Electronic Colloquium on Computational Complexity, Revision 4 of Report No. 72 (2015)

On Being Far from Far and on Dual Problems in Property Testing

Roei Tell *

October 19, 2017

Abstract

This work studies a new type of problems in property testing, called *dual problems*. For a set Π in a metric space and $\delta > 0$, denote by $\mathcal{F}_\delta(\Pi)$ the set of elements that are δ-far from Π. Then, in property testing, a δ-tester for Π is required to accept inputs from Π and reject inputs from $\mathcal{F}_\delta(\Pi)$. A natural *dual problem* is the problem of δ-testing the set of "no" instances, that is $\mathcal{F}_\delta(\Pi)$: A δ-tester for $\mathcal{F}_\delta(\Pi)$ needs to accept inputs from $\mathcal{F}_\delta(\Pi)$ and reject inputs that are δ-far from $\mathcal{F}_\delta(\Pi)$; that is, it rejects inputs from $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$. When $\Pi = \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ the dual problem is essentially equivalent to the original one, but this equality does not hold in general.

Many dual problems constitute appealing testing problems that are interesting by themselves. In this work we derive lower bounds and upper bounds on the query complexity of several classes of natural dual problems: These include dual problems of properties of functions (e.g., testing error-correcting codes and testing monotone functions), of properties of distributions (e.g., testing equivalence to a known distribution), and of various graph properties in the dense graph model and in the bounded-degree model. We also show that testing any dual problem with *one-sided error* is either trivial or requires a linear number of queries.

Keywords: Metric spaces, Property Testing, Dual Problems.

*Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel. Email: `roei.tell@weizmann.ac.il`

ISSN 1433-8092

Contents

1 Introduction

Let (Ω, Δ) be a metric space, [1] let $\Pi \subseteq \Omega$ be a set in this space, and let $\delta > 0$ be a distance parameter. A natural object that we are frequently interested in is the set of points in Ω that are δ-far from Π, denoted $\mathcal{F}_\delta(\Pi) = \{x \in \Omega : \Delta(x, \Pi) \geq \delta\}$. Viewing $\mathcal{F}_\delta$ as an operator on the power set of Ω, a natural question is what happens when applying the operator $\mathcal{F}_\delta$ twice; that is, what is the structure of sets of the form $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ for some $\Pi \subseteq \Omega$. One might mistakenly expect that for any metric space Ω, set $\Pi \subseteq \Omega$, and distance parameter $\delta > 0$ it holds that $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) = \Pi$. However, although it is always true that $\Pi \subseteq \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$, it is not necessarily true that $\Pi = \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$. Furthermore, in some spaces, most notably in the Boolean hypercube, the equality is even *typically* false (i.e., it is false for most subsets; see Section 1.1). In fact, the study of sets of the form $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ turns out to be quite complex. To the best of our knowledge, this basic question has not been explored so far.

The study of sets of the form $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ has an interesting application in theoretical computer science, specifically in the context of *property testing* (see, e.g., [Gol10]). In property testing, an ϵ-tester for $\Pi \subseteq \{0,1\}^n$ is required to accept every input in Π, with high probability, and reject every input in $\mathcal{F}_\delta(\Pi)$, with high probability, where $\delta = \epsilon \cdot n$ refers to absolute distance, and $\epsilon > 0$ refers to the relative distance. [2] This constitutes a promise problem, in which the set of "yes" instances is Π and the set of "no" instances is $\mathcal{F}_\delta(\Pi)$. One plausible question in this context is what is the relationship between the complexity of ϵ-testing the set of "yes" instances Π and the complexity of the *dual problem* of ϵ-testing the set of "no" instances $\mathcal{F}_\delta(\Pi)$. In many cases, the "far set" (i.e., $\mathcal{F}_\delta(\Pi)$) actually constitutes a natural property, making the corresponding dual problem an interesting testing problem by itself (see elaboration in Section 1.2).

For any set $\Pi \subseteq \{0,1\}^n$ and $\delta = \epsilon \cdot n$, an ϵ-tester for the dual problem of Π is required to accept every input in $\mathcal{F}_\delta(\Pi)$, with high probability, and reject every input in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$, with high probability. Indeed, if $\Pi = \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$, then the problem of ϵ-testing Π is essentially equivalent to its dual problem. We call such sets $\mathcal{F}_\delta$-closed:

Definition 1.1 *($\mathcal{F}_\delta$-closed sets). For a metric space Ω, a parameter $\delta > 0$, and a set $\Pi \subseteq \Omega$, if $\Pi = \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$, then we say that Π is $\mathcal{F}_\delta$-closed in Ω.*

However, as mentioned above, not all sets are $\mathcal{F}_\delta$-closed, and for some spaces and δ parameters, most sets are actually *not* $\mathcal{F}_\delta$-closed. Moreover, in many cases it is unfortunately non-obvious to determine whether Π is $\mathcal{F}_\delta$-closed or not.

Our contributions. In this work we introduce dual problems in property testing, motivate their study, and obtain results regarding their complexity. We show that in general, testing dual problems with *one-sided error* requires a linear number of queries, unless the problem

[1] Throughout the paper we will usually identify a metric space (Ω, Δ) with its set of elements Ω, in which case the metric is implicit and denoted by Δ.

[2] Being consistent with the property testing literature, we let $\epsilon > 0$ denote the relative (Hamming) distance. In contrast, it is more convenient to analyze the δ-far operator while referring to absolute distance (denoted by $\delta > 0$). Note that the abstract indeed used different notations, merely for simplicity of presentation.

is trivial to begin with; this stands in sharp contrast to testing standard problems with one-sided error. In addition, we determine the complexity of *several specific natural dual problems*, corresponding to well-known testing problems; these dual problems include:

- Testing whether a string is **far from being a codeword** in an error-correcting code.
- Testing whether a function is **far from being monotone**.
- Testing whether a distribution is **far from being uniform**.
- Testing whether a graph is **far from being k-colorable** in the dense graph model.
- Testing whether a graph is **far from being connected** in the bounded-degree model.
- Testing whether a graph is **far from being cycle-free** in the bounded-degree model.

Some of these dual problems are essentially equivalent to their original problems (i.e., the corresponding sets $\Pi_n \subseteq \{0,1\}^n$ are $\mathcal{F}_\delta$-closed, for $\delta = \epsilon \cdot n$; see Definition 1.3), and in these cases the query complexity of the dual is the same as the query complexity of the original. However, other dual problems mentioned above are different from the original problems (i.e., $\Pi_n \neq \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$), and sometimes even significantly different; in these cases we present a tester for the dual problem, which is different from known testers for the original problem, and sometimes also has higher query complexity. Beyond the immediate implications of these results (of determining the complexity of specific problems), their proofs typically also include structural results related to the relevant property.

Some of the specific aforementioned dual problems concern testing of *natural properties* that are of independent interest. Moreover, the general questions underlying the current work ("far-from-far" sets and dual testing problems) seem appealing both by themselves and as extensions (or "duals") of well-studied problems. In addition, as is the case when studying standard property testing questions, the study of dual problems is not only algorithmic, but in fact typically focuses on *structural features* of the property in question.

1.1 On the non-triviality of the notion of $\mathcal{F}_\delta$-closed sets

As mentioned above, one might mistakenly expect that for every Ω and δ, all sets will be $\mathcal{F}_\delta$-closed. Indeed, for any metric space Ω, taking a value of δ such that $\delta \leq \inf_{x \neq y \in \Omega}\{\Delta(x,y)\}$ ensures that all sets are trivially $\mathcal{F}_\delta$-closed, whereas taking a value of δ such that $\delta > \sup_{x,y}\{\Delta(x,y)\}$ ensures that all non-trivial subsets are *not* $\mathcal{F}_\delta$-closed (since any $\Pi \neq \emptyset$ satisfies $\mathcal{F}_\delta(\Pi) = \emptyset$ and $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) = \Omega$). However, for *any metric space* Ω and any δ in between these two values there exist both $\mathcal{F}_\delta$-closed sets and sets that are not $\mathcal{F}_\delta$-closed.

