k,\Delta}}}$, we can find this defective subgraph by t tests with probability at least $1-\delta$, where $\Delta = \Delta(H)$ and*

$$P_{k,\Delta} = \frac{1}{2k\Delta}\left(1 - \frac{1}{2\Delta}\right)^{2\Delta-1}\left(1 - \sqrt{\frac{1}{2k\Delta}}\left(1-\frac{1}{2\Delta}\right)^{\Delta-1}\right)^{2\Delta-2}.$$

Proof: By Theorem 3 and the union bound we know

$$Pr\Big(\bigcup_{1\le i<j\le m} A_{i,j}\Big) \le m^2(1-P_{k,\Delta})^t.$$

Thus, this upper bound becomes close to zero if t is large enough. It is easy to check if $t = \dfrac{\ln \frac{m^2}{\delta}}{\ln \frac{1}{1-P_{k,\Delta}}}$, then $m^2(1-P_{k,\Delta})^t = \delta$. In other words, we can distinguish between each pair of H_i and H_j with probability at least $1-\delta$ if we choose tests randomly and independently. □

If we set $\delta = \frac{1}{m}$, then for sufficiently large m, we can find the defective subgraph with $\frac{3\ln m}{\ln \frac{1}{1-P_{k,\Delta}}}$ tests with high probability.

4 Concluding remarks

In the present paper we assume that the graph G includes few edges since the Lovász Local Lemma is more powerful when the dependencies between events are rare. In the graph G with $O(n^2)$ edges, the parameter $r_G(H)$ is high, which means it is better to use the union bound. In this case, there are $1 + \lceil \dfrac{2\ln m}{\ln \frac{1}{1-P_{k,\Delta}}} \rceil$ tests that find the defective subgraph non-adaptively.

Finally, if we consider dense and sparse defective subgraph separately, we can obtain a better upper bound for the number of tests in the case of sparse defective subgraph.

Acknowledgements

The present study is a part of Hamid Kameli's Ph.D. Thesis. The author would like to express his deepest gratitude to Professor Hossein Hajiabolhassan for introducing a generalization of learning a hidden subgraph problem and for his precious comments and discussion.

References

[1] M. Aigner. Search problems on graphs. *Discrete Applied Mathematics*, 14(3):215 – 230, 1986. ISSN 0166-218X. doi: http://dx.doi.org/10.1016/0166-218X(86)90026-0. URL http://www.sciencedirect.com/science/article/pii/0166218X86900260.

[2] N. Alon and V. Asodi. Learning a hidden subgraph. *SIAM Journal on Discrete Mathematics*, 18(4): 697–712, 2005.

[3] N. Alon and J. H. Spencer. *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., Hoboken, NJ, third edition, 2008. ISBN 978-0-470-17020-5.

[4] N. Alon, R. Beigel, S. Kasif, S. Rudich, and B. Sudakov. Learning a hidden matching. *SIAM Journal on Computing*, 33(2):487–501, 2004.

[5] D. Angluin and J. Chen. Learning a hidden graph using queries per edge. *Journal of Computer and System Sciences*, 74(4):546 – 556, 2008. ISSN 0022-0000. doi: http://dx.doi.org/10.1016/j.jcss.2007.06.006. URL http://www.sciencedirect.com/science/article/pii/S0022000007000724.

[6] R. Beigel, N. Alon, S. Kasif, M. S. Apaydin, and L. Fortnow. An optimal procedure for gap closing in whole genome shotgun sequencing. In *Proceedings of the Fifth Annual International Conference on Computational Biology*, RECOMB '01, pages 22–30, New York, NY, USA, 2001. ACM. ISBN 1-58113-353-7. doi: 10.1145/369133.369152. URL http://doi.acm.org/10.1145/369133.369152.

[7] M. Bouvel, V. Grebinski, and G. Kucherov. *Combinatorial Search on Graphs Motivated by Bioinformatics Applications: A Brief Survey*, pages 16–27. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005. ISBN 978-3-540-31468-4. doi: 10.1007/11604686_2. URL http://dx.doi.org/10.1007/11604686_2.

[8] H. Chang, H.-B. Chen, H.-L. Fu, and C.-H. Shi. Reconstruction of hidden graphs and threshold group testing. *Journal of Combinatorial Optimization*, 22(2):270–281, 2011. ISSN 1382-6905. doi: 10.1007/s10878-010-9291-0. URL http://dx.doi.org/10.1007/s10878-010-9291-0.

[9] H. Chang, H.-L. Fu, and C.-H. Shih. Learning a hidden graph. *Optimization Letters*, 8(8):2341–2348, Dec 2014. ISSN 1862-4480. doi: 10.1007/s11590-014-0751-9. URL https://doi.org/10.1007/s11590-014-0751-9.

