Quantum property testing in sparse directed graphs

Simon Apers*[1], Frédéric Magniez*[1], Sayantan Sen†[2], and Dániel Szabó*[1]

[1]Université Paris Cité, CNRS, IRIF, Paris, France
[2]Centre for Quantum Technologies, National University of Singapore, Singapore

We initiate the study of quantum property testing in sparse directed graphs, and more particularly in the unidirectional model, where the algorithm is allowed to query only the outgoing edges of a vertex.

In the classical unidirectional model the problem of testing k-star-freeness, and more generally k-source-subgraph-freeness, is almost maximally hard for large k. We prove that this problem has almost quadratic advantage in the quantum setting. Moreover, we prove that this advantage is nearly tight, by showing a quantum lower bound using the method of dual polynomials on an intermediate problem for a new, property testing version of the k-collision problem that was not studied before.

To illustrate that not all problems in graph property testing admit such a quantum speedup, we consider the problem of 3-colorability in the related undirected bounded-degree model, when graphs are now undirected. This problem is maximally hard to test classically, and we show that also quantumly it requires a linear number of queries.

1 Introduction

1.1 Context

In this modern big data era, the size of inputs has grown so much that even just reading the full input has become extremely expensive computationally. To tackle this challenge, the framework of *property testing* was initiated. It focuses on designing ultrafast algorithms (also known as "testers") that read only a small part of the input, and distinguish inputs that satisfy some property from inputs that are "far" from satisfying it. As a possible use-case, when the exact computation is expensive, one can use property testing algorithms as a precursor to running the final algorithm. If the input does not pass the property testing test, we can safely reject it, without running the expensive final computation.

At the same time, the field of quantum computing has significantly influenced many computer science paradigms, including cryptography, algorithms, and large-scale data processing. This new perspective on computer science based on quantum physics has sparked many fresh research directions. This includes the topic of this work, which combines quantum computing and property testing. More specifically, we consider quantum algorithms for *graph* property testing.

*Research supported in part by the European QuantERA project QOPT (ERA-NET Cofund 2022-25), the French PEPR integrated projects EPiQ (ANR-22-PETQ-0007) and HQI (ANR-22-PNCQ-0002), and the French ANR project QUOPS (ANR-22-CE47-0003-01).

†Research supported by the National Research Foundation, Singapore and A*STAR under its Quantum Engineering Programme NRF2021-QEP2-02-P05.

Graphs are of paramount importance for instance when it comes to understanding large datasets, since they provide a natural way to represent and analyze complex relationships inside datasets. Goldreich, Goldwasser, and Ron [GGR98] were the first to consider graphs in the context of property testing. Formally, given some form of query access to an unknown graph G on N vertices, and a property $\mathcal{P}$ of interest, the goal is to distinguish with high probability if G satisfies the property $\mathcal{P}$, or whether it is "*far*" from all graphs that satisfy $\mathcal{P}$, with a suitable notion of farness. In [GGR98] the "dense" model was considered, where a graph is accessed through *adjacency queries*: for a pair of vertices (u, v), the query reveals whether (u, v) is an edge in the graph. In this model, a graph G is ε-far from satisfying $\mathcal{P}$ if one needs to add or remove at least εN^2 edges of G to obtain a graph that satisfies $\mathcal{P}$.

In a later work, Goldreich and Ron [GR02] introduced the "bounded-degree" model for testing sparse graphs, focusing on the properties of bipartiteness and expansion. In this model, a d-bounded degree graph G with N vertices is accessed by performing *neighbor queries*: for a vertex v and an integer $i \in [d]$, the query (v, i) returns either the i-th neighbor of v, or some special symbol if v has less than i neighbors. The graph G is said to be ε-far from some property $\mathcal{P}$, if one needs to add or delete at least εdN edges of G to obtain a graph that satisfies $\mathcal{P}$. Over the last two decades, there has been a significant number of works in this model, and we refer the interested reader to the survey by Goldreich [Gol10].

Some researchers have considered efficient quantum algorithms for testing both classical and quantum objects, see for instance [BFNR08, ABRW16, HLM17, BDCG$^+$20, AS19] and the survey [MdW16]. Notably, the authors in [ACL11] initiated the study of bounded degree graph property testing in the quantum model. One important result in this context is the result of [BDCG$^+$20], who proved that there can be exponential quantum advantage in the bounded degree graph model of property testing. However, as mentioned in their paper, the graph property admitting the exponential quantum advantage is not a natural one.

1.2 Property testing of directed bounded degree graphs

While all of the aforementioned works consider undirected graphs, many real-world instances (such as the world wide web) actually correspond to *directed* graphs. Consequently, Bender and Ron [BR02] introduced a model of property testing for directed graphs, focusing on the properties of acyclicity and connectivity. Following that work, we open a new research line by studying quantum algorithms for testing directed graphs. As we will see, by doing so we address new fundamental questions in the field of quantum complexity. Answering them requires using recent techniques and partially answering some new or open questions.

As described in [BR02], for bounded-degree directed graphs there are two natural query models: (i) the *unidirectional* model, where the algorithm is allowed to query the outgoing edges of a vertex, but not the incoming edges, and (ii) the *bidirectional* model, where the algorithm can query both the incoming and outgoing edges of a vertex. Interestingly, [BR02] showed that strong connectivity is testable in the bidirectional model (i.e., it can be tested with a number of queries that depends on ε but not on N), but it requires $\Omega(\sqrt{N})$ queries in the unidirectional model. Later, the testability of other graph properties like Eulerianity, vertex and edge connectivity [OR11, YI10b, FNY$^+$20, CY19] was also shown in the bidirectional model. While there is a clear distinction between the two models, Czumaj, Peng and Sohler [CPS16] showed that if a property is testable in the bidirectional model, then it has a *sublinear* (i.e., $o(N)$) query complexity in the unidirectional model.

In this work we consider a particularly important problem in the unidirectional model: the

problem of testing *subgraph-freeness*. More precisely, we examine the problem of testing "*k-source-subgraph-freeness*", where the goal is to test H-freeness for some constant-sized subgraph H with k "source components", where a source component is a strongly connected subgraph that has no incoming edges. This problem was first studied by Hellweg and Sohler [HS12], and they provided a testing algorithm that performs $O(N^{1-1/k})$ queries. They also proved a tight lower bound of $\Omega(N^{2/3})$ for the $k = 3$ case (see [HS12, Theorem 1 and Theorem 3]). Very recently, Peng and Wang [PW23] proved a matching lower bound for any constant k. In particular, they showed that $\Omega(N^{1-\frac{1}{k}})$ queries are necessary for testing k-star-freeness in the unidirectional model, for arbitrary k (see [PW23, Theorem 1.2]). Notice that asymptotically the complexity of testing k-star-freeness becomes $\Omega(N)$. This also proves that the aforementioned reduction of [CPS16] cannot be made much stronger: for the property of k-star-freeness, the separation between the query complexities in the bi- and unidirectional models is maximal, because this property can be tested using constant number of queries in the bidirectional model.

1.3 Our results and techniques

In this work, we present two lines of results for quantum property testing of graph properties.

On the one hand, we prove that the problem of testing k-source-subgraph-freeness in the unidirectional model, which is almost maximally hard for large k in the classical regime, has almost quadratic advantage in the quantum setting. Moreover, we prove that this quantum advantage is nearly tight, by showing a quantum lower bound using the method of dual polynomials.

On the other hand, we show that not all problems in graph property testing admit such a quantum speedup such as the problem of 3-colorability in the related *undirected* bounded-degree model, which requires $\Omega(N)$ queries in both classical and quantum models.

We start by describing the upper bound result for testing k-source-subgraph-freeness.

Theorem 1.1 (Restated in Theorem 3.2)**.** *The quantum query complexity of testing k-source-subgraph-freeness in the unidirectional model is $O\Big(N^{\frac{1}{2}\left(1-\frac{1}{2^k-1}\right)}\Big)$.*

In order to prove the above result, we connect it to the problem of finding k-collisions, which is one of the most-studied problems in quantum query complexity – see [LZ19] and the references therein. In that work, they also prove a matching lower bound showing that their algorithm for *finding* k-collisions is optimal. However, the property testing variant of this problem – which is what we study – could be easier, and their lower bound technique does not apply to this case. This is why we use a different method to prove a lower bound, which is almost matching.

Theorem 1.2 (Corollary of Theorem 1.4)**.** *The quantum query complexity of testing k-source-subgraph-freeness in the unidirectional model is $\tilde{\Omega}\big(N^{\frac{1}{2}\left(1-\frac{1}{k}\right)}\big)$.*

We apply the well-studied dual polynomial method for proving this lower bound. We would like to point out that [BDCG$^+$20] also used dual polynomial method for proving quantum lower bounds for other problems. However, their techniques do not directly apply for our purpose, and we make adaptations of them. Here, we consider the auxiliary problem of k-collision-freeness that we reduce to k-source-subgraph-freeness. In general, testing collision-freeness of functions is a difficult problem. In fact, both the classical works of [HS12] and [PW23] use a collision testing result from [RRSS09]. Moreover, the authors in [ABRW16] stated it as an open question if one could use a variant of the proportional moments technique of [RRSS09] for proving better quantum lower bounds. However,

it is not clear if we can use the result of [RRSS09] directly for our purpose. Instead, we adapt the dual polynomial method to our context, which we believe will be of independent interest for improving the quantum lower bounds of other problems.

Doing so, we make the first step towards extending the techniques of [RRSS09]. At the heart of their lower bound is a construction of two positive integer random variables, X_1 and X_2, with different expectations but with the following conditions on the first $k-1$ moments: $\mathbb{E}[X_1]/\mathbb{E}[X_2] = \mathbb{E}[X_1^2]/\mathbb{E}[X_2^2] = \ldots = \mathbb{E}[X_1^{k-1}]/\mathbb{E}[X_2^{k-1}]$. Such a construction leads then to a randomized query complexity lower bound of $\Omega(N^{1-\frac{1}{k}})$ for various distinguishing problems such as k-collision-freeness [PW23]. We conjecture that a similar result holds in the quantum setting with a lower bound of $\Omega\left(N^{\frac{1}{2}(1-\frac{1}{k})}\right)$. One can consider this work as a proof of this conjecture for the special case of k-collision-freeness, and we hope that it will serve as a step towards proving it in general.

Finally, even for the case of undirected graphs, we show that not all problems in graph property testing have a quantum advantage. For this we consider the property testing variant of the famous problem of 3-colorability: namely, distinguishing whether the undirected graph G can be properly colored with 3 colors, or one needs to modify a large fraction of its edges to make it 3-colorable. In the classical bounded degree setting, this problem has been studied by [BOT02], who proved a lower bound of $\Omega(N)$ queries. In this work, we present a simple argument that proves that there exists no sublinear quantum tester either for this problem. Our result is stated as follows:

Theorem 1.3 (Restated in Theorem 5.1)**.** *The quantum query complexity of testing of 3-colorability of undirected bounded-degree graphs is* $\Omega(N)$.

We will prove this theorem in Section 5. The authors in [BOT02] prove classical hardness by using the hardness of distinguishing k-wise independent functions from uniform ones. By the polynomial method (again), this problem is also hard for quantum algorithms, and so we find a similar lower bound. In [YI10a], the authors used various reductions to argue that a number of other problems are maximally hard, including Hamiltonian Path/Cycle, approximating Independent Set/Vertex Cover size etc. As a corollary of our quantum lower bound, we also obtain maximal quantum query complexity for these problems.

1.4 Related works on collision finding

The problem of *collision finding* is a ubiquitous problem in the field of algorithm theory with wide applications in cryptography. Here, given a sequence s of N integers, the goal is to find a duplicate in s. If one has the guarantee that $\Theta(N)$ elements of the sequence are duplicated, which is the case for instance when the sequence consists of random integers from $[N]$, it is well-known that classically $\Theta(\sqrt{N})$ queries are necessary and sufficient due to the birthday paradox. In the quantum model, this can be solved with query complexity $\Theta(N^{1/3})$ by the algorithm of Brassard, Høyer and Tapp [BHT98]. The matching lower bound was first stated for a specific set of hard instances known as 2-to-1 (i.e. each integer appears exactly twice or not at all) by Aaronson and Shi [AS04]. For some constant integer $k \geq 3$, those results can be further extended for finding k-collisions in a random input with suitable alphabet size, so that it contains $\Theta(N)$ k-duplicates with high probability. The classical query complexity for this problem is $\Theta(N^{1-1/k})$ [HS12, PW23], and quantumly it is $\Theta\left(N^{\frac{1}{2}\left(1-\frac{1}{2^k-1}\right)}\right)$ [LZ19]. The situation is more complex for non-random inputs.

