On Proximity Oblivious Testing*

Oded Goldreich†
Department of Computer Science
Weizmann Institute of Science
Rehovot, Israel.
oded.goldreich@weizmann.ac.il

Dana Ron‡
Department of EE–Systems
Tel-Aviv University
Ramat-Aviv, Israel.
danar@eng.tau.ac.il

February 9, 2011

Abstract

We initiate a systematic study of a special type of property testers. These testers consist of repeating a basic test for a number of times that depends on the proximity parameter, whereas the basic test is oblivious of the proximity parameter. We refer to such basic tests by the term proximity-oblivious testers.

While proximity-oblivious testers were studied before – most notably in the algebraic setting – the current study seems to be the first one to focus on graph properties. We provide a mix of positive and negative results, and in particular characterizations of the graph properties that have constant-query proximity-oblivious testers in the two standard models (i.e., the adjacency matrix and the bounded-degree models). Furthermore, we show that constant-query proximity-oblivious testers do not exist for many easily testable properties, and that even when proximity-oblivious testers exist, repeating them does not necessarily yield the best standard testers for the corresponding property.

Keywords: Property Testing, Graph Properties

*An extended abstract of this work appeared in the proceedings of the *41st Annual ACM Symposium on Theory of Computing* (*STOC*), pages 141–150, 2009.

†Partially supported by the Israel Science Foundation (grants No. 460/05 and 1041/08).

‡Partially supported by the Israel Science Foundation (grant No. 89/05 and 246/08).

Contents

1 Introduction

In the last decade, the area of property testing has attracted much attention (see, the surveys [12, 26] as well as the more recent ones [27, 28]). Loosely speaking, property testing typically refers to sub-linear time probabilistic algorithms 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; that is, the object is seen as a function and the testers get oracle access to this function (and thus may be expected to work in time that is sub-linear in the length of the object).

The foregoing description refers to the notion of "far away" objects, which in turn presumes a notion of distance between objects as well as a parameter determining when two objects are considered to be far from one another. The latter parameter is called the proximity parameter, and is often denoted ϵ; that is, one typically requires the tester to reject with high probability any object that is ϵ-far from the property.

Needless to say, in order to satisfy the aforementioned requirement, any tester (of a reasonable property) must obtain the proximity parameter as auxiliary input (and determine its actions accordingly). The question, addressed in this work, is what does the tester do with this parameter (or how does the parameter affect the actions of the tester). A very minimal effect is exhibited by testers that, based on the value of the proximity parameter, determine the number of times that a basic test is invoked, where the basic test is oblivious of the proximity parameter. For example, the celebrated linearity tester of [10] repeats a basic test that consists of selecting two random points, x and y, and probing the value of the function at the points x, y, and $x + y$. This basic test is repeated for a number of times that is inversely proportional to the proximity parameter.

Our focus is on such basic tests (i.e., basic tests that are oblivious of the proximity parameter). We call such tests proximity oblivious, and note that they are implicit in prior works; most notably in the context of testing algebraic properties (see, e.g., [29] and [22]) and testing monotonicity (e.g., [15]). In this work we initiate a general study of proximity oblivious testers, and consider a variety of questions regarding them, while focusing on testing graph properties (in two standard models). Specifically, we ask:

- Which properties have proximity oblivious tests (of small query complexity)?
- How does the detection probability of such tests grow as a function of the distance of the object from the property, and how does this relate to the query complexity of the best (standard) tester for the corresponding property.

For a precise formulation of proximity-oblivious testers and a summary of our results, see Sections 2 and 3, respectively.

Motivation: Property testing can be thought of as relating local views to global properties, where the local view is provided by the queries and the global property is the distance to a predetermined set. Proximity-oblivious testing takes this relation to an extreme by making the local view independent of the distance. In other words, it refers to the smallest local view that may provide information about the global property (i.e., the distance to a predetermined set). A major motivation for our study is that understanding a natural subclass of testers (i.e., proximity-oblivious ones) may shed light on property testing at large.

2 Definitional Treatment

In continuation to the introduction, we consider proximity-oblivious testers, and note that standard testers (which err with probability at most $1/3$)[1] may be obtained by repeating these proximity-oblivious testers for an appropriate number of times.

Definition 2.1 (vanilla version): *Let Π be a set of functions over a finite set Ω. A* proximity-oblivious tester *for Π is a probabilistic oracle machine T that satisfies the following two conditions:*

1. *The tester accepts each function in Π with probability 1; that is, for every $f \in \Pi$ it holds that $\Pr[T^f = 1] = 1$.*
2. *For some monotone function $\rho : (0,1] \to (0,1]$, each function $f \notin \Pi$ is rejected by T with probability at least $\rho(\delta_\Pi(f))$, where*

$$\delta_\Pi(f) \stackrel{\rm def}{=} \min_{g\in\Pi}\{\delta(f,g)\} \text{ and } \delta(f,g) \stackrel{\rm def}{=} \Pr_{x\in\Omega}[f(x) \neq g(x)]. \quad (1)$$

   The function ρ is called the detection probability of the tester *T.*

Indeed, we require that $\rho(\epsilon) > 0$ for every $\epsilon > 0$, whereas extending Item 2 to $f \in \Pi$ (while avoiding contradiction with Item 1) mandates extending ρ so that $\rho(0) = 0$. The requirement that ρ is monotone (i.e., monotonically increasing) does not rule out cases where the tight lower-bound is non-monotone (e.g., [7]), because ρ is not required to be tight.[2] Also, we may assume, without loss of generality, that $\rho(\epsilon) \leq \epsilon$.

We note that (as outlined in the introduction), using a proximity-oblivious tester T (as in Definition 2.1), we can obtain a standard (one-sided error) tester (of error probability at most $1/3$). Specifically, given the proximity parameter ϵ, the standard tester invokes T for $\Theta(1/\rho(\epsilon))$ times, and accepts if and only if all these invocations accept.

Note that it is natural to require one-sided error in (Item 1 of) Definition 2.1, because otherwise functions in Π may be accepted with probability that is lower than the acceptance probability of some functions that are not in Π (but are close to Π). This presupposes that Item 2 of Definition 2.1 remains intact. For a discussion of an alternative formulation, which allows two-sided error, see Section 6.3.

Definition 2.1 does not specify the query complexity of the (proximity-oblivious) tester, and indeed an oracle machine that queries the entire domain of the function qualifies as a (proximity-oblivious) tester (with detection probability $\rho(\epsilon) = 1$ for every $\epsilon > 0$). Needless to say, we are interested in (proximity-oblivious) testers that have significantly lower query complexity. To facilitate an asymptotic treatment, we refer to infinite families of finite functions, and provide the tester with the size of the function's domain.

[1] Analogously to Definition 2.1, a standard tester for a property Π is a probabilistic oracle machine T that satisfies the following conditions:

1. The tester accepts each $f \in \Pi$ with probability at least $2/3$; that is, for every $f \in \Pi$ and every $\epsilon > 0$, it holds that $\Pr[T^f(\epsilon) = 1] \geq 2/3$.
2. Given any $\epsilon > 0$ and oracle access to any f that is ϵ-far from Π (i.e., $\delta_\Pi(f) > \epsilon$), the tester rejects with probability at least $2/3$ (i.e., $\Pr[T^f(\epsilon) = 0] \geq 2/3$).

We say that the tester has one-sided error if it accepts each $f \in \Pi$ with probability 1 (i.e., for every $f \in \Pi$ and every $\epsilon > 0$, it holds that $\Pr[T^f(\epsilon) = 1] = 1$).

[2] In fact, it suffices to require that for every $x > 0$ it holds that $\rho'(x) \stackrel{\rm def}{=} \inf_{y\geq x}\{\rho(y)\} > 0$. Indeed, in such a case, ρ' is a monotonically non-decreasing lower-bound (of ρ). Furthermore, we may obtain a monotonically increasing lower-bound (of ρ) by defining $\rho''(x) \stackrel{\rm def}{=} (1 + x) \cdot \rho'(x)/2$.

Definition 2.2 (main version): *Let* $\Pi = \bigcup_{n\in\mathbb{N}} \Pi_n$*, where* Π_n *contains functions defined over the domain* $[n] \stackrel{\rm def}{=} \{1, ..., n\}$*, and let* $\rho : (0, 1] \to (0, 1]$ *be monotone. A* proximity-oblivious tester with detection probability ρ for Π *is a probabilistic oracle machine* T *that satisfies the following two conditions:*

1. *For every* $n \in \mathbb{N}$ *and* $f \in \Pi_n$*, it holds that* $\Pr[T^f(n) = 1] = 1$.
2. *For every* $n \in \mathbb{N}$ *and* $f : [n] \to \{0, 1\}^*$ *not in* Π_n*, it holds that* $\Pr[T^f(n) = 0] \geq \rho(\delta_{\Pi_n}(f))$*, where* $\delta_{\Pi_n}(f) = \min_{g\in\Pi_n}\{\delta(f, g)\}$ (as in Eq. (1)).

Definition 2.2 can be further extended so to cover also (proximity-oblivious) testers that obtain other parameters of the function being tested (e.g., a degree bound in the case of testing low-degree polynomials). Note that Definition 2.2 mandates that the detection probability is only a function of the relative distance to the property; indeed, one may relax this requirement but one should stay away from the trivial lower-bound (which corresponds to only requiring that for every $f \notin \Pi$ there exists a computation of T^f that rejects).

3 Summary of our Results

Recall that the (three-query) linearity test of [10] is actually a proximity-oblivious tester, and that its detection probability is linear (i.e., $\rho(\epsilon) = \Omega(\epsilon)$). The same holds also for several known low-degree tests (see, e.g., [29]), testers of monotonicity (e.g., [15]), and some of the results regarding locally testable codes (see [19] and the end of Section 6). In this work, we study the existence and quality (i.e., ρ) of efficient proximity-oblivious testers in other domains, most importantly in the domain of testing graph properties.

3.1 In the dense graphs model

We start (in Section 4) with the setting of testing properties of graphs in the adjacency matrix model (introduced in [16]). We consider several natural properties and show constant-query proximity-oblivious testers of optimal (up to a constant factor) detection probability. For example, we show that:

1. The set of graphs each consisting of a collection of isolated cliques has a three-query proximity-oblivious tester of quadratic detection probability (i.e., $\rho(\epsilon) = \Omega(\epsilon^2)$), whereas no constant-query proximity-oblivious tester of this property can do better (i.e., have detection probability $\rho(\epsilon) = \omega(\epsilon^2)$). We note that this property has a standard (adaptive) tester of $\widetilde{O}(1/\epsilon)$-query complexity [18, Sec. 3].
2. For every integer $c \geq 2$, the set of graphs consisting of up to c isolated cliques has a c^2-query proximity-oblivious tester, and the optimal detection probability is $\rho(\epsilon) = \Theta(\epsilon^{c/2})$. We note that these properties have a standard (non-adaptive) tester of $\widetilde{O}(1/\epsilon)$-query complexity [18, Sec. 6].

In contrast to the aforementioned positive results, we show that the set of bipartite graphs has no constant-query proximity-oblivious tester, although it does have a standard tester of poly$(1/\epsilon)$-query complexity [16, 5].

Summarizing the lessons from the foregoing examples, we note that they provide negative examples to both research projects advocated in the introduction. That is:

- There exist easily testable properties that do not have constant-query proximity oblivious tests. Indeed, this is demonstrated by the result for bipartiteness.

- For properties that do have constant-query proximity oblivious tests, the standard tester derived from the best possible proximity oblivious test is significantly inferior to some other (standard) tester. Indeed, this is demonstrated by the result for the property of being a collection of c isolated graphs, since the derived standard tester has query complexity $\Omega(\epsilon^{-c/2})$ (whereas this property has a standard $\widetilde{O}(\epsilon^{-1})$-query tester).

Addressing the first foregoing research project, we *characterize the class of graph properties having constant-query proximity-oblivious testers.*

Theorem 3.1 (loosely stated, cf. Theorem 4.7): *A graph property has constant-query proximity-oblivious testers* (in the dense graph model) *if and only if it expressible as an induced subgraph freeness property.*[3]

Indeed, this class is rather restricted when compared to the class of graph properties having a standard tester of complexity that only depends on ϵ (as characterized in [4]).

We also provide a method for determining the optimal (up to a constant factor) detection probability function of any property that has a constant-query proximity-oblivious tester (cf. Theorem 4.8). This method refers to the corresponding family of forbidden (induced) subgraphs, and the aforementioned tight quantitative results are obtained using it.

3.2 In the bounded-degree graphs model

Next (in Section 5), we turn to testing graph properties in the bounded-degree model (introduced in [17]). In this model, we also *characterize the class of graph properties having constant-query proximity-oblivious testers.* Interestingly, this class is a strict superset of the class of properties having such testers in the adjacency matrix model. We note that, also in the current model, the class of properties having constant-query proximity-oblivious testers is rather restricted when compared to the class of graph properties having a standard tester of complexity that only depends on ϵ (as explored in [17, 9]).

The characterization of the class of graph properties having constant-query proximity-oblivious testers in the bounded-degree model gives rise to a *generalized notion of subgraph freeness*, which may be of independent interest (see Definition 5.1). This notion generalizes both the notions of non-induced and induced subgraph freeness, and is more expressive than the latter. For example, the generalized notion allows to capture non-hereditary properties such as (degree) regularity. Our characterization refers to an auxiliary condition, which we term *non-propagating* (see Definition 5.3).

Theorem 3.2 (loosely stated, cf. Theorem 5.5): *A graph property has constant-query proximity-oblivious testers* (in the bounded-degree graph model) *if and only if it expressible as an general subgraph freeness property that satisfies the non-propagation condition. This class strictly contains all induced subgraph freeness properties.*

Indeed, we do not know whether every general subgraph freeness property satisfies the non-propagation condition (see Open Problem 5.8).

[3]Loosely speaking, an **induced subgraph freeness property** is a set of graphs that does not contain certain graphs as induced subgraphs. That is, such a property is determined by a finite set of finite graphs, denoted $\mathcal{F}$, and it consists of all graphs G such that no induced subgraph of G is in $\mathcal{F}$.

Focusing on induced subgraph freeness properties (which do have constant-query proximity-oblivious testers in both models), we note that *the detection probability in the bounded-degree model is a polynomial that depends on the number of connected components in the individual graphs of the forbidden family* (i.e., $\rho(\epsilon) = \Omega(\epsilon^c)$, where c is the maximum number of connected components in any forbidden graph). This is very different from the behavior in the dense graphs model, where even for $c = 1$ (i.e., connected forbidden subgraphs) the detection probability varies from linear to quadratic and to super-polynomial (i.e., $\rho(\epsilon) = \epsilon$ versus $\rho(\epsilon) = \Theta(\epsilon^2)$ versus $\rho(\epsilon) < \epsilon^{\Omega(\log(1/\epsilon))}$).

The technical angle. We comment that the techniques establishing the characterizations in the two different graph testing models are quite different (as one should expect given the different nature of the two models). In particular, as hinted above, the analysis of the bounded-degree model seems more novel.

3.3 Generic observations and discussions

Finally (in Section 6), we present a few generic observations. Specifically, we relate the existence of constant-query proximity-oblivious testers to the existence of constant-size refutations of membership (or proofs of non-membership) and certain testers that reject based on such refutations. We also shortly discuss the possibility of allowing proximity-oblivious testers two have two-sided error probability.

We note that, in the context of locally testable codes (LTCs), proximity-oblivious (codeword) testers are related to *strong* codeword tests (as in [19, Def. 2.2]), whereas standard (codeword) testers are related to the standard definition of codeword tests (termed *weak* in [19, Def. 2.1]).

4 Testing Graph Properties in the Adjacency Matrix Model

In the adjacency matrix model, an N-vertex graph $G = ([N], E)$ is represented by the Boolean function $g : [N] \times [N] \to \{0,1\}$ such that $g(u,v) = 1$ if and only if u and v are adjacent in G (i.e., $\{u,v\} \in E$). Distance between graphs is measured in terms of their aforementioned representation (i.e., as the fraction of (the number of) different matrix entries (over N^2)), but occasionally we shall use the more intuitive notion of the fraction of (the number of) edges over $\binom{N}{2}$.

Notation. For a fixed graph $G = ([N], E)$, we denote the set of neighbors of vertex $v \in [N]$ by $\Gamma(v)$; that is, $\Gamma(v) \stackrel{\text{def}}{=} \{u : \{u,v\} \in E\}$.

4.1 A few illustrative results

We start with the simple case of testing whether a graph is a clique.

Proposition 4.1 `Clique` *has a single-query proximity-oblivious tester with detection probability $\rho(\epsilon) = \epsilon$, where* `Clique` *is the set of all graphs consisting of a single clique.*

Proof: The claim follows by considering the straightforward tester that uniformly selects two random vertices, and accepts if and only if there is an edge between them. ■

Proposition 4.2 `BiClique` *has a three-query proximity-oblivious tester with detection probability $\rho(\epsilon) = \epsilon$, where* `BiClique` *is the set of all graphs consisting of a single bi-clique* (i.e., a complete bipartite graph).

The following proof may serve as a very simple demonstration of the "enforce and test" technique (see [28, Sec. 4]), which underlies the design and analysis of many testers in the dense graph model (e.g., the ones of [16]).

Proof: Consider a tester that sets $s \in [N]$ as an arbitrary vertex, selects $v, u \in [N]$ uniformly, and accepts if and only if the subgraph induced by $\{s, u, v\}$ has an even number of edges.[4]

Clearly, if G is a bi-clique then this test always accepts, because either all vertices reside on the same side (and so (s,u), (s,v), and (u,v) are all non-edges) or a single vertex is in solitude (and is thus adjacent to the other two vertices, which in turn are non-adjacent).