Theorem 1.2 *(non-triviality of the notion of $\mathcal{F}_\delta$-closed sets; see Theorem B.1). For any metric space Ω, if $\delta \in \left(\inf_{x \neq y}\{\Delta(x,y)\}, \sup_{x \neq y}\{\Delta(x,y)\}\right)$, then there exists a non-trivial $\Pi \subseteq \Omega$ that is $\mathcal{F}_\delta$-closed and a non-trivial $\Pi' \subseteq \Omega$ that is not $\mathcal{F}_\delta$-closed.*

In addition to the *existence* of sets that are not $\mathcal{F}_\delta$-closed, in some metric spaces such sets are actually the typical case, rather than the exception. Most notably, in the Boolean hypercube it holds that a $(1-o(1))$-fraction of the sets are not $\mathcal{F}_\delta$-closed. (This is the case since for a random set $\Pi \subseteq \{0,1\}^n$ and $\delta \geq 3$, with high probability it holds that $\mathcal{F}_\delta(\Pi) = \emptyset$.) Furthermore, in contrast to what one might expect, points in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ might not even be *close* to Π. In particular, there exist spaces Ω and sets $\Pi \subseteq \Omega$ such that some points in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) \setminus \Pi$ are relatively far from Π (i.e., almost δ-far from Π); such sets also exist in the Boolean hypercube (see Proposition C.1).

A study of sets of the form $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ appears in the appendices, and includes sufficient and/or necessary conditions for a set to be $\mathcal{F}_\delta$-closed, proofs of the existence and prevalence of sets that are not $\mathcal{F}_\delta$-closed, and results on the distance of points in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ from Π.

1.2 Dual problems in property testing

For a space $\Omega = \Sigma^n$, and a set $\Pi \subseteq \Sigma^n$, and $\epsilon > 0$, the standard property testing problem is the one of ϵ-testing Π, and the corresponding *dual problem* is the one of ϵ-testing $\mathcal{F}_{\epsilon \cdot n}(\Pi)$.

What is the meaning of dual testing problems? First, for *some* properties, the dual problem is an appealing property that is interesting by itself. Consider, for example, the set of distributions that are far from uniform, the set of functions that are far from monotone, or the set of graphs that are far from being connected. All these sets constitute natural properties, and one might be interested in testing them. Secondly, in general, for *every* property Π the dual problem is intuitively related to the original problem: It can be viewed as distinguishing between inputs that any ϵ-tester for Π must reject, and inputs that need to be significantly changed in order to be rejecetd by any ϵ-tester for Π. Thirdly, the *query complexity* of a testing problem and of its dual problem are related (see Observation 1.4).

Similar to standard testing problems, in dual problems we are also interested in the *asymptotic* complexity. That is, for a property $\Pi = \{\Pi_n\}_{n\in\mathbb{N}}$ such that $\Pi_n \subseteq \Sigma^n$, we seek either an asymptotic upper bound on the query complexity of ϵ-testing $\mathcal{F}_{\epsilon\cdot n}(\Pi_n)$ for *every* $\epsilon > 0$, or a lower bound for *some* value of $\epsilon > 0$. Accordingly, for a property $\Pi = \{\Pi_n\}_{n\in\mathbb{N}}$, we will refer to the *dual problem of the problem of testing* Π, or in short to *the dual problem of* Π.

Definition 1.3 *(dual problems that are equivalent to the original problems). For a set Σ, let $\Pi = \{\Pi_n\}_{n\in\mathbb{N}}$ such that $\Pi_n \subseteq \Sigma^n$. If for every sufficiently small $\epsilon > 0$ and sufficiently large n it holds that Π_n is $\mathcal{F}_{\epsilon\cdot n}$-closed, then the problem of testing Π is* equivalent to its dual problem. *Otherwise, the problem of testing Π is* different from its dual problem.

We stress that even if a standard testing problem Π is equivalent to its dual, it does not imply that the standard problem is the "dual problem of its dual". This is since the definition of dual problems is inherently different than that of standard problems, with respect to the dependence on the proximity parameter $\epsilon > 0$. In particular, in standard problems, the sets of "yes" instances $\{\Pi_n\}_{n\in\mathbb{N}}$ are fixed, and the sets of "no" instances $\{\mathcal{F}_{\epsilon\cdot n}(\Pi_n)\}_{n\in\mathbb{N}}$ depend on the proximity parameter $\epsilon > 0$; in contrast, in dual problems, both the sets of "yes" instances $\{\mathcal{F}_{\epsilon\cdot n}(\Pi_n)\}_{n\in\mathbb{N}}$ and the sets of "no" instances $\{\mathcal{F}_{\epsilon\cdot n}(\mathcal{F}_{\epsilon\cdot n}(\Pi_n))\}_{n\in\mathbb{N}}$ depend on ϵ.

The query complexity of any dual problem is closely related to the query complexity of its original problem. First, clearly, if the dual problem is equivalent to its original problem, then their query complexities are identical. Secondly, since for every set $\Pi \subseteq \Sigma^n$ and every $\delta > 0$ it holds that $\Pi \subseteq \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$, an ϵ-tester for $\mathcal{F}_{\epsilon \cdot n}(\Pi)$ always yields an ϵ-tester for Π, by complementing the output of the tester.

Observation 1.4 *(the query complexity of dual problems). The query complexity of a dual problem is lower bounded by the query complexity of its original problem.*

1.2.1 Testing dual problems with one-sided error

A preliminary general result is that testing dual problems with *one-sided error* requires a linear number of queries (regardless of whether or not the dual problem is equivalent to its original). Recall that in property testing, testers with one-sided error always accept "yes" inputs; in the case of dual problems, these are testers that always accept inputs from $\mathcal{F}_{\epsilon \cdot n}(\Pi)$.

Theorem 1.5 *(testing dual problems with one-sided error). For a set Σ, let $\Pi = \{\Pi_n\}_{n \in \mathbb{N}}$ such that $\Pi_n \subseteq \Sigma^n$. Suppose that for all sufficiently large n it holds that $\Pi_n \neq \emptyset$ and that there exist inputs that are $\Omega(n)$-far from Π_n. Then, the query complexity of testing the dual problem of Π with* one-sided error *is $\Omega(n)$.*

A consequence of Theorem 1.5 is that testing dual problems with one-sided error is either trivial or requires a linear number of queries. (Since testing a dual problem with one-sided error and query complexity $o(n)$ is possible only if the distance of every input from the property is $o(n)$, in which case both the original problem and its dual are trivial.) This fact stands in *sharp contrast to standard property testing problems*: In standard property testing, for essentially any sub-linear function $q : \mathbb{N} \to \mathbb{N}$, there exists a property of Boolean functions such that the query complexity of testing the property with one-sided error is $\Theta(q(n))$ [GKNR12].

1.2.2 Dual problems in testing properties of functions

When testing properties of functions, we identify each function $f : [n] \to \Sigma$ with its evaluation sequence, viewed as $f \in \Sigma^n$. The metric space is thus Σ^n, and the (absolute) distance between two functions is the Hamming distance between their string representations in Σ^n; equivalently, it is the number of inputs on which they disagree.

Many well-known properties of functions induce an error-correcting code with constant relative distance in Σ^n (e.g., linearity testing [BLR90] and low-degree testing [RS96]). For all such properties, the dual testing problem is equivalent to the original problem.

