On Testing Asymmetry in the Bounded Degree Graph Model

Oded Goldreich*

December 25, 2023

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 $s(n) = \Omega(\log n)$, the query complexity of ϵ-testing the set of asymmetric n-vertex graphs in which each connected component has size at most $s(n)$ is at most $O(\sqrt{n} \cdot s(n)/\epsilon)$ and at least $\Omega(\sqrt{n^{1-O(\epsilon)}/s(n)})$.

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

Preliminary versions of this work appeared as TR20-118 of *ECCC*. The current version eliminates some inaccuracies and hand-wavings that appeared in the preliminary versions.

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 [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*.

A graph is called asymmetric if the identity permutation is its only automorphism. Specifically, 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) and by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 819702).

$G' = (V', E')$ 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 $\texttt{aut}(G)$, is the set of permutations that preserve the graph G; that is, $\pi \in \texttt{aut}(G)$ 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*.*

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 obtain partial results. Our first result refers to the complexity of testing asymmetry with a proximity parameter that vanishes at a moderate rate.

Theorem 1.2 (lower bound on the query complexity of testing asymmetric graphs (in the bounded-degree graph model)): *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 $O(\log n)$.*

Considering the related problem of testing the set of asymmetric graphs with small connected components, we are able to obtain matching upper abnd lower bounds.

Theorem 1.3 (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* $\text{poly}(\log n)$ *is at most $\widetilde{O}(n^{0.5}/\epsilon)$ and at least $\widetilde{\Omega}(n^{0.5-O(\epsilon)})$. Furthermore, the upper bound holds for one-sided error testers, whereas the lower bound holds also for general* (i.e., two-sided error) *testers.*

The results 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 $\text{poly}(s(n))$. 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]

We stress that Theorems 1.2 and 1.3 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.

Conventions. Throughout this work, we consider undirected simple graphs (i.e., no self-loops and parallel edges). A graph property is a set of such 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]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 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 2.1 (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.

2.1 Establishing Theorem 1.2

We generalize the claim of Theorem 1.2 by replacing the logarithmic bound with an arbitrary function $s : \mathbb{N} \to \mathbb{N}$ such that $s(n) = \Omega((\log n)/\log\log n)$.

Theorem 2.2 (Theorem 1.2, 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]:

(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. ■

A lower bound for one-sided error testers. Using the strategy of the proof of Theorem 2.2, one can show 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* $\exp(\omega(1/\epsilon(n)))$. This can be proved by setting $s(n) = \Theta(1/\epsilon(n))$, and considering a generic n-vertex graph that contains copies of all asymmetric $s(n)$-vertex graphs (as connected components). If the graph contains only such connected components, then it is $\epsilon(n)$-far from being asymmetric, and so an $\epsilon(n)$-tester must reject it (with probability at least 2/3). However, a one-sided error tester must see two isomorphic connected components in order to reject, beacuse the n-vertex graph may contain a single copy of each of these $s(n)$-vertex graphs along with an asymmetric $(n - N_d(s))$-vertex connected component.

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).

2.2 Establishing Theorem 1.3

We generalize the claim of Theorem 1.3 by replacing the polylogarithmic size bound (on the connected components) with an arbitrary function $s : \mathbb{N} \to \mathbb{N}$ such that $s(n) = \Omega((\log n)/\log\log n)$.

Theorem 2.3 (Theorem 1.3, generalized): *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 ϵ-testing $\Pi_n^{(s)}$ is $\Omega((n/s(n))^{0.5-O(\epsilon)})$.*

   In particular, the query complexity of $(1/(3d \cdot s(n)))$-testing $\Pi_n^{(s)}$ is $\Omega((n/s(n))^{0.5})$.

2. *There exists a one-sided error ϵ-tester for $\Pi^{(s)}$ that makes $O(n^{0.5} \cdot s(n)/\epsilon)$ queries, and runs in time $\widetilde{O}(n^{0.5}/\epsilon) \cdot \mathrm{poly}(s(n))$, provided that $n \mapsto s(n)$ can be computed in time $\mathrm{poly}(s(n))$.*[2]

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 $s(n)$-vertex graphs is $\exp(O(s \log s))$, which is smaller than $n/s(n)$. Theorem 2.3 follows by combining Propositions 2.4 and 2.5, which are stated and proved next.

Proposition 2.4 (lower bound on testing $\Pi^{(s)}$ (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 ϵ-testing the set of n-vertex asymmetric graphs that consist of connected components of size at most s is $\Omega((n/s(n))^{0.5-O(\epsilon)})$.*

Proof: We use the same ingredients as in the proof of Theorem 2.2, but generalize the argument as follows. Specifically, recall that, for $s(n) = \frac{c \log_2 n}{d \log_2 \log_2 n}$, we denote by C a collection of $m = n/s(n)$ non-isomorphic $s(n)$-vertex d-regular graphs that are asymmetric expanders. Letting $t = n^{O(\epsilon)}$, the claim of the proposition follows by showing that $\Omega(\sqrt{m/t})$ 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 in G_1 as a connected component.