To analyze what happens when $G = ([N], E)$ is ϵ-far from being a bi-clique, observe that s induces a partition of the graph to its neighbors and non-neighbors (i.e., the 2-partition $(\Gamma(s), [N] \setminus \Gamma(s))$). Note that if G were a bi-clique then every vertex $w \in \Gamma(s)$ (resp., $w \in [N] \setminus \Gamma(s)$) would have satisfied $\Gamma(w) = [N] \setminus \Gamma(s)$ (resp., $\Gamma(w) = \Gamma(s)$). However, since G is ϵ-far from being a bi-clique, the sum of the number of edges in $(\Gamma(s) \times \Gamma(s)) \cup (([N] \setminus \Gamma(s)) \times ([N] \setminus \Gamma(s)))$ and the number of non-edges in $\Gamma(s) \times ([N] \setminus \Gamma(s))$ must exceed $\epsilon \cdot N^2$, and we call the corresponding vertex pairs bad. Thus, the probability that a pair (u,v) is bad is greater than ϵ, whereas each bad pair causes our tester to reject (because in the subcase that $(u,v) \in E \cap (\Gamma(s) \times \Gamma(s))$ the induced subgraph has three edges and in the other two subcases (i.e., $(u,v) \in E \cap (([N] \setminus \Gamma(s)) \times ([N] \setminus \Gamma(s)))$ and $(u,v) \in (\Gamma(s) \times ([N] \setminus \Gamma(s))) \setminus E$) the induced subgraph has a single edge). ■

Proximity-oblivious testers with $\rho(\epsilon) = o(\epsilon)$. So far, we considered proximity-oblivious testers with a linear detection probability (i.e., $\rho(\epsilon) = \Omega(\epsilon)$). We now turn to cases where ρ is polynomial but not linear. Such a natural case is provided by the graph property that corresponds to graphs that consist of a fixed number of isolated cliques. Specifically, for any fixed integer $c \geq 1$, consider the set of graphs, denoted $\mathcal{CC}^{\leq c}$, that consist of at most c isolated cliques. Note that Proposition 4.1 refers to $\mathcal{CC}^{\leq 1}$, whereas Proposition 4.2 refers to graphs that are closely related to $\mathcal{CC}^{\leq 2}$ (i.e., a graph is in $\mathcal{CC}^{\leq 2}$ if and only if its complement graph is a bi-clique). The following result refers to the case of $c \geq 3$.

Proposition 4.3 *For every constant $c \geq 3$, the property $\mathcal{CC}^{\leq c}$ has a $\binom{c+1}{2}$-query proximity-oblivious tester with detection probability $\rho(\epsilon) > \epsilon^{c+1+o(1)}$. On the other hand, $\mathcal{CC}^{\leq c}$ has no constant-query proximity-oblivious tester with detection probability $\rho(\epsilon) = \omega(\epsilon^{c/2})$.*

We note that Section 6.2 of the companion paper [18] provides a standard (non-adaptive) tester for $\mathcal{CC}^{\leq c}$ having query complexity $\widetilde{O}(1/\epsilon)$ and constant error probability. This standard tester is superior to the one obtained by repeating any proximity-oblivious tester for an adequate number of times (since for any $c \geq 3$ the number of repetitions must be $\Omega(\epsilon^{-c/2})$). We mention that the lower-bound on $\rho(\epsilon)$ provided by Proposition 4.3 can be improved (see Proposition 4.11).

Proof: The lower-bound on ρ follows from the analysis of the $\mathcal{CC}^{\leq c}$-tester that is provided in [18, Sec. 6.2]. Specifically, we refer to the fact that the analysis in [18] establishes that (with high probability) a sample of $\widetilde{O}(1/\epsilon)$ vertices (from any graph that is ϵ-far from $\mathcal{CC}^{\leq c}$) induces a subgraph not in $\mathcal{CC}^{\leq c}$. (The said analysis actually establishes something much stronger, but the foregoing suffices here.)[5] Note that any graph G' that is not in $\mathcal{CC}^{\leq c}$ contains an induced subgraph of at most $c+1$ vertices that is not in $\mathcal{CC}^{\leq c}$, because G' either has at least $c+1$ connected components (which yields an independent set of size $c+1$) or has a connected component that is not a clique

[4] We mention that in Section 6.1 of the companion paper [18] we considered a standard tester that selects $O(1/\epsilon)$ random pairs of vertices (in addition to an arbitrary s as above).

[5] Details are omitted in light of the fact that Proposition 4.11 establishes a stronger lower-bound.

(which yields three vertices that miss some edge among them). It follows that the said $\widetilde{O}(1/\epsilon)$-vertex sample contains such $c+1$ vertices. Thus, the proximity-oblivious tester that selects $c+1$ uniformly distributed vertices and accepts if and only if the induced graph is in $\mathcal{CC}^{\leq c}$ has detection probability at least $\Omega(1)/\binom{\widetilde{O}(1/\epsilon))}{c+1} > \epsilon^{c+1+o(1)}$.

For the impossibility claim (or rather the upper-bound on ρ), consider a random graph consisting of c small cliques, each of size $\sqrt{2\epsilon} \cdot N$, and a large clique of size $(1 - c\sqrt{2\epsilon}) \cdot N$. This graph is ϵ-far from $\mathcal{CC}^{\leq c}$, but the probability that any k vertices induce a subgraph that is not in $\mathcal{CC}^{\leq c}$ is upper-bounded by $\binom{k}{c} \cdot \sqrt{2\epsilon}^c$, because only subsets that contain representatives from each of the small cliques yield a subgraph not in $\mathcal{CC}^{\leq c}$. Recalling that we refer to constant-query proximity-oblivious testers (which must accept if the induced subgraph is in $\mathcal{CC}^{\leq c}$), the upper-bound follows (since $\binom{k}{c} \cdot \sqrt{2\epsilon}^c = O(\epsilon^{c/2})$ for constant k). ■

Proximity-oblivious testers with detection probability that is even smaller are provided by [1].

Proposition 4.4 (implicit in [1]): `Triangle-Freeness` *has a three-query proximity-oblivious tester with detection probability $\rho(\epsilon)$ that is the reciprocal of a tower of* $\text{poly}(1/\epsilon)$*-many exponents. On the other hand,* `Triangle-Freeness` *has no constant-query proximity-oblivious tester with detection probability $\rho(\epsilon) = \text{poly}(\epsilon)$.*

We note that [1] actually established that every constant-query proximity-oblivious tester for `Triangle-Freeness` must have detection probability $\rho(\epsilon) < \epsilon^{\Omega(\log(1/\epsilon))}$.

Easily testable properties having no proximity-oblivious testers. While bipartiteness can be tested with query-complexity that is polynomial in the reciprocal of the proximity parameter [16], this property has no constant-query proximity-oblivious tester. That is:

Proposition 4.5 *Bipartiteness has no constant-query proximity-oblivious tester.*

Proof: For every $\epsilon > 0$, consider a graph G that consists of $t \stackrel{\text{def}}{=} \sqrt{1/2\epsilon}$ sets, denoted $V_0, V_1, ..., V_{t-1}$, each of size $\sqrt{2\epsilon} \cdot N$ such that there is an edge between a pair of vertices if and only if these vertices reside in "adjacent" sets; that is, $\{u, v\}$ is an edge if and only if for some $i \in \{0, ..., t-1\}$ it holds that $u \in V_i$ and $v \in V_{(i+1) \bmod t}$. Clearly, for an odd t, the graph G is ϵ-far from being bipartite, but a proximity-oblivious tester of query complexity less than t cannot reject G (because any non-bipartite subgraph of G must contain at least t vertices). ■

4.2 Connection to induced subgraph freeness

The reader may have noticed that the proximity-oblivious testers presented so far worked by searching for a small "forbidden subgraph" in the input graph (see, e.g., the proof of Propositions 4.1, 4.2 and 4.3). In contrast, the non-existence of constant-query proximity-oblivious testers was demonstrated by proving the non-existence of constant-size "forbidden subgraphs" in all no-instances (see, indeed, the proof of Proposition 4.5). We show that this is no coincidence, since there is a close relationship between the two notions.

Definition 4.6 (induced subgraph freeness): *Let $\mathcal{F}$ be a set of graphs. A graph G is called $\mathcal{F}$-free if it contains no induced subgraph that is isomorphic to some graph in $\mathcal{F}$.*

Note that Definition 4.6 refers to *induced subgraphs*, whereas in many works the term $\mathcal{F}$-freeness means having no subgraph (not necessarily an induced one) that is in $\mathcal{F}$.

Theorem 4.7 (characterization for the dense graphs model): *Let $\Pi = \bigcup_{N\in\mathbb{N}} \Pi_N$ be a graph property such that each Π_N consists of all N-vertex graphs that satisfy Π. Then, Π has a constant-query proximity-oblivious tester if and only if there exists a constant c and an infinite sequence $\overline{\mathcal{F}} = (\mathcal{F}_N)_{N\in\mathbb{N}}$ of sets of graphs such that*

1. *each $\mathcal{F}_N$ contains graphs of size at most c; and*
2. *Π_N equals the set of N-vertex $\mathcal{F}_N$-free graphs.*

Furthermore, if membership in Π is decidable, then a computable proximity-oblivious tester yields a computable sequence of sets, and vice versa.

Note that the specific detection probability function ρ is irrelevant for the "only if" direction, which only relies on the fact that $\rho(\epsilon) > 0$ for every $\epsilon > 0$.[6] On the other hand, the opposite direction actually provides a lower-bound on the detection probability, albeit a very weak one (i.e., $\rho(\epsilon)$ is the reciprocal of a tower of poly$(1/\epsilon)$-many towers of exponents). Combining both directions, we conclude that any graph property that has a constant-query proximity-oblivious tester has such a tester with detection probability function that is lower-bounded by a specific function[7] of the proximity parameter (albeit the reciprocal of a tower of towers of exponents).

Proof: Suppose that Π has a constant-query proximity tester. By [20, Thm. 4.5] (see also [21]), every one-sided error tester of query complexity q for Π_N can be converted into a one-sided error *canonical tester* of query complexity $2q^2$, where for some $\mathcal{G}_N$ (which depends only on Π_N and q), the canonical tester uniformly selects a random set of $2q$ vertices and accepts the input graph iff the induced subgraph is in $\mathcal{G}_N$. We stress that the proof provided in [20, Sec. 4] maintains the error probability of the tester, and thus applies also to generalized (one-sided error) testers of arbitrary error probability. Thus, if Π has a q-query proximity-oblivious tester then for every N there exists a set of $2q$-vertex graphs $\mathcal{G}_N$ such that a graph is in Π_N iff each of its $2q$-vertex induced subgraphs is in $\mathcal{G}_N$. Defining $\mathcal{F}_N$ as the set of all $2q$-vertex graphs that are not in $\mathcal{G}_N$, we conclude that Π_N equals the set of N-vertex graphs that are $\mathcal{F}_N$-free.

Suppose, on the other hand, that for some constant c and a sequence of sets $(\mathcal{F}_N)_{N\in\mathbb{N}}$ of graphs it holds that each $\mathcal{F}_N$ contains graphs of size at most c and Π_N equals the set of N-vertex $\mathcal{F}_N$-free graphs. Our goal is to derive a constant-query proximity tester for Π. The case of identical sets (i.e., $\mathcal{F}_N = \mathcal{F}_{N+1}$ for every N) follows almost immediately from [3]. Specifically, [3, Thm. 6.1] implies that for every set of c-vertex graphs $\mathcal{F}$ and for every $\epsilon > 0$, there exist numbers $s(\epsilon)$ and $\delta(\epsilon)$ for which the following holds: For every graph G that is ϵ-far from being $\mathcal{F}$-free and contains at least $s(\epsilon)$ vertices, with probability at least $\delta(\epsilon)$ over the choice of a sample of size $s(\epsilon)$ the sample contains an induced copy of some graph in $\mathcal{F}$. It follows that, with probability at least $\binom{s(\epsilon)}{c}^{-1} \cdot \delta(\epsilon)$, a random set of c vertices (of such a graph G) induces a subgraph that is in $\mathcal{F}$. The argument extends the general case (of an arbitrary sequence of sets $(\mathcal{F}_N)_{N\in\mathbb{N}}$), by partitioning all integers according to the corresponding sets. This yields testers for each of the finitely many possible sets, and so the final tester will incorporate all these testers, and activate the one that suits the size of the input graph. Lastly, we note that the functions s and δ provided by [3, Thm. 6.1] satisfy $s(\epsilon)/\delta(\epsilon) = \mathrm{TT}(1/\epsilon)$, where $\mathrm{TT}(n)$ is a tower of poly(n)-many towers of exponents (with the polynomial depending only on c). ■

[6]Indeed, this holds even if the detection probability function is allowed to depend on N (as long as $\rho(N, \epsilon) > 0$ for every $\epsilon > 0$ and $N \in \mathbb{N}$).

[7]This lower-bounding function is determined based only on the aforementioned constant (number of queries).

A special case and a quantitative version. A natural special case of properties having constant-query proximity-oblivious testers is properties that correspond to sets of $\mathcal{F}$-free graphs, for arbitrary finite sets $\mathcal{F}$. Indeed, this corresponds to the special case of Theorem 4.7 in which all the sets in the sequence $\overline{\mathcal{F}}$ are identical (i.e., $\mathcal{F}_N = F_{N+1}$ for every N). In this case, the detection probability of any constant-query proximity-oblivious tester is determined by the quantity $\rho_{\mathcal{F}}$ defined next.

- For a c-vertex graph F, we denote by $\mu_F(G)$ the fraction of c-vertex subsets that induce the subgraph F in the graph G.
- For a finite set of graphs $\mathcal{F}$, we denote by $\rho_{\mathcal{F}}(\epsilon)$ the infimum of the value of $\max_{F\in\mathcal{F}}\{\mu_F(G)\}$ taken over all graphs G that are ϵ-far from being $\mathcal{F}$-free.[8]

Recall that by Theorem 4.7 (or rather by [3, Thm. 6.1]), for every $\mathcal{F}$, the function $\rho_{\mathcal{F}}$ is well-defined. Furthermore, $\rho_{\mathcal{F}}$ is lower-bounded by the reciprocal of a tower of towers of exponents. The following result asserts that the detection probability of the best possible constant-query proximity-oblivious for $\mathcal{F}$-freeness is determined by $\rho_{\mathcal{F}}$.

Theorem 4.8 *Let c be an integer and $\mathcal{F}$ be a finite set containing graphs that each has at most c vertices. Then, $\mathcal{F}$-freeness has a $\binom{c}{2}$-query proximity-oblivious tester of detection probability $\rho_{\mathcal{F}}$, whereas any constant-query proximity-oblivious tester for $\mathcal{F}$-freeness has detection probability $O(\rho_{\mathcal{F}})$.*

Proof: First note that the straightforward proximity-oblivious tester for $\mathcal{F}$-freeness (which selects a random set of c vertices and accepts if and only if it is $\mathcal{F}$-free) has detection probability $\rho_{\mathcal{F}}$.

In order to justify the upper-bound (on the detection probability of any constant-query proximity-oblivious testers) we recall that, by [20, Thm. 4.5], it suffices to consider constant-query proximity-oblivious testers that select a random set of $c' = O(1)$ vertices and accept the input N-vertex graph iff the induced subgraph is in some adequate set $\mathcal{G}_N$. We stress that this $\mathcal{G}_N$ need not complement the set $\mathcal{F}$, and in particular c' may be different from c. Still, without loss of generality, we may assume that $c' \geq c$.

Let us first assume that $\mathcal{G}_N$ does not depend on N (i.e., $\mathcal{G}_N = \mathcal{G}_{N+1}$ for every $N \geq c'$). In this case, $\mathcal{G}_N = \mathcal{G}_{c'}$ must equal the set of c'-vertex graphs that are $\mathcal{F}$-free. The reason being that a c'-vertex graph G has a unique induced subgraph with c' vertices, being the graph itself. Now, on the one hand (by the acceptance criterion of the tester), the input (c'-vertex) graph G is accepted by the tester if and only if $G \in \mathcal{G}_{c'}$, whereas on the other hand the tester is required to accept G if and only if it is $\mathcal{F}$-free.

In the general case, the sequence $(\mathcal{G}_N)_{N\in\mathbb{N}}$ may contain a finite number of possible sets (of c'-vertex graphs). For each $N \geq c'$, consider the smallest integer n such that $\mathcal{G}_N = \mathcal{G}_n$, and denoted it by $n(\mathcal{G}_N)$; that is, $n(\mathcal{G}) = \min\{n \geq c' : \mathcal{G}_n = \mathcal{G}\}$. Note that $n^* = \max\{n(\mathcal{G}_N) : N \geq c'\}$ is a constant, because there are finitely many different sets $\mathcal{G}_N$. (Indeed, in the special case (where $\mathcal{G}_N = \mathcal{G}_{N+1}$), it holds that $n^* = c'$, since $n(\mathcal{G}_N) = c'$ for every $N \geq c'$.) Now, consider a tester that, on input an N-vertex graph, accepts if and only if the subgraph induced by $n(\mathcal{G}_N)$ random vertices is in $\mathcal{G}'_N$, where $\mathcal{G}'_N$ consists of the set of all $n(\mathcal{G}_N)$-vertex graphs G' such that every c' vertices in G' induce a subgraph that is in $\mathcal{G}_N$. The detection probability of this tester (on any graph) is lower-bounded by the detection probability of the original tester, whereas the new tester never rejects graphs that were never rejected by the original tester. Thus, we can apply the analysis

[8] Indeed, in the case that $\mathcal{F}$ consists of c-vertex graphs, an alternative definition can be based on defining $\mu_{\mathcal{F}}(G)$ as the fraction of c-vertex subsets that induce in G a subgraph that belong to $\mathcal{F}$. Needless to say, these two definition are related by a factor of at most $|\mathcal{F}|$.

of the special case (of equal $\mathcal{G}_N$'s) here, and conclude that $\mathcal{G}'_N = \mathcal{G}'_{n(\mathcal{G}_N)}$ must equal the set of $n(\mathcal{G}_N)$-vertex graphs that are $\mathcal{F}$-free.