Theorem 1.6 *(testing duals of error-correcting codes). For any* error-correcting code with constant relative distance*, the problem of testing the code is equivalent to its dual problem.*

A notable example of a property of functions that does *not* induce an error-correcting code is the property of *monotone functions*, first considered for testing in [GGL+00]. For a poset $[n]$ and an ordered set Σ, a function $f : [n] \to \Sigma$ is *monotone* if for every $x, y \in [n]$ such that $x \leq y$ it holds that $f(x) \leq f(y)$. Nevertheless, the problem of testing this property is also equivalent to its dual problem:

Theorem 1.7 *(testing whether a function is far from monotone). The problem of testing* monotone Boolean functions over the Boolean hypercube *is equivalent to its dual problem.*

In fact, in Section 4 we prove a broad generalization of Theorem 1.7, as follows. For every $n \in \mathbb{N}$, consider functions from a poset $([n], \leq)$ to a range Σ_n, and assume that the width of the poset is at most $\frac{n}{2 \cdot |\Sigma_n|}$, where the width of a poset is the size of a maximum antichain in it. In this case, the problem of testing monotone functions from $[n]$ to Σ_n is equivalent to its dual problem. Note that the width requirement is quite mild: In particular, an ℓ-dimensional hypercube has size $n = 2^\ell$ and width $O(2^\ell / \sqrt{\ell}) = o(n)$.

1.2.3 Dual problems in distribution testing

Turning to *distribution testing* [BFR+13], one well-known problem is as follows: Fixing a predetermined distribution $\mathbf{D}$ over $[n]$, an ϵ-tester gets independent samples from an input distribution $\mathbf{I}$, and its task is to determine whether $\mathbf{I} = \mathbf{D}$ or $\mathbf{I}$ is ϵ-far from $\mathbf{D}$ in the ℓ_1 norm. When considering the worst-case, over all families of distributions, the distribution identity testing problem is different from its dual problem.

Proposition 1.8 *(testing whether a distribution is far from a known distribution). There exists a distribution family $\{\mathbf{D}_n\}_{n \in \mathbb{N}}$ such that the problem of testing whether an input distribution $\mathbf{I}_n$ is identical to $\mathbf{D}_n$ is different from its dual problem.*

However, for several specific (and natural) classes of distribution families, this problem is equivalent to its dual problem. In particular,

Theorem 1.9 *(testing whether a distribution is far from a predetermined distribution that has low ℓ_∞ norm). Let $\{\mathbf{D}_n\}_{n \in \mathbb{N}}$ be a family of distributions such that $lim_{n \to \infty} \|\mathbf{D}_n\|_\infty = 0$ (where $\|\mathbf{D}_n\|_\infty = \max_{i \in [n]}\{\Pr_{\mathbf{r} \sim \mathbf{D}_n}[\mathbf{r} = i]\}$). Then, the problem of testing whether an input distribution $\mathbf{I}_n$ is identical to $\mathbf{D}_n$ is equivalent to its dual problem.*

Theorem 1.9 implies that the problem of testing whether an input distribution is *far from being the uniform distribution* is equivalent to its original problem. Some distribution families that do not meet the condition of Theorem 1.9 also induce dual problems that are equivalent to their original problems: In particular, this applies to distribution families that assign $\Omega(1)$ probabilistic mass to *every* element in their support (see Proposition 5.4).

1.2.4 Dual problems in testing graph properties

When testing graph properties, we are interested in metric spaces in which the points are graphs, and the absolute distance between two graphs is the size of the symmetric difference of their edge-sets. A property of graphs is a set of graphs that is closed under taking isomorphisms of the graphs. We consider dual problems in two models of testing graph properties: The dense graph model [GGR98] and the bounded-degree model [GR02]. In both models, many well-known testing problems are different from their dual problems.

1.2.4.1 The dense graph model

In the dense graph model, an ϵ-tester queries the adjacency matrix of a graph over v vertices, and tries to determine whether the graph has some property or $\epsilon \cdot \binom{v}{2}$ edges need to be added and/or removed from the edge-set of the graph in order for it to have the property.

One well-known problem in this model is that of testing whether a graph is k-colorable (see [GGR98]). We consider the dual problem, of testing whether a graph is *far from being k-colorable*. This problem is *different* from its original problem, but its query complexity is nevertheless $O(1)$, as is the case for the original problem.

Theorem 1.10 *(testing whether a graph is far from being k-colorable). For any $k \geq 2$, the problem of testing whether a graph is k-colorable is different from its dual problem. Nevertheless,* the query complexity of the dual problem is $O(1)$.

However, unlike the complexity of the original problem, the constant in the $O(1)$ notation in Theorem 1.10 might be huge; in particular, our upper-bound has a tower-type dependence on the reciprocal of the proximity parameter. (This is the case since our proof relies on a result by Fischer and Newman [FN07], which in turn relies on Szemerédi's regularity lemma.)

The following proposition asserts that two other well-known problems in the dense graph model are different from their dual problems. The first problem is testing, for $\rho \in (0, 1)$, whether a graph on v vertices has a *clique of size* $\rho \cdot v$ (see [GGR98]). The second is the *graph isomorphism* problem (see [Fis05, FM08]): For an explicitly known graph G that is fixed in advance, the problem consists of testing whether an input graph is isomorphic to G.

Proposition 1.11 *(ρ-clique and graph isomorphism).*

1. *For any $\rho \leq \frac{1}{2}$, the problem of testing whether a graph on v vertices has a clique of size $\rho \cdot v$ is different from its dual problem.*
2. *There exist graph families $\{G_n\}_{n\in\mathbb{N}}$ such that testing whether an input graph H_n is isomorphic to G_n is different from its dual problem.*

In contrast to the dual problem of k-colorability, we do not know what is the query complexity of the two dual problems mentioned in Proposition 1.11.

1.2.4.2 The bounded-degree model

In the bounded-degree model [GR02] we are interested only in sparse graphs; in particular, we assume that the degree of every vertex in an input graph is at most d, where typically $d = O(1)$. A testing scenario in this model is as follows. Given an input graph over n vertices, we fix in advance an arbitrary ordering of the neighbors of each vertex in the graph. Then, an ϵ-tester may issue queries of the form "who is the i^{th} neighbor of $u \in [n]$?", and needs to determine whether the graph has some property or $\epsilon \cdot d \cdot n$ edges need to be added and/or removed from the edge-set of the graph in order for it to have the property.

One well-known problem in this model is that of testing whether a graph is connected (see [GR02]). We consider the dual problem, of testing whether a graph is *far from being* *connected*. Interestingly, although the dual problem is "very different" from the original one (in the sense that $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$ contains graphs that are $\Omega(n)$-far from being connected), the query complexity of the dual problem is very close to that of the original problem.

Theorem 1.12 *(testing whether a graph is far from being connected). For any $d \geq 3$, the problem of testing whether a graph is connected is different from its dual problem. Nevertheless,* the query complexity of the dual problem is $\text{poly}(1/\epsilon)$.

Another well-known problem in this model is testing cycle-free graphs (see [GR02]). We consider the dual problem, of testing whether a graph is *far from being cycle-free*.

Theorem 1.13 *(testing whether a graph is far from being cycle-free). For any $d \geq 3$, the problem of testing whether a graph is cycle-free (i.e., a forest) is different from its dual problem. Nevertheless,* the query complexity of the dual problem is $\text{poly}(1/\epsilon)$.

The well-known problem of testing bipartiteness in this model is also not equivalent to its dual problem, but we do not know what its query complexity is.

Proposition 1.14 *(testing whether a graph is far from bipartite). The problem of testing whether a graph is bipartite is different from its dual problem.*

1.3 A generalization: On being δ'-far from δ-far

So far, the dual problem of a property $\Pi = \{\Pi_n\}_{n\in\mathbb{N}}$ was defined using a *single* proximity parameter $\epsilon > 0$. This parameter $\epsilon > 0$ determines both the "yes" inputs for testing (i.e., $\mathcal{F}_{\epsilon\cdot n}(\Pi_n)$) and the distance of the "no" inputs from the "yes" inputs (i.e., it also determines $\mathcal{F}_{\epsilon\cdot n}(\mathcal{F}_{\epsilon\cdot n}(\Pi_n))$. A more general notion of dual testing problems is obtained by considering *two* proximity parameters, $\epsilon > 0$ and $\epsilon' > 0$, such that $\epsilon > 0$ determines the "yes" inputs for testing, and $\epsilon' > 0$ is the proximity parameter that determines the distance of the "no" inputs form the "yes" inputs; that is, the generalized dual problem consists of distinguishing between $\mathcal{F}_{\epsilon\cdot n}(\Pi_n)$ and $\mathcal{F}_{\epsilon'\cdot n}(\mathcal{F}_{\epsilon\cdot n}(\Pi_n))$.

