Electronic Colloquium on Computational Complexity, Revision 2 of Report No. 109 (2020)

On Testing Hamiltonicity in the Bounded Degree Graph Model

Oded Goldreich*

February 9, 2021

Abstract

We show that testing Hamiltonicity in the bounded-degree graph model requires a linear number of queries. This refers to both the path and the cycle versions of the problem, and similar results hold also for the directed analogues. In addition, we present an alternative proof for the known fact that testing Independent Set Size (in this model) requires a linear number of queries.

Note (Feb 2021): We just found out that the main result has been proved by Yuichi Yoshida and Hiro Ito more than a decade ago [YI].

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., [G17]). Much of this attention was devoted to testing graph properties in a variety of models including the dense graph model [GGR], and the bounded-degree graph model [GR02] (surveyed in [G17, Chap. 8] and [G17, Chap. 9], resp.). The current work refers to the bounded-degree model, in which graphs are represented by their incidence function 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

*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. 1041/08).

ISSN 1433-8092

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 (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 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 $\epsilon\cdot dn/2$.*

(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). Also, when stating that the query complexity is $\Omega(n)$, we mean that this bound holds for all sufficiently small $\epsilon>0$; that is, there exists a constant $\epsilon_0>0$ such that distinguishing between n-vertex graphs in Π and n-vertex graphs that are ϵ_0-far from Π requires $\Omega(n)$ queries.

For many natural graph properties, the query complexity of testing them in the bounded-degree model is known (see [G17, Sec. 9.6]). In particular, the complexity of testing Connectivity, Eulerian, Degree-Regularity, Subgraph-Freeness, and Minor-Freeness is $\mathrm{poly}(1/\epsilon)$, the complexity of testing Bipartiteness and Expansion is $\Theta(\sqrt{n})\cdot\mathrm{poly}(1/\epsilon)$, and the complexity of testing 3-Colorability and Independent Set Size is $\Theta(n)$. One property that is conspicuously missing from the foregoing list is Hamiltonicity.

Contents of this work. The main result presented in this work is that *the query complexity of testing Hamiltonicity in the bounded-degree graph model is* $\Theta(n)$. This refers both to the path and to the cycle versions, and analogous results hold also for the directed analogues. In addition, we present a full proof of the same lower bound for testing Independent Set Size (ISS), which was sketched in [BOT]. (Indeed, since our focus is on these lower bounds, we are using the more liberal definition of testing with two-sided error probability.)

The technique. Both our proofs (i.e., for Hamiltonian and ISS) use (local and gap-preserving) reductions from testing problems that are known to require linear query complexity. The reductions we use are the standard (polynomial-time) reductions that are employed in demonstrating the NP-completeness of these sets. The same holds for the lower bound for 3-Colorability presented by Bogdanov, Obata, and Trevisan [BOT]. In all cases, one needs to verify the following facts:

1. The reduction generates graphs of *bounded maximum degree.*

   This feature is often verified explicitly in the context of proving NP-completeness results, when the aim is establishing that hardness also holds for graphs of bounded degree.

2. The reduction is *local* in the sense that each incidence query to the generated graph can be answered by making few queries to the original instance (e.g., few incidence queries in the case that the original instance is also a graph).

   This feature is an expected consequence of the fact that the original reductions use simple gadgets that are connected in simple ways. Still this feature must be verified, since things that look simple are not necessarily local.

3. The reduction is *gap-preserving*; that is, object that are far from the original set are mapped to graphs that are far from the property.

   This issue also arises in the context of demonstrating the NP-hardness of approximation problems, but the gaps being preserved may be different (see [PY88] versus [GGR, Sec. 1.2.4.3]). In some cases, the standard reductions are gap-preserving, but in other cases some modifications are required (see, e.g., [PY88, P94]).

Specifically, we establish the query complexity lower bound for testing Hamiltonicity by reducing from a bounded version of Max-3SAT, for which a linear query complexity lower bound was established in [BOT]. The analogous result for ISS is obtained by reducing from 3-Colorability. (In contrast, the lower bound for testing ISS is claimed in [BOT] by referring to a reduction from a bounded version of Max-3LIN (going via Max-3SAT).)

