Electronic Colloquium on Computational Complexity, Report No. 118 (2020)

On Testing Asymmetry in the Bounded Degree Graph Model

Oded Goldreich*

August 4, 2020

Abstract

We consider the problem of testing asymmetry in the bounded-degree graph model, where a graph is called asymmetric if the identity permutation is its only automorphism. Seeking to determine the query complexity of this testing problem, we provide partial results. Considering the special case of n-vertex graphs with connected components of size at most $s(n) = \Omega(\log n)$, we show that the query complexity of ϵ-testing asymmetry (in this case) is at most $O(\sqrt{n} \cdot s(n)/\epsilon)$, whereas the query complexity of $o(1/s(n))$-testing asymmetry (in this case) is at least $\Omega(\sqrt{n/s(n)})$.

In addition, we show that testing asymmetry in the dense graph model is almost trivial.

Contents

1 Introduction

Property testing refers to probabilistic algorithms of sub-linear complexity for deciding whether a given object has a predetermined property or is far from any object having this property. Such algorithms, called testers, obtain local views of the object by *performing queries* and their performance guarantees are stated with respect to a distance measure that (combined with a distance parameter) determines which objects are considered far from the property.

In the last couple of decades, the area of property testing has attracted significant attention (see, e.g., [5]). Much of this attention was devoted to testing graph properties in a variety of models including the dense graph model [7], and the bounded-degree graph model [8] (surveyed in [5, Chap. 8] and [5, Chap. 9], resp.). We mention, without elaboration, that the known results concerning these models include both results regarding general classes of graph properties and results regarding many natural graph properties. Yet, one natural property that (to the best of our knowledge) was not considered before is *asymmetry*.

A graph is called asymmetric if the identity permutation is its only automorphism. Recall that, for a (labeled) graph $G = (V, E)$ and a bijection $\phi : V \to V'$, we denote by $\phi(G)$ the graph

*Department of Computer Science, Weizmann Institute of Science, Rehovot, Israel. E-mail: oded.goldreich@weizmann.ac.il. Partially supported by the Israel Science Foundation (grant No. 1041/18).

ISSN 1433-8092

$G' = (V', E')$ such that $E' = \{\{\phi(u), \phi(v)\} : \{u,v\} \in E\}$, and say that G' is isomorphic to G. The set of automorphisms of the graph $G = (V, E)$, denoted $\mathtt{aut}(G)$, is the set of permutations that preserve the graph G; that is, $\pi \in \mathtt{aut}(G)$ if and only if $\pi(G) = G$.

Definition 1 (asymmetric and symmetric graphs): *A graph is called* asymmetric *if its sets of automorphisms is a singleton, which consists of the trivial automorphism* (i.e., the identity permutation)*. Otherwise, the graph is called* symmetric*.*

It turns out that testing asymmetry in the dense graph model is quite trivial, because, under the corresponding distance measure, every graph is close to being asymmetric (see Section 3). Our focus is on the bounded-degree graph model, where we only establish partial results.

Theorem 2 (testing asymmetric graphs with small connected components (in the bounded-degree graph model)): *The query complexity of ϵ-testing whether an n-vertex graph is asymmetric and has connected components of size at most $s(n)$ is at most $O(n^{1/2} \cdot s(n)/\epsilon)$ and is at least $\Omega((n/s(n))^{1/2})$ provided that $\epsilon = o(1/s(n))$ and $s(n) = \Omega(\log n)$. Furthermore, the upper bound holds for one-sided error testers, whereas the lower bound holds also for general* (i.e., two-sided error) *testers.*

Note that, for $s(n) = o((\log n)/\log\log n)$, the testing problem is trivial, since the number of bounded-degree s-vertex graphs is smaller than $\exp(O(s \log s))$.[1] Indeed, the gap between the lower and upper bounds increases with s; in contrast, for $s(n) = \text{poly}(\log n)$ and $\epsilon \in [o(1/s(n)), \text{poly}(s(n))]$, we can assert that *the query complexity of ϵ-testing the set of asymmetric n-vertex graphs having connected components of size $s(n)$ is $\widetilde{\Theta}(\sqrt{n})$.*

Needless to say, we failed in our attempts to determine the query complexity of testing asymmetric graphs in the general case, and leave it as an open problem. We mention that a (less) partial state of knowledge also exists for (both versions of) the graph isomorphism testing problem.[2]