Generalized dual problems are actually more similar to standard testing problems, compared to non-generalized dual problems. This is the case since we can fix $\epsilon > 0$, and define the *generalized ϵ-dual problem* as the problem of testing the fixed property $\{\mathcal{F}_{\epsilon\cdot n}(\Pi_n)\}_{n\in\mathbb{N}}$ with an arbitrarily small proximity parameter $\epsilon' > 0$. [3] The latter definition is just the standard definition of property testing, for the fixed property $\{\mathcal{F}_{\epsilon\cdot n}(\Pi_n)\}_{n\in\mathbb{N}}$. In Section 8 we formalize this notion, and show the following (informally stated):

Theorem 1.15 *(testers for generalized dual problems; informal). For every constant $\epsilon, \epsilon' > 0$:*

1. *The query complexity of the generalized dual problem of k-colorable graphs in the dense graphs model is $F(\epsilon, \epsilon')$, for some function F that does not depend on n.*

2. *The query complexity of the generalized dual problem of connected graphs in the bounded-degree graphs model is* $\text{poly}(1/\min\{\epsilon', \epsilon\})$.

[3] When fixing $\epsilon > 0$, and letting $\epsilon' > 0$ be arbitrary, we focus mainly on the setting of $\epsilon' \leq \epsilon$. This focus is justified by the fact that the case of $\epsilon' > \epsilon$ reduces to the case of $\epsilon' \leq \epsilon$ (see Observation 8.2).

3. *The query complexity of the generalized dual problem of cycle-free graphs in the bounded-degree graphs model is* poly $(1/\min\{\epsilon', \epsilon\})$.

1.4 Our techniques

In testing specific dual problems, we rely on one of two general techniques. We either show that the dual problem is equivalent to the original problem (see Section 1.4.1); or, if the dual problem is different from the original problem, we can sometimes reduce the dual problem to the corresponding *tolerant testing* problem (see Section 1.4.2), and then either solve the latter or rely on a known solution. In fact, in all cases where we reduce the dual problem to tolerant testing, we are also able to reduce the *generalized* dual problem to tolerant testing, and thus in these cases we also obtain a solution for the generalized dual problem.

1.4.1 Dual problems that are equivalent to the original problem

The first technique is *showing that the dual problem is equivalent to the original*. For a property $\Pi = \{\Pi_n\}_{n\in\mathbb{N}}$, this requires showing that for every sufficiently large n and sufficiently small $\epsilon > 0$, the set Π_n is $\mathcal{F}_{\epsilon\cdot n}$-closed (as in Definition 1.3). While this is easy to see in the case of error-correcting codes (which yields Theorem 1.6), [4] proving this for the set of monotone functions (Theorem 1.7) and in the setting of distribution identity testing (Theorem 1.9) requires proving various *structural results* about the corresponding sets.

To prove that the set of monotone functions from a poset $[n]$ to a range Σ is $\mathcal{F}_{\epsilon\cdot n}$-closed, assuming that the width of $[n]$ is bounded (by $\frac{n}{2\cdot|\Sigma|}$) and that $\epsilon > 0$ is sufficiently small, we show the following: It is possible to modify every function f that is neither monotone nor $(\epsilon \cdot n)$-far from monotone to a function f' that is $(\epsilon \cdot n)$-far from monotone, by changing value of the function at *strictly less than $\epsilon \cdot n$ points*. The latter statement indeed implies that the set of monotone functions is $\mathcal{F}_{\epsilon\cdot n}$-closed (see Fact 2.2); we stress that the statement is far from trivial, and in particular it does not hold for a general set, and actually *characterizes* the $\mathcal{F}_{\epsilon\cdot n}$-closed sets (so any set that is not $\mathcal{F}_{\epsilon\cdot n}$-closed does not satisfy such a property; see Appendix A). To prove that such a modification (of f to f') is possible, we rely on the following lemma: For every function f as above, there exists a collection of at least $n/4$ pairs (x, y) such that all points in all pairs are distinct, and for every pair it holds that $x < y$ and $f(x) = f(y)$ (see Lemma 4.1.1). The existence of such a collection allows us to modify f to a function that is $(\epsilon \cdot n)$-far from monotone, by changing the value of the function at a single point in each of $\epsilon \cdot n - 1$ pairs. For the special case of monotone *Boolean* functions, we use an analogous structural lemma of Fischer *et al.* [FLN+02], in order to prove that the set of monotone Boolean functions is $\mathcal{F}_{\epsilon\cdot n}$-closed in a stronger sense (see Proposition 4.3).

In the setting of distribution identity testing, we fix a distribution $\mathbf{D}$ over n elements, and our goal is to show that the singleton $\{\mathbf{D}\}$, which is the property to be tested in this setting, is $\mathcal{F}_\delta$-closed. We stress that distribution testing is carried out over the simplex with the ℓ_1-norm, and in this setting it is not true that every singleton is $\mathcal{F}_\delta$-closed (see Proposition 5.1). To show that $\{\mathbf{D}\}$ is $\mathcal{F}_\delta$-closed, we show that if $\mathbf{D}$ has sufficiently small ℓ_∞ norm (or satisfies other

[4] Intuitively, this is because a code can be thought of as a collection of "isolated" singletons.

useful properties), then every distribution $\mathbf{X} \neq \mathbf{D}$ that is not δ-far from $\mathbf{D}$ (in ℓ_1-norm) can be modified to a distribution $\mathbf{Z}$ that is δ-far from $\mathbf{D}$, by changing *less than δ of the probabilistic mass of* $\mathbf{X}$ (i.e., the distance between $\mathbf{X}$ and $\mathbf{Z}$ is less than δ). As in the case of monotone functions, this allows us to deduce that the singleton $\{\mathbf{D}\}$ is $\mathcal{F}_\delta$-closed (see Fact 2.2). The modification procedure of $\mathbf{X}$ to $\mathbf{Z}$ is based on a low-level technical case-analysis that is somewhat involved; a high-level overview appears in the proof of Proposition 5.2.

1.4.2 Reductions of (generalized) dual problems to tolerant testing

The second technique is useful when the dual problem is different from the original one. In this case, for several properties we are able to prove that the dual problem *reduces to the corresponding tolerant testing problem*. Tolerant testing, introduced by Parnas, Ron, and Rubinfeld [PRR06], is the following problem: Given a set Π_n, a parameter $\delta > 0$ and $\alpha < 1$, the tolerant testing problem consists of distinguishing between inputs that are $(\alpha \cdot \delta)$-close to Π_n and inputs that are δ-far from Π_n. Reducing dual problems to tolerant testing problems is done by showing that, for some $\alpha < 1$, all points in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$ are $(\alpha \cdot \delta)$-close to Π_n. We stress that *dual problems do not reduce to tolerant testing in general* (see Proposition C.1). Thus, proving that a specific dual problem reduces to tolerant testing requires structural results about the specific property at hand. Moreover, after reducing a dual problem to tolerant testing problem, we still need to solve the tolerant testing problem itself. [5]

A general useful result is that in order to reduce the dual problem of a property Π to the corresponding tolerant testing problem, it suffices to show the following: Every input x that is neither in the property Π nor far from Π can be modified to an input x' such that $\Delta(x, x')$ is small, and $\Delta(x', \Pi) \geq \Delta(x, \Pi) + 1$; that is, x' is farther (by one unit) from Π, compared to x (see Proposition 2.5; for an example of a property Π in the dense graph model for which such a statement *does not hold*, see Proposition C.1 and Remark C.2). In fact, this approach also allows us to reduce the *generalized* dual problem to tolerant testing (see Proposition 8.3), and thus we are also able to obtain testers for the generalized dual problem.