Remarkably, the complexity of *testing* k-collision-freeness (i.e., the absence of k-collisions) is harder to settle on the lower bound side. In this work, our lower bound shows the hardness of

distinguishing inputs that have linearly many collisions from those that do not have any. For $k = 2$, the two problems have the same complexity, since intuitively the only way to distinguish is to find a collision. This can be formalized easily in the classical case. Quantumly, this is more challenging, but the lower bound in [AS04] proved the hardness of distinguishing between 2-to-1 instances and ones with no duplicate.

However, for larger k, distinguishing such inputs might be easier than finding a collision. Classically the upper bound of $O(N^{1-1/k})$ is straightforward for the finding variant. In the lower bounds of [HS12, PW23], they consider the distinguishing version, so classically the question is settled. But in the quantum setting, the upper and lower bounds of [LZ19] are tight only for finding k-collisions, and for the distinguishing variant, we are not currently aware of anything better than the $\Omega(N^{1/3})$ lower bound corresponding to the $k = 2$ case. To our knowledge, this problem has not yet been studied in the quantum setting. Here, we present an almost matching lower bound.

Theorem 1.4 (Restated in Theorem 4.1)**.** *The quantum query complexity of testing k-collision-freeness is $\tilde{\Omega}\left(N^{\frac{1}{2}\left(1-\frac{1}{k}\right)}\right)$.*

1.5 Open problems

First, there is still a gap between our lower and upper bound on the quantum query complexity of testing k-collision-freeness. In [MTZ20], the authors keep using the dual polynomial method to improve the lower bound of [BKT20] for the k-distinctness problem. They achieve this by using a slightly different dual polynomial for THR_N^k, where they allow more weight on the false positive inputs. This makes it impossible to prove the high correlation of the dual and the primal function, so they use a modified block composition. Our technique might be combined with this other approach to improve our lower bound to $\tilde{\Omega}(N^{1/2-1/(4k)})$.

As we mentioned before, the authors in [ABRW16] stated it as an open question if one could use a variant of the proportional moments result of [RRSS09] to prove optimality of quantum property testers in the unidirectional model. This work may be considered as the first attempt to generalize this technique to the quantum setting.

In [CPS16] it was proved that if a graph property can be tested with $O(1)$ queries in the bidirectional model, then it can be tested using $O(N^{1-\Omega(1)})$ queries in the unidirectional model. It would be very interesting to investigate if it also implies a quantum tester with query complexity say $O(N^{1/2-\Omega(1)})$.

2 Preliminaries

2.1 Notations and basic definitions

Let us denote $[n] = \{1, \ldots, n\}$ and $[n]_0 = \{0, \ldots, n\}$. When dealing with Boolean variables, we will usually use $b \in \{-1, 1\}$ instead of $b \in \{0, 1\}$. We can get to one from the other easily with the mapping $b \mapsto 1 - 2b$, or its inverse, which means that -1 is going to be treated as the "true" or "accepting" value. The reason for using $\{-1, 1\}$ is that when dealing with dual polynomials it is easier to use this notation.

We denote by 1^n the length-n binary vector made only of 1s, and respectively -1^n. The Hamming weight $|x|$ of $x \in \{-1, 1\}^n$ is then defined as the number of -1s in x, that is $|x| = \#\{i \in [n] : x_i = -1\}$. Let $H^n_{\leq w} = \{x \in \{-1, 1\}^n : |x| \leq w\}$ denote the set of length-n

binary vectors with Hamming weight at most w. For any $x \in \mathbb{R}$, $\mathrm{sgn}(x) = 1$ when $x \geq 0$, and -1 otherwise.

For a polynomial p, let $\deg(p)$ denote its degree. The composition $f \circ g : \{-1,1\}^{nm} \to \{-1,1\}$ of two Boolean functions $f : \{-1,1\}^n \to \{-1,1\}$ and $g : \{-1,1\}^m \to \{-1,1\}$ is defined as $(f \circ g)(x) = f(g(x_1), \ldots, g(x_n))$ where $x = (x_1, \ldots, x_n)$ with each $x_i \in \{-1,1\}^m$.

A directed graph or digraph $G = (V, E)$ is a pair of a vertex set V and an edge set E. The latter consists of directed edges that are ordered pairs of vertices: we say that $(u, v) \in E$ is directed from u to v where $u, v \in V$. We say that there is a directed path from $s = v_0$ to $t = v_{l+1}$ (with $s, t \in V$) if there exists an integer l and vertices $v_1, \ldots, v_l \in V$ such that $\forall i \in [l]_0 : \ (v_i, v_{i+1}) \in E$. A digraph $G = (V, E)$ is called *strongly connected* if for every $u \in V$ and $v \in V \setminus \{u\}$, there exists a directed path from u to v. A *subgraph* of a graph $G = (V, E)$ is any graph $G' = (V', E')$ satisfying $V' \subseteq V$, $E' \subseteq E$ and $E' \subseteq V' \times V'$.

Finally, throughout this work, notations $O(\cdot)$ and $\Omega(\cdot)$ will be hiding the dependencies on parameters ε, k and d that we consider to be constants.

2.2 Query complexity

In query complexity, we consider inputs $x \in \Sigma^I$ over a finite alphabet Σ and indexed by a set I. They are not given explicitly to the algorithm. Instead, the algorithm has query access to an input oracle O_x encoding x, where the algorithm queries the character x_i of x, for any $i \in I$. Quantumly, the query access is described by the unitary operator $O_x |i\rangle |z\rangle = |i\rangle |z \oplus x_i\rangle$, for $z \in \Sigma$ and $i \in I$, where $\oplus$ is usually the bit-wise exclusive-OR operation up to some binary encoding of the elements of Σ. But our lower bound technique applies to any reversible operation $\oplus$.

Query complexity measures the minimum number of queries that an algorithm has to make in order to decide whether a property $P : D_P \subseteq \Sigma^I \to \{-1,1\}$ is satisfied, for an arbitrary input $x \in D_P$. Since the work of [BBC+01], it has been known that in the Boolean case (i.e. when $\Sigma = \{-1,1\}$), the acceptance probability of a T-query bounded-error quantum algorithm is a multivariate polynomial (in all the x_i's) of degree at most $2T$.

In this work, two kinds of inputs are going to play a crucial role. When the input is a sequence $s = (s_1, \ldots, s_N)$ of positive integers $\leq R$, then $I = [N]$ and $\Sigma = [R]$.

In the undirected bounded-degree graph model, we have query access to the adjacency list of an undirected graph $G = (V, E)$ with maximum degree d, represented as an oracle $\mathcal{O}_G : V \times [d] \to V \cup \{\perp\}$. Then we can set $I = V \times [d]$ and $\Sigma = V \cup \{\perp\}$, such that for any $v \in V$ and $i \in [d]$, we have the following:

$$\mathcal{O}_G(v, i) = \begin{cases} w, & \text{if } w \in V \text{ is the } i\text{-th neighbor of } v; \\ \perp, & \text{if } \deg(v) < i. \end{cases}$$

For bounded-degree directed graphs, there exist two query models. In the *bidirectional model*, we have access to both the outgoing and incoming edges of each vertex. Correspondingly, it is imposed that both the in- and out-degrees of a vertex are bounded by d. In the *unidirectional model*, we can only make queries to the adjacency list of the outgoing edges, and we impose only that the out-degrees of a vertex are bounded by d. Since in this work the primary focus will be on the latter model, let us formally define it below.

In the unidirectional bounded-degree graph model, we have query access to the adjacency list of a digraph $G = (V, E)$ where the out-degree of every vertex is at most d_{out}: for all $v \in V$: $\deg_{\mathrm{out}}(v) \leq d_{\mathrm{out}}$. This access is represented as an oracle $\mathcal{O}_G^{\mathrm{out}} : V \times [d_{\mathrm{out}}] \to V \cup \{\perp\}$. Then we can

set $I = V \times [d_{\text{out}}]$ and $\Sigma = V \cup \{\perp\}$, such that for any $v \in V$ and $i \in [d_{\text{out}}]$, we have the following:

$$\mathcal{O}_G^{\text{out}}(v,i) = \begin{cases} w, & \text{if } w \in V \text{ is the } i\text{-th out-neighbor of } v; \\ \perp, & \deg_{\text{out}}(v) < i. \end{cases}$$

For completeness, we note that in some of the previous work on the unidirectional model they do impose the degree bound on both the out- and in-degree [CPS16]. This is mostly because this makes for an easier comparison between the uni- and bidirectional models, as this way they allow the same set of graphs. In this work we only assume that only the out-degrees are bounded by d.

2.3 Property testing

In decision problems, the algorithm has to decide if the input satisfies a property or not. In the case of property testing the question is relaxed: the algorithm has to distinguish inputs that satisfy the property from those that are "far" (according to some distance measure) from any input that satisfies it.

The choice of distance measure usually depends on the query model considered. As discussed before, the general query access can be viewed as black box access to the input $x \in \Sigma^I$ where querying an index $i \in I$ reveals $x_i \in \Sigma$. This way, the distance of two objects is described as the proportion of positions where they differ:

$$x \text{ is } \varepsilon\text{-far from } y \iff |\{i \in I : x_i \neq y_i\}| \geq \varepsilon |I|.$$

Applying this to the case of bounded-degree graphs degree bound d and query access to the adjacency list, the distance of two graphs is the number of edges where they differ divided by $|V|d$. The distance of an object x from a property P is the minimum distance between x and any object that satisfies P.

Definition 2.1 (Property Testing)**.** *Let $0 < \varepsilon < 1$ be a constant. An algorithm $\mathcal{A}$ is an* ε-tester *for the property $\mathcal{P}$ if*

1. *For all $x \in \mathcal{P}$:* $\Pr[\mathcal{A}(x) = \text{accept}] \geq 2/3$*;*
2. *For all x that are ε-far from $\mathcal{P}$:* $\Pr[\mathcal{A}(x) = \text{accept}] \leq 1/3$.

Notice that no restriction is given on the acceptance probability of the algorithm for inputs that do not satisfy $\mathcal{P}$ but are ε-close to it.

2.4 Problem definitions

We now define the problems we study and argue about certain relations between them. While the problems are phrased as decision problems, ultimately we will care about the quantum query complexity for testing the corresponding properties. The complexity is going to be parameterised by a parameter k. Moreover, the parameter k, the degree bound d and the parameter ε, are all considered to be constants throughout this paper.

Let us start with some definitions that will be useful to define our problems precisely.

Definition 2.2 (Source component)**.** *Let $G = (V, E)$ be a digraph. A set $S \subseteq V$ is called a* source component *if it induces a strongly connected subgraph in G, and in G there is no edge from $V \setminus S$ to S.*

Definition 2.3 (k-star)**.** *A k-star is a digraph on $k+1$ vertices and k edges with one center vertex, and k source vertices connected to the center vertex.*

We will now state the decision variant of several problems. The "property" corresponding to a decision problem is the set of inputs that should be accepted in the decision problem.

k-Source-Subgraph-Freeness

Parameter: Graph H of constant size with at most k source components

Query access: d-bounded out-degree directed graph G on N vertices (unidirectional model)

Task: Accept iff G is H-free, that is, no subgraph of G is isomorphic to H

In [HS12, PW23], the authors examine the classical query complexity of testing k-source-subgraph-freeness. They consider the bounded-degree unidirectional model, albeit with a bound on both the in- and out-degrees.