2 In the bounded-degree graph model

In the bounded-degree model, graphs are represented by their incidence functions and distances are measured as the ratio of the number of differing incidences over the maximal number of edges. Specifically, for a degree bound $d \in \mathbb{N}$, we represent a graph $G = ([n], E)$ of maximum degree d by the incidence function $g : [n] \times [d] \to [n] \cup \{0\}$ such that $g(v, i)$ indicates the i^{th} neighbor of v (where $g(v, i) = 0$ indicates that v has less than i neighbors). The distance between the graphs $G = ([n], E)$ and $G' = ([n], E')$ is defined as the symmetric difference between E and E' over $dn/2$, and oracle access to a graph means oracle access to its incidence function.

[1] This implies that an n-vertex graph that consists of connected components of size at most $s(n) = o((\log n)/\log\log n)$ must have a few identical components, and is thus symmetric.

[2] For all sufficiently small $\epsilon > 0$, the following is known regarding ϵ-testing graph isomorphism in the bounded-degree graph model [6].

1. The query complexity of *ϵ-testing isomorphism to a fixed n-vertex graph* is $\widetilde{\Omega}(n^{1/2})$.
2. The query complexity of *ϵ-testing isomorphism between two n-vertex graphs* is $\widetilde{\Omega}(n^{2/3})$.

The lower bounds are shown by using graphs that have connected components of size $\text{poly}(\log n)$, and in this case the lower bounds are tight [6]. We mention that, unlike Theorem 2, one-sided error testing of isomorphism (even to a fixed graph) has linear query complexity [6, Thm. 2.5].

Definition 3 (testing graph properties in the bounded-degree graph model): *For a fixed degree bound d, a* tester *for a graph property Π is a probabilistic oracle machine that, on input parameters n and ϵ, and oracle access to* (the incidence function of) *an n-vertex graph $G = ([n], E)$ of maximum degree d, outputs a binary verdict that satisfies the following two conditions.*

1. *If $G \in \Pi$, then the tester accepts with probability at least $2/3$.*
2. *If G is ϵ-far from Π, then the tester accepts with probability at most $1/3$, where G is* ϵ-far *from Π if for every n-vertex graph $G' = ([n], E') \in \Pi$ of maximum degree d it holds that the symmetric difference between E and E' has cardinality that is greater than $\epsilon \cdot dn/2$.*

If the tester accepts every graph in Π with probability 1, then we say that it has one-sided error*; otherwise, we say that it has* two-sided error.

(Throughout this work, we consider undirected simple graphs (i.e., no self-loops and parallel edges).)

The query complexity of a tester for Π is a function (of the parameters d, n and ϵ) that represents the number of queries made by the tester on the worst-case n-vertex graph of maximum degree d, when given the proximity parameter ϵ. Fixing d, we typically ignore its effect on the complexity (equiv., treat d as a hidden constant). The query complexity of $\epsilon(n)$-testing Π is defined as the query complexity of testing when the proximity parameter is set to $\epsilon(n)$; that is, we say that the query complexity of $\epsilon(n)$-testing Π is at least $Q(n)$ if distinguishing between n-vertex graphs in Π and n-vertex graphs that are $\epsilon(n)$-far from Π requires at least $Q(n)$ queries.

Theorem 4 (Theorem 2, restated): *For $s : \mathbb{N} \to \mathbb{N}$, let $\Pi^{(s)} = \bigcup_{n\in\mathbb{N}} \Pi_n^{(s)}$ such that $\Pi_n^{(s)}$ is the set of asymmetric n-vertex graphs that have connected components of size at most $s(n)$. Then, for every degree bound $d \geq 3$, the following holds.*

1. *If $s(n) = \Omega((\log n)/\log\log n)$, then the query complexity of $(1/(3d \cdot s(n)))$-testing $\Pi_n^{(s)}$* (in the bounded-degree graph model) *is $\Omega((n/s(n))^{1/2})$.*
2. *There exists a one-sided error ϵ-tester for $\Pi^{(s)}$* (in the bounded-degree graph model) *that makes $O(n^{1/2} \cdot s(n)/\epsilon)$ queries, and runs in time $\widetilde{O}(n^{1/2}/\epsilon) \cdot \mathrm{poly}(s(n))$.*

We stress that Part 1 holds also for two-sided error testers. Recall that, for $s(n) = o((\log n)/\log\log n)$, the testing problem is trivial, since the number of bounded-degree s-vertex graphs. Theorem 4 follows by combining Propositions 5 and 6, which are stated and proved next.