Computational complexity versus property testing. The foregoing results and the way in which the lower bounds are proved beg the conjecture that any NP-complete problem regarding bounded-degree graphs is hard to test in the bounded-degree graph model. This is obviously wrong.

Consider, for example, the artificial problem of testing whether an n-vertex graph has a simple path of length $\sqrt{n}$. Although this set of bounded-degree graphs is NP-complete, it is almost trivial to test it in the bounded-degree graph model: Specifically, if $\epsilon \geq 2/\sqrt{n}$, then we accept (without making any query), and otherwise (i.e., $\epsilon < 2/\sqrt{n}$) we recover the entire graph (by making $O(n) = O(1/\epsilon^2)$ queries) and decide accordingly. The point is that any n-vertex graph is $\frac{\sqrt{n}}{dn/2}$-close to having a path of length $\sqrt{n}$, where $d \geq 1$ is the degree bound.

Alternatively, consider the set consisting of all bounded-degree graphs that are either Hamiltonian or contain an isolated vertex. Although this set is NP-complete, via a standard reduction that has an image that contains only connected graphs, it is almost trivial to test this set in the bounded-degree graph model (since any n-vertex graph is $\frac{d}{dn/2}$-close to this set).

In contrast, we observe that *there exists sets of bounded-degree graphs that are recognizable in polynomial-time and yet are extremely hard to test in the bounded-degree graph model.* This follows from the fact that the local reduction from testing 3LIN (mod 2) to testing 3-Colorability used by Bogdanov, Obata, and Trevisan [BOT] is invertible in polynomial-time (which is a common feature of reductions used in the context of NP-completeness proofs).[1] Indeed, their reduction actually demonstrates that the set of (3-colorable) graphs that are obtained by applying this reduction to satisfiable 3LIN (mod 2) instances is hard to test (i.e., requires linear query complexity in the bounded-degree graph model).[2]

[1] Of course, 3LIN (i.e., the satisfiability of linear equations (with three varaiables each) over GF(2)) is easily solvable in polynomial-time. Nevertheless, Bogdanov *et al.* [BOT] use a reduction of 3LIN to 3-Colorability (via 3SAT) that originates in the theory of NP-completeness in order to reduce between the testing problems.

[2] Like almost all reductions of this type, the analysis of the reduction actually refers to the promise problem induced by the image of the reduction (i.e., the image of both the yes- and no-instances).

We seize this opportunity to call attention to a celebrated computational problem whose testing complexity (in the bounded-degree graph model) is not fully understood: We refer to Graph Isomorphism, which is polynomial-time solvable for graphs of bounded-degree [L82]. As for testing Graph Isomorphism in the bounded-degree graph model, the following is known [G19].

1. The query complexity of *testing isomorphism to a fixed n-vertex graph* is $\widetilde{\Omega}(n^{1/2})$.
2. The query complexity of *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 $\mathrm{poly}(\log n)$, and in this case the lower bounds are tight [G19].

2 Testing Hamiltonicity

We first show that the directed version of the problem is extremely hard to test, where we refer to a bounded-degree model for directed graphs in which one can query both for outgoing and incoming edges (see [BR02] or [G17, Sec. 9.7.2]). Specifically, a directed graph $G = ([n], E)$ of degree bound d is represented by the incidence functions $g_{\mathtt{out}}, g_{\mathtt{in}} : [n] \times [d] \to [n] \cup \{0\}$ such that $g_{\mathtt{out}}(v,i)$ (resp., $g_{\mathtt{in}}(v,i)$) indicates the vertex incident at the i^{th} outgoing (resp., incoming) edge of v (where $g_{\mathtt{out}}(v,i) = 0$ (resp., $g_{\mathtt{in}}(v,i) = 0$) indicates that v has less than i outgoing (resp., incoming) edges).

Theorem 2 (testing directed Hamiltonicity is extremely hard): *Let $\mathcal{DH}$ be the set of directed graphs having a directed Hamiltonian path. For every constant $d \geq 3$, testing $\mathcal{DH}$ in the foregoing bounded-degree model* (with degree bound d) *requires a linear number of queries.*

Proof: We use a (local gap-preserving) reduction from the problem of distinguishing between satisfiable 3CNF formulae and 3CNF formulae that are far from being satisfied in the sense that any assignment to their variables violates many of their clauses. Specifically, our starting point is the following result of Bogdanov, Obata, and Trevisan [BOT].