The case of k-colorability in the dense graph model is relatively straightforward: We show that every graph G that is neither k-colorable nor $(\epsilon \cdot n)$-far from k-colorable can be modified to a graph G' that is farther from being k-colorable, by adding $\binom{k+1}{2}$ edges to G. This is possible because every such graph G contains an independent set on $k+1$ vertices (since G is not $(\epsilon \cdot n)$-far from k-colorable, and assuming that $\epsilon > 0$ is sufficiently small; see Lemma 6.2.1), and adding a clique on these $k+1$ vertices yields G' that is farther from being k-colorable (see the proof of Proposition 6.2). We can then rely on the solution of Fischer and Newman [FN07] for the tolerant testing problem of k-colorability (see Theorem 6.4).

In the case of connectivity in the bounded-degree model (with degree bound d), we first prove a general result that *fully characterizes* the distance of a graph from being connected, in terms of the number of various types of connected components in the graph. Specifically, we say that a vertex has k free degrees if the degree of the vertex in the graph is $d-k$, and we say

[5] We note that there exist cases in which the tolerant testing problem is *significantly more difficult* than the dual problem. For example, according to Theorem 1.9, the complexity of testing whether a distribution is far from uniform is $\Theta(\sqrt{n})$; however, the results of Valiant and Valiant [VV11] imply that the complexity of the corresponding tolerant testing problem is $\Theta(n/\log(n))$.

that a connected component (resp., the entire graph) has k free degrees if the sum of the free degrees of the vertices in the component (resp., in the graph) is k. We show that the distance of a graph from being connected is essentially a linear function of the number of connected components of various free degrees in the graph, as well as the number of free degrees in the entire graph (for precise details see Lemma 7.4 in Section 7.1.1).

Relying on this characterization, we then show how to modify every graph G that is neither connected nor far from being connected to a graph G' that is farther from being connected, by modifying only $O(d)$ edges in G (see the proof of Proposition 7.6). Finally, we construct a *tolerant tester for connectivity* in the bounded-degree model with query complexity $O(1)$. This tester essentially estimates the distance of a graph from being connected, up to an additive error of $O(\epsilon \cdot n)$, using $O(\epsilon^{-3} \cdot d)$ queries. The construction relies on our aforementioned characterization of the latter distance, and is a variation on the algorithm of Chazelle, Rubinfeld, and Trevisan [CRT05] for estimating the number of connected components in a graph (for a high-level description and precise details, see Section 7.1.3 and Theorem 7.7).

The last setting is the dual problem of testing cycle-free graphs in the bounded-degree model. In this case, similarly to the previous proofs, we show how to modify every graph G that is not far from being cycle-free to a graph G' that is farther from being cycle-free, by adding only two or three edges to the graph. Marko and Ron [MR06] already noted that the distance of a graph G on n vertices from being cycle-free equals $|E(G)| + |C(G)| - n$, where $|C(G)|$ is the number of connected components in G. Our modification of G to G' is based on a simple case analysis, and consists of either adding a triangle to the graph, or connecting two vertices that are both in the same connected component (see Claim 7.9.1). We then rely on the tolerant tester for this problem by Marko and Ron [MR06].

2 Preliminaries

2.1 Metric spaces

Throughout the paper we denote by Ω a set with at least two elements, and we usually assume that it is equipped with a metric $\Delta : \Omega^2 \to [0, \infty)$, such that (Ω, Δ) is a metric space. We will usually use shorthand notation, and identify the metric space (Ω, Δ) with its set of elements Ω, and the metric Δ will be implicit. We call a metric space Ω graphical when Ω is the vertex-set of a connected undirected graph, such that for any $x, y \in \Omega$ it holds that $\Delta(x, y)$ is the length of a shortest path between x and y.

A special case of a graphical metric space is the Boolean hypercube, equipped with the Hamming distance. We denote the n-dimensional Boolean hypercube by H_n, and for $x, y \in H_n$ we denote by $\mathtt{sd}(x, y)$ the symmetric difference between x and y; that is, $\mathtt{sd}(x, y) = \{i \in [n] : x_i \neq y_i\}$. Then $\Delta(x, y) = |\mathtt{sd}(x, y)|$. Also, for every $x \in H_n$, we denote by $\|x\|_1$ the Hamming weight of x.

For any set $\Pi \subseteq \Omega$, we denote its complement by $\overline{\Pi} \stackrel{\text{def}}{=} \{x \in \Omega : x \notin \Pi\}$. Also, for any $x \in \Omega$ and $\delta > 0$ we denote the closed radius-δ ball around x by $B[x, \delta] \stackrel{\text{def}}{=} \{y : \Delta(x, y) \leq \delta\}$ and the open radius-δ ball around x by $B[x, \delta) \stackrel{\text{def}}{=} \{y : \Delta(x, y) < \delta\}$.

2.2 The "δ-far" operator

Abusing the notation Δ, for $x \in \Omega$ and non-empty $\Pi \subseteq \Omega$ we let $\Delta(x, \Pi) \stackrel{\text{def}}{=} \inf_{p \in \Pi}\{\Delta(x, p)\}$. If $\Delta(x, \Pi) \geq \delta$ then we say that x is δ-far from Π. For any space Ω and $\delta > 0$, we define the *δ-far operator* $\mathcal{F}_\delta : \mathcal{P}(\Omega) \to \mathcal{P}(\Omega)$ by $\mathcal{F}_\delta(\Pi) \stackrel{\text{def}}{=} \{x : \Delta(x, \Pi) \geq \delta\}$ for any non-empty $\Pi \subseteq \Omega$, and $\mathcal{F}_\delta(\emptyset) \stackrel{\text{def}}{=} \Omega$; that is, $\mathcal{F}_\delta(\Pi)$ is the set of elements that are δ-far from Π.

2.3 Property testing

In property testing, we assume that $\Omega = \Sigma^n$, for an arbitrary set Σ, and $n \in \mathbb{N}$. To avoid confusion, throughout the paper we will denote the (relative) proximity parameter for testing by $\epsilon > 0$, whereas the absolute distance between inputs will be denoted by $\delta > 0$. Indeed, in this case $\delta = \epsilon \cdot n$.

Definition 2.1 *(property testing). For a set Σ, a property $\Pi = \{\Pi_n\}_{n \in \mathbb{N}}$ such that $\Pi_n \subseteq \Sigma^n$, and parameter $\epsilon > 0$, an ϵ-*tester for Π *is a probabilistic algorithm T that gets oracle access to $x \in \Sigma^n$, in the sense that for any $i \in [n]$ it can query for the i^{th} symbol of x, and satisfies the following two conditions:*

1. *If $x \in \Pi_n$ then* $\Pr[T^x(1^n) = 1] \geq \frac{2}{3}$.
2. *If $x \in \mathcal{F}_{\epsilon \cdot n}(\Pi_n)$ then* $\Pr[T^x(1^n) = 0] \geq \frac{2}{3}$.

The query complexity *of an ϵ-tester T for Π is a function $q : \mathbb{N} \to \mathbb{N}$, such that for every $n \in \mathbb{N}$ it holds that $q(n)$ is the maximal number, over any $x \in \Sigma^n$ and internal coin tosses of T, of oracle queries that T makes. The* query complexity of ϵ-testing Π *is a function $q : \mathbb{N} \to \mathbb{N}$ such that for every $n \in \mathbb{N}$ it holds that $q(n)$ is the minimum, over all query complexities q' of ϵ-testers for Π, of $q'(n)$.*

We will sometimes slightly abuse Definition 2.1, by referring to ϵ-testers for $\Pi \subseteq \Sigma^n$, where n is a generic integer (instead of referring to ϵ-testers for an infinite sequence $\Pi = \{\Pi_n\}_{n \in \mathbb{N}}$).