It follows that the aforementioned tester rejects an input N-vertex graph G if and only if it has selected a random set of $n(\mathcal{G}_N) = O(1)$ vertices such that the induced subgraph is not $\mathcal{F}$-free. The probability of the latter event is upper-bounded by $\sum_{F\in\mathcal{F}} \binom{n(\mathcal{G}_N)}{|V(F)|} \cdot \mu_F(G)$, where $V(F)$ denotes the vertex set of the graph F. Recalling that $\mathcal{F}$ is finite and $n(\mathcal{G}_N) \leq n^* = O(1)$, it follows that this tester has detection probability $O(\rho_{\mathcal{F}})$. ■

In light of Theorem 4.8, the study of the detection probability of constant-query proximity-testers for natural properties that have such testers (i.e., $\mathcal{F}$-freeness), reduces to the study of the corresponding quantities $\rho_{\mathcal{F}}$ for various $\mathcal{F}$. A few examples follow.

1. The property `Clique` (see Proposition 4.1) corresponds to the set of $\{I_2\}$-free graphs, where I_2 denotes an independent set of two vertices. Needless to say, $\rho_{\{I_2\}}(\epsilon) = \epsilon$.

   Similarly $\rho_{\{P_2\}}(\epsilon) = \epsilon$, where P_2 denotes a single edge (which may be viewed as a path of two vertices).

2. Denoting by $\mathcal{CC}$ (standing for Clique Collection) the set of graphs consisting of a collection of (any number of) isolated cliques, we note that $\mathcal{CC}$ equals the set of $\{P_3\}$-free graphs, where P_3 denotes a three-vertex graph with exactly two edges (i.e., a path of three vertices). We show (in Proposition 4.10) that $\rho_{\{P_3\}}(\epsilon) = \Theta(\epsilon^2)$.

3. Recall that $\mathcal{CC}^{\leq c}$ is the set of graphs consisting of a collection of at most c isolated cliques (see Proposition 4.3). Note that $\mathcal{CC}^{\leq c}$ equals the set of $\{P_3, I_{c+1}\}$-free graphs, where I_{c+1} denotes an independent set of $c+1$ vertices. Combining Theorem 4.8 and Proposition 4.3,[9] it follows that $\rho_{\{P_3,I_{c+1}\}}(\epsilon) = O(\epsilon^{c/2})$ for every $c \geq 3$. We show (in Proposition 4.11) that $\rho_{\{P_3,I_{c+1}\}}(\epsilon) = \Omega(\epsilon^{c/2})$.

   Note that Proposition 4.2 implies that $\rho_{\{P_3,I_3\}}(\epsilon) = \Omega(\epsilon)$, because `BiClique` consists of graphs whose complement graph is in $\mathcal{CC}^{\leq 2}$. Clearly, $\rho_{\{P_3,I_3\}}(\epsilon) = O(\epsilon)$.

4. Recall that Proposition 4.4 refers to `Triangle-Freeness`, which corresponds to $\{C_3\}$-freeness where C_3 is the three-vertex cycle. Recall that [1] established that $\rho_{\{C_3\}}$ is a super-polynomial function, whereas $\rho_{\{C_3\}}$ was known to be lower-bounded by the reciprocal of a tower of exponents.

We mention that the work of [6] provides a characterization of the class of graphs F for which ρ_F is lower-bounded by a polynomial (i.e., $\rho_F(\epsilon) \geq \text{poly}(\epsilon)$). In particular, their results imply that ρ_F is lower-bounded by a polynomial only for at most seven graphs (i.e., the graphs P_2, P_3, P_4, C_4 and their complements). The foregoing discussion begs to try to extend their study to finite sets of graphs; that is, *for every finite set of graphs $\mathcal{F}$, determine the behavior of $\rho_{\mathcal{F}}$*. In particular:

Open Problem 4.9 *Determine the class of sets of graphs $\mathcal{F}$ for which $\rho_{\mathcal{F}}$ is lower-bounded by a polynomial.*

[9] Actually, the proof of Proposition 4.3 directly implies upper (and lower) bounds on $\rho_{\{P_3,I_{c+1}\}}$.

4.3 The detection probability of Clique Collection (i.e., $\rho_{\{P_3\}}(\epsilon) = \Theta(\epsilon^2)$)

Recall that (by Theorem 4.7) $\mathcal{CC}$ has a constant-query proximity-oblivious tester, since $\mathcal{CC}$ corresponds to $\{P_3\}$-freeness. Furthermore, by Theorem 4.8, the detection probability of the best possible constant-query proximity-oblivious for $\mathcal{CC}$ equals $\Theta(\rho_{\{P_3\}})$.

Proposition 4.10 (the best detection probability for $\mathcal{CC}$): $\rho_{\{P_3\}}(\epsilon) = \Theta(\epsilon^2)$.

Proposition 4.10 follows from Section 4 in the companion paper [18]; specifically, the upper bound (on $\rho_{\{P_3\}}$) uses the graphs of [18, Sec. 4.1] (which are ϵ-far from $\mathcal{CC}$), whereas the lower bound follows from the basic parts of Claims 4.3.1 and 4.3.2 in [18, Sec. 4.2]. For sake of self-containment, we provide a full proof below (where the aforementioned basic parts appear as Claims 4.10.1 and 4.10.2, respectively). We note that the following proof is significantly simpler than the analysis in [18, Sec. 4].

We mention that the constant-query proximity-oblivious tester resulting from Proposition 4.10 yields a standard (non-adaptive) tester of query complexity $O(\epsilon^{-2})$, which improves over the $\widetilde{O}(\epsilon^{-2})$ bound of [6, Thm. 2] (which, in turn, is based on inspecting the subgraph induced by a random set of $O(\epsilon^{-1}\log(1/\epsilon))$ vertices). However, in [18, Sec. 4.2] we present an alternative (non-adaptive) tester of query complexity $\widetilde{O}(\epsilon^{-4/3})$, and in [18, Sec. 3] we present an adaptive tester of query complexity $\widetilde{O}(\epsilon^{-1})$.

Proof: The proof adapts ideas from the study of non-adaptive testers for $\mathcal{CC}$, conducted in the companion paper [18]. For the upper-bound consider an N-vertex graph G consisting of $(6\epsilon)^{-1}$ connected components, each being a bi-clique with $3\epsilon N$ vertices on each side. The graph G is ϵ-far from $\mathcal{CC}$, but $\mu_{\{P_3\}}(G) \leq (6\epsilon)^2$, because a copy of P_3 must contain three vertices in the same connected component.

For the lower-bound we consider an arbitrary graph $G = ([N], E)$ that is ϵ-far from $\mathcal{CC}$. Let $G' = ([N], E')$ be a graph in $\mathcal{CC}$ that is closest to G, and let $(V_1, ..., V_t)$ be its partition into cliques. For the sake of simplicity, we shall refer to the V_i's as cliques, even though they are not (necessarily) cliques in G, and we shall refer to the partition $(V_1, ..., V_t)$ as the *best possible partition for G*. Two main observations regarding this partition follow.

Observation 1: For every $i \in [t]$ and every $S \subseteq V_i$, it holds that $|E \cap (S \times (V_i \setminus S))| \geq |S \times (V_i \setminus S)|/2$, since otherwise replacing the clique V_i by two cliques, S and $V_i \setminus S$, yields a better partition for G.

Observation 2: For every $i \neq j \in [t]$, it holds that $|E \cap (V_i \times V_j)| \leq |V_i \times V_j|/2$, since otherwise replacing the two cliques V_i and V_j by a single clique $V_i \cup V_j$ yields a better partition for G.

Now, since G is ϵ-far from $\mathcal{CC}$, either G misses at least $\frac{\epsilon}{2} \cdot \binom{N}{2}$ edges within these V_i's or G has at least $\frac{\epsilon}{2} \cdot \binom{N}{2}$ superfluous edges between distinct V_i's. We show that in either case, with probability at least $\Omega(\epsilon^2)$, three uniformly selected vertices induce the subgraph P_3. We call such a triplet a witness.

The pivot of the analysis is relating the fraction of bad vertex pairs (i.e., either missing "internal" edges or superfluous "external" edges) to the fraction of witnesses. Specifically, we shall show that the existence of $\frac{\epsilon}{2} \cdot \binom{N}{2}$ missing internal edges (resp., $\frac{\epsilon}{2} \cdot \binom{N}{2}$ superfluous "external" edges) implies the existence of $\Omega(\epsilon^2 N^3)$ witnesses. The following notation will be useful: for every $i \in [t]$ and $v \in [N]$ (not necessarily in V_i), we denote by $\Gamma_i(v)$ the set of neighbors of v in V_i, and $\overline{\Gamma}_i(v) \stackrel{\text{def}}{=} V_i \setminus (\Gamma_i(v) \cup \{v\})$.

We first consider the case in which at least $\frac{\epsilon}{2}\cdot\binom{N}{2}$ internal edges are missing (i.e., $\sum_{i\in[t]}\sum_{v\in V_i}|\overline{\Gamma}_i(v)| > \epsilon\cdot\binom{N}{2}$). Note that every triple (v,u,w) such that $u\in\overline{\Gamma}_i(v)$, $w\in\Gamma_i(v)$ and $\{u,w\}\in E$ is a witness. Using Observation 1, we have for each $v\in V_i$:

1. $|\Gamma_i(v)|\geq|\overline{\Gamma}_i(v)|$; and
2. the density of edges between $\overline{\Gamma}_i(v)$ and $\Gamma_i(v)$ is at least 1/2.

Thus, for $v\in V_i$, the number of witnesses that contain v is at least $|\Gamma_i(v)|\cdot|\overline{\Gamma}_i(v)|/2 \geq |\overline{\Gamma}_i(v)|^2/2$. It follows that the total number of witnesses is lower-bounded by

$$\frac{1}{2}\cdot\sum_{i\in[t]}\sum_{v\in V_i}|\overline{\Gamma}_i(v)|^2 \;\geq\; \frac{1}{2}\cdot N\cdot\left(\frac{\sum_{i\in[t]}\sum_{v\in V_i}|\overline{\Gamma}_i(v)|}{N}\right)^2 \tag{2}$$

which is lower-bounded by $\Omega(\epsilon^2N^3)$ as desired. For sake of reference, we highlight the following claim, which was established above.

Claim 4.10.1 *For every $v\in V_i$, the number of witnesses containing v is $\Omega(|\overline{\Gamma}_i(v)|^2)$.*

We now turn to the case in which there are at least $\frac{\epsilon}{2}\cdot\binom{N}{2}$ superfluous "external" edges; that is, in this case $\sum_{v\in[N]}|\Gamma'(v)|>\epsilon\cdot\binom{N}{2}$, where for every $v\in V_i$ we define $\Gamma'(v)\stackrel{\rm def}{=}\bigcup_{j\neq i}\Gamma_j(v)$. In this case, we shall show that the number of witnesses that contain each specific $v\in[N]$ is $\Omega(|\Gamma'(v)|^2)$, and the claim (regarding the total number of witnesses) will follow as in the previous case. Thus, it is left to establish the following.

Claim 4.10.2 *The number of witnesses containing v is $\Omega(|\Gamma'(v)|^2)$.*

Proof: In addition to the notations $\Gamma_i(v)=\Gamma(v)\cap V_i$, $\overline{\Gamma}_i(v)=V_i\setminus(\Gamma(v)\cup\{v\})$, and $\Gamma'(v)=\bigcup_{j:v\notin V_j}\Gamma_j(v)$, we shall use the notation $E(V',V'')\stackrel{\rm def}{=}\{(v',v'')\in(V'\times V''):\{v',v''\}\in E\}$. The proof will proceed via a case analysis, which refers to an arbitrary $i\in[t]$ and $v\in V_i$.

Case 1: Much of $\Gamma'(v)$ is contained in a single V_j; that is, there exists an index j such that $|\Gamma_j(v)|>|\Gamma'(v)|/10$. Fixing such an index j, we distinguish two subcases regarding the fraction of V_j that is not covered by $\Gamma'(v)$ (i.e., the relative density of $\overline{\Gamma}_j(v)$ in V_j).

Case 1.1: **$|\overline{\Gamma}_j(v)|\geq|V_j|/10$.** In this case the claim follows by considering most of the possible choices of $u\in\Gamma_j(v)$ and $w\in\overline{\Gamma}_j(v)$. Specifically, by Observation 1, $|E(\Gamma_j(v),\overline{\Gamma}_j(v))|$ is lower-bounded by $|\Gamma_j(v)|\cdot|\overline{\Gamma}_j(v)|/2$, and so at least half of the triples in $T_v\stackrel{\rm def}{=}\{(v,u,w):(u,w)\in\Gamma_j(v)\times\overline{\Gamma}_j(v)\}$ are witnesses (i.e., $(u,w)\in E$ and $(v,u)\in E$, but $(v,w)\notin E$), whereas $|T_v|=|\Gamma_j(v)|\cdot|\overline{\Gamma}_j(v)|=\Omega(|\Gamma'(v)|^2)$ (because $|\overline{\Gamma}_j(v)|\geq|V_j|/10\geq|\Gamma_j(v)|/10$ and $|\Gamma_j(v)|>|\Gamma'(v)|/10$).

Case 1.2: **$|\overline{\Gamma}_j(v)|\leq|V_j|/10$ (i.e., $|\Gamma_j(v)|\geq 0.9|V_j|$).** We first note that $|V_i|>|\Gamma'(v)|/20$, because otherwise we would obtain a better partition by moving the vertex v from V_i to V_j (since $|\Gamma_i(v)|\leq|V_i|$ whereas $|\Gamma_j(v)|-|\overline{\Gamma}_j(v)|\geq 0.8|V_j|$ and $|V_j|\geq|\Gamma_j(v)|\geq|\Gamma'(v)|/10$). We consider two subcases regarding the cardinality of the set $\Gamma_i(v)$:

1. If $|\Gamma_i(v)| \geq 0.9 \cdot |V_i|$, then the claim follows by considering a constant fraction of the possible choices of $u \in \Gamma_j(v)$ and $w \in \Gamma_i(v)$. Specifically, using Observation 2, it holds that

$$\begin{aligned} |E(\Gamma_j(v), \Gamma_i(v)| &\leq |E(V_j, V_i)| && (3)\\ &\leq \frac{1}{2} \cdot |V_j| \cdot |V_i| && (4)\\ &\leq \frac{1}{2} \cdot \frac{|\Gamma_j(v)|}{0.9} \cdot \frac{|\Gamma_i(v)|}{0.9} && (5)\\ &< 0.7 \cdot |\Gamma_j(v)| \cdot |\Gamma_i(v)|, && (6) \end{aligned}$$

where the second inequality uses $|\Gamma_j(v)| \geq 0.9|V_j|$ and $|\Gamma_i(v)| \geq 0.9|V_i|$. We obtain at least $(1-0.7) \cdot |\Gamma_j(v)| \cdot |\Gamma_i(v)|$ pairs $(u, w) \in (\Gamma_j(v) \times \Gamma_i(v)) \setminus E$ (and the corresponding triples (v, u, w) are witnesses). Using $|\Gamma_j(v)| \geq |\Gamma'(v)|/10$ and $|\Gamma_i(v)| \geq 0.9|V_i| = \Omega(\Gamma'(v)|)$, we lower-bound the said number by $\Omega(|\Gamma'(v)|^2)$.

2. If $|\Gamma_i(v)| \leq 0.9 \cdot |V_i|$, then we have many missing internal edges inside V_i with v as an endpoint (i.e., $|\overline{\Gamma}_i(v)| = \Omega(\Gamma'(v)|)$), and we invoke the corresponding analysis (as in the case of $\sum_{i\in[t]} \sum_{v\in V_i} |\overline{\Gamma}_i(v)| \geq \epsilon \cdot \binom{N}{2}$). Specifically, we obtain $\frac{1}{2} \cdot |\Gamma_i(v)| \cdot |\overline{\Gamma}_i(v)|$ witnesses (corresponding to edges $\{u, w\}$ such that $u \in \overline{\Gamma}_i(v)$ and $w \in \Gamma_i(v)$), and using the subcase hypothesis (and Observation 1) this number is lower-bounded by $\frac{1}{2} \cdot 0.5|V_i| \cdot 0.1|V_i|$, which is lower-bounded by $\Omega(|\Gamma'(v)|^2)$ (since $|V_i| > |\Gamma'(v)|/20$ holds in Case 1.2).

This completes the treatment of Case 1.2.