Proposition 5 (lower bound on testing asymmetric graphs (in the bounded-degree graph model)): *For every $d \geq 3$ and any $s : \mathbb{N} \to \mathbb{N}$ such that $s(n) = \Omega((\log n)/\log\log n)$, the query complexity of $(1/(3d \cdot s(n)))$-testing whether an n-vertex graph is asymmetric is $\Omega((n/s(n))^{1/2})$. This holds even if it is guaranteed that the tested graph consists of connected components of size at most $s(n)$.*

Proof: We use the following facts, proved in [2, 3]: (F1) most d-regular s-vertex graphs are asymmetric, and (F2) their number exceeds $N_d(s) = \Omega(s/d!)^{ds/2}$. Note that (F1) even also if we require the graphs to be connected, since most d-regular graphs are actually expanders. Hence, for $s(n) = \frac{c \log_2 n}{d \log_2 \log_2 n}$ it holds that $\frac{N_d(s(n)}{s(n)!} > 2^{(0.5d-1)c\log_2 n - o(\log n)}$, which is larger than n when $c > 2/(d-2)$. It follows that there exists a collection, denoted C, of $m = n/s(n)$ non-isomorphic $s(n)$-vertex graphs that are asymmetric and connected. The claim of the proposition follows by showing that $\Omega(\sqrt{m})$ queries are necessary for distinguish the following two distributions:

1. A random isomorphic copy of the n-vertex graph that contains connected components from each graph in C.

2. A random isomorphic copy of the n-vertex graph that contains two connected components from each graph in C', where C' is a random $m/2$-subset of C.

Note that each graph in the support of the first distribution is asymmetric, whereas each graph in the support of the second distribution is $(1/(3d \cdot s(n)))$-far being asymmetric. The latter claim holds because making such a graph asymmetric requires modifying the incidence of at least one vertex in at least $m/2$ of its connected components, which amounts to at least $\frac{m}{4} = \frac{n}{4s(n)} > \frac{1}{3d \cdot s(n)} \cdot dn/2$ edge-modifications.

The fact that $\Omega(\sqrt{m})$ queries are necessary to distinguish the foregoing two distributions is proved by the "birthday" argument. Specifically, when making q queries to a graph drawn from the second distribution, we encounter vertices in two different connected components that are isomorphic to the same graph (in C) with probability at most $\binom{q}{2}/m$. Whenever this event does not occur, the answers are distributed identically to the way they are distributed when querying a graph drawn from the first distribution. ■

Proposition 6 (upper bound on testing asymmetric graphs (in the bounded-degree graph model)): *For every $d \geq 3$, there exists a one-sided error tester of query complexity $O(n^{1/2} \cdot s/\epsilon)$ for the set of n-vertex asymmetric graphs that consist of connected components of size at most s. Furthermore, the running time of the tester is $\widetilde{O}(n^{1/2}/\epsilon) \cdot \mathrm{poly}(s)$.*

Proof: On input parameters n, s and $\epsilon > 0$, and oracle access to a graph $G = ([n], E)$, the algorithm proceeds as follows.

1. It selects uniformly at random $m = O(\sqrt{n}/\epsilon)$ vertices $v_1, ..., v_m \in [n]$.

2. For each $i \in [m]$, the algorithm starts a (e.g., BFS) exploration of the connected component in which v_i resides, and halts rejecting if it discovers a connected component having more than s vertices.

3. If for some $i \in [m]$, the connected component explored from v_i is symmetric, then the algorithm halts rejecting.

4. If for some $i, j \in [m]$, the connected components explored from v_i and v_j are different but isomorphic (i.e., v_i does not reside in the same connected component as v_j but these two connected components are isomorphic), then the algorithm halts rejecting.

If the algorithm did not reject, then it accepts.

The query complexity of this algorithm is $O(m \cdot s)$, while its running time is dominated by Steps 3 and 4. Observe, however, that Steps 3 and 4 can be implemented in time $\widetilde{O}(m) \cdot \mathrm{poly}(s)$ by using the canonical labeling algorithm (for bounded-degree graphs) of [1] (along with a sorting algorithm).