2.4 Sets of the form $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ and $\mathcal{F}_\delta$-closed sets: Useful tools

We state several properties of sets of the form $\mathcal{F}_\delta(\mathcal{F}_\delta\Pi))$ and of $\mathcal{F}_\delta$-closed sets, which will be useful for us throughout the paper. The proofs of these properties, as well as the general study of sets of the form $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$, is deferred to the appendices. One basic charaterization of $\mathcal{F}_\delta$-closed sets that will be very useful for us is the following:

Fact 2.2 *(Item 2 of Theorem A.2). For any Ω, $\delta > 0$, and $\Pi \subseteq \Omega$, it holds that Π is $\mathcal{F}_\delta$-closed if and only if for every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ there exists $z \in \mathcal{F}_\delta(\Pi)$ such that $\Delta(z, x) < \delta$.*

Loosely speaking, a necessary condition for a set Π in a graphical space to be $\mathcal{F}_\delta$-closed is that it does not "enclose" some vertex $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ from "all sides". In particular, the following proposition shows that if Π is $\mathcal{F}_\delta$-closed, then every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ is connected to $\mathcal{F}_\delta(\Pi)$ via a path that does not intersect Π (nor any vertex that is adjacent to Π).

Proposition 2.3 *(see Proposition A.3). For a graphical Ω and $\delta \geq 2$, let $\Pi \subseteq \Omega$ be an $\mathcal{F}_\delta$-*closed set. *Then, for every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$, there exists a path $x = v_0, v_1, ..., v_l = z$ such that $z \in \mathcal{F}_\delta(\Pi)$, and for every $i \in [l]$ it holds that $\Delta(v_i, \Pi) \geq 2$.*

We now present a *sufficient* condition for a set in a graphical metric space to be $\mathcal{F}_\delta$-closed. Loosely speaking, a set is *strongly $\mathcal{F}_\delta$-closed* if for any $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ there exist a neighbor x' of x that is farther from Π than x itself.

Definition 2.4 *(strongly $\mathcal{F}_\delta$-closed sets). For a graphical Ω and $\delta > 0$, a set $\Pi \subseteq \Omega$ is* strongly $\mathcal{F}_\delta$-closed *if and only if for every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ there exists a neighbor x' of x such that $\Delta(x', \Pi) = \Delta(x, \Pi) + 1$.*

As is implied by the name, every strongly $\mathcal{F}_\delta$-closed set is also $\mathcal{F}_\delta$-closed (see Proposition A.6). The following condition, which also applies in *graphical metric spaces*, allows us to upper-bound the distance of vertices in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ from Π.

Proposition 2.5 *(see Proposition C.4). Let Ω be a graphical space, let $\Pi \subseteq \Omega$, let $\delta \geq 2$, and let $\delta' \leq \delta$. If there exists an integer m such that for every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ there exists x' satisfying $\Delta(x, x') \leq m$ and $\Delta(x', \Pi) > \Delta(x, \Pi)$, then $\mathcal{F}_{\delta'}(\mathcal{F}_\delta(\Pi)) \subseteq \left\{y : \Delta(y, \Pi) \leq \delta - \frac{\delta'}{m}\right\}$.*

In the context of property testing we are interested in a sequence $\Pi = \{\Pi_n\}_{n \in \mathbb{N}}$, and in distances of the form $\delta = \epsilon \cdot n$, for a constant $\epsilon > 0$. In this context, when Proposition 2.5 holds with parameters $\delta = \epsilon \cdot n$ and $\delta' = \epsilon' \cdot n$ and $m = O(1)$, it follows that the distance of vertices in $\mathcal{F}_{\epsilon' \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n))$ from Π_n is bounded away from $\epsilon \cdot n$ by $\delta'/m = \frac{\epsilon' \cdot n}{m} = \Omega(n)$. In this case the dual problem of Π_n reduces to the corresponding tolerant testing problem.

3 General results regarding the query complexity of dual problems

In this section we state and prove several *preliminary general results* about the query complexity of dual problems. Specifically, we show a lower bound on testing dual problems with one-sided error, and we observe that dual problems of error-correcting code are equivalent to the original problems.

3.1 Testing dual problems with one-sided error

We first need to extend Definition 2.1, by defining two special types of testers. We then extend Observation 1.4 (which asserted that testing a dual problem is at least as difficult as testing the original problem) such that it will also apply to these two types of testers.

Definition 3.1. *For any ϵ-tester T as in Definition 2.1,*

1. *If the probability in Condition (1) of Definition 2.1 (i.e., the probability that inputs in Π are accepted) is 1, then we say that T has* one-sided error.
2. *If the probability in Condition (2) of Definition 2.1 (i.e., the probability that inputs in $\mathcal{F}_{\epsilon \cdot n}(\Pi)$ are rejected) is 1, then we say that T has* perfect soundness.

Observation 3.2 *(Observation 1.4, extended). The query complexity of a dual problem is lower bounded by the query complexity of its original problem. Moreover, the query complexity of testing a dual problem with one-sided error (resp., with perfect soundness) is lower bounded by the query complexity of testing the original problem with perfect soundness (resp., with one-sided error).*

We also need the following proposition, which appeared in our previous technical report [Tel14, Apdx. A].

Proposition 3.3 *(testing standard problems with perfect soundness). For a set Σ, let $\Pi = \{\Pi_n\}_{n\in\mathbb{N}}$ such that $\Pi_n \subseteq \Sigma^n$. Suppose that for all sufficiently large n it holds that $\Pi_n \neq \emptyset$ and that there exist inputs that are $\Omega(n)$-far from Π_n. Then, the query complexity of testing Π with* perfect soundness *is $\Omega(n)$.*

Proof. The main claim needed for the proof is the following, which asserts that the existence of an $(\epsilon/2)$-tester with perfect soundness and query complexity $q(n)$ implies that every input is $(q(n) + (\epsilon/2) \cdot n)$-close to the property.

Claim 3.3.1. *For Π as in the hypothesis and any $\epsilon > 0$, if there exists an ϵ-tester for Π* with perfect soundness *and query complexity q, then for a sufficiently large n and every $z \in \Sigma^n$ it holds that $\Delta(z, \Pi_n) < q(n) + \epsilon \cdot n$.*

Proof. Let $\epsilon > 0$, and assume that there exists an ϵ-tester T for Π with perfect soundness and query complexity q. By the hypothesis, for a sufficiently large n it holds that $\Pi_n \neq \emptyset$, and hence there exists $x \in \Pi_n$. Now, there exist random coins r such that the residual deterministic tester $T^x(1^n, r)$ (i.e., the deterministic tester obtained by fixing random coins r) accepts after making $q(n)$ queries. Denote the coordinates of these $q(n)$ queries by $(i_1, i_2, ..., i_{q(n)})$, where we assume for simplicity and without loss of generality that T always makes exactly q queries.

Note that every $z' \in \Sigma^n$ such that $(z'_{i_1}, z'_{i_2}, ..., z'_{i_{q(n)}}) = (x_{i_1}, x_{i_2}, ..., x_{i_{q(n)}})$ is accepted by the residual deterministic tester with random coins r. Since T has perfect soundness, this implies that every such z' satisfies $\Delta(z', \Pi_n) < \epsilon \cdot n$ (since inputs that are $(\epsilon \cdot n)$-far must be rejected with probability 1). Hence, for any $z \in \Sigma^n$, by changing the $q(n)$ coordinates $(z_{i_1}, z_{i_2}, ..., z_{i_{q(n)}})$ to equal $(x_{i_1}, x_{i_2}, ..., x_{i_{q(n)}})$, we obtain a string z' such that $\Delta(z', \Pi_n) < \epsilon \cdot n$. This implies that every $z \in \Sigma^n$ satisfies $\Delta(z, \Pi_n) \leq \Delta(z, z') + \Delta(z', \Pi_n) < q(n) + \epsilon \cdot n$. □

Now, by the hypothesis, for some $\epsilon > 0$ and any sufficiently large n there exists $z \in \Sigma^n$ such that $\Delta(z, \Pi_n) \geq \epsilon \cdot n$. For $\epsilon' < \epsilon$, let T be an ϵ'-tester with perfect soundness for Π, and denote its query complexity by q. Then, by Claim 3.3.1,

$$\epsilon \cdot n \leq \Delta(z, \Pi_n) \leq q(n) + \epsilon' \cdot n$$

which implies that $q(n) = \Omega(n)$. ■

Theorem 1.5 follows immediately by combining Observation 3.2 and Proposition 3.3. It follows that *dual problems* can be tested with one-sided error and $o(n)$ queries only if the distance of every input from the property is $o(n)$. However, in this case both the original problem and its dual are trivial to begin with.