Case 2: No single V_j contains much of $\Gamma'(v)$; that is, for every $j \in [t]$ it holds that $|\Gamma_j(v)| \leq |\Gamma'(v)|/10$. As in Case 1, we consider two subcases regarding the relative part of V_j covered by $\Gamma'(v)$, but in the current case we consider a partition of the set $J \stackrel{\text{def}}{=} \{j : |\Gamma_j(v)| \geq 1\}$ and distinguish cases regarding the intersection of $\Gamma'(v)$ with the sets V_j in each part. Specifically, we let $J' \stackrel{\text{def}}{=} \{j : |\Gamma_j(v)| > 0.9|V_j|\}$, and consider the following two subcases.

Case 2.1: $\sum_{j\in J'} |\Gamma_j(v)| \geq 0.5 \cdot |\Gamma'(v)|$**.** In this case J' has cardinality at least five (since $|\Gamma_j(v)| \leq 0.1 \cdot |\Gamma'(v)|$ for every j). Let $C_v = \bigcup_{j\in J'} \Gamma_j(v)$, and note that the vertices in C_v belong to several cliques V_j. We shall show that the case hypothesis implies that there are many missing edges between pairs of vertices in C_v. Intuitively this holds because C_v essentially covers $\bigcup_{j\in J'} V_j$, whereas (by Observation 2) for any $j_1 \neq j_2$ there are many non-edges in $V_{j_1} \times V_{j_2}$. This ensures that we have many witnesses of the form (v, u, w), where $u, w \in C_v$ and $\{u, w\} \notin E$. Details follow.

For every $j_1 \neq j_2 \in J'$, by Observation 2 (and since $|\Gamma_j(v)| > 0.9|V_j|$ for every $j \in J'$), it holds that

$$|E(\Gamma_{j_1}(v), \Gamma_{j_2}(v))| \leq \frac{1}{2} \cdot |V_{j_1}| \cdot |V_{j_2}| < 0.7 \cdot |\Gamma_{j_1}(v)| \cdot |\Gamma_{j_2}(v)|\,. \qquad (7)$$

Therefore the number of non-edges between pairs in C_v is lower-bounded by

$$\begin{aligned} &\sum_{j_1\neq j_2\in J'} (1-0.7) \cdot |\Gamma_{j_1}(v)| \cdot |\Gamma_{j_2}(v)| && (8)\\ &= 0.3 \cdot \left(\sum_{j_1,j_2\in J'} |\Gamma_{j_1}(v)| \cdot |\Gamma_{j_2}(v)| - \sum_{j\in J'} |\Gamma_j(v)|^2 \right) && (9)\\ &\geq 0.3 \cdot \left((0.5 \cdot |\Gamma'(v)|)^2 - 0.1 \cdot |\Gamma'(v)|^2 \right) && (10) \end{aligned}$$

where the last inequality is due to the case hypotheses (i.e., $\sum_{j\in J'}|\Gamma_j(v)|\geq 0.5\cdot|\Gamma'(v)|$ and $|\Gamma_j(v)|\leq 0.1\cdot|\Gamma'(v)|$). Thus, $|(C_v\times C_v)\setminus E|>0.04\cdot|\Gamma'(v)|^2$, and the claim follows.

Case 2.2: $\sum_{j\in J\setminus J'}|\Gamma_j(v)|\geq 0.5\cdot|\Gamma'(v)|$. Let $J''\stackrel{\rm def}{=}J\setminus J'=\{j:1\leq|\Gamma_j(v)|\leq 0.9|V_j|\}$, and note that for $j\in J''$ (as considered in this case) it may be that $|\Gamma_j(v)|\ll|V_j|$ and consequently for $j_1\neq j_2\in J''$ it may hold that $E(\Gamma_{j_1}(v),\Gamma_{j_2}(v))\approx|\Gamma_{j_1}(v)|\cdot|\Gamma_{j_2}(v)|$. More generally, redefining $C_v\stackrel{\rm def}{=}\bigcup_{j\in J''}\Gamma_j(v)$, it may be that $|E(C_v,C_v)|\approx\binom{|C_v|}{2}$, and so the approach of Case 2.1 may not work in general (although it will work in the first subcase). Thus, letting $J'''\stackrel{\rm def}{=}\{j\in J'':|V_j|\leq|\Gamma'(v)|/10\}$, we consider two subcases:

1. If $\sum_{j\in J'''}|\Gamma_j(v)|\geq 0.4\cdot|\Gamma'(v)|$ then we redefine $C_v\stackrel{\rm def}{=}\bigcup_{j\in J'''}\Gamma_j(v)$ and show that $|E(C_v,C_v)|\leq 0.99\cdot\binom{|C_v|}{2}$. This is the case because otherwise we obtain a contradiction to the optimality of the partition (by replacing the sub-partition $(V_j)_{j\in J'''}$ with $(C_v,(V_j\setminus C_v)_{j\in J'''})$).Thus, we have reached a situation as in Case 2.1, and we proceed as in that case.
2. If $\sum_{j\in J''\setminus J'''}|\Gamma_j(v)|\geq 0.1\cdot|\Gamma'(v)|$ then we proceed similarly to Case 1.1. Specifically, for each $j\in J''\setminus J'''$, we note that the density of edges in $\Gamma_j(v)\times\overline{\Gamma}_j(v)$ is at least one half, whereas $|\overline{\Gamma}_j(v)|\geq 0.1|V_j|\geq 0.1\cdot 0.1\cdot|\Gamma'(v)|$ (by $j\in J''$ and $j\notin J'''$, respectively). Thus, the number of witnesses (v,u,w) such that $(u,w)\in\Gamma_j(v)\times\overline{\Gamma}_j(v)$ (and $\{u,w\}\in E$) is at least
$$\sum_{j\in J''\setminus J'''}\frac{|\Gamma_j(v)\times\overline{\Gamma}_j(v)|}{2}\geq\sum_{j\in J''\setminus J'''}\frac{|\Gamma_j(v)|\cdot|\Gamma'(v)|}{200}\qquad(11)$$
which is $\Omega(|\Gamma'(v)|^2)$ by the subcase hypothesis.

These completes the treatment of Case 2.2.

Thus, a lower bound of $\Omega(|\Gamma'(v)|^2)$ was proved in all cases, and the claim follows. ■

This completes the proof of the entire proposition. ■

4.4 An improved result for $\mathcal{CC}^{\leq c}$ (i.e., $\rho_{\{P_3,I_{c+1}\}}(\epsilon)=\Omega(\epsilon^{c/2})$)

Recall that, for every constant $c\geq 3$, Proposition 4.3 established that the property $\mathcal{CC}^{\leq c}$ has a constant-query proximity-oblivious tester with $\rho(\epsilon)>\epsilon^{c+1+o(1)}$ (whereas any constant-query proximity-oblivious tester for $\mathcal{CC}^{\leq c}$ must satisfy $\rho(\epsilon)=O(\epsilon^{c/2})$). In this section we improve the lower-bound on ρ, and in fact obtain a tight result. By Theorem 4.8, it suffices to prove that $\rho_{\{P_3,I_{c+1}\}}(\epsilon)=\Omega(\epsilon)^{c/2}$, since $\mathcal{CC}^{\leq c}$ corresponds to $\{P_3,I_{c+1}\}$-freeness.

Proposition 4.11 (the best detection probability for $\mathcal{CC}^{\leq c}$): *For every integer $c\geq 3$, it holds that $\rho_{\{P_3,I_{c+1}\}}(\epsilon)=\Omega(\epsilon)^{c/2}$.*

The proof builds on the first part of the analysis of the $\mathcal{CC}^{\leq c}$-tester that is provided in [18, Sec. 6.2]. Actually, we shall modify also this part, and thus we provide a self-contained description of the entire argument.

Proof: Suppose that $G=([N],E)$ is an N-vertex graph that is ϵ-far from $\mathcal{CC}^{\leq c}$. As a mental experiment, we consider a uniformly distributed set of $\Theta(\epsilon^{-1/2})$ vertices of G, denoted S. We shall show that, for a typical S (i.e., for most choices of S) and for a uniformly selected vertex v, with probability $\Omega(\epsilon)$, the subgraph induced by $S\cup\{v\}$ is not in $\mathcal{CC}^{\le c}$. In such a case, the said subgraph contains $c+1$ vertices that induce a subgraph not in $\mathcal{CC}^{\le c}$. That is, for a typical S, with probability at least $\min\{|S|^{-(c+1)},\Omega(\epsilon)\cdot|S|^{-c}\}=\Omega(\epsilon)^{(c+2)/2}$ either a sample of $c+1$ vertices in S or a sample of c vertices in S and a single vertex v in $[N]$ yields an induced subgraph that is not in $\mathcal{CC}^{\le c}$ (i.e., is not $\{P_3,I_{c+1}\}$-free). Thus, $\mu_{\{P_3,I_{c+1}\}}(G)=\Omega(\epsilon)^{(c+2)/2}$, and it follows that $\rho_{\{P_3,I_{c+1}\}}(\epsilon)=\Omega(\epsilon)^{(c+2)/2}$. The proposition will follow by a somewhat more refined analysis.

We think of S as being selected in $c+1$ phases, where in phase t, a new uniform sample S^t, of $\Theta(\epsilon^{-1/2})$ vertices, is selected (recall that c is a constant). Intuitively, the objective of the first c phases is to yield a partition of all the graph vertices into at most $c+1$ subsets in a way that facilitates finding evidence of the fact that the original graph is not in $\mathcal{CC}^{\le c}$. For example, one main part of the argument is showing that, with high (constant) probability, it is either the case that the set of vertices with no neighbors in S is of size $O(\epsilon^{1/2}\cdot N)$ or S contains an independent set of size $c+1$ (and we are done). Let us elaborate on the way this assertion is proved.

Intuitively, with high (constant) probability, if the number of vertices that do not have any neighbor among the vertices selected so far is relatively big, then we obtain such a vertex in the next phase. Indeed, if the set of vertices with no neighbors in S is of size $\Omega(\epsilon^{1/2}\cdot N)$, then after each of the first c phases it is the case that the number of vertices that do not have any neighbor among the vertices selected so far is relatively big. Thus, we should have been quite unlucky not to obtain such a vertex in each of the following phases. Assuming that we are not unlucky, S does contain an independent set of size $c+1$, and it follows that $\mu_{\{P_3,I_{c+1}\}}(G)=\Omega(|S|^{-(c+1)})=\Omega(\epsilon)^{-(c+1)/2}$. However, a closer look at the situation reveals that we can select such an independent set (in S) by selecting an arbitrary vertex in S^1, and then selecting an adequate vertex in each S^t for each $t=2,...,c+1$ (i.e., a vertex of S^t that has no neighbors in $\bigcup_{k=1}^{t-1}S^k$). It follows that $\mu_{\{P_3,I_{c+1}\}}(G)=\Omega(\prod_{t=2}^{c+1}|S^t|^{-1})=\Omega(\epsilon)^{c/2}$. Note that the argument applies also if it only holds that the set of vertices with no neighbors in $S^{\le c}\stackrel{\rm def}{=}\bigcup_{k=1}^{c}S^k$ is of size $\Omega(\epsilon^{1/2}\cdot N)$. Let us generalize this argument further.

Claim 4.11.1 *For $s>2c$, suppose that a graph $G'=([s],E')$ is not in $\mathcal{CC}^{\le c}$. Then, with probability greater than $s^{-c}/2$, a uniformly selected set of $c+1$ vertices induces in G' a subgraph that is not in $\mathcal{CC}^{\le c}$.*

Proof: If G' contains an induced copy of P_3, then three uniformly selected vertices hit it with probability at least $s^{-3}\ge s^{-c}$, since $c\ge 3$. Otherwise (i.e., if $G'\notin\mathcal{CC}^{\le c}$ contains no induced copy of P_3), it must be the case that G' is a collection of at least $c+1$ isolated cliques. We arbitrarily cluster these cliques into $c+1$ sets, and consider the probability that a sample of $c+1$ vertices hits a vertex in each of these $c+1$ sets. This probability is lower-bounded by $\prod_{i=1}^{c+1}x_i$ subject to $\sum_{i=1}^{c+1}x_i=1$ and $x_i\ge 1/s$ for every i. The minimum is obtained at $x_1=\cdots=x_c=1/s$, and the claim follows. ■

We now turn to defining the $(c+1)$-partition (of the graph vertices) that arises from the sample S. For each $1\le t\le c+1$, let $S^{\le t}=\bigcup_{k=1}^{t}S^k$. If for any $1\le t\le c$, the subgraph induced by $S^{\le t}$ is not a collection of at most c cliques, then we are done (by Claim 4.11.1). Otherwise, let $C_1^t,...,C_{c^t}^t$ denote the $c^t\le c$ cliques in the subgraph induced by $S^{\le t}$. For each $1\le t\le c$, we define the following partition of the set of all graph vertices (i.e., $[N]$):

$$V_j^t \stackrel{\rm def}{=} \{v\ :\ \Gamma(v)\cap S^{\le t}=C_j^t\} \quad \text{for } 1\le j\le c^t\,, \tag{12}$$

$$R_0^t \stackrel{\rm def}{=} \{v\ :\ \Gamma(v)\cap S^{\le t}=\emptyset\} \tag{13}$$

$$R_1^t \stackrel{\rm def}{=} V \setminus \Big(R_0^t \cup \Big(\bigcup_{1 \le j \le c^t} V_j\Big)\Big) . \tag{14}$$

That is, for $1 \le j \le c^t$, the subset V_j^t consists of the vertices that neighbor all vertices in C_j^t and no other vertex in $S^{\le t}$, the subset R_0^t consists of all vertices that have no neighbor in $S^{\le t}$, and R_1^t consists of all vertices that either neighbor only some of the vertices in one of the cliques C_j^t, but not all, or that have neighbors in more than one of the cliques. Observe that $V_j^{t+1} \subseteq V_j^t$ and $R_0^{t+1} \subseteq R_0^t$ while $R_1^{t+1} \supseteq R_1^t$.

Given the above notation, we make two observations. The first observation is that, for any $1 \le t \le c$, if S^{t+1} contains some vertex in R_1^t, then the subgraph induced by $S^{\le(t+1)}$ is not a collection of (at most c) cliques, and so we are done (again, by Claim 4.11.1). It follows that if $|R_1^t| > \frac{1}{4}\epsilon^{1/2}N$, for some $t \le c$, then we are done (because with high probability S^{t+1} will contain some vertex in R_1^t). The second observation is that if S^{t+1} contains some vertex in R_0^t, then $c^{t+1} \ge c^t + 1$. Note that as long as $|R_0^t| > \frac{1}{4}\epsilon^{1/2}N$, the probability that S^{t+1} does not contain any vertex in R_0^t is at a small constant. Therefore, either $|R_0^c| \le \frac{1}{4}\epsilon^{1/2}N$, or we are done (because with high probability S^{t+1} will contain a vertex from each R_0^t (for $t = 1, ..., c$), which together with S^1 induce a subgraph that is not in $\mathcal{CC}^{\le c}$).

In light of the foregoing paragraph, from this point on, we assume that the subgraph induced by $S^{\le(c+1)}$ is a collection of at most c cliques, that $|R_1^c| \le \frac{1}{4}\epsilon^{1/2}N$ and that $|R_0^c| \le \frac{1}{4}\epsilon^{1/2}N$. To simplify the notation, we use the shorthand R_0 for R_0^c, and R_1 for R_1^c, the shorthand c' for c^c, and the shorthand V_j for V_j^c (resp., C_j for C_j^c). We also denote $R_0 \cup R_1$ by R.

Recall that $G = ([N], E)$ is ϵ-far from $\mathcal{CC}^{\le c}$. This means that for every partition of the graph vertices into at most c subsets, the total number of vertex pairs that "violate the partition" (i.e., either both vertices belong to the same subset but do not have an edge between them or they belong to different subsets but do have an edge between them) is greater than ϵN^2. In particular, this holds for the partition that we shall define next. We consider a partition, denoted $(\widetilde{V}_j)_{j \in \{0,1,...,c'\}}$, where for every $j \in [c']$ it holds that $V_j \subseteq \widetilde{V}_j$, while the vertices in R are partitioned as follows. Each vertex $v \in R_1$ is placed in an arbitrary $\widetilde{V}_j$ such that v has some neighbor in C_j. If $c' < c$ then R_0 is defined as $\widetilde{V}_0$, and otherwise R_0 is placed in $\widetilde{V}_1$ (i.e., in an arbitrary $\widetilde{V}_j$).

Note that the total number of vertex pairs in $R \times R$ is at most $\frac{1}{4}\epsilon N^2$, since $|R| \le \frac{1}{2}\epsilon^{1/2}N$. Recalling that G is ϵ-far from $\mathcal{CC}^{\le c}$, it follows that (at least) one of the following three events must hold:

1. There are at least $\frac{1}{4}\epsilon N^2$ missing edges between pairs of vertices that belong to the same subset $\widetilde{V}_j$ such that these pairs have no element in R_0 and at most one element in R_1. That is, the current case refers to pairs $(u, v) \in \bigcup_{j=1}^{c'}(\widetilde{V}_j \times \widetilde{V}_j)$ such that $\{u, v\} \notin E$ and $\{u, v\} \cap R_0 = \emptyset$ and $|\{u, v\} \cap R_1| \le 1$.

2. There are at least $\frac{1}{4}\epsilon N^2$ superfluous edges between pairs of vertices that belong to different subsets $\widetilde{V}_j$ and $\widetilde{V}_k$ and have at most one element in R. That is, the current case refers to pairs $(u, v) \in \bigcup_{j \ne k \in \{0,1,...,c'\}}(\widetilde{V}_j \times \widetilde{V}_k)$ such that $|\{u, v\} \cap R| \le 1$.