For proving a lower bound, we will look at a special case of the main problem: k-star-freeness. Notice that a k-star has k source components, hence a lower bound for this problem implies the same lower bound for the more general k-source-subgraph-freeness problem.

k-Star-Freeness

Parameter: Integer $k \geq 2$

Query access: d-bounded out-degree directed graph G on N vertices (unidirectional model)

Task: Accept iff G is k-star-free, that is, no subgraph of G is isomorphic to the k-star

For the lower bound on k-star-freeness testing, we are going to use as a "helper problem" the decision variant of the k-collision problem.

k-Collision-Freeness

Parameter: Integer $k \geq 2$

Query access: Sequence of integers $s = (s_1, \ldots, s_N) \in [R]^N$

Task: Accept iff s is k-collision-free, i.e. there is no $i_1, \ldots, i_k \in [N]$ with $s_{i_1} = \cdots = s_{i_k}$

As discussed in the introduction (Section 1.4), very little was known about the property testing version of this problem prior to this work. We only know the that the complexity is $\Theta(N^{1/3})$ when $k = 2$, and it is between $\Omega(N^{1/3})$ and $O\left(N^{\frac{1}{2}\left(1-\frac{1}{2^k-1}\right)}\right)$ for larger k.

Reduction from k-collision-freeness to k-star-freeness Now we are going to prove that testing k-collision-freeness can be reduced to testing k-star-freeness (or more generally to testing k-source-subgraph-freeness). Thus, a lower bound on testing k-collision-freeness yields a lower bound on testing k-source-subgraph-freeness. Also, an algorithm for testing k-source-subgraph-freeness yields an upper bound on testing k-collision-freeness.

While the proof goes similarly to [HS12, Theorem 3], our reduction is not identical because we have a slightly different "helper problem". Since they consider that the in-degree of vertices to be bounded as well, for the collision problem, they assume that the sequence does not contain any collision of size larger than k (defined as *k-occurrence-freeness*).

Proposition 2.4. *The problem of ε-testing k-collision-freeness of a sequence from $[R]^N$ can be reduced to $\frac{\varepsilon N}{d(N+R)}$-testing k-star-freeness of an $(N+R)$-vertex sparse directed graph with out-degree bound $d \geq 1$.*

Proof. Let us assume that we have an algorithm that solves the k-star-freeness testing problem on graphs with out-degree bound $d \geq 1$, and we want to use it to test k-collision-freeness of a sequence $s = (s_1, \ldots, s_N) \in [R]^N$. We construct a digraph G that has N outer vertices $u_1, \ldots, u_N$ and R inner vertices $v_1, \ldots, v_R$; edges only exist from the outer vertices towards the inner ones such that u_i is connected to v_j iff $s_i = j$. Observe that the maximum out-degree in G is 1, so its out-degree is bounded by d for any $d \geq 1$.

It is clear that s is k-collision-free iff G is k-star-free. On the other hand, if s is ε-far from k-collision-freeness, it implies that more than εN edges have to be deleted in G to make it k-star-free. Thus G is $\varepsilon' = \frac{\varepsilon N}{d(N+R)}$-far from k-star-freeness. □

3 Quantum algorithm for testing subgraph-freeness

In this section, we prove that there is a quantum speedup for testing H-freeness in directed graphs with d-bounded out-degree, for any graph H that has k source components. For large but constant k, the speedup is near-quadratic. The algorithm relies on the following simple claim. It shows that if G is far from being H-free then it contains many (not necessarily disjoint) copies of H. This problem was studied in [GR02] in the classical setting.

Proposition 3.1. *Assume that an N-vertex graph G with d-bounded out-degree is ε-far from H-freeness. Then for each source component of H, there are $\Omega(\varepsilon N)$ vertices that are part of that component in an H-subgraph in G.*

Proof. Since G is ε-far from H-freeness, more than $\varepsilon N d$ edges need to be removed to get an H-free graph. Thus, for each of the k source components of H, there are more than $\varepsilon N d$ edges in G for which there is an H-subgraph where the edge starts from a vertex of the source component. Otherwise, deleting these (at most $\varepsilon N d$) edges would make that source component disconnected from the rest of H in all the H-instances, making G free from having H as a subgraph. This, in turn, implies that for each of the source components of H, there are at least εN vertices that are part of the source component in some H-subgraph in G, since every vertex can have at most d outgoing edges. □

To illustrate the algorithm, we first consider the $k = 2$ case to build our intuition. Our algorithm takes inspiration from the BHT algorithm for collision finding [BHT98]. In the following, we set $m = |V(H)|$, number of vertices in H. We use the shorthand BFS for breadth-first search.

1. Randomly sample a vertex subset $\mathcal{S}$ in G of size $t = \Theta(N^{1/3})$. Perform a depth-m BFS from every vertex $v \in \mathcal{S}$.

2. Perform a Grover search over the remaining vertices $V \backslash \mathcal{S}$ to look for a vertex that – after doing a depth-m BFS from it – completes an H-instance with one of the sampled vertices.

3. If any occurrence of H in G is found, output **Reject**. Otherwise, output **Accept**.

Note that if G is H-free, then the above algorithm will always accept. Now we need to argue that if G is ε-far from being H-free, then with constant probability the above algorithm will find a copy of H and thus reject.

To do so, pick one source component of H. By Proposition 3.1, with high probability a constant fraction of the t vertices in $\mathcal{S}$ are part of that source component in an H-subgraph in G. For such vertices, the BFS in step 1 will discover the entire source component, as well as all other vertices reachable from that source component in H. Then, in step 2 we effectively search for a vertex that is in the remaining source component of such an instance of H that we already partly discovered. Again by Proposition 3.1, with high probability there are $\Omega(t)$ many of them. By doing a BFS from that vertex, we uncover the resulting H-subgraph. This proves correctness of the algorithm.

Finally, we bound the algorithm's query complexity. Step 1 only makes $O(t) = O(N^{1/3})$ many queries. For step 2, we argued that there are $\Omega(t)$ many "good" vertices, and hence Grover search will make $O(\sqrt{N/t}) = O(N^{1/3})$ quantum queries.

We are now ready to state our general upper bound result. The algorithm and proof follow the same lines as the $k = 2$ case.

Theorem 3.2 (Restatement of Theorem 1.1)**.** *Let H be a digraph of constant size with k source components. The quantum query complexity of testing H-freeness of an N-vertex graph with bounded out-degree in the unidirectional model is $O\Big(N^{\frac{1}{2}\left(1-\frac{1}{2^k-1}\right)}\Big)$.*

Proof. In order to extend the $k = 2$ case described above to larger k, we first try to find many partial H-instances with $k-1$ source components found, and then extend one of them to a complete H-instance. We present a brief description of our algorithm below, where m is the number of vertices of H:

1. Randomly sample a vertex subset $\mathcal{S}_1$ in G of size t_1. Perform a depth-m BFS from every vertex $s \in \mathcal{S}_1$.

2. For iterations $i = 2$ to $k-1$, do the following:

   (a) Perform a Grover search t_i times on the vertices $V \setminus \mathcal{S}_{i-1}$ to find a set $\mathcal{S}_i$ of t_i many partial H-instances that cover i of its source components.

   (b) Set $\mathcal{S}_{i+1} = \mathcal{S}_{i-1} \cup \mathcal{S}_i$.

3. Perform Grover search on $V \backslash \mathcal{S}_k$ to find a complete H-instance.

4. If any occurrence of H in G is found, output **Reject** and terminate the algorithm. Otherwise, output **Accept**.

The correctness proof is same as the $k = 2$ case, and is omitted. To bound the query complexity, note that the first Grover search finds t_2 partial H-instances with 2 of its source components found, which takes $O(t_2\sqrt{N/t_1})$ queries. Similarly, for i-th iteration, the algorithm performs $O(t_i\sqrt{N/t_{i-1}})$ quantum queries for every $i \in [k-1]$. Finally, finding one complete H-instance costs $O(\sqrt{N/t_{k-1}})$ queries. Thus the total query complexity is $O(t_1 + \sum_{i=1}^{k-1} t_{i+1}\sqrt{N/t_i})$ with $t_k = 1$. Similar to the multi-collision algorithm in [LZ19, Section 3], we can equate all terms by setting $t_i = \Theta\left(N^{\frac{2^{k-i}-1}{2^k-1}}\right)$, which yields the claimed quantum query complexity $O\left(N^{\frac{1}{2}\left(1-\frac{1}{2^k-1}\right)}\right)$. □

4 Quantum lower bound

As discussed in Section 2, we are going to prove a lower bound on the problem of testing k-collision-freeness.

Theorem 4.1 (Restatement of Theorem 1.4)**.** *Let $k \geq 3$ and $0 < \varepsilon < 1/(4^{k-1}\lceil 20(2k)^{k/2}\rceil)$ be constants. Let N be a large enough positive integer. Then the quantum query complexity of the ε-testing of k-collision-freeness of a sequence of integers $S = (s_1, \ldots, s_N) \in [N]^N$ with parameter ε is $\Omega(N^{1/2-1/(2k)}/\ln^2 N)$.*

The proof of the theorem is at the end of Section 4.3. Observe that Theorem 1.2 is implied by Theorem 4.1 and the reduction in Proposition 2.4. Our proof mostly follows the structure of [BKT20, Section 6.1], and in particular it uses the notion of dual polynomial for non-Boolean partial symmetric functions. Our main technical contribution in this section is the proof of Lemma 4.18, because the corresponding proof in [BKT20] crucially relies on a fact that does not hold for our problem. We will discuss it in detail below.

In the following, we first state some general results related to the polynomial method for non-Boolean functions, then we use these results for our problem to state the exact statement that we prove in the technical part.

4.1 The (dual) polynomial method

For Boolean functions We consider a property on Boolean vectors as a function $f : D \subseteq \{-1,1\}^n \to \{-1,1\}$. Since the work of [BBC+01] it has been known that the acceptance probability $p(x)$ of a T-query bounded-error quantum algorithm on input $x \in D$ is a polynomial of degree at most $2T$. Thus, the polynomial $1 - 2p(x)$ must be a good approximation of f. Observe that $1 - 2p(x)$ remains bounded outside D since $p(x)$ remains a probability defined by the algorithm, with no constraint.

In order to formalize this, we first define the notion of approximate degree of a Boolean function, and then relate it to its query complexity.

Definition 4.2 (Approximate bounded degree)**.** *Let $f : D \subseteq \{-1,1\}^n \to \{-1,1\}$ and $\delta > 0$. A polynomial $p : \{-1,1\}^n \to \mathbb{R}$ δ-approximates f on D if*

$$\forall x \in D : \ |f(x) - p(x)| < \delta \quad \text{and} \quad \forall x \in \{-1,1\}^n \setminus D : \ |p(x)| < 1 + \delta.$$

Moreover, the δ-approximate bounded degree $\mathrm{bdeg}_\delta(f)$ of f on D is the smallest degree of such a polynomial.

Lemma 4.3 ([BBC+01, AAI+16]). *Let $f : D \subseteq \{-1,1\}^n \to \{-1,1\}$ and $\delta > 0$. If a quantum algorithm computes f on D with error δ using T queries then there is a polynomial p of degree at most $2T$ that 2δ-approximates f on D.*

In particular, this implies that the quantum query complexity for computing f with error δ is $\mathrm{bdeg}_{2\delta}(f)/2$, and so we will focus on proving lower bounds on the approximate bounded degree.

We now turn to a dual characterization of this polynomial approximation. This method of dual polynomials dates back to [She11, SZ09] for initially studying communication complexity. Below we refer to some results stated in [BKT20] for studying query complexity.

Definition 4.4 (Pure high degree). *A function $\psi : \{-1,1\}^n \to \mathbb{R}$ has pure high degree at least Δ if for every polynomial $p : \{-1,1\}^n \to \mathbb{R}$ with $\deg(p) < \Delta$ it satisfies $\sum_{x\in\{-1,1\}^n} p(x)\psi(x) = 0$. We denote this as $\mathrm{phd}(\psi) \geq \Delta$.*

One can observe that $\mathrm{phd}(\psi) \geq \Delta$ is equivalent to the fact that all the monomials of ψ are of degree at least Δ. Then by weak LP duality we get the following result.