3.2 Testing dual problems of error-correcting codes

In an n-dimensional hypercube, a code $\Pi = \{\Pi_n\}_{n\in\mathbb{N}}$ has constant relative distance $\zeta > 0$ if for every $n \in \mathbb{N}$ it holds that $\min_{x\neq y\in\Pi_n}\{\Delta(x,y)\} \geq \zeta \cdot n$. The following proposition implies that for any code Π with constant relative distance $\zeta > 0$, and any $\epsilon \leq \frac{\zeta}{2}$, it holds that Π_n is (strongly) $\mathcal{F}_{\epsilon\cdot n}$-closed.

Proposition 3.4 *(codes with constant relative distance are $\mathcal{F}_{\epsilon\cdot n}$-closed). Let $\Pi_n \subseteq \Sigma^n$ and $\zeta > 0$ such that $|\Pi_n| \geq 2$, and $\min_{x\neq y\in\Pi_n}\{\Delta(x,y)\} \geq \zeta \cdot n$. Then, for any $\epsilon \leq \frac{\zeta}{2}$, it holds that Π_n is strongly $\mathcal{F}_{\epsilon\cdot n}$-closed.*

Proof. Let $\epsilon \leq \frac{\zeta}{2}$ and let $\delta = \epsilon \cdot n$. Let $x \notin \Pi_n \cup \mathcal{F}_\delta(\Pi_n)$, and note that x is in the $(\delta - 1)$-neighborhood of exactly one $p \in \Pi_n$. By changing x in one location $i \in [n]$ such that $x_i = p_i$, we obtain a neighbor x' of x such that either $x' \in \mathcal{F}_\delta(\Pi_n)$ (and $\Delta(x,\Pi_n) = \delta - 1$), or x' is still in the $(\delta - 1)$-neighborhood of p, in which case $\Delta(x',\Pi_n) = \Delta(x',p) = \Delta(x,p) + 1$. ■

Theorem 1.6 follows as a corollary of Proposition 3.4. Note that in cases where these problems involve testing Boolean functions over $\{0,1\}^\ell$, the generated error-correcting code is in $\{0,1\}^{2^\ell}$. In particular, according to Theorem 1.5, the corresponding dual problems cannot be tested with one-sided error and $o(2^\ell)$ queries.

4 Testing functions that are far from monotone

In this section we prove (a generalization of) Theorem 1.7, which asserts that the dual problem of monotonicity testing is, in many settings, equivalent to the original problem.

Let us first recall the setting for monotonicity testing, and introduce some notation. Let $[n]$ be a partially ordered set, [6] and let Σ be an ordered set. A function $f : [n] \to \Sigma$ is *monotone* if for every $x,y \in [n]$ such that $x \leq y$, it holds that $f(x) \leq f(y)$. Throughout the section, we identify every function $f : [n] \to \Sigma$ with a corresponding string $f \in \Sigma^n$. Recall the following standard definitions: An antichain in a poset is a set of elements in the poset that are pairwise incomparable; and the width of a poset is the size of a maximum antichain in it.

4.1 Monotone functions yield $\mathcal{F}_\delta$-closed sets

The main result needed to prove Theorem 1.7 is the following, which asserts that for many posets $[n]$ and ranges Σ, the set of monotone functions $[n] \to \Sigma$ is $\mathcal{F}_\delta$-closed.

Proposition 4.1 *(the set of monotone functions is $\mathcal{F}_\delta$-closed). Let $[n]$ be a partially ordered set, and let Σ be a finite ordered set such that the width of $[n]$ is at most $\frac{n}{2\cdot|\Sigma|}$. Then, for every $\delta < \frac{n}{4}$, the set of monotone functions from $[n]$ to Σ is $\mathcal{F}_\delta$-closed.*

Proof. For a sufficiently large $n \in \mathbb{N}$, denote the set of monotone functions from $[n]$ to Σ by $\Pi_n \subseteq \Sigma^n$, and let $\delta < \frac{n}{4}$. To show that Π_n is $\mathcal{F}_\delta$-closed, we rely on Fact 2.2: For every $f \notin \Pi_n \cup \mathcal{F}_\delta(\Pi_n)$, we show a function $h \in \mathcal{F}_\delta(\Pi_n)$ such that $\Delta(f,h) < \delta$.

[6] Similar to metric spaces, we usually identify a partially ordered set $([n],\leq)$ with its set of elements $[n]$, and the order relation is implicit and denoted by $\leq$.

High-level overview. First, we define some terminology that we will need. For any $f : [n] \to \Sigma$, we call $(x, y) \in [n] \times [n]$ a *violating pair* for f if $x < y$ and $f(x) > f(y)$. Observe that f is monotone if and only if there are no violating pairs for f. Also, we call $(x, y) \in [n] \times [n]$ a *flat pair* for f if $x < y$ and $f(x) = f(y)$. A collection of *disjoint violating pairs* for f is a collection $\mathcal{V}$ of violating pairs such that for every $(x_1, y_1) \neq (x_2, y_2) \in \mathcal{V}$ it holds that x_1, x_2, y_1, y_2 are distinct. A collection of *disjoint flat pairs* is defined analogously.

The proof idea is as follows. Let $f \notin \Pi_n \cup \mathcal{F}_\delta(\Pi_n)$. First, let us assume that there exists a collection $\mathcal{C}$ of δ disjoint pairs in $[n]$, such that one pair in $\mathcal{C}$ is violating for f, and the other $\delta - 1$ pairs are flat for f. Then, observe that for every flat pair in $\mathcal{C}$, we can change the value of f at one input in the pair, thereby turning it into a violating pair (i.e., for a pair (x, y), if $f(x) = f(y) = \max_{\sigma\in\Sigma}\{\sigma\}$, we can set $f(y)$ to be any other $\sigma \in \Sigma$, and otherwise, we can set $f(x) = \max_{\sigma\in\Sigma}\{\sigma\}$). Thus, by changing the value of f on one input in each flat pair in $\mathcal{C}$, we obtain $h \in \Sigma^n$ such that $\Delta(h, f) = |\mathcal{C}| - 1 = \delta - 1$ and that $\mathcal{C}$ is a collection of disjoint violating pairs for h of size δ. The proposition follows since a function h that has a collection of δ disjoint violating pairs satisfies $\Delta(h, \Pi_n) \geq \delta$ (see Claim 4.1.3).

To prove that the collection $\mathcal{C}$ (of $\delta - 1$ flat pairs and one violating pair) exists, we use the fact that the width of $[n]$ is bounded. In particular, we show that there exists a collection $\mathcal{T}$ of $\frac{n}{4}$ disjoint flat pairs for f (see Lemma 4.1.1). Since $f \notin \Pi_n$, there exists at least one violating pair (x, y) for f. This pair shares a common element with at most two pairs in $\mathcal{T}$. Using the fact that $\delta \leq \frac{n}{4} - 1$, it follows that there exists $\mathcal{T}' \subseteq \mathcal{T}$ such that $\mathcal{C} = \mathcal{T}' \cup \{(x, y)\}$ is a collection of disjoint pairs, and $|\mathcal{T}'| \geq |\mathcal{T}| - 2 = \frac{n}{4} - 2 \geq \delta - 1$. To conclude, note that the pair $(x, y) \in \mathcal{C}$ is violating for f, and that all other pairs in $\mathcal{C}$ are flat.