Claim 2.1 (implicit in [BOT, Sec. 6]): *Suppose that 3CNF formulae with n variables and m clauses are represented by functions $L : [m] \times [3] \to [n] \times \{\pm 1\}$ such that $L(i,j) = (k, \sigma)$ if the j^{th}* literal *in the i^{th} clause is an occurrence of the variable x_k and the sign of this literal is σ. Then, for some universal constant c, distinguishing between the following two types of 3CNF formulae requires $\Omega(n) = \Omega(m)$ queries to the corresponding function.*

1. *Satisfiable 3CNF formulae in which every variable appears in at most c clauses.*
2. *3CNF formulae in which every variable appears in at most c clauses and every assignment satisfies less than 90% of the clauses.*

The claim holds even when the potential distinguisher is also given oracle access to a function $C : [n] \times [c] \to [m] \cup \{0\}$ such that $C(i,j)$ is the index of the clause *that contains the j^{th} occurrence of the variable x_i, where $C(i,j) = 0$ indicates that variable x_i appears in less than j clauses* (i.e., $C(i,j) = k$ iff $L(k, \ell) = (i, \sigma)$ for some $\ell \in [3]$ and $\sigma \in \{\pm 1\}$).

(Actually, as indicated in [GR20, Clm. 3.2], the foregoing claim holds even when the "structure" of the formula is fixed (i.e., C is universal) and the actual input consists merely of the signs of the various variable occurrences (i.e., the σ's).)

We observe that the standard polynomial-time reduction of `3SAT` to `Hamiltonicity` is suitable for our purposes; specifically, it is local and preserves gaps (in the sense detailed below). But first, let us recall the specific reduction that we have in mind. Given $L : [m] \times [3] \to [n] \times \{\pm 1\}$ and $C : [n] \times [c] \to [m] \cup \{0\}$ as above, we consider a directed (bounded-degree) graph $G = (V, E)$ consisting of two designated vertices (denoted s and t), m clause-vertices (denoted $c_1, ..., c_m$), and n variable-gadgets (described next), that are connected as follows.

- Each variable-gadget consists of $3c + 3$ vertices that form a path of edges in both directions; specifically, we denote the vertices of the i^{th} gadget by $v_{i,1}, ..., v_{i,3c+3}$, and connected $v_{i,j}$ and $v_{i,j+1}$ by a pair of anti-parallel edges.

- The first and last vertices of each variable-gadget are connected by directed edges to the first and last vertices of the next gadget; that is, we have edges going from both $v_{i,1}$ and $v_{i,3c+3}$ to both $v_{i+1,1}$ and $v_{i+1,3c+3}$.

- For every $i \in [n]$ and $j \in [c]$, if $C(i, j) = k$, then $v_{i,3j}$ and $v_{i,3j+1}$ are connected to c_k by edges that are directed as follows. If the i^{th} variable occurs positively in the k^{th} clause (i.e., $L(k, \ell) = (i, +1)$ for some $\ell \in [3]$), then the first edge is directed from $v_{i,3j}$ to c_k and the second edge is directed from c_k to $v_{i,3j+1}$; otherwise (i.e., $L(k, \ell) = (i, -1)$ (indicating a negative occurrence)), the first edge is directed from c_k to $v_{i,3j}$ and the second edge is directed from $v_{i,3j+1}$ to c_k.

  Vertices $v_{i,3j}$ and $v_{i,3j+1}$ are called the ports of the $C(i, j)^{\text{th}}$ clause (equiv., of the corresponding vertex $c_{C(i,j)}$) in the i^{th} gadget. The vertex $v_{i,3j-1}$ is not connected to any clause-vertex; it is only connected to $v_{i,3j-2}$ and $v_{i,3j}$.

- Directed edges go from the (source) vertex s to the first and last vertices of the first gadget (i.e., to $v_{1,1}$ and $v_{1,3c+3}$), and directed edges go from the first and last vertices of the last gadget (i.e., from $v_{n,1}$ and $v_{n,3c+3}$) to the (terminal) vertex t.

