On Testing Asymmetry in the Bounded Degree Graph Model

Oded Goldreich*

February 10, 2025

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 two partial results.

1. The query complexity of $O(1/\log n)$-testing asymmetry of n-vertex graphs is $\widetilde{\Omega}(\sqrt{n/\log n})$, even if the tested graph is guaranteed to consist of connected components of size $O(\log n)$.
2. For every $\epsilon : \mathbb{N} \to [0,1]$ such that $\epsilon(n) = \omega((\log\log n)/\log n)$, the query complexity of one-sided error $\epsilon(n)$-testing asymmetry of n-vertex graphs is at least $\exp(O(1/\epsilon(n)))$.

In addition, we show that testing asymmetry in the dense graph model is almost trivial, because (in this model) every graph is close to being asymmetric.

Preliminary versions of this work appeared as TR20-118 of *ECCC*. The current version is drastically different: It retains the main results of the previous versions, but omits several other results, because – in retrospect – we find these results distracting. Instead, we emphasize the fact that our results, regarding the bounded-degree graph model, are highly unsatisfactory and leave much to be understood.

1 Overview

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 three 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 [6] and the bounded-degree graph model [7] (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*.

*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) and by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 819702).

The definition of asymmetric graphs is based on the definition of graph isomorphism. Specifically, for a (labeled) graph $G = (V, E)$ and a bijection $\phi : V \to V'$, we denote by $\phi(G)$ the graph $G' = (V', E')$ 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 1.1 (asymmetric and symmetric graphs): *A graph is called* asymmetric *if its set of automorphisms is a singleton, which consists of the trivial automorphism* (i.e., the identity permutation)*. Otherwise, the graph is called* symmetric*.*

Although the title of this paper draws attention to the bounded-degree graph model, only part of the paper (i.e., Section 2) studies testing asymmetric graphs in that model. Yet, our focus is on the bounded-degree graph model, where we obtain very partial results. A possible contribution of this paper is pointing out this sour state of affairs and calling for a better understanding of the complexity of testing asymmetry in the bounded-degree graph model. In contrast, we discard the problem of testing asymmetric graphs in the dense graph model, showing that it is almost trivial, because all graphs are close to be asymmetric in that model (see Section 3).

We review both sections in the following corresponding subsections. But before going so, we review a couple of basic conventions. First, throughout this work, we consider undirected simple graphs (i.e., no self-loops and parallel edges). Second, we discuss graph properties, which are each a set of graphs that is closed under isomorphism; that is, Π is a graph property if for every graph $G = (V, E)$ and any bijection $\pi : V \to V'$ it holds that $G \in \Pi$ if and only if $\pi(G) \in \Pi$.

1.1 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.

Definition 1.2 (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*.*

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 1.3 (lower bound on the query complexity of testing asymmetric graphs (in the bounded-degree graph model)): *For every $d \geq 3$, the query complexity of $O(1/\log n)$-testing asymmetry of n-vertex graphs is at least $\widetilde{\Omega}(n^{0.5})$. Furthermore, this holds even if the tested graph is guaranteed to consist of connected components of size $O(\log n)$.*

We stress that this result holds also for two-sided error testers. The result generalize to graphs with connected components of size at most $s(n) = \Omega(\log n)$, but in that case the gap between the upper and lower bounds is $\mathrm{poly}(s(n))$; see Theorem 2.1. The proof relies on the fact that the number of asymmetric bounded-degree s-vertex graphs is $\exp(\Omega(s \log s))$. Hence, using $o(\sqrt{n/s(n)})$ queries one cannot distinguish the following two distributions:

1. A random n-vertex graph that consists of $n/s(n)$ *different* asymmetric $s(n)$-vertex connected components;

2. A random n-vertex graph that consists of $n/s(n)$ asymmetric $s(n)$-vertex connected components such that *each connected component is isomorphic to exactly one other connected component*.

(See details in Section 2.)

In contrast, for $s(n) = o((\log n)/\log\log n)$, the (two-sided error) testing problem is trivial, because the number of bounded-degree s-vertex graphs is upper-bounded by $\exp(O(s \log s))$. 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 contain some isomorphic components, and is thus symmetric.