2. A random isomorphic copy of the n-vertex graph that consists of t copies of each of m/t graphs selected at random in C; that is, we first select a random m/t-subset of C, denoted C', and take a random isomorphic copy of the n-vertex graph $G_{C'}$ that consists of t copies of each graph in C'.

Note that the first distribution is defined exactly as in the proof of Theorem 2.2, whereas the second distribution in the latter proof corresponds to the special case of $t = 2$. Recalling that each graph in the support of the first distribution is asymmetric (and so in $\Pi^{(s)}$), the key observation here is that *each graph in the support of the second distribution is ϵ-far from* $\Pi^{(s)}$. This observation relies on the setting of $t = n^{O(\epsilon)}$.

[2]Actaully, it suffices to require that $n \mapsto s(n)$ can be computed in time $\mathrm{poly}(s(n)) \cdot \sqrt{n}$.

Intuitively, making $G_{C'}$ be a member of $\Pi^{(s)}$ requires modifying almost all its connected components. Furthermore, $G_{C'}$ contains t copies of each $s(n)$-vertex graph in C', and these t copies have to be modified to different graphs. Using $t = n^{O(\epsilon)}$, it follows that there are at most $o(t)$ graphs that are $(1-o(1))\cdot\epsilon$-close to a fixed $s(n)$-vertex graph. Hence, modifying each component of $G_{C'}$ such that the resulting graph is in $\Pi^{(s)}$ requires more than $\epsilon dn/2$ edge modifications. The proof is completed by first showing that the distance of $G_{C'}$ to $\Pi^{(s)}$ is essentially reflected by such a modification (which presevres the connected components of $G_{C'}$).

In light of the above, we prove the foregoing observation in two steps. In the first step we show that *if $G_{C'}$ is ϵ-close to $\Pi^{(s)}$, then it can be made asymmetric by making $O(\epsilon dn)$ edge modifications, while preserving the partition of its vertices to connected components.* However, as shown in the second step, in this case the cost of the modification is $\omega(\epsilon dn)$, because the t connected components of $G_{c'}$ that are copies of the same $s(n)$-vertex graph in C' must be mapped to different $s(n)$-vertex graphs.

Claim 2.4.1 (reduction to a special case): *If $G_{C'}$ is ϵ-close to $\Pi^{(s)}$, then $G_{C'}$ can be made asymmetric by making $O(\epsilon dn)$ edge modifications, while preserving the partition of its vertices to connected components.*

Proof: We use the following two facts.

1. The graph $G_{C'}$ consists of connected components that are each an s-vertex expander graph, whereas each graph in $\Pi^{(s)}$ consists of connected components that are each of size at most s.
2. The cost (in edges) of splitting a connected component of $G_{C'}$ into two disconnected parts is at least a constant factor of the cost of modifying its smaller part arbitrarily, since each connected component is an expander (per Fact 1).

We shall use these facts in order to argue that, up to a constant factor, the distance between $G_{C'}$ and $\Pi^{(s)}$ can obtained by a mapping that sends connected components of $G_{C'}$ to connected components of a graph in $\Pi^{(s)}$.

We start by recalling that the hypothesis that $G_{C'} = ([n], E)$ is ϵ-close to $\Pi^{(s)}$ means that there exist $G' \in \Pi^{(s)}$ and a bijection $\pi : [n] \to [n]$ such that $\pi(G_{C'})$ and G' differ on at most $\epsilon \cdot dn/2$ edges. Fixing these G' and π, we say that a connected component of $G_{C'}$ is good if a strict majority of its vertices are mapped by π to the same connected component of G'.

Suppose first that all connected components of $G_{C'}$ are good. In this case, we consider the one-to-one function $\mu : [m] \to [n]$ that maps the connected components of $G_{C'}$ to (a subset of) the connected components of G' while keeping together the majority of vertices in each component; that is, $\mu(i) = j$ if a strict majority of the vertices of the $i^{\rm th}$ component of $G_{C'}$ are mapped (by π) to the $j^{\rm th}$ component of G'. Now, let $\pi' : [n] \to [n]$ be the bijection obtained from π by relocating the minority vertices of each connected component such that the vertices of the $i^{\rm th}$ component (of $G_{C'}$) that are not mapped to the $\mu(i)^{\rm th}$ component (of G') by π are mapped to it by π' (or rather π' maps these vertices to the $\mu(i)^{\rm th}$ component of some graph in $\Pi^{(s)}$). Then, in comparison to the distance between $\pi(G_{C'})$ and G', this relocation increases the distance between $\pi'(G_{C'})$ and $\Pi^{(s)}$ by an additive term of at most d times the number of minority vertices, but this is compensated by a decrease that is due to avoiding the splitting of these connected components (which by Fact 2 is a constant factor of the added term).