3. There are at least $\frac{1}{4}\epsilon N^2$ missing edges between pairs of vertices that belong to the same subset $\widetilde{V}_j$ but have exactly one endpoint in R_0 and no endpoint in R_1; that is, pairs in $(R_0 \cap \widetilde{V}_1) \times V_1$. (Recall that R_0 was placed either in $\widetilde{V}_0$ or in $\widetilde{V}_1$, whereas $V_0 = \emptyset$; hence, $\bigcup_{k=0}^{c'}((R_0 \cap \widetilde{V}_k) \times V_k)$ equals $(R_0 \cap \widetilde{V}_1) \times V_1$.)

We shall show that in each of these three cases, with probability at least $\Omega(\epsilon^{c/2})$, a uniformly selected set of $c+1$ vertices induces a subgraph that is not in $\mathcal{CC}^{\leq c}$.

Case 1. Recall that this case refers to missing edges within some $\widetilde{V}_j$, where $j \in [c']$, such that at least one endpoint of such an edge is not in R (and none is in R_0). In this case, with probability at least $\epsilon/4$, a uniformly distributed pair $(u,v) \in [N] \times [N]$ hits such a missing edge (i.e., in particular, $(u,v) \notin E$ and $u,v \in \widetilde{V}_j$ for some $j \in [k]$). Assume, without loss of generality, that $u \in V_j$ (i.e., $u \notin R$), and let w be an arbitrary neighbor of $v \in \widetilde{V}_j$ in C_j (which is guaranteed to exist since $v \in \widetilde{V}_j \setminus R_0$, whereas $v \in R_1$ is placed in $\widetilde{V}_j$ only if it has neighbors in C_j). Recall that w is also a neighbor of u (since $u \in V_j$ neighbors all vertices in C_j). Hence, selecting uniformly a vertex in S, we hit this w with probability $1/|S|$. It follows that if we select uniformly and independently three vertices in $[N]$, then, with probability $\frac{\epsilon}{4} \cdot \frac{\Omega(1)}{|S|} = \Omega(\epsilon^{3/2})$, we obtain a triple (u,v,w) such that $(u,v) \notin E$ whereas $(u,w),(v,w) \in E$.

Case 2. Recall that this case refers to superfluous edges between some $\widetilde{V}_j$ and $\widetilde{V}_k$, where $j \neq k \in \{0,1,...,c'\}$, such that at least one endpoint of such an edge is not in R. In this case, with probability at least $\epsilon/4$, a uniformly distributed pair $(u,v) \in [N] \times [N]$ hits such a superfluous edge (i.e., in particular, $(u,v) \in E$ and $(u,v) \in \bigcup_{j \neq k}(\widetilde{V}_j \times \widetilde{V}_k)$). Assume, without loss of generality, that $u \in V_j$ and $v \in \widetilde{V}_k$, where v may be in R (and even in R_0). If $v \in \widetilde{V}_k \setminus R_0$ then we let w be an arbitrary neighbor of v in C_k, and note that w is not a neighbor of u (since $u \in V_j$ neighbors no vertex in C_k). Otherwise (i.e., $v \in R_0$), let $w \in C_j$ be an arbitrary non-neighbor of v, and note that w is a neighbor of u (since $u \in V_j$). Thus, either way, w is a neighbor of exactly one of the two vertices u and v, and selecting uniformly a vertex in S, we hit w with probability $1/|S|$. It follows that if we select uniformly and independently three vertices in $[N]$, then, with probability $\Omega(\epsilon^{3/2})$, we obtain a triple (u,v,w) such that $(u,v) \in E$ whereas $(u,w) \notin E$ if and only if $(v,w) \in E$.

Case 3. Recall that this case refers to missing edges between vertices of R_0 and vertices of V_1 (i.e., the part V_j to which R_0 was added). It follows that $c' = c$ and that $|R_0| > \epsilon N/4$. Thus, we can obtain an independent set of size $c+1$ by selecting one vertex from R_0 and a vertex from each of the sets $C_1,...,C_c$. The probability that a uniformly selected sample of $c+1$ vertices yields such a set is at least

$$\Pr[S \text{ is good}] \cdot \frac{\epsilon}{4} \cdot \prod_{k=1}^{c} \frac{|C_k|}{|S|} \quad > \quad \frac{\epsilon}{5} \cdot \min_{\substack{x_1,...,x_c \geq |S|^{-1} \\ \sum_{k=1}^{c} x_k = 1}} \left\{ \prod_{k=1}^{c} x_k \right\} \tag{15}$$

$$> \quad \frac{\epsilon}{6} \cdot |S|^{-(c-1)} \tag{16}$$

which yields the lower-bound of $\Omega(\epsilon)^{(c+1)/2}$. To obtain a better bound, we modify the argument a little.

Suppose that for every j such that $|V_j| \geq \epsilon^{1/2}N$ it holds that $\frac{|C_j|}{|S|} \geq \frac{1}{2} \cdot \frac{|V_j|}{N}$. (This assumption will be justified at the end of the proof.) Then, we modify the construction (of the partition $(\widetilde{V}_j)$) such that in the case of $c = c'$ the set R_0 is placed in the smallest set V_j (rather than in an arbitrary set V_j). Turning back to Case 3, we recall that in this case there are $\epsilon N^2/4$ missing edges between R_0 and V_j, and it follows that $|R_0| \cdot |V_j| \geq \epsilon N^2/4$. Recalling that $|R_0| \leq \epsilon^{1/2}N/4$, we have $|V_j| \geq \epsilon^{1/2}N$ and it follows that $\frac{|R_0|}{N} \cdot \frac{|C_j|}{|S|} \geq \frac{\epsilon}{8}$ (because $\frac{|C_j|}{|S|} \geq \frac{|V_j|}{2N}$). Note that we can obtain an independent set of size $c+1$ by selecting a pair from $R_0 \times C_j$ and a vertex from each of the other $c-1$ sets C_k's, and recall that the largest C_k must have size at least $|S|/3c$ (because $|C_k|/|S| \geq |V_k|/2N \geq (1-\epsilon^{1/2})/2c$). The probability that a uniformly selected sample of $c+1$

vertices yields such a set is at least

$$\Pr[S \text{ is good}] \cdot \frac{\epsilon}{8} \cdot \prod_{k\in[c]\setminus\{j\}} \frac{|C_k|}{|S|} \;>\; \frac{\epsilon}{9} \cdot \min_{\substack{x_1,\dots,x_{c-2} \geq |S|^{-1} \\ x_{c-1} \geq 1/3c}} \left\{ \prod_{k=1}^{c-1} x_k \right\} \tag{17}$$

$$> \frac{\epsilon}{27c} \cdot |S|^{-(c-2)} \tag{18}$$

which yields the lower-bound of $\Omega(\epsilon)^{c/2}$.

It remains to deal with the assumption that $|C_j|/|S| \geq |V_j|/2N$ for every j such that $|V_j| \geq \epsilon^{1/2}N$. To this end we add one more phase in the choice of S (where we think of this phase as taking place before phase $c+1$ that was used in the foregoing discussion to bound $|R|$). Let S' denote the vertices selected in the first c phases and let S'' be the vertices selected in the additional phase, where $|S''| = 4|S'|$. Let $C'_1, \dots, C'_{c'}$ be the cliques in the subgraph induced by S', and for each $1 \leq j \leq c'$ let V'_j be the vertices that neighbor all vertices in C'_j and no other vertices in S'. In the sample S'', let $C''_j = S'' \cap V'_j$. By a multiplicative Chernoff bound, with high probability over the choice of S'', it holds that $|C''_j|/|S''| \geq (3/4)|V'_j|/N$ for every j such that $|V'_j| \geq \epsilon^{1/2}N$. Assuming that this is in fact the case, we define $C_j = C'_j \cup C''_j$ and $V_j = \{v : \Gamma(v) \cap (S' \cup S'') = C_j\}$.

If there is any new clique in S'', then it corresponds to a small set of vertices (since the set of vertices that do not belong to any V'_j is small).[10] Using the fact that S is the union of S', S'' and the sample selected in phase $c+1$, we have $|S| < (3/2)|S''|$ (since $|S''| = 4|S'|$ and $|S'| = c \cdot (|S| - |S'| - |S''|)$) and $|C_j|/|S| \geq (3/4)|C''_j|/|S''| \geq (3/4) \cdot (3/4)|V'_j|/N$. Using $V_j \subseteq V'_j$, we get that $|C_j|/|S| > |V_j|/2N$ for every $|V_j| \geq \epsilon^{1/2}N$. ■

5 Testing Graph Properties in the Bounded-Degree Model

The bounded-degree model refers to a fixed degree bound, denoted $d \geq 2$. An N-vertex graph $G = ([N], E)$ (of maximum degree d) is represented in this model by a function $g : [N] \times [d] \to \{0, 1, \dots, N\}$ such that $g(v, i) = u \in [N]$ if u is the i^{th} neighbor of v and $g(v, i) = 0$ if v has less than i neighbors.[11] Distance between graphs is measured in terms of their aforementioned representation (i.e., as the fraction of (the number of) different array entries (over dN)), but occasionally we shall use the more intuitive notion of the fraction of (the number of) edges over $dN/2$.

It turns out that, in the current model, constant-query proximity-oblivious testers exist for all graph properties that have such testers in the adjacency matrix model. However, in the current model, the scope of constant-query proximity-oblivious testers extends somewhat beyond the former. Specifically, while in the adjacency matrix model such testers exist for any "induced subgraph freeness" property, the current model also allows testing properties that correspond to a generalized notion of subgraph freeness, which includes properties that are not hereditary (e.g., the set of graphs in which each vertex has at least three neighbors).

5.1 Generalized subgraph freeness

The generalized notion of subgraph freeness defined next is pivotal to proximity-oblivious testing in the bounded-degree model. Intuitively, the definition refers to forbidden patterns that are captured

[10] Indeed, the sizes of the sets V'_j behave as the sizes of the sets V_j, which were analyzed in the beginning of this proof.

[11] We assume here that the neighbors of v appear in arbitrary order in the sequence $g(v,1), \dots, g(v, \deg(v))$, where $\deg(v) \stackrel{\text{def}}{=} |\{i : g(v,i) \neq 0\}|$.

by graphs that are augmented by a three-way marking of their vertices (where the markings are "full", "semi-full', and "partial"). What is forbidden, is embeddings of these graphs in larger graphs (i.e., in the graphs to which the property refers) that satisfy certain conditions (depending on the marking). Firstly, edges (regardless of the marking of their endpoints) in the marked graph should be mapped (in such an embedding) to edges of the large graph. Secondly, pairs non-adjacent vertices that are not both marked "partial" must be mapped to non-adjacent vertices (in the large graph). Finally, any vertex marked "full" must be mapped to a vertex that has no neighbors outside of the range of the mapping. Thus, while the "partial" and "semi-full" markings imposes conditions regarding the range of the mapping, the "full" marking imposes conditions that extend beyond that range. See illustration in Figure 1.

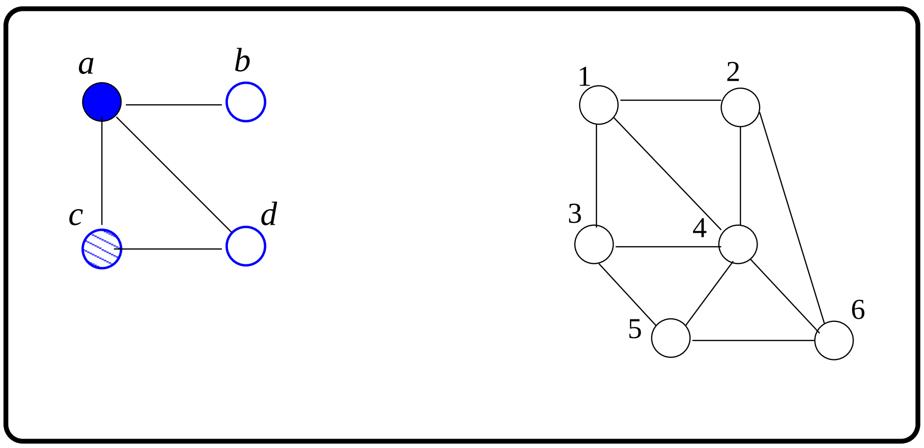

Figure 1: The 4-vertex marked graph is embedded in the 6-vertex graph such that the full vertex a is mapped to 1, the semi-full vertex c is mapped to 3, and the partial vertices b and d are mapped to 2 and 4, respectively.

Definition 5.1 (generalized subgraph freeness): *A* marked graph *is a graph with each vertex marked as either* full *or* semi-full *or* partial. *Such a marked graph $F = ([n], E_F)$ can be* embedded *in a graph $G = ([N], E_G)$ if there exists a 1-1 mapping $f : [n] \to [N]$ such that for every $v \in [n]$ the following three conditions hold:*

1. *If v is marked full, then f yields a bijection between the set of neighbors of v in F and the set of neighbors of $f(v)$ in G. That is, in this case $\Gamma_G(f(v)) = f(\Gamma_F(v))$, where for $H \in \{F, G\}$ we denote $\Gamma_H(x) \stackrel{\rm def}{=} \{w : \{x, w\} \in E_H\}$, and for $S \subseteq [N]$ we denote $f(S) \stackrel{\rm def}{=} \{f(u) : u \in S\}$.*

2. *If v is marked semi-full, then f yields a bijection between the set of neighbors of v in F and the set of neighbors of $f(v)$ in the subgraph of G induced by $f([n])$. That is, in this case $\Gamma_G(f(v)) \cap f([n]) = f(\Gamma_F(v))$.*

3. *If v is marked partial, then f yields an injection of the set of neighbors of v in F to the set of neighbors of $f(v)$ in G. That is, in this case $\Gamma_G(f(v)) \supseteq f(\Gamma_F(v))$.*

Such f is called an embedding of F in G. *The graph G is called* F-free *if F cannot be embedded in G* (i.e., there is no embedding of F in G). *For a set of marked graphs $\mathcal{F}$, a graph G is called* $\mathcal{F}$-free *if for every $F \in \mathcal{F}$ the graph G is F-free.*

Indeed, the standard notion of non-induced subgraph freeness is a special case of generalized subgraph freeness, obtained by considering the corresponding marked graph in which all vertices are marked partial. Similarly, the notion of induced subgraph freeness (as in Definition 4.6) is a special case of generalized subgraph freeness (as in Definition 5.1), obtained by considering the corresponding marked graph in which all vertices are marked semi-full. Introducing vertices that are marked full adds a new type of constraint; specifically, this constraint mandates the non-existence of neighbors that are outside the marked subgraph. For example, using vertices that are marked full it is possible to disallow certain degrees in the graph. Thus, *the generalized notion of subgraph freeness includes properties that are not hereditary* (e.g., regular graphs), *whereas induced and non-induced subgraph freeness are hereditary.*

We mention that the notion of generalized subgraph freeness remains as expressive when disallowing either semi-full or partial markings (see appendix). When allowing the consideration of a different set of marked graphs (of a constant size) for each size of graphs in the property, we obtain the following notion.

Definition 5.2 (local properties): *Let $\Pi = \bigcup_{N\in\mathbb{N}} \Pi_N$ be a graph property such that each Π_N consists of all N-vertex graphs that satisfy Π. The property Π is called* local *if there exists an integer s and an infinite sequence $\overline{\mathcal{F}} = (\mathcal{F}_N)_{N\in\mathbb{N}}$ such that for every N the following conditions hold:*

1. *$\mathcal{F}_N$ is a set of marked graphs, each of size at most s;*
2. *Π_N equals the set of N-vertex graphs that are $\mathcal{F}_N$-free.*

In such a case we say that Π is $\overline{\mathcal{F}}$-local.

We note that induced subgraph freeness (in the sense of Theorem 4.7) implies locality (in the sense of Definition 5.2); that is, for every sequence $\overline{\mathcal{F}}$ as in Theorem 4.7, the corresponding property Π is local.

5.2 The non-propagating condition

Although it may seem that all local properties have a constant-query proximity-oblivious tester (in the current model), the claim only holds for local properties that satisfy the following non-propagating condition.

Definition 5.3 (the non-propagating condition): *Let $\overline{\mathcal{F}} = (\mathcal{F}_N)_{N\in\mathbb{N}}$ be a sequence of sets of marked graphs as in Definition 5.2.*

- *For a graph $G = ([N], E)$, we say that a subset $B \subset [N]$* covers *$\mathcal{F}_N$ in G if for every marked graph $F \in \mathcal{F}_N$ and every embedding of F in G, at least one vertex of F is mapped to a vertex in B.*

  (Recall that, for $F = ([n], E')$, an embedding of F in G is a 1-1 mapping $f : [n] \to [N]$ that satisfies the three conditions in Definition 5.1. The foregoing if-statement asserts that for any such embedding f there exists $v \in [n]$ such that $f(v) \in B$.)