Let $\Pi = \Pi_n^{(s)}$ denote the set of n-vertex asymmetric graphs that consist of connected components of size at most s. Evidently, the algorithm accepts each graph in Π with probability 1. On the other hand, if G is ϵ-far from Π, then one of the following three cases must hold.[3]

***Case 1*:** At least $\epsilon n/6$ of its vertices reside in connected components of size greater than s.

In this case, Step 2 of the algorithm rejects (w.h.p.).

***Case 2*:** At least $\epsilon n/6$ of its vertices reside in connected components of size at most s that are symmetric.

In this case, Step 3 of the algorithm rejects (w.h.p.).

***Case 3*:** At least $n' = \epsilon n/6$ of its vertices reside in connected components of size at most s that are asymmetric, and yet the graph induced by the set S of these n' vertices is $\Omega(n\epsilon/n')$-far from being asymmetric.

In other words, letting G_S denote the subgraph of G induced by S, consider the graph G' that results by augmenting G_S with $(n-n')/s$ connected components (each of size s) that are neither symmetric nor isomorphic to any other connected component (where the existence of such a collection of s-vertex graphs has been established in the first paragraph of the proof of Proposition 5). Then, G' is $\epsilon/3$-far from being asymmetric.

Let $C_1, ..., C_{m'}$ denote the connected components of G_S, and recall that each C_i has at most s vertices. Consider the equivalence relation, denoted $\equiv$, defined by graph isomorphism (over the set of C_i's); that is, $C_i \equiv C_j$ if and only if C_i is isomorphic to C_j. Let n_k denote *the number of k-vertex connected components that reside in equivalence classes that has more than a single C_i*; that is,

$$n_k = |\{i \in [m'] : |C_i| = k \ \& \ \exists j \neq i \text{ s.t. } C_i \equiv C_j\}|$$

where $|C_i|$ denotes the number of vertices in C_i. Then, $\sum_{k\in[s]} n_k \cdot k \geq \epsilon n/6$ must hold, because otherwise G' is $\epsilon/3$-close to being asymmetric; to see this, replace the connected components in the *non-singleton equivalence classes* by asymmetric connected components of size s that are not isomorphic to any other connected component (cf. Footnote 3).

Now, if we take a sample of $\Theta(\epsilon^{-1}\sqrt{n})$ vertices, then it is very likely that $\Theta(\sqrt{n})$ of these vertices hit connected components in the non-singleton equivalence classes. Recalling that $\sum_{k\in[s]} n_k \cdot k \leq n$, we infer that this sample is likely to hit two different elements of the same class. This holds because, with high probability, we are likely to have several classes hit by at least two samples, and with probability at least $1/2$ each of these pairs of samples hit different C_i's in the relevant class.

Hence, in each of these cases, the algorithm rejects with high probability, which establishes our claim. ■

On testing the set of symmetric graphs. We mention that testing the set of symmetric graphs is almost trivial; specifically, the query complexity is 0 if $\epsilon \geq 4/n$, and $dn = O(d/\epsilon)$ otherwise. This is the case because, with respect to a degree bound d, every n-vertex graph is $\frac{2d}{dn/2}$-close to being symmetric (e.g., by making two vertices isolated).

[3] Note that if Cases 1 and 2 do not hold, then G is $2\epsilon/3$-close to a graph G' in which the exceptional (large or asymmetric) connected components are replaced by small connected components that are asymmetric and not isomorphic to all other connected components, which implies that G satisfies Case 3 (i.e., G' is $\epsilon/3$-far from being asymmetric). Recall that such a collection of s-vertex graphs does exist (see the proof of Proposition 5).

3 In the dense graph model

In the dense graph model, a graph $G = ([n], E)$ is represented by its adjacency predicate, $g : [n] \times [n] \to \{0, 1\}$, such that $g(u, v) = 1$ if and only if $\{u, v\} \in E$. The distance between the graphs $G = ([n], E)$ and $G' = ([n], E')$ is defined as the symmetric difference between E and E' over $\binom{n}{2}$, and oracle access to a graph means oracle access to its adjacency predicate.