Theorem 4.5. *[BKT20, Proposition 2.3] Let $f : D \subseteq \{-1,1\}^n \to \{-1,1\}$ and $\delta > 0$. Then $\mathrm{bdeg}_\delta(f) \geq \Delta$ iff there exists a function $\psi : \{-1,1\}^n \to \mathbb{R}$ such that*

$$\sum_{x\in D} \psi(x)f(x) - \sum_{x\in\{-1,1\}^n\setminus D} |\psi(x)| > \delta; \tag{1}$$

$$\|\psi\|_1 = \sum_{x\in\{-1,1\}^n} |\psi(x)| = 1; \tag{2}$$

$$\mathrm{phd}(\psi) \geq \Delta. \tag{3}$$

Now we are going to discuss how to extend these results to non-boolean functions, which is the interesting case for us.

For non-Boolean partial symmetric functions We now consider a property of a sequence of integers as a function $F : D \subseteq [R]_0^N \to \{-1,1\}$. The symbol 0 will play a special role that will be exhibited later on. Unfortunately one cannot just take the polynomial of those integers. The standard approach (see [Aar02]) is to encode $s = (s_1, \ldots, s_N) \in [R]_0^N$ into binary variables $x = (x_{i,j})_{i\in[N],j\in[R]_0} \in \{-1,1\}^{N(R+1)}$ encoding whether $s_i = j$ as follows: $x_{i,j} = -1$ if $s_i = j$, and $x_{i,j} = 1$ otherwise. Let $H_b^{N(R+1)} \subseteq \{-1,1\}^{N(R+1)}$ be the set of all possible encodings of vectors s, that is for every $i \in [N]$ there is exactly one $j \in [R]_0$ such that $x_{i,j} = -1$.

This way we can represent F as a function $F_b : D_b \to \{-1,1\}$ where $D_b \subseteq H_b^{N(R+1)}$ is the set of valid encodings of D. More precisely, each $x \in D_b$ satisfies two constraints: (1) $x \in H_b^{N(R+1)}$; and (2) x encodes some $s \in D$. Since only inputs $x \in H_b^{N(R+1)}$ correspond to possible input sequences of an algorithm, the polynomials derived from a quantum query algorithm might not be bounded outside of that set. This implies a slight modification on the definition of approximate degree, in order to relate it to query complexity as in [Aar02].

But before doing this, we are going to relax the constraints on the domain D_b in the case of symmetric functions, while we decrease its dimension. When F is *symmetric* (i.e. $F(s) = F(s \circ \pi_N)$ for any permutation π_N of $[N]$), one can instead define a function $F_{\leq N}$ with weaker constraints by removing the variables corresponding to the symbol 0. Define $H_{\leq N}^{NR}$ as the set of length-(NR)

binary vectors with Hamming weight at most N. Given any $x \in H^{NR}_{\leq N}$, we define its frequency vector $z(x) = (z_0, z_1, \ldots, z_R)$ with $z_j = \#\{i : x_{ij} = -1\}$, for $1 \leq j \leq R$, and $z_0 = N - z_1 - \ldots - z_R$. From the vector $z(x)$, one can define a valid sequence of integers $s(x) \in [R]^N_0$: it can be any sequence from $[R]^N_0$ that has frequency vector $z(x)$. Now we can define $F_{\leq N}$ on domain $D_{\leq N}$ as

$$D_{\leq N} = \{x \in H^{NR}_{\leq N} : s(x) \in D\} \quad \text{and} \quad F_{\leq N}(x) = F(s(x)).$$

In fact, for the special case of total symmetric functions F, we can transform F_b on $H^{N(R+1)}_b$ to $F_{\leq N}$ on $H^{NR}_{\leq N}$ due to the symmetry of F.

In [Amb05] it was proved implicitly that for symmetric F, both F_b and $F_{\leq N}$ variants are equally hard to approximate by polynomials. We now define the appropriate notion of approximate degree for $F_{\leq N}$ and relate it to the query complexity of F as in [BKT20, Theorem 6.5].

Definition 4.6 (Double-promise approximate degree)**.** *Let $F : D \subseteq [R]^N_0 \to \{-1, 1\}$ be symmetric and $\delta > 0$. Define $H^{NR}_{\leq N} \subseteq \{-1, 1\}^{NR}$ and $F_{\leq N} : D_{\leq N} \subseteq H^{NR}_{\leq N} \to \{-1, 1\}$ as above. A polynomial $p : \{-1, 1\}^{NR} \to \mathbb{R}$* double-promise δ-approximates *F on D if*

$$\forall x \in D_{\leq N} : \ |F_{\leq N}(x) - p(x)| < \delta \quad \text{and} \quad \forall x \in H^{NR}_{\leq N} \setminus D_{\leq N} : \ |p(x)| < 1 + \delta.$$

Moreover, the double-promise δ-approximate degree $\mathrm{dpdeg}_\delta(F_{\leq N})$ *of $F_{\leq N}$ on $D_{\leq N}$ is the smallest degree of such a polynomial.*

Lemma 4.7 ([Aar02, Amb05],[BKT20, Theorem 6.5])**.** *Let $F : D \subseteq [R]^N_0 \to \{-1, 1\}$ be symmetric and $\delta > 0$. Define $H^{NR}_{\leq N} \subseteq \{-1, 1\}^{NR}$ and $F_{\leq N} : D_{\leq N} \subseteq H^{NR}_{\leq N} \to \{-1, 1\}$ as above. If a quantum algorithm computes F on D with error δ using T queries then there is a polynomial p of degree at most $2T$ that double-promise 2δ-approximates $F_{\leq N}$ on $D_{\leq N}$.*

As for the Boolean case, this implies that a quantum algorithm computing F with error δ must make at least $\mathrm{dpdeg}_{2\delta}(F_{\leq N})/2$ queries. We can now also take the dual of this characterization.

Theorem 4.8 ([BKT20, Proposition 6.6])**.** *Let $F : D \subseteq [R]^N_0 \to \{-1, 1\}$ be symmetric. Define $F_{\leq N} : D_{\leq N} \to \{-1, 1\}$ as above. Then $\mathrm{dpdeg}_\delta(F_{\leq N}) \geq \Delta$ iff there exists a function $\psi : \{-1, 1\}^{NR} \to \mathbb{R}$ such that*

$$\forall x \in \{-1, 1\}^{NR} \setminus H^{NR}_{\leq N}, \quad \psi(x) = 0; \tag{4}$$

$$\sum_{x \in D_{\leq N}} \psi(x) F^{\leq N}(x) - \sum_{x \in H^{NR}_{\leq N} \setminus D_{\leq N}} |\psi(x)| > \delta; \tag{5}$$

$$\|\psi\|_1 = 1 \quad \text{and} \quad \mathrm{phd}(\psi) \geq \Delta. \tag{6}$$

4.2 Preparation

Technically, the problem we use in the proof of Theorem 4.1 is slightly more restricted than k-collision-freeness: we want to distinguish no k-collision from many distinct collisions of size at least k.

Definition 4.9 (Collision function)**.** *Let $\gamma \in (0, 1)$. The symmetric function $\mathrm{Collision}^{k,\gamma}_{N,R} : D_{\mathrm{Collision}^{k,\gamma}_{N,R}} \subset [R]^N \to \{-1, 1\}$ is defined by $\mathrm{Collision}^{k,\gamma}_{N,R}(s) = -1$ if no integer occurs at least k times in s, $\mathrm{Collision}^{k,\gamma}_{N,R}(s) = 1$ if there are more than γR distinct integers that occur at least k times in s, and it is undefined otherwise.*

Notice that this problem is not a property testing problem, as the outcome is not determined based on the distance between inputs. Nevertheless, it is a valid promise problem and a special case of testing k-collision-freeness, that we use to prove a lower bound on the other problems of interest.

To prove a bound on Collision, we will actually relate it to the composition of two more elementary functions, where by composition we mean $(f \circ g)(x) = f(g(x_1), \ldots, g(x_n))$, for appropriate functions f and g. Define (i) the threshold function $\mathrm{THR}_N^k : \{-1,1\}^N \to \{-1,1\}$ which is -1 if the input bitstring contains at least k many -1s, and it is 1 otherwise; and (2) the gap version of OR, that is $\mathrm{GapOR}_R^\gamma : D_{\mathrm{GapOR}_R^\gamma} \subset \{-1,1\}^R \to \{-1,1\}$ which takes value 1 if the input is 1^R, -1 if the input contains at least γR many -1s, and is undefined otherwise. We show that the double-promise approximate degree of $\mathrm{GapOR}_R^\gamma \circ \mathrm{THR}_N^k$ lower bounds the quantum query complexity of the collision problem.

Lemma 4.10. *Let $k \geq 3$, $0 < \gamma < 1$, $\delta > 0$ and $c > 2$ be constants such that $N/c \leq R \leq N/2$. If the double-promise δ-approximate degree of $\mathrm{GapOR}_R^\gamma \circ \mathrm{THR}_N^k$ on domain further restricted to $H_{\leq N}^{NR}$ is at least Δ, then every quantum algorithm computing $\mathrm{Collision}_{N,N}^{k,\gamma/c}$ with error $\delta/2$ must require at least $\Delta/2$ queries.*

Before proving this lemma, we prove some helper propositions. In order to apply the dual polynomial method for partial symmetric functions, we start by proving that $\mathrm{Collision}_{N,R'}^{k,\gamma'}$ is at least as hard as a very similar problem. We introduce a "dummy-augmented" version $\mathrm{dCollision}_{N,R}^{k,\gamma} : D_{\mathrm{dCollision}_{N,R}^{k,\gamma}} \subseteq [R]_0^N \to \{-1,1\}$ of the problem $\mathrm{Collision}_{N,R}^{k,\gamma}$ for the purpose of proving Lemma 4.10, where now the input sequence can have integer 0, but those 0s are just ignored when they occur. We show that it is enough to prove a lower bound for this second version.

Proposition 4.11. *Let $k \geq 3$, $0 < \gamma < 1$ and $c > 2$ be constants such that $N/c \leq R \leq N/2$. Then $\mathrm{dCollision}_{N,R}^{k,\gamma}$ can be reduced to $\mathrm{Collision}_{N,N}^{k,\gamma/c}$.*

Proof. An input to $\mathrm{dCollision}_{N,R}^{k,\gamma}$ is a sequence $s = (s_1, \ldots, s_N)$ where each $s_i \in [R]_0$. Let us define a family of functions T_i that map from $[R]_0$ to $[R']$ for $R' = R + \lceil N/2 \rceil$: $T_i(s) = s$ if $s > 0$ and $T_i(0) = R + \lceil i/2 \rceil$.

Notice that $(s_1, \ldots, s_N)$ is free from k-collisions (ignoring collisions of the dummy character 0) if and only if $(T_1(s_1), \ldots, T_N(s_N))$ is free from k-collisions, i.e. new k-collisions cannot be created by this transformation (only 2-collisions but we assume $k \geq 3$).

On the other hand, if $(s_1, \ldots, s_N)$ contains more than γR distinct k-collisions, then so does $(T_1(s_1), \ldots, T_N(s_N))$. Since $\gamma R \geq (\gamma/c)N$, $\mathrm{Collision}_{N,N}^{k,\gamma/c}$ will reject. □

The following proposition relates dCollision to GapOR ∘ THR.

Proposition 4.12. *The domain of $\mathrm{GapOR}_R^\gamma \circ \mathrm{THR}_N^k$ is*

$$D_{\mathrm{GapOR}_R^\gamma \circ \mathrm{THR}_N^k} = \{x \in \{-1,1\}^{NR} : (\mathrm{THR}_N^k(x_1), \ldots, \mathrm{THR}_N^k(x_R)) \in H_{\geq \gamma R}^R \cup \{1^R\}\}.$$

where $x = (x_1, \ldots, x_R)$ with each $x_i \in \{-1,1\}^N$.

The domain of $(\mathrm{dCollision}_{N,R}^{k,\gamma})^{\leq N}$ is

$$D_{(\mathrm{dCollision}_{N,R}^{k,\gamma})^{\leq N}} = H_{\leq N}^{NR} \cap D_{\mathrm{GapOR}_R^\gamma \circ \mathrm{THR}_N^k}.$$

Moreover, restricted to the latter domain they are the same function:

$$(\mathrm{dCollision}^{k,\gamma}_{N,R})^{\leq N} = \mathrm{GapOR}^{\gamma}_{R} \circ \mathrm{THR}^{k}_{N}.$$

We are now ready to give the proof of Lemma 4.10.