We now turn to the general case (where not all connected components of $G_{C'}$ are good). In this case, we consider a one-to-one mapping $\mu : [m] \to [n]$ such that if the $i^{\rm th}$ connected component of

$G_{C'}$ is good, then $\mu(i)$ equals the index of the connected component of G' that contains the majority of the vertices of the i^{th} connected component of $G_{C'}$. We define π' accordingly; that is, π' maps all vertices of the i^{th} connected component of $G_{C'}$ to the $\mu(i)^{\text{th}}$ connected components of some graph in $\Pi^{(s)}$. We bound the effect of moving from π to π' on the good connected components as in the special case. Turning to the non-good connected components, we observe that we can afford to modify all incidences in these components, because each of these components is split (by π) into two large disconnected parts, which means that their contribution to the symmetric difference between the graphs $\pi(G_{C'})$ and G' is also $\Omega(s(n))$ per component. ■

Claim 2.4.2 (analyzing the special case): *If $t = \omega(n^{4c\epsilon'})$, then making $G_{C'}$ asymmetric while preserving the partition of its vertices to connected components, requires $(1 - o(1)) \cdot \epsilon' dn$ edge modifications.*

Proof: We first observe that making the graph $G_{C'}$ asymmetric, while preserving the partition of its vertices to connected components, requires mapping the t copies of each graph in C' (which appear in different connected components in $G_{C'}$) to t different $s(n)$-vertex graphs. The point is that the number of s-vertex d-regular graphs that are ϵ'-close to a given graph is smaller than $\binom{ds}{2\epsilon' \cdot ds} \cdot (s+1)^{2\epsilon' \cdot ds} < (ds)^{4\epsilon' ds} < n^{4c\epsilon'}$, where the last inequality is due to $s = \frac{c \log_2 n}{\log_2 \log_2 n}$. Using $n^{4c\epsilon'} = o(t)$, it follows that making $G_{C'}$ asymmetric (while preserving its connected components) requires modifying at least $2\epsilon' \cdot ds$ incidences in almost each of its connected components (i.e., in $t - 1 - o(t)$ out of the t copies of each graph in C'). In total, this amounts to more than $(1 - o(1)) \cdot m \cdot \epsilon' \cdot ds = (1 - o(1)) \cdot \epsilon' \cdot dn$ edge modifications. ■

Conclusion. Using $\epsilon' = O(\epsilon)$ in Claim 2.4.2, and combining the two claim, it follows that *$G_{C'}$ is ϵ-far from $\Pi^{(s)}$*: Specifically, if $G_{C'}$ were ϵ-close to $\Pi^{(s)}$, then (by Claim 2.4.1) it could be made asymmetric while maintaining its connected components by making $O(\epsilon dn)$ edge modifications, in contradiction to Claim 2.4.2.

The fact that $\Omega(\sqrt{m/t})$ queries are necessary to distinguish the foregoing two distributions is proved by the "birthday" argument (as in the proof of Theorem 2.2). 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 $\binom{q}{2}/|C'|$, whereas $|C'| = m/t$. ■

Proposition 2.5 (upper bound on testing $\Pi^{(s)}$ (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 $\widetilde{O}(n^{1/2}/\epsilon) \cdot \text{poly}(s)$.*

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 $\widetilde{O}(m) \cdot \text{poly}(s)$ 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.

***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 most $\epsilon n/3$ of its vertices reside in connected components that are either of size exceeding s or are symmetric.

We shall show that, in this case, Step 4 of the algorithm rejects (w.h.p.).

Let S denote the set of all other vertices (i.e., vertices that reside in asymmetric connected components of size at most s), and let G_S denote the subgraph of G induced by S. Consider the graph G' that results by augmenting G_S with $(n - |S|)/s$ 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 Theorem 2.2). Recalling that G is ϵ-far from Π and $n - |S| \leq \epsilon n/3$, it follows that G' is $\epsilon/3$-far from being asymmetric. We shall show that, in this case, Step 4 of the algorithm rejects G' (w.h.p.), and it follows that it rejects G (w.h.p.).

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 (see the foregoing comment regarding the existence of such a collection).

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 equivalence 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. ■

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 3.1 (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 3.2), whereas in the second case one can afford to retrieve the entire graph.

Proposition 3.2 (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 3.2.1 (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 3.2.2 (uniqueness of the subgraph induced by S): *For every ℓ-subset S fixed in Step 1 of Construction 3.2.1, 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. 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 by 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 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

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

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 $|I| \geq (\ell - |\mathtt{FP}(\pi)|)/3$. 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.

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

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

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

which is $o(1)$. The claim follows. ■

Conclusion. Using Claim 3.2.2, 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 3.2.2.)

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, 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.