Needless to say, Thorem 1.3 tells us nothing about the query complexity of $\epsilon(n)$-testing asymmetry of n-vertex graphs when $\epsilon(n) = \omega((\log\log n)/\log n)$. Restricting ourselves to one-sided error testers, the following resuult tell us that in that case the query complexity is at least exponential in the proximity parameter (i.e., ϵ).

Theorem 1.4 (lower bound on the query complexity of one-sided error testing asymmetric graphs (in the bounded-degree graph model)): *For every $d \geq 3$ and every $\epsilon : \mathbb{N} \to [0,1]$ such that $\epsilon(n) = \omega((\log\log n)/\log n)$, the query complexity of one-sided error $\epsilon(n)$-testing asymmetric n-vertex graphs is at least $\exp(\Omega(1/\epsilon(n))\log(1/\epsilon(n)))$. Furthermore, this holds even if the tested graph is guaranteed to consist of connected components of size $O(1/\epsilon(n))$.*

The proof, also presented in Section 2, follows the basic strategy of Theorem 1.3 with a twist that capitalizes on the one-sided error condition. Note that in both cases the lower bound is (slightly) super-exponential in the size of the connected components. We warn, however, that the size of the connected components is inversely proportional to the proximity parameter. This dependence is inherent in light of the following result.

Proposition 1.5 (on graphs that consists of asymmetric connected components (in the bounded-degree graph model)): *For every $d \geq 3$ and every $s : \mathbb{N} \to \mathbb{N}$, every n-vertex d-regular graph that consists of connected components that are asymmetric $s(n)$-vertex graphs is $O(1/s(n))$-close to being asymmetric.*

(The proof is also presented in Section 2.)

We stress that Theorems 1.3 and 1.4 leave open the question of providing reasonable estimates for the query complexity of ϵ-testing asymmetric n-vertex graphs (in the bounded-degree graph model), when $\epsilon > 0$ is a constant. Specifically, it may be that the query complexity of ϵ-testing asymmetric n-vertex graphs is $f(\epsilon)$ for some function $f : (0,1] \to \mathbb{N}$ (e.g., $f(\epsilon) = \exp(\widetilde{\Theta}(1/\epsilon))$), but it may also be that this complexity must depend on n. Furthermore, it may be that $\Omega(1)$-testing asymmetric graphs requires linear query complexity.

On testing the set of symmetric graphs.

We mention that testing the set of symmetric graphs (in the bounded-degree model) 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).

1.2 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 1.6 (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. As stated upfront, it turns out that testing the set of asymmetric graphs in the dense graph model is almost trivial; specifically,

Theorem 1.7 (testing asymmetric graphs in the dense graph model): *The query complexity of testing asymmetry graphs is 0 if $\epsilon > O((\log n)/n)$, and $\widetilde{O}(1/\epsilon^2)$ otherwise.*

This holds because in the first case (i.e., $\epsilon > O((\log n)/n)$), every n-vertex graph is ϵ-close to being asymmetric (see Proposition 3.1), whereas in the second case one can afford to retrieve the entire graph (since $\binom{n}{2} = \widetilde{O}(1/\epsilon^2)$).

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 every 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.)

2 In the Bounded-Degree Graph Model

In this section we present proofs of the results that were stated in Section 1.1. Starting with Theorem 1.3, we first generalize its claim by replacing the logarithmic bound with an arbitrary function $s : \mathbb{N} \to \mathbb{N}$ such that $s(n) = \Omega\left(\frac{\log n}{\log\log n}\right)$.

Theorem 2.1 (Theorem 1.3, generalized): *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 $s(n)$.*

We stress that the bound holds also for two-sided error testers.

Proof: We use the following facts, proved in [2, 3] (for every $d \geq 3$):

(F1): Most d-regular s-vertex graphs are asymmetric,

(F2): The number of d-regular s-vertex graphs is $N_d(s) = \Omega(s/d)^{ds/2}$.