- *We say that $\overline{\mathcal{F}}$ is* non-propagating *if there exists a* (monotonically non-decreasing) *function $\tau : (0, 1] \to (0, 1]$ such that the following two conditions hold.*

  1. *For every $\epsilon > 0$ there exists $\beta > 0$ such that $\tau(\beta) < \epsilon$.*

2. *For every graph $G = ([N], E)$ and every $B \subset [N]$ such that B covers $\mathcal{F}_N$ in G, either G is $\tau(|B|/N)$-close to being $\mathcal{F}_N$-free or there are no N-vertex graphs that are $\mathcal{F}_N$-free.*[12]

A local property Π is non-propagating *if there exists a non-propagating sequence $\overline{\mathcal{F}}$* (as above) *such that Π is $\overline{\mathcal{F}}$-local.*

Intuitively, non-propagation means that the elimination of all possible embeddings of $\mathcal{F}$ in G, which necessarily use vertices in B, does not require modifying G "much beyond" B. For example, the set of graphs that have no isolated vertices constitutes a local property that is non-propagating (see the proof of Part 3 of Proposition 5.4). Indeed, it is natural to consider functions τ of the form $\tau(\beta) = O(\beta)$, but Definition 5.3 allows arbitrary functions τ (which may depend arbitrarily on $\overline{\mathcal{F}}$). In contrast to what one might naturally conjecture, as shown in Proposition 5.4, not all sequences of (sets of) marked graphs are non-propagating. On the other hand, the local properties that correspond to induced subgraph freeness (as in Theorem 4.7) are non-propagating. Indeed, the question of whether or not every local property is non-propagating remains open (see Section 5.4). We stress that a property may be local with respect to several different sequences of (sets of) marked graphs, where some of these sequences may be non-propagating and the other not. We also note that the issue of non-propagation arises in the (strong) lower bound for testing properties that can be defined by 3CNF formula [8] as well as in the orientation model for testing (e.g., [14]).

Proposition 5.4 (on satisfying the non-propagating condition):

1. (negative): *For every $d \geq 3$, there exists a sequence of sets of marked graphs $\overline{\mathcal{F}} = (\mathcal{F}_N)_{N\in\mathbb{N}}$ as in Definition 5.2 that does not satisfy the non-propagating condition.*

2. (positive – induced subgraph freeness): *For every sequence of sets of graphs $\overline{\mathcal{F}} = (\mathcal{F}_N)_{N\in\mathbb{N}}$ as in Theorem 4.7, the property of being $\overline{\mathcal{F}}$-free*[13] *is local and non-propagating; that is, there exists a sequence of sets of* marked *graphs $\overline{\mathcal{F}}' = (\mathcal{F}'_N)_{N\in\mathbb{N}}$ as in Definition 5.2 such that (1)* induced *subgraph freeness w.r.t $\overline{\mathcal{F}}$ is equivalent to* generalized *subgraph freeness w.r.t $\overline{\mathcal{F}}'$, and (2) $\overline{\mathcal{F}}'$ is non-propagating.*

3. (positive – non-hereditary properties): *There exist non-hereditary properties that are local and non-propagating. For example, the set of regular graphs constitutes such a property.*

Proof: We start by proving Part 1 (i.e., the negative claim). Consider a set $\mathcal{F}$ consisting of $\lfloor d/2 \rfloor + 1$ marked graphs that effectively impose the following two constraints (on $\mathcal{F}$-free graphs): (1) either there are no isolated vertices or all vertices are isolated, and (2) each non-isolated vertex has an odd degree. Specifically, the set $\mathcal{F}$ consists of the following two types of marked graphs: (see Figure 2):

1. A marked graph consisting of three vertices with a single edge connecting two vertices that are both marked partial, and an isolated vertex that is marked full. (This forbidden graph mandates that if the target graph contains any isolated vertex then it cannot contain any edges.)

[12] Indeed, it is more natural to disallow the latter possibility in the definition, but this would have made our exposition somewhat more cumbersome.

[13] That is, we refer to the set $\Pi = \bigcup_{N\in\mathbb{N}} \Pi_N$ such that each Π_N consists of all N-vertex graphs that are $\mathcal{F}_N$-free, where here we refer to induced subgraph freeness.

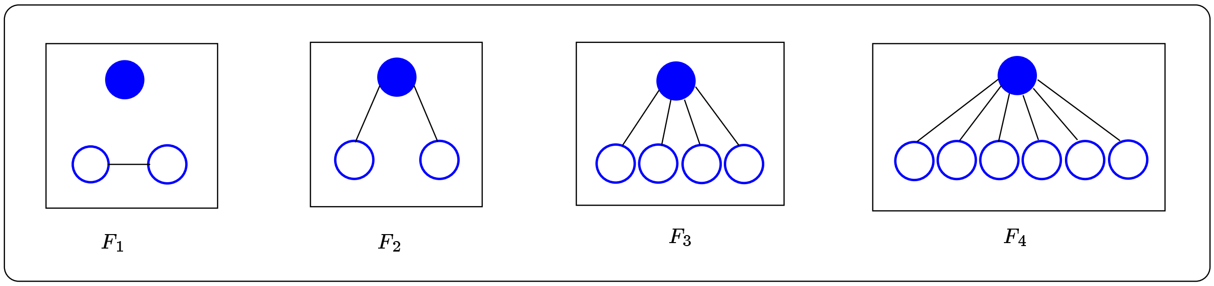

Figure 2: The forbidden marked graphs for the case $d = 7$ in Part 1 of the proof of Proposition 5.4. The graph F_1 is of the first type, and the graphs F_2, F_3, F_4 are of the second type.

2. For every even $i \in \{2, ..., d\}$, we have a graph with a single vertex marked full having i neighbors marked partial and having no other edges. (This set of forbidden graphs mandates that each non-isolated vertex has an odd degree.)

Note that if N is odd, then the only N-vertex graph that is $\mathcal{F}$-free is a set of N isolated vertices.[14] However (for odd N), consider any graph G that consists of a single isolated vertex and $N-1$ vertices that have odd degrees (e.g., G may consists of a single isolated vertex and a 3-regular $(N-1)$-vertex graph). Then, G contains only one vertex (i.e., the isolated vertex) that must appear in the image of any embedding of some $F \in \mathcal{F}$ in G. Thus, we obtain an infinite sequence of graphs that are $\Omega(1)$-far from being $\mathcal{F}$-free, whereas only one vertex (in each of these graphs) must be contained in any embedding of some $F \in \mathcal{F}$ in this graph. Indeed, this proves that $\mathcal{F}$ (or rather $\overline{\mathcal{F}} = (\mathcal{F}_N)_{N\in\mathbb{N}}$ such that $\mathcal{F}_N = \mathcal{F}$ for every $N \in \mathbb{N}$) does not satisfy the non-propagating condition (because we need $\tau(1/N) = \Omega(1)$, whereas $\lim_{N\to\infty} \tau(1/N)$ must equal zero).

Turning to Part 2 (i.e., the positive claim regarding induced subgraph freeness), we consider an arbitrary set of (unmarked) graphs $\mathcal{F}$ and the set of N-vertex graphs that are $\mathcal{F}$-free (as per Definition 4.6). As noted before, this property (or set) is local, because induced subgraph freeness can be emulated by generalized subgraph freeness. Specifically, for each $F \in \mathcal{F}$, we introduce a corresponding marked graph $F' \in \mathcal{F}'$ such that the graph F' is obtained from F by marking all vertices as semi-full. It follows that, for every $\overline{\mathcal{F}} = (\mathcal{F}_N)_{N\in\mathbb{N}}$ as in the proposition's hypothesis, the corresponding induced subgraph freeness property (i.e., $\overline{\mathcal{F}}$-freeness) is $\overline{\mathcal{F}}'$-local, where $\overline{\mathcal{F}}' = (\mathcal{F}'_N)_{N\in\mathbb{N}}$ is such that $\mathcal{F}'_N$ is obtained from $\mathcal{F}_N$ by the foregoing procedure.

The main point of Part 2 is proving that the sequence $\overline{\mathcal{F}}' = (\mathcal{F}'_N)_{N\in\mathbb{N}}$ is non-propagating. Let $G = ([N], E)$ and $B \subset [N]$ be as in Definition 5.3 (i.e., B covers $\mathcal{F}'_N$ in G). It follows that the subgraph of G induced by $[N] \setminus B$, denoted $G|_{[N]\setminus B}$, is $\mathcal{F}_N$-free (because if $G|_{[N]\setminus B}$ contains an induced subgraph that is isomorphic to $F \in \mathcal{F}_N$, then this isomorphism yields an embedding of the corresponding $F' \in \mathcal{F}'_N$ in G such that no vertex of F' is mapped to a vertex in B). We may assume, without loss of generality, that $|B| < N - 2ds$, where s is the maximum size of a graph in $\mathcal{F}_N$ (since otherwise non-propagation holds trivially, assuming $N > 4ds$). Using the fact that $G|_{[N]\setminus B}$ is $\mathcal{F}_N$-free, we claim that *the subgraph, denoted G', that results from G by turning B into an*

[14]Note that, for odd N, this set of graphs (i.e., the set of graphs consisting of isolated vertices) is $\mathcal{F}'$-free with respect to a non-propagating $\mathcal{F}'$ that contains a single graph that forbids any edges (i.e., the graph consists of a single edge with both endpoints marked partial). Thus, the current difficulty can be bypassed by using the general formalism, which refers to a sequence of sets of forbidden graphs (i.e., we may consider the sequence $(\mathcal{F}_N)_{N\in\mathbb{N}}$, where $\mathcal{F}_N = \mathcal{F}$ if N is even and $\mathcal{F}_N = \mathcal{F}'$ otherwise).

independent set is $\mathcal{F}_N$-free. This claim follows by considering an arbitrary s-vertex subset, S, and noting that if S induces a subgraph of G' that is in $\mathcal{F}_N$ then $S' \stackrel{\rm def}{=} S \setminus B$ combined with $r = s - |S'|$ adequate vertices induce the same subgraph in $G|_{[N]\setminus B}$: Pick r vertices in $[N] \setminus (B \cup S')$ such that in G these vertices constitute an independent set that neighbors no vertex in S'.[15] Thus, G is $2(|B|/N)$-close to being $\mathcal{F}_N$-free (which is the same as being $\mathcal{F}'_N$-free). It follows that $\overline{\mathcal{F}}'$ satisfies the non-propagating condition (with $\tau(\beta) = 2\beta$).

Finally, we turn to Part 3 (i.e., the positive claim regarding non-hereditary properties). Consider, for example, the set of graphs that contain no isolated vertices, which coincides with the set of graphs that are I-free where I is the marked graph that consists of a single (isolated) vertex that is marked full. Clearly, this set is not hereditary. To see that $\{I\}$ is non-propagating, consider any graph $G = ([N], E)$ and $B \subset [N]$ as in Definition 5.3 (i.e., every embedding of I in G maps the single vertex of I to a vertex in B). It follows that $[N] \setminus B$ contains no isolated vertices, and so G is $(|B|/dN)$-close to being I-free. Thus, $\{I\}$ satisfies the non-propagating condition (with $\tau(\beta) = \beta/d$).

Lastly, we show that the set of regular graphs, which is also non-hereditary, is local and non-propagating. To see that this set constitutes a local property, consider a set of marked graphs $\mathcal{F}_N$ that forbids two vertices of different degrees; a typical member of this set consists of two vertices marked full that are connected to a different number of vertices marked partial (in addition to, possibly, an edge between the two 'full' vertices). In addition, if N is odd then we also forbid odd degrees. To see that this sequence of sets $\overline{\mathcal{F}} = (\mathcal{F}_N)_{N\in\mathbb{N}}$ of marked graphs is non-propagating, consider any graph $G = ([N], E)$ and $B \subset [N]$ as in Definition 5.3 (i.e., every embedding of some $F \in \mathcal{F}_N$ in G maps some vertex of F to a vertex in B). Letting $C \stackrel{\rm def}{=} B \cup \bigcup_{v\in B} \Gamma(v)$ be the set of all vertices that are either in B or neighbor it, we note that all vertices in $[N] \setminus C$ have the same degree. Intuitively, G can be made regular by only modifying edges that are incident at C. This is easy to see if we allow multiple edges, and essentially holds also otherwise.[16] It follows that $\overline{\mathcal{F}}$ satisfies the non-propagating condition (with $\tau(\beta) = O(d\beta)$). ■

5.3 The characterization

We now turn to the main result of the current section.

Theorem 5.5 (characterization for the bounded-degree graphs model): *A graph property Π has a constant-query proximity-oblivious tester if and only if Π is local and non-propagating.*

Unlike in the case of Theorem 4.7 (see Footnote 6), here we rely on the fact that the detection probability function depends only on the proximity parameter. We stress that the class of properties having constant-query proximity-oblivious tester is a strict superset of the class of properties that refer to induced subgraph freeness.

Proof: We start by showing that any non-propagating local graph property Π has a constant-query proximity-oblivious tester. Suppose that Π is $\overline{\mathcal{F}}$-local, where $\overline{\mathcal{F}} = (\mathcal{F}_N)_{N\in\mathbb{N}}$, and let c and r be upper bounds on the number of connected components and the radius of each connected

[15]Such r vertices exist, because $[N] \setminus (B \cup S')$ contains at least $(N - |B|) - (d+1)|S'|$ vertices that do not neighbor S', and such a set contains an independent set of size $\frac{N-|B|-(d+1)|S'|}{d} > r$.

[16]Replacing each pair of edges in $C \times ([N] \setminus C)$ by a single edge between the endpoints in $[N] \setminus C$, we maintain the degree of vertices in $[N] \setminus C$ while leaving at most one edge in $C \times ([N] \setminus C)$. Replacing the subgraph induced by C by an adequate subgraph, we obtain the desired regular graph. Finally, multiple edges can be eliminated as follows. Suppose that we wish to eliminate an edge that connects u and v. Then, we select an edge (u', v') such that (u, u') and (v, v') and not edges, and omit the edges (u, v) and (u', v') while adding the edges (u, u') and (v, v').

component (in each graph in $\mathcal{F}_N$), respectively. We consider the following tester T (for Π):[17] on input an N-vertex graph G, the tester selects at random c start vertices $v_1, ..., v_c \in [N]$, performs a BFS of depth $r+1$ starting at each v_i, and accepts if and only if the subgraph explored in these c executions of BFS is $\mathcal{F}_N$-free. More precisely, T accepts unless there is an embedding of some $F \in \mathcal{F}_N$ in the said subgraph such that each vertex of F is mapped to a vertex of G that is at distance at most r from some v_i. (The extra level of the BFS is used in order to identify all edges incident at vertices that reside in level r).[18]

Clearly, T always accepts any N-vertex graph that is $\mathcal{F}_N$-free. In the analysis of T's detection probability (of graphs that are not $\mathcal{F}_N$-free), we shall consider a more relaxed rejection criterion that checks, for every $F \in \mathcal{F}_N$, whether the $i^{\rm th}$ connected component of F can be embedded in the subgraph explored in the $i^{\rm th}$ BFS such that some vertex of this component is mapped to v_i (i.e., the $i^{\rm th}$ start vertex). Thus, we refer to an embedding that maps the $i^{\rm th}$ connected component of F to the r-neighborhood of v_i, where the r-neighborhood of a vertex v in G is defined as follows. It is the graph that is isomorphic to the subgraph of G that contains all the vertices that are at distance at most $r+1$ from v and all edges that are incident at vertices that is at distance at most r from v. The vertices in this graph are unlabeled, and the vertex corresponding to v is the designated *center* of the graph. It will be instructive to consider a function (depending on G) that assigns each vertex $v \in [N]$ its r-neighborhood.

Towards analyzing the detection probability of T, let us consider the following simplified property testing problem referring to functions from $[N]$ to $[m]$. The property, denoted $\mathcal{P}$, is defined by a fixed set of (forbidden) sequences $\overline{\mathrm{F}} \subseteq [m]^c$ such that a function $f : [N] \to [m]$ is in $\mathcal{P}$ if, for every $v_1, ..., v_c \in [N]$, it holds that $(f(v_1), ..., f(v_c)) \notin \overline{\mathrm{F}}$. We analyze the straightforward tester that selects uniformly $v_1, ..., v_c \in [N]$ and accepts if and only if $(f(v_1), ..., f(v_c)) \notin \overline{\mathrm{F}}$. Suppose that f is ϵ-far from $\mathcal{P}$ (and that $\epsilon N > cm$), and let $V \stackrel{\rm def}{=} \{v : \Pr_{r\in[N]}[f(r) = f(v)] \geq \epsilon/m\}$ denote the set of ("typical") points that are assigned values that appear relatively frequently. Then, f restricted to V is not in $\mathcal{P}$, because otherwise we can modify f on $[N] \setminus V$ (using arbitrary values in $\{f(v) : v \in V\}$) and obtain a function in $\mathcal{P}$ that is ϵ-close to f. It follows that there exist $v_1, ..., v_c \in V$ such that $(f(v_1), ..., f(v_c)) \in \overline{\mathrm{F}}$, and it follows that

$$\begin{aligned} \Pr_{u_1,...,u_c\in[N]}[(f(u_1), ..., f(u_c)) \in \overline{\mathrm{F}}] &\geq \Pr_{u_1,...,u_c\in[N]}[(\forall i \in [c])\, f(u_i) = f(v_i)] && (19)\\ &\geq \left(\min_{v\in V}\left\{\Pr_{r\in[N]}[f(r) = f(v)]\right\}\right)^c && (20)\end{aligned}$$

which is lower-bounded by $(\epsilon/m)^c$.

The foregoing paragraph suggests to define a function f such that $f(v)$ describes the r-neighborhood of vertex v in G. However, the current situation is more complex because the r-neighborhoods of the various vertices in G are related, and thus modifying f at one vertex may require modifying it in many other vertices. This is where the non-propagating condition comes into play. Indeed, in the following analysis we shall refer to the function τ provided by the non-propagating condition. We shall also assume that $\Pi_N \neq \emptyset$ (and rely on the convention that if $\Pi_N = \emptyset$ then T rejects without making any queries).