Proof of Lemma 4.10. Using Proposition 4.11, instead of $\mathrm{Collision}^{k,\gamma/c}_{N,N}$ we can consider $\mathrm{dCollision}^{k,\gamma}_{N,R}$ (with the appropriate parameters) to show a lower bound. By Proposition 4.12, we can use Lemma 4.7 to relate the query complexity of $\mathrm{dCollision}^{k,\gamma}_{N,R}$ to the double-promise degree of $\mathrm{GapOR}^{\gamma}_{R} \circ \mathrm{THR}^{k}_{N}$ with domain further restricted to $H^{NR}_{\leq N}$. □

4.3 Main lower bound

Let us fix $f = (\mathrm{GapOR}^{\gamma}_{R} \circ \mathrm{THR}^{k}_{N})$ with domain $D = D_{(\mathrm{GapOR}^{\gamma}_{R} \circ \mathrm{THR}^{k}_{N})}$ (See Proposition 4.12). For technical reasons, in the rest of the section we fix $k \geq 3$ and $N = \lceil 20(2k)^{k/2} \rceil R$.[1]

We first define a construction used in order to compose dual polynomials.

Definition 4.13 (Dual block composition)**.** *The dual block composition of two functions $\phi : \{-1,1\}^n \to \mathbb{R}$ and $\psi : \{-1,1\}^m \to \mathbb{R}$ is a function $\phi \star \psi : \{-1,1\}^{nm} \to \mathbb{R}$ defined as*

$$(\phi \star \psi)(x) = 2^n \, \phi(\mathrm{sgn}(\psi(x_1)), \ldots, \mathrm{sgn}(\psi(x_n))) \prod_{i \in [n]} |\psi(x_i)|$$

where $x = (x_1, \ldots, x_n)$ and $x_i \in \{-1,1\}^m$, for $i \in [n]$.

This subsection is dedicated to the proof of the following lemma which, together with Lemma 4.10, implies Theorem 4.1. Observe that we have to zero out the support of the dual polynomial outside of $H^{NR}_{\leq N}$, since our target domain is not D but $D \cap H^{NR}_{\leq N}$ in Lemma 4.10.

Lemma 4.14. *Let $N = \lceil 20(2k)^{k/2} \rceil R$ and $0 < \gamma < 1/4^{k-1}$. Then there exists a function $\zeta : \{-1,1\}^{NR} \to \mathbb{R}$ such that*

$$\forall x \in \{-1,1\}^{NR} \setminus H^{NR}_{\leq N}, \quad \zeta(x) = 0; \tag{7}$$

$$\sum_{x \in H^{NR}_{\leq N} \cap D} \zeta(x) f(x) - \sum_{x \in H^{NR}_{\leq N} \setminus D} |\zeta(x)| > 2/3; \tag{8}$$

$$\|\zeta\|_1 = 1 \quad \text{and} \quad \mathrm{phd}(\zeta) \in \Omega\left(\sqrt{N^{1-1/k}} / \ln^2 N\right). \tag{9}$$

Proof. The construction of ζ starts by *block composing* (Definition 4.13) two dual polynomials ϕ, ψ, one for $\mathrm{GapOR}^{\gamma}_{R}$ and one for THR^{k}_{N}. The dual polynomial ϕ for $\mathrm{GapOR}^{\gamma}_{R}$ is given by Proposition 4.15. The dual polynomial ψ for THR^{k}_{N} is given by the first part of Lemma 4.16.

The block composition $\phi \star \psi$ is a good candidate for the dual polynomial of f. Indeed, Lemma 4.18 shows that it satisfies Equation (8), showing correlation at least $9/10 > 2/3$. One could also check that it satisfies Equation (9). Nonetheless it does not satisfy Equation (7).

We can now use the second part of Lemma 4.16 to argue that there exists another dual polynomial ζ that satisfies Equation (7) and Equation (9). Moreover, this ζ is close to $\phi \star \psi$ so that it also satisfies Equation (8), with the weaker but sufficient correlation $9/10 - 2/9 > 2/3$. This concludes the proof. □

[1] These parameters are used in [BKT20] to prove Proposition A.4, which we will use.

As we have seen, the previous proof relies on the following results. The first one is direct and we omit its proof.

Proposition 4.15. *Let $\phi : \{-1,1\}^R \to \mathbb{R}$ be such that $\phi(-1^R) = -1/2$, $\phi(1^R) = 1/2$, and $\phi(z) = 0$ for all $z \in \{-1,1\}^R \setminus \{-1^R, 1^R\}$. Then $\|\phi\|_1 = 1$, $\mathrm{phd}(\phi) \geq 1$, and*

$$\sum_{x \in \{-1,1\}^R} \phi(x)\mathrm{OR}(x) = 1.$$

The second lemma is the rephrasing of several scattered results in [BKT20] that we unify into one statement for more clarity. For the sake of completeness, we explain how to prove it in Appendix A.2.

Lemma 4.16. *Let $N = \lceil 20(2k)^{k/2} \rceil R$ and $\phi : \{-1,1\}^R \to \mathbb{R}$ from Proposition 4.15. Then there exists $\psi : \{-1,1\}^N \to \mathbb{R}$ and $\zeta : \{-1,1\}^{NR} \to \mathbb{R}$ such that*

1. *$\|\psi\|_1 = 1$, $\mathrm{phd}(\psi) \geq c_1 \sqrt{k^{-1} N^{1-1/k}}$,*

$$\sum_{x \in D_+} |\psi(x)| \leq \frac{1}{48N}, \quad \text{and} \quad \sum_{x \in D_-} |\psi(x)| \leq \frac{1}{2} - \frac{2}{4^k},$$

   where $D_+ = \{x \in \{-1,1\}^N : \psi(x) > 0, \mathrm{THR}_N^k(x) = -1\}$ and $D_- = \{x \in \{-1,1\}^N : \psi(x) < 0, \mathrm{THR}_N^k(x) = 1\}$.

2. *$\|\zeta\|_1 = 1$, $\mathrm{phd}(\zeta) = \Omega(\sqrt{N^{1-1/k}}/\ln^2 N)$, $\|\zeta - \phi \star \psi\|_1 \leq 2/9$, and $\zeta(x) = 0$ for all $x \in \{-1,1\}^{NR} \setminus H^{NR}_{\leq N}$.*

For the next lemma, we will use the following proposition, which was implicitly used in the proofs of [BKT20, Propositions 5.5 and 5.6], but not stated in this general form. We include its proof in Appendix A.3. By convention, we denote $D_{+1} = D_+$ and $D_{-1} = D_-$.

Proposition 4.17. *Let $S \subseteq \{-1,1\}^{NR}$. Let $g : \{-1,1\}^R \to \{-1,1\}$, $h : \{-1,1\}^N \to \{-1,1\}$, $\phi : \{-1,1\}^R \to \mathbb{R}$. Let $\psi : \{-1,1\}^N \to \mathbb{R}$ be such that $\|\psi\|_1 = 1$ and $\sum_{x \in \{-1,1\}^N} \psi(x) = 0$. Then the following hold.*

1. *When λ denotes the probability mass function $\lambda(u) = |\psi(u)|$:*

$$\sum_{x \in S} |(\phi \star \psi)(x)| = \sum_{z \in \{-1,1\}^R} |\phi(z)| \cdot \Pr_{x \sim \lambda^{\otimes R}}[x \in S | (\dots, \mathrm{sgn}(\psi(x_i)), \dots) = z];$$

2. *When $\mu_i^{z_i}$ denotes the probability mass function on $\{-1,1\}$ (parameterized by $z_i \in \{-1,1\}$) such that $\mu_i^{z_i}(-1) = 2\sum_{x \in D_{z_i}} |\psi(x)|$, and $\mu = \mu^z = \mu_1^{z_1} \otimes \dots \otimes \mu_R^{z_R}$ the independent product distribution on $\{-1,1\}^R$:*

$$\sum_{x \in \{-1,1\}^{NR}} (\phi \star \psi)(x) \cdot (g \circ h)(x) = \sum_{z \in \{-1,1\}^R} \phi(z) \cdot \mathop{\mathbb{E}}_{y \sim \mu}[g(\dots, y_i z_i, \dots)].$$

Finally we are ready to prove the last missing statement, which is our main technical contribution to this part. The proof of [BKT20, Lemma 6.9] does not apply directly to this problem: they use the fact that the dual polynomial ψ of their inner function (OR) has one sided error, which is not the case here.

As now we focus on the composed function f (and the dual composition $\phi \star \psi$), the domain is not restricted to small Hamming weight inputs anymore.

Lemma 4.18. *Let $N = \lceil 20(2k)^{k/2} \rceil R$ and $0 < \gamma < 1/4^{k-1}$. Functions ϕ from Proposition 4.15 and ψ from Lemma 4.16 satisfy*

$$\sum_{x \in D} (\phi \star \psi)(x) \cdot f(x) - \sum_{x \in \{-1,1\}^{NR} \setminus D} |(\phi \star \psi)(x)| \geq 9/10.$$

Proof. We rewrite the left hand side by manipulating the sets we consider in the sums, and then we will bound separately the terms we get.

$$\begin{aligned}
&\sum_{x \in D} (\phi \star \psi)(x) \cdot f(x) - \sum_{x \in \{-1,1\}^{NR} \setminus D} |(\phi \star \psi)(x)| \\
&= \sum_{x \in \{-1,1\}^{NR}} (\phi \star \psi)(x) \cdot (\mathrm{OR} \circ \mathrm{THR}_N^k)(x) \\
&\quad - \left(\sum_{x \in \{-1,1\}^{NR} \setminus D} (\phi \star \psi)(x) \cdot (\mathrm{OR} \circ \mathrm{THR}_N^k)(x) + \sum_{x \in \{-1,1\}^{NR} \setminus D} |(\phi \star \psi)(x)| \right) \\
&\geq \sum_{x \in \{-1,1\}^{NR}} (\phi \star \psi)(x) \cdot (\mathrm{OR} \circ \mathrm{THR}_N^k)(x) - 2 \sum_{x \in \{-1,1\}^{NR} \setminus D} |(\phi \star \psi)(x)|
\end{aligned}$$

We first lower bound the first term.

Claim 4.19.

$$\sum_{x \in \{-1,1\}^{NR}} (\phi \star \psi)(x) \cdot (\mathrm{OR} \circ \mathrm{THR}_N^k)(x) \geq 1 - e^{-\frac{R}{4^{k-1}}} - \frac{R}{48N}.$$

Proof of claim. Using Item 2 of Proposition 4.17, the left hand side can be written as

$$\sum_{z \in \{-1,1\}^R} \phi(z) \cdot \mathop{\mathbb{E}}_{y \sim \mu} [\mathrm{OR}(\dots, y_i z_i, \dots)].$$

Recall that $\phi(z) = 0$ when z is anything but -1^R or 1^R, so only two terms are left to study.

If $z = -1^R$, using Item 1 of Lemma 4.16 each y_i is -1 with probability $\leq 1 - 1/4^{k-1}$ and 1 with probability $\geq 1/4^{k-1}$. If there is any $y_i = 1$ then the value of the OR is still -1. The probability of this event is $\geq 1-(1-1/4^{k-1})^R \geq 1-e^{-\frac{R}{4^{k-1}}}$. So the expected value is $\leq (-1)(1-e^{-\frac{R}{4^{k-1}}})+e^{-\frac{R}{4^{k-1}}} = -1+2e^{-\frac{R}{4^{k-1}}}$. Since in this case $\phi(-1^R) = -1/2$, the contribution to the sum is at most $1/2-e^{-\frac{R}{4^{k-1}}}$.

If $z = 1^R$, then, using Item 1 of Lemma 4.16 again, each y_i is -1 with probability $\leq 1/(48N)$. If any y_i is -1 then the value of the OR becomes -1. The union bound tells us that the probability of this is $\leq R/(48N)$, so the expected value is at least $-R/(48N) + 1 - R/(48N) = 1 - R/(24N)$. Multiplied by $\phi(1^R) = 1/2$ the contribution is at least $1/2 - R/(48N)$. Thus, the first term can be lower bounded by $1 - e^{-\frac{R}{4^{k-1}}} - \frac{R}{48N}$. ◇

Now we bound the second term.