Note that $|V| = 2+(3c+3)\cdot n+m$, and that each vertex has at most three incoming (resp., outgoing) edges. Hence, we may use any constant $d \geq 3$. Recalling that $n \leq 3m$, we have $|V| < 10(c + 1)m$.

We note that the foregoing reduction constitutes a local reduction from the problem considered in Claim 2.1 to testing $\mathcal{DH}$, where the locality condition is due to the fact that we can determine incidences in the resulting graph G by making $O(1)$ queries to L and C. In particular, $v_{i,3j}$ and $v_{i,3j+1}$ are always incident to $c_{C(i,j)}$, and the direction of these edge is determined by the value of σ such that $L(C(i, j), \ell) = (i, \sigma)$, where $\ell \in [3]$. Hence, answering an incidence query regarding $v_{i,j}$ amounts to a single query to C and three queries to L. Likewise, c_k is always incident to vertices in the gadgets of the variables indicated in $L(k, 1), L(k, 2)$ and $L(k, 3)$, whereas the incident vertices (in these gadgets) can be determined by querying all the vertices of these gadgets (where, actually, c queries per gadget suffice).

The standard proof of the validity of the foregoing reduction asserts that G has an Hamiltonian path if and only if the original formula is satisfiable, but we need to strengthen the negative direction and show that G is far from Hamiltonian if each assignment violates more than 10% of the clauses.

As a warm-up, we recall the argument for the positive direction: Assuming that $\tau : [n] \to \{0,1\}$ is a truth assignment that satisfies the original formula, we present the following Hamiltonian path in G. Essentially, the path traverses all gadgets one after the other, where the $i^{\rm th}$ gadget is traversed in the direction determined by the value of $\tau(i)$, and while possibly taking detours to visit some of the clause-vertices. Specifically, the $i^{\rm th}$ gadget is traversed from $v_{i,1}$ to $v_{i,3c+3}$ if $\tau(i) = 1$, and in the opposite direction otherwise. If x_i is the first variable that satisfies the $k^{\rm th}$ clause, then we avoid traversing the edge that connects the ports of c_k in the $i^{\rm th}$ variable-gadget (i.e., the edge connecting $v_{i,3j}$ and $v_{i,3j+1}$ such that $C(i,j) = k$), and take a detour via c_k instead. (Note that if x_i appears positively in the $k^{\rm th}$ clause, then $\tau(i) = 1$ must hold, and we traverse the $i^{\rm th}$ gadget from $v_{i,1}$ to $v_{i,3c+3}$, which means that the detour has the form $v_{i,3j} \to c_k \to v_{i,3j+1}$; otherwise, we traverse the $i^{\rm th}$ gadget from $v_{i,3c+3}$ to $v_{i,1}$, and the detour has the form $v_{i,3j+1} \to c_k \to v_{i,3j}$.)

We now establish the opposite direction, or rather its contrapositive; that is, assuming that the graph G is close to being Hamiltonian, we show the existence of an assignment to the original formula that satisfies more than 90% of its clauses. Suppose that G is δ-close to a graph G' that is Hamiltonian, and consider a Hamiltonian path P' in G'. Omitting from G' all edges that are not in $G = (V,E)$, we obtain a collection $\mathcal{P}$ of at most $\delta \cdot d|V|$ vertex-disjoint paths that use only edges of G and cover all vertices of G' (equiv., all vertices of G). We shall see that these paths can be used very much as a single path is used in the original proof of validity, except that here we only obtain an assignment that satisfies a $1 - O(\delta)$ fraction of the clauses (rather than all clauses), but this suffices.

Fixing $\mathcal{P}$, we say that a gadget is good (in $\mathcal{P}$) *if there exists a path $P \in \mathcal{P}$ that passes through all vertices of the gadget such that if the path uses one of the edges that connect the gadget to a clause-vertex, then it uses both these edges.* (Specifically, suppose that $v_{i,3j}$ and $v_{i,3j+1}$ are connected to c_k in G (via one edge directed to c_k and one edge that is directed from c_k), then the edge connecting $v_{i,3j}$ and c_k is on the path if and only if the edge connecting $v_{i,3j+1}$ and c_k is on the path.) The key observation, proved next, is that the number of gadgets that are not good (i.e., bad gadgets) can be upper-bounded in terms of $|\mathcal{P}|$.