Fixing any $\epsilon > 0$, let $\beta > 0$ be a relatively large number such that $\tau(\beta) < \epsilon$ (e.g., $\beta = \sup_{\tau(x)<\epsilon}\{x\}/2$). The number of vertices at distance at most $r+1$ from any vertex in a graph

[17] The foregoing description refers to the case that $\Pi_N \neq \emptyset$; otherwise, T just reject without making any queries.

[18] Needless to say, we need to identify edges that connect pairs of vertices that reside at level r. Furthermore, we also need to identify edges that connect vertices at level r with vertices at level $r+1$, or rather to verify that no such edges exist for certain vertices. This is important in case the embedding maps a vertex marked full to level r.

of maximum degree d is at most $\sum_{i=0}^{r+1} d^i < 2d^{r+1}$. By the definition of the r-neighborhood of a vertex, the number of values that the r-neighborhood can take is upper bounded by $2^{\binom{2d^{r+1}}{2}} \cdot 2d^{r+1}$ (where the first term in the product corresponds to the number of (unlabeled) graphs over $2d^{r+1}$ vertices, and the second term correspond to the choice of the center vertex). This expression is upper bounded by $2^{d^{3r}}$. Hence, for $m \stackrel{\rm def}{=} 2^{d^{3r}}$, in any graph and for every $\delta \geq 0$, at most a δ fraction of the vertices have an r-neighborhood that occurs in less than a δ/m fraction of the vertices. Now, consider any N and any N-vertex graph $G = ([N], E)$ that is ϵ-far from Π, and let B denote the set of vertices that have an r-neighborhoods that occurs in less than $\beta N/m$ vertices. By the aforementioned observation, $|B| \leq \beta N$. We claim that there exist c vertices $v_1, ..., v_c \in ([N] \setminus B)$ and a marked graph $F \in \mathcal{F}_N$ that can be embedded in G such that the following holds. For every $i \leq c_F$ some vertex of the $i^{\rm th}$ connected component of F is mapped to v_i, where $c_F \leq c$ denotes the number of connected components in F. This claim holds because otherwise, for every $F \in \mathcal{F}_N$, every embedding of F in G must map some vertex of F to a vertex in B. By the non-propagating condition this implies that the graph G is $\tau(|B|/N)$-close to Π_N, whereas $\tau(|B|/N) < \epsilon$ (in contradiction to G being ϵ-far from Π_N). Using the claim it follows that some $F \in \mathcal{F}_N$ can be embedded in G so that for each i the $i^{\rm th}$ connected component of F is mapped inside the r-neighborhood of some $v_i \in ([N] \setminus B)$, and thus T rejects if it selects this sequence (i.e., $v_1, ..., v_c$) of start vertices. Recalling that $[N] \setminus B$ contains only vertices with an r-neighborhood that occurs in many (i.e., $\beta N/m$) vertices, we proceed as in the foregoing warm-up (regarding generic functions from $[N]$ to $[m]$). Specifically, the probability that c uniformly selected vertices of G have this specific forbidden sequence of r-neighborhoods (as the aforementioned $v_1, ..., v_c$) is at least $(\beta/m)^c$. Recalling that T rejects when seeing this sequence of r-neighborhoods, we are done (i.e., we showed that any graph that is ϵ-far from Π is rejected with probability at least $(\sup_{\tau(x)<\epsilon}\{x\}/2m)^c$).

We now turn to showing that any property that has a constant-query proximity-oblivious tester is indeed local and non-propagating. We start by providing canonical testers for the current model, where the canonization process resembles (but is different from) the process applied in the adjacency matrix model (see Theorem 4.7, which uses [20, Thm. 4.5]). Needless to say, unlike in the latter model, we have no hope to obtain non-adaptive testers (cf. [25]). Still, we may obtain a relaxed notion of non-adaptivity (i.e., a notion of "indirect non-adaptivity"), like the one implicit in the following definition.

Definition 5.5.1 (canonical testers in the bounded-degree model): *A probabilistic oracle machine M is called* canonical *if, on input N and oracle access to $g : [N] \times [d] \to \{0, 1, ..., N\}$, the machine M behaves as follows.*

1. *For some predetermined function $s : \mathbb{N} \to \mathbb{N}$, the machines selects uniformly a set S of $s(N)$ elements in $[N]$.*

2. *For some predetermined function $\ell : \mathbb{N} \to \mathbb{N}$, the machine conducts a $\ell(N)$-step BFS from each vertex in S. That is, for every $v \in S$, and every $t = 1, ..., \ell(N)$ and $i_1, ..., i_t \in [d]$, the machine obtains the value $g(v, i_1, ..., i_t)$, where $g(v, i_1, ..., i_t) \stackrel{\rm def}{=} g(w, i_t)$ if $w = g(v, i_1, ..., i_{t-1}) \neq 0$ and $g(v, i_1, ..., i_t) \stackrel{\rm def}{=} 0$ otherwise. Indeed, if $w = g(v, i_1, ..., i_{t-1}) \neq 0$, then the value $g(v, i_1, ..., i_t)$ is obtained by making the query (w, i_t).*

3. *The machine M decides according to N and the subgraph of G explored by it. Specifically, M's decision depends on a fixed set of marked graphs, denoted $\mathcal{F}_N$, such that M accepts if and only if no $F \in \mathcal{F}_N$ appears in the explored subgraph of G. That is, G is accepted if there*

is no embedding of any $F \in \mathcal{F}_N$ (in G) that maps each vertex of F to a vertex that is at distance at most $\ell(N)$ from one of the $s(N)$ start vertices.

Indeed, the tester T presented in the first part of the proof is canonical (with constant s and ℓ). Our point, however, is that any tester can be converted into a canonical one. Unlike in the adjacency matrix model (cf. [20]), the current transformation incurs an exponential blow-up in the query complexity. Since we aim to apply this canonization transformation to (constant-query) proximity-oblivious testers, we state the transformation for generalized testers allowing arbitrary rejection probabilities of arbitrary no-instances.

Claim 5.5.2 *Let T be a generalized one-sided error tester of query complexity q for a property Π of graphs of maximum degree d. Then, Π has a canonical tester of query complexity $Q \stackrel{\rm def}{=} \widetilde{O}(d^q)$ that always accepts any graph in Π and rejects any graph G not in Π with probability that is lower-bounded by the probability that T rejects G.*

Proof: The core of the desired transformation is obtained by an adequate adaptation of the transformation provided in [20, Sec. 4]. Analogously to [20, Sec. 4.1], we first convert T into a tester T' that makes all queries as postulated in Steps 1 and 2 of Definition 5.5.1, while setting s and ℓ to equal q. After acting as postulated in these two canonical steps, the tester T' emulates the execution of T while answering its queries as follows. When T makes a query (v,i) such that v did not appear in any prior query or answer, the tester T' allocates to v the next unused vertex u in the initial sample S, and otherwise T' just uses the allocation determined before; that is, if v did not appear before then T' defines $\pi(v) = u$ and otherwise T' just uses the value $\pi(v)$ defined before. The answer provided by T' to the query (v,i) of T is $\pi^{-1}(g(\pi(v),i))$ if the latter is defined, and otherwise the answer is defined as a new random value r (different from all queries made by T and all answers given to T) and $\pi(r)$ is defined to equal $g(\pi(v),i)$. If $\pi(r)$ is in S then (in the future) it will be considered used.

Note that all the values $g(\cdot,\cdot)$ used by T' in the foregoing process are values that appear in one of the BFS executions (i.e., we use $g(u,i)$ for either $u \in S$ or for some u that appeared as an answer to some prior query (w,j), i.e., $u = g(w,j)$). On the other hand, the execution of T' on input G corresponds to an execution of T on a random isomorphic copy of G (where the isomorphism is provided by the permutation π, which is selected on-the-fly by T').

Next, analogously to [20, Sec. 4.2], we note that, without loss of generality, the decision of T' is sample-oblivious and label-oblivious; that is, the decision depends only on the edges (and non-edges) among the explored vertices (i.e., the underlying subgraph explored by the BFS executions), and not on the actual labels of these vertices in G. This is proved by making T' accept with probability that equals the average of all relevant probabilities (i.e., the acceptance probabilities that are associated with each of the possible relabellings of the subgraph), and observing that the probability that the resulting T' accepts G equals the probability that the original T' accepted a random isomorphic copy of G. Note that the decision of the resulting T' may still depends on an identification of the $s(N)$ initial vertices (from which the corresponding BFS executions were started), but it does not depend on the labels of these (or any other) vertices.[19]

Finally, we use the fact that T' has one-sided error in order to make the final decision deterministic as well as invariant under the identification of the $s(N)$ initial vertices. Firstly, as in [20,

[19]Indeed, the identity of the start vertex (of an exploration) need not be uniquely determined by the subgraph explored in an ℓ-step BFS, even when ℓ is known. Consider, for example, a 4-step BFS yielding the subgraph that consists of the edges $\{0,1\},\{1,2\},\{1,3\},\{2,3\},\{1,4\},\{4,5\},\{5,6\}$. Note that the corresponding 4-step BFS exploration could have been initiated at vertex 0 as well as either at vertex 2 (or 3) or at vertex 6.

Sec. 4.2], we note that if T' rejects with non-zero probability when seeing a particular subgraph of G then it must be the case that G is not in Π, and hence we may modify T' such that it rejects with probability 1 in this case. Similarly, we may extend the rejection criterion by omitting the identification of the $s(N)$ initial vertices (but maintaining the distinction between vertices whose neighborhood was fully explored and those discovered in the last step of one of the BFS executions). That is, if T' rejects with one identification of the initial vertices then the resulting tester will reject when seeing the same subgraph with any other possible identification of the initial vertices. Thus, the final decision of the resulting tester only depends on the marked graph that it sees in its exploration, where vertices are marked partial if and only if they were discovered at the last step of one of the BFS executions (and are marked full otherwise). Indeed, this tester is canonical, and the claim follows. ■

Applying Claim 5.5.2 to any constant-query proximity-oblivious tester for Π, we obtain a canonical tester of constant query complexity. Letting $\overline{\mathcal{F}} = (\mathcal{F}_N)_{N\in\mathbb{N}}$ be the sequence of sets of marked graphs used by (Step 3 of) this tester, we claim that, for every N and every N-vertex graph G, it holds that $G \in \Pi$ if and only if G is $\mathcal{F}_N$-free. The claim follows by noting that $G \in \Pi$ if and only if the canonical tester accepts it with probability 1, which happens if and only if G is $\mathcal{F}_N$-free (by the description of the canonical tester and the definition of generalized subgraph freeness). It follows that Π is local (and, in fact, it is $\overline{\mathcal{F}}$-local).

It is left to prove that $\overline{\mathcal{F}}$ is non-propagating. We shall refer to the canonical tester derived above, and specifically to its detection probability function ρ (which equals the detection probability function of the constant-query proximity-oblivious tester of the hypothesis). Let us denote the query complexity of the canonical tester by q. We define $\tau : (0,1] \to (0,1]$ so that $\tau(\beta)$ equals a "relatively small" $\alpha \in (0,1]$ that satisfies $\rho(\alpha) > q\beta$ (e.g., $\tau(\beta) = 2\inf_{\rho(x)>q\beta}\{x\}$ if $\rho(1/2) > q\beta$ and $\tau(\beta) = 1$ otherwise). Note that, indeed, for every $\epsilon > 0$ there exists $\beta > 0$ such that $\tau(\beta) < \epsilon$. We shall show that $\overline{\mathcal{F}}$ satisfies the non-propagating condition with respect to this function τ. For any N, consider any graph $G = ([N], E)$ and any $B \subset [N]$ such that every embedding of any $F \in \mathcal{F}_N$ in G maps some vertex of F to B. Assume, towards the contradiction, that G is $\tau(|B|/N)$-far from Π_N (while $\Pi_N \neq \emptyset$), where Π_N denotes the set of N-vertex graphs that are $\mathcal{F}_N$-free. Then, the canonical tester must reject G with probability at least $\rho(\tau(|B|/N))$. On the other hand, the canonical tester may reject G only if one of the vertices that it visits resides in B. Since each vertex is visited with probability at most q/N, it holds that $\rho(\tau(|B|/N)) \leq q \cdot |B|/N$, which contradicts the definition of τ (i.e., $\rho(\tau(\beta)) > q\beta$).[20] ■

A quantitative version. We note that the proof of Theorem 5.5 provides a rather tight relation between the optimal detection probability of constant-query proximity-oblivious testers and the function τ used in the definition of the non-propagating condition (cf., Definition 5.3). Specifically, these two functions are roughly inverses of one another; for example, polynomial detection probability (i.e., $\rho(\epsilon) = \epsilon^{O(1)}$) correspond to constant-root functions (i.e., $\tau(\beta) = \beta^{\Omega(1)}$), whereas exponential detection probability (i.e., $\rho(\epsilon) = 2^{-O(1/\epsilon)}$) correspond to logarithmic functions (i.e., $\tau(\beta) = O(1/\log(1/\beta))$). A closer look at the proof of Theorem 5.5 also yields the following corollary.

Corollary 5.6 *For every sequence of graphs $\overline{\mathcal{F}} = (\mathcal{F}_N)_{N\in\mathbb{N}}$ as in Theorem 4.7, the property of being $\overline{\mathcal{F}}$-free has a constant-query proximity-oblivious tester of polynomial detection probability function* (i.e., $\rho(\epsilon) \geq \mathrm{poly}(\epsilon)$). *Furthermore, the degree of this polynomial equals the maximum number of connected components in a graph in $\overline{\mathcal{F}}$.*

[20]Indeed, we assumed that $\tau(\beta) < 1$, and the claim hold vacuously otherwise.

We note that the said dependency is optimal. Consider, for example, the graph F that consists of $c < d$ connected components such that the $i^{\rm th}$ component consists of a single vertex marked full that is connected to i vertices marked partial. Then, the set of $\{F\}$-free graphs consists of graphs whose degree distribution does not contain the entire set $[c]$ (i.e., for any $\{F\}$-free graph G there exists $i \in [c]$ such that no vertex in G has degree i). On the other hand, a constant-query proximity tester for this set has detection probability $\rho(\epsilon) = O(\epsilon)^c$, because an N-vertex graph that is ϵ-far from this set may have ϵN vertices of each problematic degree (whereas we should see all problematic degrees when rejecting).

Proof: As shown in the proof of Proposition 5.4, this property is local and non-propagating with $\tau(\beta) = O(\beta)$. Let c denote an upper bound on the number of connected components in any graph in $\overline{\mathcal{F}}$, and let r denote a corresponding bound on the radius of such components. Then, the first part of the proof of Theorem 5.5 implies that this property has a $2d^{r+1}$-query proximity-oblivious tester with detection probability $\rho(\epsilon) > (\beta/\exp(d^{3(r+1)}))^c$, where $\beta = \Omega(\epsilon)$ satisfies $\tau(\beta) < \epsilon$. The claim follows. ■

Easily testable properties having no proximity-oblivious testers. While connectivity can be tested with query-complexity that is inversely proportional to the proximity parameter [17], this property has no constant-query proximity-oblivious tester. That is:

Proposition 5.7 *Connectivity has no constant-query proximity-oblivious tester. Furthermore, connectivity is not a local property.*

Proof: Let $\mathcal{F}$ be a set of marked graphs as in Definition 5.1, and suppose that the largest graph in $\mathcal{F}$ has n vertices. We shall show that, for every $N \geq 2n+4$, the set of connected N-vertex graphs does not coincide with the set of N-vertex graphs that are $\mathcal{F}$-free. Consider, towards the contradiction, a graph G that consists of two isolated cycles, each of size at least $n+2$. If G is $\mathcal{F}$-free then we are done (since G is not connected). On the other hand, if G is not $\mathcal{F}$-free, then we consider an embedding of some $F \in \mathcal{F}$ in G, and note that each cycle contains at least one pair of adjacent vertices that are not in the image of this embedding (i.e., let (u_i, v_i) denote such a pair on the $i^{\rm th}$ cycle). Then, by switching edges between the two cycles, we obtain an N-vertex cycle that is still not $\mathcal{F}$-free (i.e., replace the edges (u_1, v_1) and (u_2, v_2) by the edges (u_1, u_2) and (v_1, v_2)), and so we are done. ■

5.4 Conclusion

We end this section by explicitly stating the main problem left open.

Open Problem 5.8 (are all local properties non-propagating?) *Let $\overline{\mathcal{F}} = (\mathcal{F}_N)_{N\in\mathbb{N}}$ be an arbitrary sequence of sets of marked graphs as in Definition 5.2. Is it the case that there exists another such sequence $\overline{\mathcal{F}}' = (\mathcal{F}'_N)_{N\in\mathbb{N}}$ that is non-propagating and defines the same property* (i.e., for every N and any N-vertex graph G it holds that G is $\mathcal{F}_N$-free if and only if G is $\mathcal{F}'_N$-free)*?*

Note that $\mathcal{F}'_N$ must depend on N even if $\mathcal{F}_N$ does not depend on N (i.e., $\mathcal{F}_N = \mathcal{F}$ for a fixed $\mathcal{F}$ and all N).[21] Recall that a property may be local with respect to several different sequences of (sets

[21] Consider the set $\mathcal{F}$ used in the proof of Part 1 of Proposition 5.4, and let $\mathcal{F}'$ be an arbitrary set of marked graphs such that every graph is $\mathcal{F}'$-free if and only if it is $\mathcal{F}$-free. Then, a graph G' with an even number of vertices that are each of odd degree is $\mathcal{F}'$-free. On the other hand, augmenting G' with a single isolated vertex, we obtain a graph G that is $\Omega(1)$-far from being $\mathcal{F}'$-free and yet only one vertex (i.e., the isolated vertex) must be contained in the image of any embedding of any $F' \in \mathcal{F}'$ in the graph G.

of) marked graphs, where some of these sequences may be non-propagating and the other not (cf. the proof of Part 1 of Proposition 5.4).