Definition 7 (testing graph properties in the dense graph model): *A* tester *for a graph property* Π *is a probabilistic oracle machine that, on input parameters n and ϵ, and oracle access to* (the adjacency predicate of) *an n-vertex graph $G = ([n], E)$, outputs a binary verdict that satisfies the following two conditions.*

1. *If $G \in \Pi$, then the tester accepts with probability at least* $2/3$*.*
2. *If G is ϵ-far from Π, then the tester accepts with probability at most* $1/3$*, where G is* ϵ-far *from Π if for every n-vertex graph $G' = ([n], E') \in \Pi$ it holds that the symmetric difference between E and E' has cardinality that is greater than $\epsilon \cdot \binom{n}{2}$.*

The query complexity of a tester for Π is a function (of the parameters n and ϵ) that represents the number of queries made by the tester on the worst-case n-vertex graph, when given the proximity parameter ϵ. In this section, we show that testing the set of asymmetric graphs in the dense graph model is almost trivial; specifically, the query complexity is 0 if $\epsilon > O((\log n)/n)$, and $\binom{n}{2} = \widetilde{O}(1/\epsilon^2)$ otherwise. This holds because in the first case (i.e., $\epsilon > O((\log n)/n)$), all n-vertex graphs are ϵ-close to being asymmetric (see Proposition 8), whereas in the second case one can afford to retrieve the entire graph.

Proposition 8 (all graphs are close to being asymmetric): *In the dense graph model, every n-vertex graph G is $\frac{O(\log n)}{n}$-close to being asymmetric.*

Proof: Given an arbitrary graph $G = ([n], E)$, we construct a random variant of it, denoted G', by re-randomizing $O(n \log n)$ of its adjacencies, and show that (w.h.p.) the resulting graph is asymmetric. Specifically, we consider the following "randomized" version.

Construction 8.1 (construction of G'): *Given an arbitrary graph $G = ([n], E)$, we proceed as follows.*

1. *Select an arbitrary subset, S, of $\ell = O(\log n)$ vertices in G.*
2. *Replace the subgraph of G induced by S with a random ℓ-vertex graph.*
3. *Replace the bipartite subgraph that connects S and $[n] \setminus S$ by a random bipartite graph; that is, for each $s \in S$ and $v \in [n] \setminus S$, the edge $\{s, v\}$ in contained in the resulting graph G' with probability* $1/2$*.*

We shall first show that, with very high probability, the subgraph of G' induced by S is not isomorphic to the subgraph of G' that is induced by any other ℓ-subset.

Claim 8.2 (uniqueness os S): *For every ℓ-subset S fixed in Step 1 of Construction 8.1, with high probability over Steps 2 and 3, for every ℓ-subset $S' \neq S$ of $[n]$, the subgraph of G' induced by S' is not isomorphic to the subgraph of G' induced by S.*

Proof: The case of $S' \cap S = \emptyset$ is easy, because in this case the subgraph of G' induced by S' is fixed is Step 1 (since it equals the subgraph of G induced by S'), whereas a random ℓ-vertex graph (as selected in Step 2) is isomorphic to this fixed graph with probability at most $(\ell!) \cdot 2^{-\binom{\ell}{2}} \ll \binom{n}{\ell}^{-1}$, where the inequality uses a sufficiently large $\ell = O(\log n)$. Hence, we can afford to take a union bound over all ℓ-subsets that are disjoint of S. However, for sets that are not disjoint of S, the foregoing probability bound does not hold, and a more careful analysis is called for. Nevertheless, the foregoing analysis does provide a good warm-up towards the rest.

First, for each ℓ-set $S' \subset [n]$ such that $S' \neq S$, we shall upper-bound the probability that the subgraphs of G' induced by S and S' are isomorphic as a function of $|S \cap S'|$. For every bijection $\pi : S \to S'$, let $\mathtt{FP}(\pi) \stackrel{\rm def}{=} \{v \in S : \pi(v) = v\}$ denote the set of fixed-points of π, and note that $|\mathtt{FP}(\pi)| \leq \ell - 1$ (since $S \neq S'$). Now, letting G_R denote the subgraph of G induced by R, we claim that the probability that there exists a bijection $\pi : S \to S'$ such that $\pi(G'_S) = G'_{S'}$ is upper-bounded by