Note that (F1) holds even if we require the graphs to be connected, since most d-regular graphs are actually expanders. Also, for some constant c and $s(n) = \frac{c\log_2 n}{d\log_2\log_2 n}$ 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 d-regular graphs that are asymmetric and connected. The theorem 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 G_1 that consists of copies of all graphs in C; that is, G_1 consists of m connected components such that each graph in C appears as a connected component.

2. A random isomorphic copy of an n-vertex graph that consists of two copies of each of $m/2$ graphs selected at random in C; that is, we first select a random $m/2$-subset of C, denoted C', and take a random isomorphic copy of the n-vertex graph $G_{C'}$ that consists of two copies of each graph in 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 from being asymmetric. The latter claim holds because making $G_{C'}$ 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 $\binom{q}{2}/|C'|$, where $|C'| = m/2$. 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. ■

Proof of Theorem 1.4.

Using the strategy of the proof of Theorem 2.1, we prove Theorem 1.4, which asserts that *for any $\epsilon : \mathbb{N} \to [0,1]$ such that $\epsilon(n) = o((\log\log n)/\log n)$, the query complexity of $\epsilon(n)$-testing with one-sided error whether an n-vertex graph is asymmetric is* $\exp(\omega(1/\epsilon(n)))$.

Setting $s(n) = \Theta(1/\epsilon(n))$, recall that the number of d-regular asymmetric $s(n)$-vertex graphs is $(1 - o(1)) \cdot N_d(s) = \Theta(s/d)^{ds/2}$. Now, we consider a generic n-vertex graph that contains copies of all asymmetric $s(n)$-vertex graphs (as connected components). Specifically, we consider two cases:

1. The graph contains only $s(n)$-vertex connected components. In this case, as shown in the proof of Theorem 2.1, the graph is $(1/3d \cdot s(n))$-far from being asymmetric.[1] Hence, an $\epsilon(n)$-tester must reject this graph (with probability at least 2/3).

2. The tested n-vertex graph contains a single copy of each of the asymmetric $s(n)$-vertex graphs along with an asymmetric M-vertex connected component, where $M = (n - ((1 - o(1)) \cdot N_d(s(n))/s(n))) = n - o(n)$. In this case, the tested graph is asymmetric, and a one-sided tester must accept it with probability 1.

Thus, when testing these types of graphs, a one-sided error tester must see two isomorphic connected components in order to reject, which means that its query complexity must be $\Omega(\sqrt{N_d(s(n))}) = \exp(\Omega(s(n)\log s(n)))$. ■

Proof of Proposition 1.5.

Lastly, we prove Proposition 1.5, which asserts that *for every $s : \mathbb{N} \to \mathbb{N}$, every n-vertex d-regular graph that consists of connected components that are asymmetric $s(n)$-vertex graphs is $O(1/s(n))$-close to being asymmetric.*

The basic idea is to arrange the connected components in a sequence, pick two vertices in each connected component and connect the second vertex of the i^{th} component to the first vertex of the $i{+}1^{\text{st}}$ component. To simplify the analysis, we may delete all edges from $n/s(n)^2$ of the components and use the "freed" vertices to connect the other components via 2-paths rather than by edges. More importantly, we need to orient the "super path"; we can do this by connecting a single vertex to the first vertex of the first component and an $(s(n) - 1)$-vertex path to the second vertex of the last component. ■

[1] Recall that this follows from the fact that almost all its connected components are isomorphic to other connected components, because $N_d(s(n)) = o(n/s(n))$, and so the incidence of at least one vertex in each of almost all components must be modified.

3 In the Dense Graph Model

As stated in Section 1.2, Theorem 1.7 follows immediately from the following result.

Proposition 3.1 (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 of G.