A related challenge is to determine relatively tight bounds on the function τ corresponding to various non-propagating local properties. In particular, can τ always be linear?

6 Concluding Comments

In this section we present some generic observations and discuss a couple of issues.

6.1 Generic observations

An obvious condition for the existence of a constant-query proximity-oblivious tester for a particular property is the existence of constant-size "refutations" for the property.

Definition 6.1 (refutations): *For* $\Pi = \bigcup_{n\in\mathbb{N}} \Pi_n$ *as in Definition 2.2, the sequence* $((x_1, y_1), ..., (x_q, y_q))$ *is called a* refutations for membership in Π_n *if for every* $f \in \Pi_n$ *there exists* $j \in [q]$ *such that* $f(x_j) \neq y_j$. *For* $s : \mathbb{N} \to \mathbb{N}$, *we say that* Π has size-s refutations *if for every* $n \in \mathbb{N}$ *and every* $f : [n] \to \{0,1\}^*$ *that is not in* Π *there exists a sequence* $x_1, ..., x_{s(n)}$ *such that* $((x_1, f(x_1)), ..., (x_{s(n)}, f(x_{s(n)}))$ *is a refutations for membership in* Π_n.

Theorem 6.2 *For* $s : \mathbb{N} \to \mathbb{N}$, *if a property* Π (as in Definition 2.2) *has an* s*-query proximity-oblivious tester, then it has size-*s *refutations.*

Like in the case of Theorem 4.7 (see Footnote 6), we only rely on the fact that every function not in Π must be rejected with positive probability (and we don't require this probability to be solely a function of the distance of this function from Π). We note that the proof of Proposition 4.5 implicitly used the statement in Theorem 6.2 (for constant s and for the special case of bipartitness), and Proposition 5.7 could have been proved using the theorem.

Proof: Using $\rho(\epsilon) > 0$ for every $\epsilon > 0$, it follows that the proximity-oblivious tester must reject any $f \notin \Pi$ with positive probability. Fixing an arbitrary $f : [n] \to \{0,1\}^*$ that is not in Π_n, let $x_1, ..., x_q \in [n]$ be a sequence of queries made by the tester when rejecting f. Note that the one-sided error of the tester implies that $((x_1, f(x_1)), ..., (x_q, f(x_q))$ is a refutation for membership in Π_n. The theorem follows. ■

Discussion. We stress that (unlike Theorem 4.7) Theorem 6.2 only establishes a necessary condition, and recall that this condition is not sufficient (see a dramatic demonstration in [8]).[22] Indeed, the existence of a constant-query proximity-oblivious tester (for property Π) depends not only on the existence of refutations (for membership in Π) but also on the ability to find such witnesses with probability related to the distance of the function from the property (while making a constant number of queries to the function). In the context of testing bounded-degree graphs (cf. Section 5) these qualities were linked to the non-propagating condition. This link was based on the existence of a canonical testers in the latter context, whereas such testers may not exist in general. Still, in the general setting, constant-query proximity-oblivious testers are implied by standard non-adaptive testers that rely on finding constant-size refutations.

[22] Recall that [8] presents a property that has constant-size refutations but no (standard) testers of sub-linear query complexity (even when fixing a sufficiently small constant value of the proximity parameter). It follows that this property has no proximity-oblivious testers of sub-linear (let alone constant) query complexity.

Theorem 6.3 *A property Π as in Definition 2.2 has a constant-query proximity-oblivious tester if Π has a standard tester T* (of error probability 1/3) *that satisfies the following three conditions:*

1. *T is non-adaptive;*
2. *T has query complexity, denoted $q : (0, 1] \to \mathbb{N}$, that only depends on the proximity parameter; and*
3. *For some fixed $s \in \mathbb{N}$, the tester T rejects if and only if it finds size-s refutations.*

Furthermore, assuming that q is monotonically non-increasing, the resulting proximity-oblivious tester makes s queries and has detection probability at least $\rho(\epsilon) = \Omega(q(\epsilon/2)^{-s} \cdot \epsilon)$.

Indeed, an observation similar to Theorem 6.3 underlies the proof of the positive part of Proposition 4.3. (In the latter proof we use the fact that the standard tester is further restricted and derived a stronger bound on ρ.) We note that in the case of properties of functions with a constant size range (e.g., Boolean functions), any adaptive tester can be transformed into a non-adaptive tester with an exponential blow-up in the query complexity. Hence, a variant of Theorem 6.3 holds for adaptive testers as well.

Proof: On input n and oracle access to $f : [n] \to \{0, 1\}^*$, the proximity-oblivious tester, T', proceeds as follows. First, T' selects $i \in \{1, ..., \lceil \log_2 n \rceil\}$ at random such that the value i is selected with probability 2^{-i}, and invokes (the query-generating algorithm of) T with the proximity parameter 2^{-i}. Thus, T' obtains a random set of queries that T issues (non-adaptively, on proximity parameter 2^{-i}). Denoting this set by $Q = \{x_1, ..., x_{q(2^{-i})}\} \subset [n]$, the proximity-oblivious tester selects a random s-subset of Q, and queries f on these indices. Finally, T' rejects if and only if the corresponding sequence of s queries and answers constitutes a refutation for membership in Π.

Clearly, T' never rejects any $f \in \Pi$. Towards analyzing the detection probability of T', let δ denote the distance of $f : [n] \to \{0, 1\}^*$ from Π_n. Then, T' selected $i = \lceil \log_2(1/\delta) \rceil$ with probability $\Omega(\delta)$, and conditioned on this event, with probability at least 2/3, the set of queries Q combined with the corresponding answers (of f) contains a size-s refutation. In this case, a uniformly selected set of s elements in Q yields a refutation with probability at least $|Q|^{-s} = q(2^{-i})^{-s} \geq q(\delta/2)^{-s}$. ■

Discussion. Needless to say, Theorem 6.3 is applicable to many property testers, since searching (non-adaptively) for a refutation is a natural way in which one-sided error testers proceed. Examples include testers for properties such as d-dimensional Euclidean metrics [23], singletons [24], and juntas [13], and various clustering problems (cf. [2]). We note that Theorem 6.3 is applicable also in case the query complexity of the original tester as well as the size of the refutation may depend on the function's domain (i.e., $[n]$), but in this case we obtain a relaxed notion of proximity-oblivious testing in which the detection probability may depend on the function's domain. That is, if the original tester makes $q(n, \epsilon)$ to any function over $[n]$ and searches for size-$s(n)$ refutations, then we obtain a relaxed proximity-oblivious tester that makes $s(n)$ queries and has detection probability at least $\rho(n, \epsilon) = \Omega(q(n, \epsilon/2)^{-s(n)} \cdot \epsilon)$.

6.2 The case of locally testable codes

The notion of proximity-oblivious testing was discussed in the context of locally testable codes (LTCs), which are error-correcting codes augmented by efficient codeword testers (i.e., testers for the property of being a codeword). Specifically, proximity-oblivious (codeword) testers (with linear detection probability function) correspond to the definition of *strong* codeword tests as in [19, Def. 2.2], whereas a restricted form of standard (codeword) testers correspond to the standard definition of codeword tests (called *weak* in [19, Def. 2.1]). We mention that while the main results of [19] refer to strong codeword tests, most subsequent work (including [11, Sec. 8]) refer to (weak) codeword tests. It is indeed an open problem whether the parameters of [11, Cor. 8.8] (i.e., constant query complexity and one-over-polylogarithmic rate) can be obtained with respect to strong codeword testing. That is:

Open Problem 6.4 *Do some error-correcting codes of constant relative distance and one-over-polylogarithmic rate have constant-query proximity-oblivious codeword testers?*

On the other hand, proximity-oblivious testers may provide a setting in which one may establish inherent limitations on codeword testing. Specifically, we conjecture that error-correcting codes of constant relative distance that have constant-query proximity-oblivious codeword testers must have rate that is inferior to arbitrary error-correcting codes of the same relative distance.

6.3 Two-sided error probability POT

Throughout this paper we considered proximity-oblivious testers (POTs) that always accept functions having the property. As commented in Section 2, it is easier to define the notion of proximity-oblivious testers in this setting (i.e., the setting of one-sided error probability). Still, one can also define a meaningful notion of two-sided error probability proximity-oblivious testers (POTs) by generalizing Definition 2.2 as follows:

Definition 6.5 (Definition 2.2, generalized): *Let* $\Pi = \bigcup_{n \in \mathbb{N}} \Pi_n$ *and* $\rho : (0,1] \to (0,1]$ *be as in Definition 2.2. A* two-sided error POT with detection probability ρ for Π *is a probabilistic oracle machine* T *that satisfies the following two conditions, with respect to a constant* $c \in (0,1]$*:*

1. *For every* $n \in \mathbb{N}$ *and* $f \in \Pi_n$*, it holds that* $\Pr[T^f(n)=1] \geq c$.
2. *For every* $n \in \mathbb{N}$ *and* $f : [n] \to \{0,1\}^*$ *not in* Π_n*, it holds that* $\Pr[T^f(n)=1] \leq c - \rho(\delta_{\Pi_n}(f))$*, where* $\delta_{\Pi_n}(f) = \min_{g \in \Pi_n}\{\delta(f,g)\}$ (as in Eq. (1)).

The constant c *is called the* threshold probability.

Indeed, Definition 2.2 is obtained as a special case by letting $c = 1$. Furthermore, for every $c \in (0,1]$, every property Π having a one-sided error POT also has a two-sided error POT that accepts every function in Π with threshold probability c (e.g., consider a generalized POT that activates the standard POT with probability c and rejects otherwise).

We note that two-sided error POTs exist also for properties that have no standard POT. A straightforward example is the property of Boolean functions that have at least a τ fraction of 1-values, for a constant $\tau \in (0,1)$. A more telling example refers to the set of Boolean function having a fraction of 1-values that is at least τ_1 but at most τ_2, for $0 < \tau_1 < \tau_2 < 1$. Assuming, without loss of generality, that $\tau_1 + \tau_2 \geq 1$, this property has a two-sided error POT that selects uniformly two samples in the function's domain, obtains the function values on them, and accept with probability p_i if the sum of the answers equals i, where $p_0 = 0$, $p_1 = 1$ and $p_2 = 2(\tau_1 + \tau_2 - 1)/(\tau_1 + \tau_2)$.

Additional results will be reported in a forthcoming work.

Acknowledgments

We are grateful to Salil Vadhan for preliminary discussions regarding the subject of this work. We also thank Ran Raz and Asaf Shapira for helpful discussions. Lastly, we thank Ronitt Rubinfeld and an anonymous referee for asking about the possibility of two-sided error POTs (as discussed in Section 6.3).

References

[1] N. Alon. Testing subgraphs of large graphs. *Random Structures and Algorithms*, 21:359–370, 2002.

[2] N. Alon, S. Dar, M. Parnas, and D. Ron. Testing of clustering. *SIAM Journal on Discrete Math*, 16(3):393–417, 2003.

[3] N. Alon, E. Fischer, M. Krivelevich, and M. Szegedy. Efficient testing of large graphs. *Combinatorica*, 20:451–476, 2000.

[4] N. Alon, E. Fischer, I. Newman, and A. Shapira. A combinatorial characterization of the testable graph properties: It's all about regularity. In *Proceedings of the Thirty-Eighth Annual ACM Symposium on the Theory of Computing*, pages 251–260, 2006.

[5] N. Alon and M. Krivelevich. Testing k-colorability. *SIAM Journal on Discrete Math*, 15(2):211–227, 2002.

[6] N. Alon and A. Shapira. A characterization of easily testable induced subgraphs. *Combinatorics Probability and Computing*, 15:791–805, 2006.

[7] M. Bellare, D. Coppersmith, J. Håstad, M. Kiwi, and M. Sudan. Linearity testing over characteristic two. *IEEE Transactions on Information Theory*, 42(6):1781–1795, 1996.

[8] E. Ben-Sasson, P. Harsha, and S. Raskhodnikova. 3CNF properties are hard to test. *SIAM Journal on Computing*, 35(1):1–21, 2005.

[9] I. Benjamini, O. Schramm, and A. Shapira. Every minor-closed property of sparse graphs is testable. In *Proceedings of the Fourtieth Annual ACM Symposium on the Theory of Computing*, pages 393–402, 2008.

[10] M. Blum, M. Luby, and R. Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of the ACM*, 47:549–595, 1993.

[11] I. Dinur. The pcp theorem by gap amplification. *Journal of the ACM*, 54(3), 2007. Article 12.

[12] E. Fischer. The art of uninformed decisions: A primer to property testing. *Bulletin of the European Association for Theoretical Computer Science*, 75:97–126, 2001.

[13] E. Fischer, G. Kindler, D. Ron, S. Safra, and S. Samorodnitsky. Testing juntas. *Journal of Computer and System Sciences*, 68(4):753–787, 2004.

[14] E. Fischer, O. Lachish, A. Matsliah, I. Newman, and O. Yahalom. On the query complexity of testing orientations for being Eulerian. In *Proceedings of the Twelvth International Workshop on Randomization and Computation (RANDOM)*, pages 401–415, 2008. To appear in *Transactions on Algorithms*.

[15] O. Goldreich, S. Goldwasser, E. Lehman, D. Ron, and A. Samordinsky. Testing monotonicity. *Combinatorica*, 20(3):301–337, 2000.

[16] O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. *Journal of the ACM*, 45(4):653–750, 1998.

[17] O. Goldreich and D. Ron. Property testing in bounded degree graphs. *Algorithmica*, pages 302–343, 2002.

[18] O. Goldreich and D. Ron. Algorithmic aspects of property testing in the dense graphs model. *ECCC*, TR08-039, 2008.

[19] O. Goldreich and M. Sudan. Locally testable codes and PCPs of almost linear length. *Journal of the ACM*, 53(4):558–655, 2006.

[20] O. Goldreich and L. Trevisan. Three theorems regarding testing graph properties. *Random Structures and Algorithms*, 23(1):23–57, 2003.

[21] O. Goldreich and L. Trevisan. Errata to [20]. Available from `http://www.wisdom.weizmann.ac.il/~oded/p_ttt.html`, 2005.

[22] T. Kaufman and M. Sudan. Algebraic property testing: The role of invariance. In *Proceedings of the Fourtieth Annual ACM Symposium on the Theory of Computing*, pages 403–412, 2008.

[23] M. Parnas and D. Ron. Testing metric properties. *Information and Computation*, 187(2):155–195, 2003.

[24] M. Parnas, D. Ron, and A. Samorodnitsky. Testing basic boolean formulae. *SIAM Journal on Discrete Math*, 16(1):20–46, 2002.

[25] S. Raskhodnikova and A. Smith. A note on adaptivity in testing properties of bounded degree graphs. Technical Report TR06-089, Electronic Colloquium on Computational Complexity (ECCC), 2006. Available from `http://www.eccc.uni-trier.de/eccc/`.

[26] D. Ron. Property testing. In *Handbook on Randomization, Volume II*, pages 597–649, 2001. Editors: S. Rajasekaran, P. M. Pardalos, J. H. Reif and J. D. P. Rolim.

[27] D. Ron. Property testing: A learning theory perspective. *Foundations and Trends in Machine Learning*, 1(3):307–402, 2008.

[28] D. Ron. Algorithmic and analysis techniques in property testing. *Foundations and Trends in TCS*, 5(2):73–205, 2009.

[29] R. Rubinfeld and M. Sudan. Robust characterization of polynomials with applications to program testing. *SIAM Journal on Computing*, 25(2):252–271, 1996.

A Alternative Definitions of Generalized Subgraph Freeness

In this appendix we show that the notion of generalized subgraph freeness (as in Definition 5.1) remains as expressive when disallowing either semi-full or partial markings.

The emulation of partial markings by semi-full markings is analogous to the emulation of non-induced subgraph freeness by induced subgraph freeness. That is, every graph $F = ([n], E_F)$, containing a vertex v that is marked partial can be replaced by a collection of graphs $F' = ([n], E'_F)$ such that E'_F contains E_F as well as some additional edges incident at v, and v is marked semi-full.

On the other hand, the effect of a marked graph containing semi-full vertices can be emulated by a set of marked graphs in which the corresponding vertices are marked full but are connected to some auxiliary vertices marked partial. Specifically, each marked graph $F \in \mathcal{F}$ is replaced by a corresponding set of marked graphs such that each F' in this set is as follows. (Note that by the first emulation, we may assume without loss of generality that F contains no vertices marked partial.) The vertex-set of F' consists of the vertices of F, which are all marked full, and a set of auxiliary vertices, which are all marked partial. All edges of F are edges in F', and in addition F' contains some edges with at least one endpoint that is marked partial (representing a vertex outside F). Without loss of generality, we only add edges with exactly one endpoint marked partial (and the other endpoint marked full). Thus, F' consists of a copy of F augmented by an arbitrary bipartite graph with vertices of F (marked full) on one side and auxiliary vertices (marked partial) on the other side. Without loss of generality, we only include a vertex that is marked partial if it is adjacent to some vertex marked full. All marked graphs F' that can be obtained in the foregoing manner are included in the derived set of marked graphs $\mathcal{F}'$. Thus, bearing in mind that all graphs have maximum degree at most d, we replace each marked graph in $\mathcal{F}$ by a finite set of marked graphs.