$$\begin{aligned} & \sum_{\pi: S \stackrel{1\text{-}1}{\to} S'} \min\left(2^{-|\mathtt{FP}(\pi)| \cdot (\ell - |\mathtt{FP}(\pi)|)/2}, 2^{-\binom{(\ell - |\mathtt{FP}(\pi)|)/2}{2}}\right) & (1) \\ \leq & \sum_{f \in \{0,\ldots,|S \cap S'|\}} \frac{\ell!}{f!} \cdot 2^{-\max(4 \cdot f \cdot (\ell - f), (\ell - f) \cdot (\ell - f - 1))/8} & \\ < & \frac{\ell!}{|S \cap S'|!} \cdot 2^{-\Omega((\ell - |S \cap S'|) \cdot \ell)} & (2) \end{aligned}$$

where f represents the size of $\mathtt{FP}(\pi)$. To justify the upper bound claimed in Eq. (1), consider an arbitrary bijection $\pi : S \to S'$, and identify a set $I \subseteq S \setminus \mathtt{FP}(\pi)$ such that $\pi(I) \cap I = \emptyset$ and $|I| \geq (\ell - |\mathtt{FP}(\pi)|)/2$. Letting $e_{G'}(u, v) = 1$ if $\{u, v\}$ is an edge in G' and $e_{G'}(u, v) = 0$ otherwise, observe that $\pi(G'_S) = G'_{S'}$ if and only if $e_{\pi(G')}(\pi(u), \pi(v)) = e_{G'}(\pi(u), \pi(v))$ for every $\{u, v\} \in \binom{S}{2}$. Noting that $e_{\pi(G')}(\pi(u), \pi(v)) = e_{G'}(u, v)$, the first bound in Eq. (1) is justified by

$$\begin{aligned} & \Pr_{G'}\left[\forall (u, v) \in \binom{S}{2} : \quad e_{\pi(G')}(\pi(u), \pi(v)) = e_{G'}(\pi(u), \pi(v))\right] \\ \leq & \Pr_{G'}\left[\forall (u, v) \in \mathtt{FP}(\pi) \times I : \quad e_{G'}(u, v) = e_{G'}(\pi(u), \pi(v))\right] \\ = & \prod_{(u,v) \in \mathtt{FP}(\pi) \times I} \Pr_{G'}\left[e_{G'}(u, v) = e_{G'}(u, \pi(v))\right] \\ = & 2^{-|\mathtt{FP}(\pi)| \cdot |I|} \\ \leq & 2^{-|\mathtt{FP}(\pi)| \cdot (\ell - |\mathtt{FP}(\pi)|)/2} \end{aligned}$$

where the equalities are due to the disjointness of the sets $\mathtt{FP}(\pi) \times I$ and $\mathtt{FP}(\pi) \times \pi(I)$ (to the fact that $\pi(u) = u$ for every $u \in \mathtt{FP}(\pi)$), and to the fact that the incidences of all vertices in $\mathtt{FP}(\pi) \subseteq S$ are random. Similarly, we justify the second bound in Eq. (1) by

$$\begin{aligned} & \Pr_{G'}\left[\forall \{u, v\} \in \binom{S}{2} : \quad e_{\pi(G')}(\pi(u), \pi(v)) = e_{G'}(\pi(u), \pi(v))\right] \\ \leq & \Pr_{G'}\left[\forall \{u, v\} \in \binom{I}{2} : \quad e_{G'}(u, v) = e_{G'}(\pi(u), \pi(v))\right] \end{aligned}$$

$$
\begin{aligned}
&= \prod_{\{u,v\}\in\binom{I}{2}} \Pr_{G'}\left[e_{G'}(u,v) = e_{G'}(\pi(u),\pi(v))\right] \\
&= 2^{-\binom{|I|}{2}} \\
&\leq 2^{-\binom{(\ell-|\mathtt{FP}(\pi)|)/2}{2}}
\end{aligned}
$$

where the equalities are due to the disjointness of the sets $\binom{I}{2}$ and $\binom{\pi(I)}{2}$, and to the fact that the incidences of all vertices in $I \subseteq S \setminus \mathtt{FP}(\pi) \subseteq S$ are random.