Claim 4.20.

$$2 \sum_{x \in \{-1,1\}^{NR} \setminus D} |(\phi \star \psi)(x)| < e^{-2R\left(\frac{1}{4^{k-1}} - \gamma\right)^2}.$$

Proof of claim. By Item 1 of Proposition 4.17 with $S = \{-1,1\}^{NR} \setminus D$, the term can be written as follows,

$$2 \sum_{x \in \{-1,1\}^{NR} \setminus D} |(\phi \star \psi)(x)| = 2 \sum_{z \in \{-1,1\}^{R}} |\phi(z)| \cdot \Pr_{x \sim \lambda^{\otimes R}}[x \notin D \mid (\dots, \mathrm{sgn}(\psi(x_i)), \dots) = z],$$

which, using that $|\phi(z)| = 1/2$ when z is -1^R or 1^R and 0 otherwise, collapses to

$$= \Pr_{x \sim \lambda^{\otimes R}}[x \notin D \mid (\dots, \mathrm{sgn}(\psi(x_i)), z \dots) = -1^R] + \Pr_{x \sim \lambda^{\otimes R}}[x \notin D \mid (\dots, \mathrm{sgn}(\psi(x_i)), \dots) = 1^R].$$

In order to bound these two terms we introduce 0/1-variables r_i and q_i, for $i \in [R]$, related to the false positive and false negative inputs. Define $r_i = 1$ if $\mathrm{THR}_N^k(x_i) = -1$ and $\mathrm{sgn}(\psi(x_i)) = 1$, and otherwise $r_i = 0$. Similarly, $q_i = 1$ if $\mathrm{THR}_N^k(x_i) = 1$ and $\mathrm{sgn}(\psi(x_i)) = -1$, and otherwise $q_i = 0$.

Let us focus on the first term. If we sample x_i from the conditional distribution $(\lambda | \mathrm{sgn}(\psi(x_i)) = 1)$, then $\Pr[r_i = 1] = \Pr[\mathrm{THR}_N^k(x_i) = -1 | \mathrm{sgn}(\psi(x_i)) = 1] = 2\sum_{x_i \in D_+} |\psi(x_i)| \leq 1/(24N)$ (in the last step we used Item 1 of Lemma 4.16). Thus we can upper bounded the probability that an input does not satisfy the promise of GapOR_R^γ (i.e. that it is not in D) knowing that all the predictions are 1. It means that it contains at least 1 but less than γR many -1s, so this many predictions are false positive, which can be expressed by the r_i variables. In the last step below we use the union bound.

$$\Pr\left[x \notin D \mid \forall i \in [R]\ \mathrm{sgn}(\psi(x_i)) = 1\right] = \Pr\left[1 \leq \sum_{i \in [R]} r_i < \gamma R\right] \leq \Pr\left[1 \leq \sum_{i \in [R]} r_i\right] \leq \frac{R}{24N}.$$

Similarly, for the second term, if we sample x_i from the conditional distribution $(\lambda | \mathrm{sgn}(\psi(x_i)) = -1)$, then $\Pr[q_i = 1] = \Pr[\mathrm{THR}_N^k(x_i) = 1 | \mathrm{sgn}(\psi(x_i)) = -1] = 2\sum_{x_i \in D_-} |\psi(x_i)| \leq 1 - \frac{1}{4^{k-1}}$ (for the last step we used Item 1 of Lemma 4.16 again).

Then, similarly to the first term, we can upper bound the probability. Now in the last step we use the Chernoff bound, which introduces the constraint $\gamma < \frac{1}{4^{k-1}}$.

$$\Pr\left[x \notin D \mid \forall i \in [R]\ \mathrm{sgn}(\psi(x_i)) = -1\right] \leq \Pr\left[(1-\gamma)R < \sum_{i \in [R]} q_i\right] < e^{-2R\left(\frac{1}{4^{k-1}} - \gamma\right)^2}.$$

$\diamond$

Putting together the two bounds, we obtain

$$\sum_{x \in D} (\phi \star \psi)(x) \cdot f(x) - \sum_{x \in \{-1,1\}^{NR} \setminus D} |(\phi \star \psi)(x)| \geq 1 - \frac{R}{16N} - e^{-\frac{R}{4^{k-1}}} - e^{-2R(\frac{1}{4^{k-1}} - \gamma)^2}.$$

When k and $1/4^{k-1} - \gamma$ are positive constants and $R \in \Theta(N)$, this is larger than 9/10 (for large enough N). $\square$

Finally, we can conclude the proof of Theorem 4.1.

Proof of Theorem 4.1. By Lemma 4.14 there is a dual polynomial for $\mathrm{GapOR}_R^\gamma \circ \mathrm{THR}_N^k$ of pure high degree $\Omega(\sqrt{N^{1-1/k}}/\ln^2 N)$ that is only supported on $H_{\leq N}^{NR}$. By Theorem 4.8 this means that the double-promise δ-approximate degree of $\mathrm{GapOR}_R^\gamma \circ \mathrm{THR}_N^k$ with domain restricted to $H_{\leq N}^{NR}$ is $\Omega(\sqrt{N^{1-1/k}}/\ln^2 N)$. Using Lemma 4.10 with $c = \lceil 20(2k)^{k/2} \rceil$ we obtain that the bounded-error quantum query complexity of $\mathrm{Collision}_{N,N}^{k,\gamma'}$ is $\Omega(\sqrt{N^{1-1/k}}/\ln^2 N)$ if $\gamma' = \gamma/c < 1/(4^{k-1}\lceil 20(2k)^{k/2} \rceil)$. This implies the same lower bound on testing k-collision-freeness with $\varepsilon = \gamma'$ as Collision is just a more restricted version of the same problem. □

5 Testing 3-colorability

In this section, we will prove that the problem of testing 3-colorability in bounded degree graphs remains maximally hard-to-test in the quantum setting. Our lower bound proof will roughly follow the same approach as that of [BOT02]. See [BY22, Section 5.6] also for a reference.

Theorem 5.1 (Restatement of Theorem 1.3)**.** *Let G be an unknown undirected N-vertex graph with maximum degree d, and $\varepsilon \in (0,1)$ be a parameter. Given quantum query access to G in the undirected bounded-degree graph model, in order to distinguish if G is 3-colorable, or if it is ε-far from being 3-colorable, $\Omega(N)$ quantum queries are necessary.*

Let us start with the notion of k-wise independent string.

Definition 5.2 (k-wise independent string)**.** *A string $s = (s_1, \ldots, s_N) \in \{0,1\}^N$ is said to be k-wise independent if for any set of k-indices $i_1, i_2 \ldots, i_k$, the probability of any particular assignment $(b_{i_1}, b_{i_2}, \ldots, b_{i_k}) \in \{0,1\}^k$ to the indices $i_1, i_2 \ldots, i_k$ is equal to $1/2^k$.*

In order to prove the above theorem, we will be using the following observation from [ADW22], which states that distinguishing between uniformly random string and ℓ-wise independent string, for an appropriate integer ℓ, is hard for quantum algorithms.

Proposition 5.3 (Fact 1 from [ADW22])**.** *The output distribution of a quantum algorithm making q queries to a uniformly random string is identical to the same algorithm making q queries to a $2q$-wise independent string.*

To prove the lower bound of 3-colorability, the authors in [BOT02] studied another problem called $\mathrm{E}(3,c)$LIN-2, a problem related to deciding the satisfiability of a system of linear equations. Then the authors designed a reduction to 3-colorability from $\mathrm{E}(3,c)$LIN-2, which finally proves the linear query complexity lower bound for testing 3-colorability. We will also follow a similar approach here. Let us first formally define the problem of $\mathrm{E}(3,c)$LIN-2, where below $\mathbb{F}_2$ denotes the field with two elements.

Definition 5.4 ($\mathrm{E}(3,c)$LIN-2)**.** *Let $\mathcal{E}$ be a system of linear equations with N variables from $\mathbb{F}_2$, where there are 3 variables in each equation, and each variable occurs in at most c equations. This system $\mathcal{E}$ is represented as a matrix and we have query access to its entries. Given a parameter $\alpha \in (0,1)$, the goal is to distinguish if $\mathcal{E}$ is satisfiable, or at least an α-fraction of the equations need to be modified to make $\mathcal{E}$ satisfiable.*

The authors in [BOT02] proved the following lemma, which states that there exists a system of linear equations (equivalently a matrix), such that any constant fraction of the rows of this matrix are linearly independent. The authors proved this using hypergraph constructions.

Lemma 5.5 (Theorem 8 from [BOT02])**.** *For every $c > 0$, there exists a $\delta > 0$ such that for every N, there exists a matrix $A \in \{0,1\}^{cN \times N}$ with cN rows and N columns such that the following conditions hold:*

1. *Each row of A has exactly three non-zero entries.*
2. *Each column of A has exactly $3c$ non-zero entries.*
3. *Every collection of $\delta \cdot N$ rows of A is linearly independent.*

Using the existence of the matrix A corresponding to Lemma 5.5, the authors in [BOT02] used Yao's minimax lower bound technique to prove a linear lower bound for testing $\mathrm{E}(3,c)$LIN-2. More formally, they designed a pair of hard-to-distinguish distributions D_{yes} and D_{no}, such that unless $\Omega(N)$ queries are performed, no algorithm can distinguish between them. Since we will be using the same approach, we formally define D_{yes} and D_{no} below.

5.1 Testing $\mathbf{E}(3,c)$LIN-2

Lemma 5.6. *Given a matrix A (similar to the matrix mentioned in Lemma 5.5), and query access to an unknown vector y, in order to distinguish if there exists another vector x such that $Ax = y$, or for any vector x, only a constant ε fraction of the constraints encoded by A are satisfied for some $\varepsilon \in (0,1)$, $\Omega(N)$ queries are necessary.*

Construction of hard distributions: The hard-to-distinguish distributions D_{yes} and D_{no} are as follows:

1. D_{yes}**:** Choose a vector $z \in \{0,1\}^N$ uniformly at random, and set $Ay = z$.
2. D_{no}**:** Choose two vectors $y, z \in \{0,1\}^N$ uniformly at random, independently of each other, and set $Ay = z$.

Now we have the following lemma describing the properties of D_{yes} and D_{no}.

Lemma 5.7.

1. *The system of linear equations corresponding to D_{yes} are satisfiable.*
2. *The system of linear equations corresponding to D_{no} are $(1/2 - \alpha)$-far from being satisfiable for every $\alpha > 0$.*

Proof. The system of linear equations corresponding to D_{yes} are satisfiable since we can set $y = z$. On the other hand, the system of linear equations corresponding to D_{no} is far from being satisfiable with high probability. □

Now let us state the following lemma from [BOT02], which states that the inputs drawn from D_{yes} and D_{no} will remain indistinguishable unless $\Omega(N)$ queries are performed.

Lemma 5.8 (Lemma 19 of [BOT02]). *For every $\alpha > 0$, there are constants c and $\delta > 0$ such that every algorithm that distinguishes satisfiable instances of $E(3,c)$LIN-2 with N variables from instances that are $(1/2-\alpha)$-far from satisfiable must have classical query complexity at least δN.*

The key insight that is used to prove the above lemma is that applying Lemma 5.5, any $\delta \cdot N$ rows of A are linearly independent, thus any subset of $\delta \cdot N$ entries of $Ay = z$ will look uniformly random. Hence z is "*k-wise independent*" with $k = \delta \cdot N$. We will not formally prove the above lemma here, please refer to [BOT02] for formal proof.

5.2 Quantum lower bound for testing E$(3,c)$LIN-2

Now we will use Proposition 5.3 to prove a linear query quantum lower bound for testing E$(3,c)$LIN-2. The idea is that unless any quantum algorithm performs $k/2$ queries, it is hard to distinguish a uniformly random vector from a k-wise independent vector. This implies an $\Omega(k) = \Omega(N)$ query lower bound in the quantum setting. Formally, we have a quantum variant of Lemma 5.8 stated below.

Lemma 5.9. *For every $\alpha > 0$ there are constants c and $\delta > 0$ such that every algorithm that distinguishes satisfiable instances of $E(3,c)$LIN-2 with N variables from instances that are $(1/2-\alpha)$-far from satisfiable must have quantum query complexity at least δN.*

As mentioned before, using the same reduction between E$(3,c)$LIN-2 and 3-colorability from [BOT02, Section 5], we conclude that $\Omega(N)$ quantum queries are necessary to test 3-colorability in the bounded degree model, thereby proving Theorem 5.1.