Claim 2.2 (bad gadgets): *If a vertex-gadget is bad* (i.e., not good)*, then it contains a vertex that is an endpoint of some path in $\mathcal{P}$.*

Proof: Suppose that the $i^{\rm th}$ gadget is bad. The easy case is that the path $P \in \mathcal{P}$ that covers $v_{i,1}$ does not cover all vertices of the gadget but does satisfy the condition regarding the use of edges that connect the gadget to clause-vertices. In this case we take the minimum j such that $v_{i,j+1}$ is not on P, and observe that if $j > 1$ then $v_{i,j}$ is an endpoint of P, and otherwise (i.e., $j = 1$) vertex $v_{i,2}$ is an endpoint of some other path in $\mathcal{P}$.

The other case is that P does not satisfy the condition regarding the use of edges that connect the gadget to clause-vertices. Suppose, without loss of generality, that the edge connecting $v_{i,3j}$ and c_k is on P, but the edge connecting $v_{i,3j+1}$ and c_k is not on P. Still $v_{i,3j+1}$ must be covered by some path $P' \in \mathcal{P}$ (possibly $P' = P$), and in this case we show that either $v_{i,3j+1}$ is an endpoint of P' or $v_{i,3j-1}$ is an endpoint of some path in $\mathcal{P}$. We first note that $v_{i,3j+1}$ is connected (in G) only to the vertices $v_{i,3j}, v_{i,3j+2}$ and c_k, whereas P uses the edge connecting $v_{i,3j}$ and c_k but not the edge connecting $v_{i,3j+1}$ and c_k. Now, we consider two cases.

1. If $P' \neq P$, then P' can use only one edge incident to $v_{i,3j+1}$, since P covers the other neighbors (i.e., $v_{i,3j}$ and c_k) whereas the paths are vertex-disjoint. Hence, $v_{i,3j+1}$ is an endpoint of P'.

2. Otherwise (i.e., $P' = P$) either $v_{i,3j+1}$ is an endpoint of P' or $v_{i,3j+1}$ is connected by P' to both $v_{i,3j}$ and $v_{i,3j+2}$. Recalling that $v_{i,3j}$ is connected by P to c_k, it follows that $v_{i,3j-1}$ cannot be connected to $v_{i,3j}$ by a path in $\mathcal{P}$, which implies that $v_{i,3j-1}$ must be an endpoint of some path in $\mathcal{P}$ (since $v_{i,3j-1}$ is connected in G only to $v_{i,3j-2}$ and $v_{i,3j}$).

Hence, in all cases we showed that the $i^{\rm th}$ gadget contains an endpoiunt of some path in $\mathcal{P}$. ■

We conclude that at most $B \le 2 \cdot |\mathcal{P}| \le 2 \cdot \delta \cdot d|V|$ of the gadgets are bad. Actually, the sum of the number of bad gadgets (i.e., B) and the number B' of clause-vertices that are isolated in $\mathcal{P}$ (i.e., constitute trivial paths in $\mathcal{P}$) is at most $2d\delta|V|$. Defining an assignment to the variables according to the direction in which paths traverse good gadgets, we obtain a partial assignment that satisfies all clauses whose clause-vertices are covered by paths that are incident at good gadgets. Since there are at most $c \cdot B$ clauses that are incident at bad gadgets, we infer that at most $cB + B' \le c \cdot 2d\delta|V|$ of the clauses are not satisfied. Recalling that $|V| < 10(c+1)m$, it follows that the original formula has an assignment that satisfies more than a $1 - c' \cdot \delta$ fraction of the clauses, where $c' = 2cd \cdot 10(c+1)$. Using $\delta = 0.1/c'$, the theorem follows (for any constant $d \ge 3$). ■

Remark 3 (directed Hamiltonicity in the sense of directed cycle): *The proof of Theorem 2 can be easily adapted in order to establish the same result for the set of directed graphs having Hamiltonian cycles. All that is needed is to augment the construction by an edge leading from the terminal vertex t to the source vertex s.*

We now turn to establish the claim for the undirected case.