[10] D. Deng, D. Stinson, and R. Wei. The Lovász local lemma and its applications to some combinatorial arrays. *Designs, Codes and Cryptography*, 32(1-3):121–134, 2004. ISSN 0925-1022. doi: 10.1023/B:DESI.0000029217.97956.26. URL http://dx.doi.org/10.1023/B%3ADESI.0000029217.97956.26.

[11] R. Dorfman. The detection of defective members of large populations. *Ann. Math. Statist.*, 14(4): 436–440, 12 1943. doi: 10.1214/aoms/1177731363. URL http://dx.doi.org/10.1214/aoms/1177731363.

[12] D. Du and F. Hwang. *Combinatorial Group Testing and Its Applications*. Applied Mathematics. World Scientific, 2000. ISBN 9789810241070. URL `http://books.google.com/books?id=KW5-CyUUOggC`.

[13] D. Du and F. Hwang. *Pooling Designs and Nonadaptive Group Testing: Important Tools for DNA Sequencing*. Series on applied mathematics. World Scientific, 2006. ISBN 9789812568229. URL `http://books.google.com/books?id=AJcoAQAAMAAJ`.

[14] V. Grebinski and G. Kucherov. Reconstructing a hamiltonian cycle by querying the graph: Application to DNA physical mapping. *Discrete Applied Mathematics*, 88(1–3):147 – 165, 1998. ISSN 0166-218X. doi: http://dx.doi.org/10.1016/S0166-218X(98)00070-5. URL `http://www.sciencedirect.com/science/article/pii/S0166218X98000705`.

[15] V. Grebinski and G. Kucherov. Optimal reconstruction of graphs under the additive model. *Algorithmica*, 28(1):104–124, 2000. ISSN 1432-0541. doi: 10.1007/s004530010033. URL `http://dx.doi.org/10.1007/s004530010033`.

[16] J. J. Hein. An optimal algorithm to reconstruct trees from additive distance data. *Bulletin of Mathematical Biology*, 51(5):597–603, 1989. ISSN 1522-9602. doi: 10.1007/BF02459968. URL `http://dx.doi.org/10.1007/BF02459968`.

[17] H.-B. C. F. K. Hwang. A survey on nonadaptive group testing algorithms through the angle of decoding. *Journal of Combinatorial Optimization*, 15, 01 2008. doi: 10.1007/s10878-007-9083-3. URL `http://libgen.org/scimag/index.php?s=10.1007/s10878-007-9083-3`.

[18] V. King, L. Zhang, and Y. Zhou. On the complexity of distance-based evolutionary tree reconstruction. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '03, pages 444–453, Philadelphia, PA, USA, 2003. Society for Industrial and Applied Mathematics. ISBN 0-89871-538-5. URL `http://dl.acm.org/citation.cfm?id=644108.644179`.

[19] A. J. Macula and L. J. Popyack. A group testing method for finding patterns in data. *Discrete Applied Mathematics*, 144(1–2):149 – 157, 2004. ISSN 0166-218X. doi: http://dx.doi.org/10.1016/j.dam.2003.07.009. URL `http://www.sciencedirect.com/science/article/pii/S0166218X04001933`.

[20] H. Q. Ngo and D.-Z. Du. A survey on combinatorial group testing algorithms with applications to DNA library screening. In *Discrete mathematical problems with medical applications*, volume 55 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 171–182. Amer. Math. Soc., 2000.

[21] L. Reyzin and N. Srivastava. *Algorithmic Learning Theory: 18th International Conference, ALT 2007, Sendai, Japan, October 1-4, 2007. Proceedings*, chapter Learning and Verifying Graphs Using Queries with a Focus on Edge Counting, pages 285–297. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. ISBN 978-3-540-75225-7. doi: 10.1007/978-3-540-75225-7_24. URL `http://dx.doi.org/10.1007/978-3-540-75225-7_24`.

[22] L. Reyzin and N. Srivastava. On the longest path algorithm for reconstructing trees from distance matrices. *Inf. Process. Lett.*, 101(3):98–100, Feb. 2007. ISSN 0020-0190. doi: 10.1016/j.ipl.2006.08.013. URL http://dx.doi.org/10.1016/j.ipl.2006.08.013.

[23] A. Sorokin, A. Lapidus, V. Capuano, N. Galleron, P. Pujic, and S. D. Ehrlich. A new approach using multiplex long accurate pcr and yeast artificial chromosomes for bacterial chromosome mapping and sequencing. *Genome Research*, 6(5):448–453, 1996. doi: 10.1101/gr.6.5.448. URL http://genome.cshlp.org/content/6/5/448.abstract.