5.3 Other maximal hard-to-test problems

As we mentioned in the introduction, there are several other problems in the bounded degree graph model, which are maximally hard-to-test. Moreover, their lower bounds stem from similar ideas as the E$(3,c)$LIN-2 and 3-colorability lower bounds, as mentioned in [YI10a, Gol20]. Following the same path as in the previous subsection, we also obtain $\Omega(N)$ quantum query lower bounds for all these problems. For brevity, we only present the theorem statements below and omit their proofs.

Theorem 5.10 (Hamiltonian Path/Cycle). *Given quantum query access to an unknown undirected d-bounded degree N-vertex graph G for some integer d, and a parameter $\varepsilon \in (0,1)$, in order to distinguish if G has a (directed) Hamiltonian path/cycle or ε-far from having a (directed) Hamiltonian path/cycle, $\Omega(N)$ quantum queries are necessary.*

Theorem 5.11 (Approximating Independent Set/Vertex Cover size). *Given query access to an unknown undirected d-bounded degree N-vertex graph G for some integer d, and a parameter $\varepsilon \in (0,1)$, approximating the independent set size/vertex cover of G, $\Omega(N)$ quantum queries are necessary.*

6 Conclusion

We provided new algorithms and lower bounds in the so far unexplored field of quantum property testing of directed bounded degree graphs. On the way, we revisited the well-known problem of collision finding in a new, property testing setting.

In particular, we used the dual polynomial method to obtain the first step for adapting the proportional moments technique of [RRSS09] (for proving randomized lower bounds) to the quantum setting. Indeed, the classical lower bounds of [HS12] and [PW23] for testing subgraph-freeness use the results of [RRSS09] for a collision-related problem. Recently, the authors in [ABRW16] stated it as an open question if one could use a variant of the proportional moments technique of [RRSS09] for proving better quantum lower bounds. This remains an interesting open question, but we hope that this work will serve as a step towards obtaining this new quantum lower bound technique.

A last important point is the general framework of the dual polynomial method we made, which will hopefully initiate future research in the community.

References

[AAI+16] Scott Aaronson, Andris Ambainis, Janis Iraids, Martins Kokainis, and Juris Smotrovs. Polynomials, quantum query complexity, and grothendieck's inequality. In *Proceedings of the 31st Conference on Computational Complexity (CCC)*, pages 25:1–25:19, 2016.

[Aar02] Scott Aaronson. Quantum lower bound for the collision problem. In *Proceedings of 34th Annual ACM Symposium on Theory of Computing (STOC)*, pages 635–642, 2002.

[ABRW16] Andris Ambainis, Aleksandrs Belovs, Oded Regev, and Ronald de Wolf. Efficient quantum algorithms for (gapped) group testing and junta testing. In *Proceedings of the 27th annual Symposium on Discrete Algorithms (SODA)*, pages 903–922, 2016.

[ACL11] Andris Ambainis, Andrew M Childs, and Yi-Kai Liu. Quantum property testing for bounded-degree graphs. In *Proceedings of the 14th International workshop on Approximation, Randomization, and Combinatorial optimization: algorithms and techniques (APPROX-RANDOM)*, pages 365–376, 2011.

[ADW22] Simon Apers and Ronald De Wolf. Quantum speedup for graph sparsification, cut approximation, and laplacian solving. *SIAM Journal on Computing (SICOMP)*, 51(6):1703–1742, 2022.

[Amb05] Andris Ambainis. Polynomial degree and lower bounds in quantum complexity: Collision and element distinctness with small range. *Theory of Computing (TOC)*, 1(3):37–46, 2005.

[AS04] Scott Aaronson and Yaoyun Shi. Quantum lower bounds for the collision and the element distinctness problems. *Journal of the ACM (JACM)*, 51(4):595–605, 2004.

[AS19] Simon Apers and Alain Sarlette. Quantum fast-forwarding: Markov chains and graph property testing. *Quantum Information & Computation*, 19(3&4):181–213, 2019.

[BBC+01] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. Quantum lower bounds by polynomials. *Journal of the ACM (JACM)*, 48(4):778–797, 2001.

[BDCG+20] Shalev Ben-David, Andrew M Childs, András Gilyén, William Kretschmer, Supartha Podder, and Daochen Wang. Symmetries, graph properties, and quantum speedups.

In *Proceedings of the 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 649–660, 2020.

[BFNR08] Harry Buhrman, Lance Fortnow, Ilan Newman, and Hein Röhrig. Quantum property testing. *SIAM Journal on Computing (SICOMP)*, 37(5):1387–1400, 2008.

[BHT98] Gilles Brassard, Peter Høyer, and Alain Tapp. Quantum cryptanalysis of hash and claw-free functions. In *Proceedings of the 3rd Latin American Symposium on Theoretical Informatics (LATIN)*, pages 163–169, 1998.

[BKT20] Mark Bun, Robin Kothari, and Justin Thaler. The polynomial method strikes back: Tight quantum query bounds via dual polynomials. *Theory of Computing (TOC)*, 16(10):1–71, 2020.

[BOT02] Andrej Bogdanov, Kenji Obata, and Luca Trevisan. A lower bound for testing 3-colorability in bounded-degree graphs. In *Proceedings of the 43rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 93–102, 2002.

[BR02] Michael A Bender and Dana Ron. Testing properties of directed graphs: acyclicity and connectivity. *Random Structures & Algorithms (RSA)*, 20(2):184–205, 2002.

[BY22] Arnab Bhattacharyya and Yuichi Yoshida. *Property Testing - Problems and Techniques*. Springer, 2022.

[CPS16] Artur Czumaj, Pan Peng, and Christian Sohler. Relating two property testing models for bounded degree directed graphs. In *Proceedings of the 48th annual Symposium on Theory of Computing (STOC)*, pages 1033–1045, 2016.

[CY19] Hubie Chen and Yuichi Yoshida. Testability of homomorphism inadmissibility: Property testing meets database theory. In *Proceedings of the 38th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems (PODS)*, pages 365–382, 2019.

[FNY+20] Sebastian Forster, Danupon Nanongkai, Liu Yang, Thatchaphol Saranurak, and Sorrachai Yingchareonthawornchai. Computing and testing small connectivity in near-linear time and queries via fast local cut algorithms. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2046–2065, 2020.

[GGR98] Oded Goldreich, Shari Goldwasser, and Dana Ron. Property testing and its connection to learning and approximation. *Journal of the ACM (JACM)*, 45(4):653–750, 1998.

[Gol10] Oded Goldreich. Introduction to testing graph properties. *Property testing: current research and surveys*, pages 105–141, 2010.

[Gol20] Oded Goldreich. On testing hamiltonicity in the bounded degree graph model. In *Electronic Colloquium on Computational Complexity (ECCC)*, page 109, 2020.

[GR02] Oded Goldreich and Dana Ron. Property testing in bounded degree graphs. *Algorithmica*, 32:302–343, 2002.

[HLM17] Aram W Harrow, Cedric Yen-Yu Lin, and Ashley Montanaro. Sequential measurements, disturbance and property testing. In *Proceedings of the 28th Annual Symposium on Discrete Algorithms (SODA)*, pages 1598–1611, 2017.

[HS12] Frank Hellweg and Christian Sohler. Property testing in sparse directed graphs: Strong connectivity and subgraph-freeness. In *Proceedings of the 20th Annual European Symposium on Algorithms (ESA)*, pages 599–610. Springer, 2012.

[LZ19] Qipeng Liu and Mark Zhandry. On finding quantum multi-collisions. In *Proceedings of the 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT)*, pages 189–218, 2019.

[MdW16] Ashley Montanaro and Ronald de Wolf. A survey of quantum property testing. *Theory of Computing (TOC)*, 7:1–81, 2016.

[MTZ20] Nikhil S. Mande, Justin Thaler, and Shuchen Zhu. Improved Approximate Degree Bounds for k-Distinctness. In *Proceedings of the 15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC)*, pages 2:1–2:22, 2020.

[OR11] Yaron Orenstein and Dana Ron. Testing eulerianity and connectivity in directed sparse graphs. *Theoretical Computer Science (TCS)*, 412(45):6390–6408, 2011.

[PW23] Pan Peng and Yuyang Wang. An optimal separation between two property testing models for bounded degree directed graphs. In *Proceedings of the 50th International Colloquium on Automata, Languages, and Programming (ICALP)*, pages 96:1–96:16, 2023.

[RRSS09] Sofya Raskhodnikova, Dana Ron, Amir Shpilka, and Adam Smith. Strong lower bounds for approximating distribution support size and the distinct elements problem. *SIAM Journal on Computing (SICOMP)*, 39(3):813–842, 2009.

[She11] Alexander A. Sherstov. The pattern matrix method. *SIAM Journal on Computing (SICOMP)*, 40(6):1969–2000, 2011.

[SZ09] Yaoyun Shi and Yufan Zhu. Quantum communication complexity of block-composed functions. *Quantum Information & Computation*, 9(5&6):444–460, 2009.

[YI10a] Yuichi Yoshida and Hiro Ito. Query-number preserving reductions and linear lower bounds for testing. *IEICE transactions on Information and Systems*, 93(2):233–240, 2010.

[YI10b] Yuichi Yoshida and Hiro Ito. Testing k-edge-connectivity of digraphs. *Journal of Systems Science and Complexity*, 23(1):91–101, 2010.

A Deferred material from Section 4

A.1 Dual polynomial for THR

Definition A.1. *Let $M \in \mathbb{N}$ and $\alpha, \beta > 0$. We say that a function $\omega : [M]_0 \to \mathbb{R}$ satisfies the (α, β)-decay condition if $\sum_{t\in[M]_0} \omega(t) = 0$, $\sum_{t\in[M]_0} |\omega(t)| = 1$ and $|\omega(t)| \leq \alpha e^{-\beta t}/t^2$.*

In [BKT20, Section 5.1] the authors define a dual polynomial ψ of THR_N^k in the following way. Let $k, N \in \mathbb{N}$, and T an integer such that $k \leq T \leq N$. Let $c = 2k\lceil N^{1/k} \rceil$ and $m = \lfloor \sqrt{T/c} \rfloor$. Define set $S = \{1, 2, \ldots, k\} \cup \{ci^2 : 0 \leq i \leq m\}$. Define a univariate polynomial

$$\omega(t) = \frac{(-1)^{t+T-m+1}}{T!} \binom{T}{t} \prod_{r \in [T]_0 \setminus S} (t - r).$$

Then let $\psi : \{-1, 1\}^N \to \mathbb{R}$ be $\psi(x) = \omega(|x|)/\binom{N}{|x|}$ for $x \in H^N_{\leq T}$ and $\psi(x) = 0$ otherwise.

They show that ψ and ω have the following properties.

Proposition A.2. *[BKT20, Proposition 5.4] Let ω and ψ be the polynomials defined above. Then the following are true.*

1. $\sum_{x \in D_+} |\psi(x)| \leq \frac{1}{48N}$;
2. $\sum_{x \in D_-} |\psi(x)| \leq \frac{1}{2} - \frac{2}{4^k}$;
3. $\|\psi\|_1 = 1$;
4. $\mathrm{phd}(\psi) \geq c_1 \sqrt{k^{-1} T N^{-1/k}}$;
5. *ω satisfies the (α, β)-decay condition with $\alpha = (2k)^k$ and $\beta = c_2/\sqrt{kTN^{1/k}}$.*

Here D_+ and D_- denote the set of false positives and that of false negatives respectively if ψ is considered as a hypothesis for THR_N^k, i.e. $D_+ = \{x \in \{-1, 1\}^N : \psi(x) > 0, \mathrm{THR}_N^k(x) = -1\}$ and $D_- = \{x \in \{-1, 1\}^N : \psi(x) < 0, \mathrm{THR}_N^k(x) = 1\}$.

A.2 Proof of Lemma 4.16

We start with two propositions from [BKT20] that we are going to use. The first one is about the properties of the dual block composition.