Theorem 4 (testing Hamiltonicity is extremely hard): *Let $\mathcal{H}$ be the set of Hamiltonian graphs either in the sense of having a Hamiltonian path or in the sense of having a Hamiltonian cycle. For every constant $d \ge 4$, testing $\mathcal{H}$ in the standard bounded-degree model* (with degree bound d) *requires a linear number of queries.*

Proof: We use a (local gap-preserving) reduction from $\mathcal{DH}$ and invoke either Theorem 2 or Remark 3. Again, the standard polynomial-time reduction will do. Recall that this reduction replaces each vertex v in the directed graph by a three-vertex path, denoted $(v_{\tt in}, v_{\tt mid}, v_{\tt out})$, while replacing directed edges going into v (resp., out of v) by edges incident at $v_{\tt in}$ (resp., at $v_{\tt out}$); that is, the directed edge from v to u is replaced by an (undirected) edge connecting $v_{\tt out}$ and $u_{\tt in}$. Hence, given oracle access to the incidence functions of a directed graph $\vec{G} = (\vec{V}, \vec{E})$ of maximum in-degree and out-degree d, we locally construct an undirected graph $G = (V, E)$ of maximum degree $d + 1$ (e.g., the $(i+1)^{\rm st}$ neighbor of $v_{\tt out}$ is $u_{\tt in}$ if the $i^{\rm th}$ outgoing edge of v goes to u).

To see that this standard reduction is actually gap-preserving we need to show that if $\vec{G} = (\vec{V}, \vec{E})$ is far from being Hamiltonian (in the directed sense) then so is $G = (V, E)$ (in the undirected sense). We prove the contrapositive. Suppose that G is δ-close to G' that is Hamiltonian. Then, G' has at most $N = \delta \cdot (d+1)|V|/2 = \delta \cdot (d+1) \cdot 3|\vec{V}|/2$ edges that are not in G. This means that the vertices of G are covered by a collection of at most N vertex-disjoint paths. Using the correspondence between the directed edges of $\vec{G}$ and the edges connecting three-vertex paths in G, which implies a correspondence between simple paths in G and simple directed paths in $\vec{G}$, we infer that the vertices of $\vec{G}$ are covered by a collection of at most N vertex-disjoint directed paths. Hence, it suffices to add at most N directed edges to $\vec{G}$ in order to obtain a Hamiltonian graph, whereas at most $2N$ edges may need to be removed in order to maintain the degree bound. Thus, $\vec{G}$ is δ'-close to being Hamiltonian, where $\delta' = \frac{3N}{d|\vec{V}|} = \frac{9(d+1)\cdot\delta}{d/2} = \frac{18(d+1)}{d} \cdot \delta$. ■

3 Testing Independent Set Size

As mentioned in the introduction, the following result is implicit in [BOT].

Theorem 5 (testing independent set size is extremely hard): *Let $\mathcal{IS}$ be the set of graphs having an independent set that contains at least one third of the vertices. For all sufficiently large constant d, testing $\mathcal{IS}$ in the bounded-degree model* (with degree bound d) *requires a linear number of queries.*

Specifically, by [BOT, Thm. 22] approximating Minimum Vertex Cover to within a factor of 7/6 requires a linear number of queries, where the (sketched) proof uses a reduction from a bounded version of Max-3LIN. Even a superficial look at the reduction reveals the fact that the hard instances have vertex cover (and independent sets) of constant density. Hence, Theorem 5 follows by padding. We believe that the following proof is more appealing.

Proof: We use the fact that, in this very model, for sufficientrly large constant d, testing 3-Colorability requires a linear number of queries [BOT], and (locally) reduce testing 3-Colorability to testing $\mathcal{IS}$. The reduction is quite straightforward: Given a graph $G = ([n], E)$, we construct a graph $G' = (V', E')$ that consists of three copies of the graph G as well as edges connected the corresponding vertices in these copies. Specifically, $V' = \{1, 2, 3\} \times [n]$ and

$$\begin{aligned} E' \stackrel{\text{def}}{=} & \bigcup_{i\in\{1,2,3\}} \{\{(i,u),(i,v)\} : \{u,v\}\in E\} \\ & \cup \bigcup_{v\in[n]} \{\{(i,v),(j,v)\} : i\neq j\in\{1,2,3\}\}\,. \end{aligned}$$

(Hence, if G has maximum degree d, then G' has maximum degree $d + 2$.)