Construction 3.1.1 (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 3.1.2 (uniqueness of the subgraph induced by S): *For every ℓ-subset S fixed in Step 1 of Construction 3.1.1, 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 in 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. Unfortunately, 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 by S' are isomorphic, as a function of $|S \cap S'|$. For every bijection $\pi : S \to S'$, let $\mathtt{FP}(\pi) \stackrel{\text{def}}{=} \{v \in S : \pi(v) = v\}$ 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 shall show that the probability that there exists a bijection $\pi : S \to S'$ such that $\pi(G'_S) = G'_{S'}$ is upper-bounded by

$$\sum_{\pi : S \stackrel{1\text{-}1}{\to} S'} \min\left(2^{-|\mathtt{FP}(\pi)| \cdot (\ell - |\mathtt{FP}(\pi)|)/3}, 2^{-\binom{(\ell - |\mathtt{FP}(\pi)|)/3}{2}}\right) \tag{1}$$

and observe that Eq. (1) equals

$$\begin{aligned}
&\sum_{f\in\{0,...,|S\cap S'|\}} \sum_{\pi:|\mathtt{FP}(\pi)|=f} \min\left(2^{-f\cdot(\ell-f)/3}, 2^{-\binom{(\ell-f)/3}{2}}\right)\\
&\leq \sum_{f\in\{0,...,|S\cap S'|\}} \frac{\ell!}{f!}\cdot 2^{-\max(6\cdot f\cdot(\ell-f),(\ell-f)\cdot(\ell-f-1))/18}\\
&\leq \sum_{f\in\{0,...,|S\cap S'|\}} \frac{\ell!}{f!}\cdot 2^{-(\ell-f)\cdot(6f+(\ell-f-1)))/36}\\
&< \frac{\ell!}{|S\cap S'|!}\cdot 2^{-\Omega((\ell-|S\cap S'|)\cdot\ell)} \qquad (2)
\end{aligned}$$

To justify the upper bound claimed in Eq. (1), we fix an arbitrary bijection $\pi : S \to S'$ and consider the directed graph defined by π. Observing that this subgraph consists of directed cycles (of vertices of S) and paths (ending at a vertex of S'), we identify a set $I \subseteq S\setminus \mathtt{FP}(\pi)$ such that $\pi(I)\cap I = \emptyset$ and $|I| \geq (\ell - |\mathtt{FP}(\pi)|)/3$. 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

$$\begin{aligned}
&\Pr_{G'}\left[\forall(u,v)\in\binom{S}{2}: \quad e_{\pi(G')}(\pi(u),\pi(v)) = e_{G'}(\pi(u),\pi(v))\right]\\
&\leq \Pr_{G'}\left[\forall(u,v)\in\mathtt{FP}(\pi)\times I: \quad e_{G'}(u,v) = e_{G'}(\pi(u),\pi(v))\right]\\
&= \prod_{(u,v)\in\mathtt{FP}(\pi)\times I} \Pr_{G'}\left[e_{G'}(u,v) = e_{G'}(u,\pi(v))\right]\\
&= 2^{-|\mathtt{FP}(\pi)|\cdot|I|}\\
&\leq 2^{-|\mathtt{FP}(\pi)|\cdot(\ell-|\mathtt{FP}(\pi)|)/3}
\end{aligned}$$

where the first equality is due to the disjointness of the sets $\mathtt{FP}(\pi)\times I$ and $\mathtt{FP}(\pi)\times\pi(I)$ (which in turn follows from $\pi(I)\cap I = \emptyset$), and the second equality is due 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

$$\begin{aligned}
&\Pr_{G'}\left[\forall\{u,v\}\in\binom{S}{2}: \quad e_{\pi(G')}(\pi(u),\pi(v)) = e_{G'}(\pi(u),\pi(v))\right]\\
&\leq \Pr_{G'}\left[\forall\{u,v\}\in\binom{I}{2}: \quad e_{G'}(u,v) = e_{G'}(\pi(u),\pi(v))\right]\\
&= \prod_{\{u,v\}\in\binom{I}{2}} \Pr_{G'}\left[e_{G'}(u,v) = e_{G'}(\pi(u),\pi(v))\right]\\
&= 2^{-\binom{|I|}{2}}\\
&\leq 2^{-\binom{(\ell-|\mathtt{FP}(\pi)|)/3}{2}}
\end{aligned}$$

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. This completes the justificaytion of Eq. (1).

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

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

where the index i represents the size of the intersection of S' with S. Using a sufficiently large $\ell = O(\log n)$, we upper-bound Eq. (3) by

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

which is $o(1)$. The claim follows. □

Conclusion. Using Claim 3.1.2, 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 3.1.2.)
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. ■

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, 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.

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