Proposition A.3. *[BKT20, Proposition 2.20] Let $\phi : \{-1, 1\}^n \to \mathbb{R}$, $\psi : \{-1, 1\}^m \to \mathbb{R}$. The dual block composition has the following properties.*

1. *If $\|\phi\|_1 = 1$, $\|\psi\|_1 = 1$ and $\langle \phi, 1^m \rangle = 0$ then $\|\phi \star \psi\|_1 = 1$.*
2. *If $\mathrm{phd}(\phi) \geq \Delta$ and $\mathrm{phd}(\psi) \geq \Delta'$ then $\mathrm{phd}(\phi \star \psi) \geq \Delta \cdot \Delta'$.*

The second one proves the existence of the final dual polynomial ζ, that is close to the "almost good" block composition $\phi \star \psi$, given that some conditions are satisfied.

Proposition A.4. *[BKT20, Proposition 2.22] Let $R \in \mathbb{N}$ sufficiently large and $M \leq R$. Let $\phi : \{-1, 1\}^R \to \mathbb{R}$ with $\|\phi\|_1 = 1$, and let $\omega : [M]_0 \to \mathbb{R}$ satisfy the (α, β)-decay condition with some $1 \leq \alpha \leq R^2$ and $4 \ln^2 R/(\sqrt{\alpha} R) \leq \beta \leq 1$. Let $N = \lceil 20\sqrt{\alpha} \rceil R$ and $\psi : \{-1, 1\}^N \to \mathbb{R}$ be defined as $\psi(x) = \omega(|x|)/\binom{N}{|x|}$. Let $\Delta < N$ be such that $\mathrm{phd}(\phi \star \psi) \geq \Delta$. Then there exist a $\Delta' \geq \beta\sqrt{\alpha} R/(4 \ln^2 R)$ and a function $\zeta : \{-1, 1\}^{NR} \to \mathbb{R}$ such that*

1. $\mathrm{phd}(\zeta) \geq \min\{\Delta, \Delta'\}$;

2. $\|\zeta - \phi \star \psi\|_1 \leq 2/9$;

3. $\|\zeta\|_1 = 1$;

4. $\forall x \in \{-1,1\}^{NR}$ *with* $|x| > N$ $\zeta(x) = 0$.

We now restate Lemma 4.16 before proving it.

Lemma 4.16. *Let* $N = \lceil 20(2k)^{k/2} \rceil R$ *and* $\phi : \{-1,1\}^R \to \mathbb{R}$ *from Proposition 4.15. Then there exists* $\psi : \{-1,1\}^N \to \mathbb{R}$ *and* $\zeta : \{-1,1\}^{NR} \to \mathbb{R}$ *such that*

1. $\|\psi\|_1 = 1$, $\mathrm{phd}(\psi) \geq c_1\sqrt{k^{-1}N^{1-1/k}}$,

$$\sum_{x \in D_+} |\psi(x)| \leq \frac{1}{48N}, \quad \text{and} \quad \sum_{x \in D_-} |\psi(x)| \leq \frac{1}{2} - \frac{2}{4^k},$$

where $D_+ = \{x \in \{-1,1\}^N : \psi(x) > 0, \mathrm{THR}_N^k(x) = -1\}$ *and* $D_- = \{x \in \{-1,1\}^N : \psi(x) < 0, \mathrm{THR}_N^k(x) = 1\}$.

2. $\|\zeta\|_1 = 1$, $\mathrm{phd}(\zeta) = \Omega(\sqrt{N^{1-1/k}}/\ln^2 N)$, $\|\zeta - \phi \star \psi\|_1 \leq 2/9$, *and* $\zeta(x) = 0$ *for all* $x \in \{-1,1\}^{NR} \setminus H^{NR}_{\leq N}$.

Proof. With the construction of ψ described in Appendix A.1, Proposition A.2 (with $T = N$) ensures that Item 1 is satisfied. This way, we know that $\|\psi\|_1 = 1$, and that $\mathrm{phd}(\psi) \geq c_1\sqrt{k^{-1}N^{1-1/k}}$. From Proposition 4.15 we know that $\|\phi\|_1 = 1$ and $\mathrm{phd}(\phi) \geq 1$. Using item 1 of Proposition A.3 we obtain $\|\phi \star \psi\|_1 = 1$, and using Item 2 we get $\mathrm{phd}(\phi \star \psi) \geq c_1\sqrt{k^{-1}N^{1-1/k}}$.

From Proposition A.2 we know that the function ω that is used to define ψ satisfies the (α, β)-decay condition for some constant $\alpha = (2k)^k$ and $\beta = c_2/\sqrt{kN^{1+1/k}}$.

This way, we can use Proposition A.4 to obtain the function ζ we wanted for Item 2. Indeed, our functions ψ and ϕ satisfy all the conditions of the lemma with pure high degree lower bounded by $\Delta = c_1\sqrt{k^{-1}N^{1-1/k}}$; and with our parameters α, β we obtain $\Delta' = c_2(2k)^{k/2}R/(4\ln^2(R)\sqrt{kN^{1+1/k}}) \in \Omega(\sqrt{N^{1-1/k}}/\ln^2 N)$.

□

A.3 Proof of Proposition 4.17

Let us restate the proposition that we are going to prove.

Proposition 4.17. *Let* $S \subseteq \{-1,1\}^{NR}$. *Let* $g : \{-1,1\}^R \to \{-1,1\}$, $h : \{-1,1\}^N \to \{-1,1\}$, $\phi : \{-1,1\}^R \to \mathbb{R}$. *Let* $\psi : \{-1,1\}^N \to \mathbb{R}$ *be such that* $\|\psi\|_1 = 1$ *and* $\sum_{x \in \{-1,1\}^N} \psi(x) = 0$. *Then the following hold.*

1. *When* λ *denotes the probability mass function* $\lambda(u) = |\psi(u)|$*:*

$$\sum_{x \in S} |(\phi \star \psi)(x)| = \sum_{z \in \{-1,1\}^R} |\phi(z)| \cdot \Pr_{x \sim \lambda^{\otimes R}}[x \in S | (\dots, \mathrm{sgn}(\psi(x_i)), \dots) = z];$$

2. *When $\mu_i^{z_i}$ denotes the probability mass function on $\{-1,1\}$ (parameterized by $z_i \in \{-1,1\}$) such that $\mu_i^{z_i}(-1) = 2\sum_{x \in D_{z_i}} |\psi(x)|$, and $\mu = \mu^z = \mu_1^{z_1} \otimes \ldots \otimes \mu_R^{z_R}$ the independent product distribution on $\{-1,1\}^R$:*

$$\sum_{x \in \{-1,1\}^{NR}} (\phi \star \psi)(x) \cdot (g \circ h)(x) = \sum_{z \in \{-1,1\}^R} \phi(z) \cdot \mathop{\mathbb{E}}_{y \sim \mu} [g(\ldots, y_i z_i, \ldots)].$$

Proof. Remember that λ denotes the probability mass function $\lambda(u) = |\psi(u)|$ for $u \in \{-1,1\}^N$. We will need the following claim.

Claim A.5.

$$\Pr_{u \sim \lambda}[\psi(u) > 0] = \Pr_{u \sim \lambda}[\psi(u) < 0] = \frac{1}{2}.$$

Proof of claim. We know that $\sum_u \psi(u) = 0$. Thus $\sum_{u:\psi(u)>0} |\psi(u)| - \sum_{u:\psi(u)>0} |\psi(u)|$. We then conclude using that $\|\psi\|_1 = 1$. ◇

First part of Proposition 4.17 Below, we first apply the definition of the dual block composition (and the fact that 2^R and $\prod_{i \in [R]} |\psi(x_i)|$ are positive). Then we use the definition of λ which ensures that $\prod_{i \in [R]} |\psi(x_i)|$ is the probability of getting $x = (\ldots, x_i, \ldots)$ when sampling independently R times from distribution λ.

$$\begin{aligned}
\sum_{x \in S} |(\phi \star \psi)(x)| &= 2^R \sum_{x \in \{-1,1\}^{NR}} \left(\prod_{i \in [R]} |\psi(x_i)| \right) \cdot |\phi(\ldots, \mathrm{sgn}(\psi(x_i)), \ldots)| \cdot \mathbb{I}[x \in S] \\
&= 2^R \cdot \mathbb{E}_{x \sim \lambda^{\otimes R}} \left[|\phi(\ldots, \mathrm{sgn}(\psi(x_i)), \ldots)| \cdot \mathbb{I}[x \in S] \right]
\end{aligned}$$

We introduce new variables z_i that will be compared to $\mathrm{sgn}(\psi(x_i))$. Using Claim A.5, the probability of picking a $z \in \{-1,1\}^R$ from the uniform distribution such that z corresponds to the vector of the signs is $\frac{1}{2^R}$. Thus previous term can be rewritten as

$$\begin{aligned}
& 2^R \sum_{z \in \{-1,1\}^R} |\phi(z)| \cdot \Pr_{x \sim \lambda^{\otimes R}}[x \in S \wedge (\ldots, \mathrm{sgn}(\psi(x_i)), \ldots) = z] \\
= & \sum_{z \in \{-1,1\}^R} |\phi(z)| \cdot \Pr_{x \sim \lambda^{\otimes R}}[x \in S \mid (\ldots, \mathrm{sgn}(\psi(x_i)), \ldots) = z]
\end{aligned}$$

which completes the proof.

Second part of Proposition 4.17 Remember that λ denotes the probability mass function $\lambda(u) = |\psi(u)|$ for $u \in \{-1,1\}^N$. Just like in the proof of the first item,

$$\sum_{x \in \{-1,1\}^{NR}} (\phi \star \psi)(x) \cdot (g \circ h)(x) = \sum_{z \in \{-1,1\}^R} \phi(z) \cdot \mathop{\mathbb{E}}_{x \sim \lambda^{\otimes R}} [(g \circ h)(x) \mid (\ldots, \mathrm{sgn}(\psi(x_i)), \ldots) = z] .$$

Using Claim A.5, we can first notice that for any $b \in \{-1,1\}$, the probability that an x_i sampled from λ is a false b (i.e. false positive if $b = 1$ and false negative if $b = -1$) is as follows, where by

convention $D_{+1} = D_+$ and $D_{-1} = D_-$:

$$\begin{aligned}\Pr_{x_i\sim\lambda}\left[h(x_i) \neq \operatorname{sgn}(\psi(x_i)) \mid \operatorname{sgn}(\psi(x_i)) = b\right] &= \sum_{x_i\in D_b} \Pr_{x_i\sim\lambda}\left[\text{sampling } x_i \mid \operatorname{sgn}(\psi(x_i)) = b\right] \\ &= 2\sum_{x_i\in D_b} |\psi(x_i)|.\end{aligned}$$

Therefore, if $z_i = \operatorname{sgn}(\psi(x_i))$ and x_i is a false z_i, it means that z_i should be flipped to get $h(x_i)$. Let $y_i \in \{-1, 1\}$ denote whether we flip z_i. As x_i is a false z_i with probability $2\sum_{x_i\in D_{z_i}} |\psi(x_i)|$, this is the probability with which we should flip z_i, i.e. the probability that $y_i = -1$.

Thus for any $z \in \{-1, 1\}^R$, the vector $(\dots, h(x_i), \dots)$ with $x \sim \lambda^{\otimes R}$ conditioned on $(\dots, \operatorname{sgn}(\psi(x_i)), \dots) = z$ is identically distributed with $(\dots, z_i y_i, \dots)$ where y_i are random bit-flips according to $\mu_i^{z_i}$: $y_i = -1$ with probability $2\sum_{x_i\in D_{z_i}} |\psi(x_i)|$ and $y_i = 1$ otherwise.

Now we can finish the proof:

$$\begin{aligned}&\sum_{z\in\{-1,1\}^R} \phi(z) \cdot \mathop{\mathbb{E}}_{x\sim\lambda^{\otimes R}} [(g \circ h)(x) \mid (\dots, \operatorname{sgn}(\psi(x_i)), \dots) = z] \\ = &\sum_{z\in\{-1,1\}^R} \phi(z) \cdot \mathop{\mathbb{E}}_{y\sim\mu} [g(\dots, z_i y_i, \dots)].\end{aligned}$$

□