Combining Eq. (1)&(2) with a union bound over all ℓ-subsets $S' \subset [n]$ that are different from S, we upper-bound the probability that the subgraphs of G' induced by S and by some other ℓ-set are isomorphic by

$$
\sum_{S'\in\binom{[n]}{\ell}\setminus\{S\}} \frac{\ell!}{|S\cap S'|!} \cdot 2^{-\Omega((\ell-|S\cap S'|)\cdot\ell)} = \sum_{i\in\{0,\ldots,\ell-1\}} \binom{\ell}{i} \cdot \binom{n-i}{\ell-i} \cdot \frac{\ell!}{i!} \cdot 2^{-\Omega((\ell-i)\cdot\ell)} \tag{3}
$$

where the index i represents the size of the intersection with S. Using a sufficiently large $\ell = O(\log n)$, we have

$$
\begin{aligned}
\sum_{i\in\{0,\ldots,\ell-1\}} \binom{\ell}{i} \cdot \binom{n-i}{\ell-i} \cdot \frac{\ell!}{i!} \cdot 2^{-\Omega((\ell-i)\cdot\ell)} &= \sum_{i\in\{0,\ldots,\ell-1\}} \binom{\ell}{i}^2 \cdot \binom{n-i}{\ell-i} \cdot \frac{(n-i)!}{(n-\ell)!} \cdot 2^{-\Omega((\ell-i)\cdot\ell)} \\
&< \sum_{i\in\{0,\ldots,\ell-1\}} n^{\ell-i} \cdot \binom{\ell}{i}^2 \cdot 2^{-\Omega((\ell-i)\cdot\ell)} \\
&< \ell \cdot \max_{i\in\{0,\ldots,\ell-1\}} \left\{ n^{\ell-i} \cdot \binom{\ell}{i}^2 \cdot 2^{-\Omega((\ell-i)\cdot\ell)} \right\} \\
&= \ell \cdot \left(n \cdot \ell^2 \cdot 2^{-\Omega(\ell)} \right)
\end{aligned}
$$

which is $o(1)$. The claim follows. ■

Conclusion. Using Claim 8.2, we claim that (w.h.p.) the graph G' is asymmetric. This holds because each of the following claims holds with high probability.

1. Any automorphism of the graph G' maps the set S to itself.

   (Indeed, this is due to Claim 8.2.)

2. The subgraph of G' induced by S is asymmetric.

   (Recall that by [4], almost all ℓ-vertex graphs are asymmetric.)

3. Any vertex $v \in [n] \setminus S$ has a different "neighborhood pattern" with respect to S; that is, for every $u \neq v \in [n] \setminus S$, there exists $w \in S$ such that $\{u,w\}$ is an edge in G' if and only if $\{v,w\}$ is not an edge in G'.

By combining Conditions 1 and 2, it follows that any automorphism of the graph G' maps each vertex $w \in S$ to itself, whereas by Condition 3 such an isomorphism must map each $v \in [n] \setminus S$ to itself. Hence, the claim (that G' is asymmetric) follows, and the proposition follows by noting that G' is $\frac{\ell \cdot n}{n^2}$-close to G. ■

On testing the set of symmetric graphs. We mention that testing the set of symmetric graphs is also almost-trivial; specifically, the query complexity is 0 if $\epsilon \geq 1/n$, and $\binom{n}{2} = O(1/\epsilon^2)$ otherwise. This is the case because each n-vertex graph is $\frac{1}{n}$-close to being symmetric, since by [4, Thm. 1] any n-vertex graph can be made symmetric by modifying the edge relation of at most $\frac{n-1}{2}$ vertex-pairs. (Note that an upper bound of $n-1$ is obvious by picking two vertices u and v, and modifying the neighborhood of u to equal that of v.)

References

[1] L. Babai and E.M. Luks. Canonical Labeling of Graphs. In *15th ACM Symposium on the Theory of Computing*, pages 171–183, 1983.

[2] B. Bollobas. Distinguishing Vertices of Random Graphs. *North-Holland Mathematics Studies*, Vol. 62, pages 33–49, 1982.

[3] B. Bollobas. The Asymptotic Number of Unlabelled Regular Graphs. *J. Lond. Math. Soc.*, Vol. 26, pages 201–206, 1982.

[4] P. Erdos and A. Renyi. Asymmetric Graphs. *Acta Mathematica Hungarica*, Vol. 14 (3), pages 295–315, 1963.

[5] O. Goldreich. *Introduction to Property Testing.* Cambridge University Press, 2017.

[6] O. Goldreich. Testing Isomorphism in the Bounded-Degree Graph Model. *ECCC*, TR19-102, 2019.

[7] O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, pages 653–750, July 1998. Extended abstract in *37th FOCS*, 1996.

[8] O. Goldreich and D. Ron. Property Testing in Bounded Degree Graphs. *Algorithmica*, Vol. 32 (2), pages 302–343, 2002.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il