Given a tester for $\mathcal{IS}$, we obtain a tester for 3-Colorability by invoking the former tester and emulating the graph G' in a straightforward manner; that is, queries regarding the incidences of a vertex $(i, v) \in V'$ are answered by querying G on the incidences of vertex $v \in [n]$.

Evidently, if G is 3-colorable, then $G' \in \mathcal{IS}$; specifically, if $\chi : [n] \to \{1, 2, 3\}$ is a legal 3-coloring of G, then $\{(i, v) : v\in[n] \wedge \chi(v)=i\}$ is an n-vertex independent set in G'. Hence, we focus on showing that if G is ϵ-far from being 3-colorable, then G' is $\Omega(\epsilon)$-far from $\mathcal{IS}$. We prove the contrapositive. That is, assuming that G' is δ-close to $\mathcal{IS}$, we shall prove that G is $O(\delta)$-close to being 3-colorable.

Let G'' be a graph that is δ-close to G' and has an independent set of size n, denoted S''. Then, the number of edges in the subgraph of G' induced by S'' is at most $\delta\cdot(d+2)\cdot|V'|/2 = 3\delta(d+2)n/2$, and twice this number upper-bounds the number of vertices in S'' that are incident to an edge of E'. Letting S' denote the set of vertices in S'' that are not incident to an edge in E', we infer that S' is an independent set in G', whereas $|S'| \geq (1 - \delta \cdot 3(d + 2)) \cdot n$. Observing that every vertex of G contains at most one copy in S' (i.e., $|S' \cap \{(v, i) : i\in\{1, 2, 3\}\}| \leq 1$ for every $v \in [n]$), we obtain a 3-coloring χ of the subgraph of G that is induced by $S = \{v : \exists i \text{ s.t. } (i, v) \in S'\}$; specifically, $\chi(v) = i$ if and only if $(i, v) \in S'$. (Indeed, each $\chi^{-1}(i)$ is an independent set.) Omitting all edges that are incident at $[n] \setminus S$, we infer that G is δ'-close to 3-coloring, where $\delta' = \frac{d\cdot(n-|S|)}{d/2} = 2 \cdot (n - |S'|) \leq 2 \cdot \delta \cdot 3(d + 2) = 6(d + 2)\delta$. The claim follows. ■

Acknowledgements

I am grateful to Dana Ron for useful discussions.

References

[BR02] Michael Bender and Dana Ron. Testing Properties of Directed Graphs: Acyclicity and Connectivity. *Random Structures & Algorithms*, Vol. 20 (2), pages 184–205, 2002.

[BOT] Andrej Bogdanov, Kenji Obata, and Luca Trevisan. A Lower Bound for Testing 3-Colorability in Bounded-Degree Graphs. In *43rd IEEE Symposium on Foundations of Computer Science*, pages 93–102, 2002.

[G17] Oded Goldreich. *Introduction to Property Testing.* Cambridge University Press, 2017.

[G19] Oded Goldreich. Testing Isomorphism in the Bounded-Degree Graph Model. *ECCC*, TR19-102, 2019.

[GGR] Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, pages 653–750, July 1998.

[GR02] Oded Goldreich and Dana Ron. Property Testing in Bounded Degree Graphs. *Algorithmica*, Vol. 32 (2), pages 302–343, 2002.

[GR20] Oded Goldreich and Dana Ron. The Subgraph Testing Model. *TOCT*, to appear. Available from the authors' webpages.

[L82] Eugene Luks. Isomorphism of Graphs of Bounded Valence can be Tested in Polynomial Time. *Journal of Computer and System Science*, Vol. 25 (1), pages 42–65, 1982.

[PY88] Christos Papadimitriou and Mihalis Yannakakis. Optimization, Approximation, and Complexity Classes. In *20th ACM Symposium on the Theory of Computing*, pages 229–234, 1988.

[P94] Erez Petrank. The Hardness of Approximation: Gap Location. *Computational Complexity*, Vol. 4, pages 133–157, 1994.

[YI] Yuichi Yoshida and Hiro Ito. Query-number Preserving Reductions and Linear Lower Bounds For Testing. *IEICE Trans. Inf. Syst.*, Vol. 93-D (2), pages 233–240, 2010.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il