The actual proof. Let $f \notin \Pi_n \cup \mathcal{F}_\delta(\Pi_n)$. The following lemma is used as the main step towards establishing (in Corollary 4.1.2) that there exists a collection $\mathcal{C}$ of δ disjoint pairs in $[n]$ such that one of these pairs is a violating pair for f, and the other $\delta - 1$ pairs are flat pairs for f.

Lemma 4.1.1. *Let $[n]$ be a poset and Σ be an ordered set such that the width of $[n]$ is at most $\frac{n}{2\cdot|\Sigma|}$. Then, for every $f : [n] \to \Sigma$, there exists a collection of disjoint flat pairs for f of size at least $\frac{n}{4}$.*

Proof. By Dilworth's theorem [Dil50], and since the width of $[n]$ is at most $\frac{n}{2\cdot|\Sigma|}$, there exists a partition of $[n]$ into at most $\frac{n}{2\cdot|\Sigma|}$ monotone chains; that is, there exists a collection $\mathcal{M}$ such that $|\mathcal{M}| \leq \frac{n}{2\cdot|\Sigma|}$ that satisfies the following two conditions:

1. Every $c \in \mathcal{M}$ is a sequence $c = (x_1, ..., x_{n_c}) \subseteq [n]$ such that for every $i \in [n_c - 1]$ it holds that $x_i < x_{i+1}$.
2. $\mathcal{M}$ is a partition of $[n]$, in the sense that every $x \in [n]$ appears in exactly one monotone chain $c \in \mathcal{M}$.

For a fixed function f, we construct a corresponding collection $\mathcal{T}$ of disjoint flat pairs for f as follows. We go over the chains in $\mathcal{M}$, in an arbitrary order, and collect disjoint flat pairs for f, which we add to $\mathcal{T}$, while processing each chain separately. For any fixed chain

$c \in \mathcal{M}$, we partition c into $|\Sigma|$ (non-consecutive) sub-chains such that f is constant on each sub-chain; that is, the partition of c is the collection $\{c_\sigma\}_{\sigma \in \Sigma}$ such that for every $\sigma \in \Sigma$ it holds that $c_\sigma = \{x \in c : f(x) = \sigma\}$. Note that each of the sub-chains is a "monochromatic" chain, and thus, every pair of elements in each sub-chain constitutes a flat pair. Accordingly, we now try to partition every sub-chain into pairs of elements (failing to pair at most one element in each sub-chain), and add these pairs to $\mathcal{T}$.

Since we only insert flat pairs to $\mathcal{T}$, and since $\mathcal{M}$ is a partition of the poset, the set $\mathcal{T}$ is a collection of disjoint flat pairs. In addition, for every fixed chain $c \in \mathcal{M}$, we fail to pair at most $|\Sigma|$ elements (i.e., at most one element per sub-chain). Therefore, for every chain $c \in \mathcal{M}$, we collect at least $\frac{1}{2} \cdot (|c| - |\Sigma|)$ flat pairs for $\mathcal{T}$. Overall, we get at least

$$\sum_{c \in \mathcal{M}} \frac{1}{2} \cdot (|c| - |\Sigma|) = \frac{1}{2} \cdot (n - |\Sigma| \cdot |\mathcal{M}|) \geq \frac{n}{4}$$

disjoint flat pairs for $\mathcal{T}$. $\square$

Corollary 4.1.2. *Let $[n]$, Σ and δ be as in Proposition 4.1. Then, for every $f \notin \Pi_n$, there exists a collection $\mathcal{C}$ of δ disjoint pairs in $[n]$ such that one pair in $\mathcal{C}$ is a violating pair for f, and the other $\delta - 1$ pairs are flat pairs for f.*

Proof. Since $f \notin \Pi_n$, there exists a violating pair (x, y) for f. Relying on Lemma 4.1.1, there exists a collection $\mathcal{T}$ of flat pairs for f such that $|\mathcal{T}| \geq \frac{n}{4} \geq \delta + 1$. Since there are at most two pairs in $\mathcal{T}$ that share a common element with (x, y), there exists a sub-collection $\mathcal{T}' \subseteq \mathcal{T}$ such that $|\mathcal{T}'| = \delta - 1$ and $\mathcal{C} = \mathcal{T}' \cup \{(x, y)\}$ is a collection as required. $\square$

Let $\mathcal{C}$ be a collection of disjoint pairs for f, as in Corollary 4.1.2. Observe that we can turn every flat pair $(x, y) \in \mathcal{C}$ into a violating pair, by modifying the value of f at one input. By doing so, we obtain a function h such that $\Delta(f, h) = |\mathcal{C}| - 1 = \delta - 1$ and $\mathcal{C}$ is a collection of disjoint violating pairs for h of size δ. The proposition will follow from the following claim.

Claim 4.1.3. *For $h : [n] \to \Sigma$, if there exists a collection $\mathcal{C}$ of disjoint violating pairs for h having size ρ, then $\Delta(h, \Pi_n) \geq \rho$.* [7]

Proof. Let $g \in \Pi_n$ such that $\Delta(h, g) = \Delta(h, \Pi_n)$. If there exists a pair $(x, y) \in \mathcal{C}$ such that $h(x) = g(x)$ and $h(y) = g(y)$, then (x, y) is a violating pair for g, which contradicts $g \in \Pi_n$. Hence, the symmetric difference between h and g includes at least one element from each pair in $\mathcal{C}$. Since the pairs in $\mathcal{C}$ are disjoint, we get that $\Delta(h, \Pi_n) = \Delta(h, g) \geq |\mathcal{C}|$. $\square$

Thus, it holds that $h \in \mathcal{F}_\delta(\Pi_n)$. ■

Proposition 4.1 implies the following:

[7] A related claim was proved in [GGL+00, Prop 3]. However, they considered Boolean functions over the hypercube, and defined violating pairs differently.

Theorem 4.2 *(Theorem 1.7, extended). Let $\{P_n\}_{n\in\mathbb{N}}$ be a family of posets such that $P_n = ([n], \leq_n)$ for every $n \in \mathbb{N}$, and let $\{\Sigma_n\}_{n\in\mathbb{N}}$ be a family of ordered sets. Assume that for all sufficiently large n, the width of P_n is at most $\frac{n}{2\cdot|\Sigma_n|}$. Then, the problem of testing monotone functions from P_n to Σ_n is equivalent to its dual problem.*

In the special case of functions over the domain of the Boolean hypercube $\{0,1\}^\ell$, where $2^\ell = n$, Theorem 4.2 applies when the range satisfies $|\Sigma| \leq \sqrt{\ell}/2$. This is the case since, by Sperner's theorem, the width of the ℓ-dimensional hypercube, which has the element-set $[n] = [2^\ell]$, is $\binom{\ell}{\lfloor \ell/2 \rfloor}$ (and hence if $|\Sigma| \leq \sqrt{\ell}/2$, we get that the width satisfies $\binom{\ell}{\lfloor \ell/2 \rfloor} < \frac{n}{\sqrt{\ell}} \leq \frac{n}{2\cdot|\Sigma|}$). Thus, Theorem 1.7 follows from Theorem 4.2 as a special case.

In addition, the proof of Proposition 4.1 shows that for a poset P_n and a range Σ_n as in Theorem 4.2, there always exist functions that are $\Omega(n)$-far from being monotone. Thus (using Theorem 1.5), testing the dual problem with one-sided error requires $\Omega(n)$ queries.

4.2 Monotone Boolean functions yield strongly $\mathcal{F}_\delta$-closed sets

We now show that in the case of $|\Sigma| = 2$ (i.e., for Boolean functions over a poset $[n]$), the set of monotone functions is actually *strongly* $\mathcal{F}_\delta$-closed. This fact will be used in Section 8, and is also interesting combinatorially: It asserts that any Boolean function that is not too far from being monotone can be made farther from monotone by changing its value at a single input. The proof idea is similar to the proof of Proposition 4.1, but we will use an additional lemma, which is specific for Boolean functions, and was proved in [FLN+02].

Proposition 4.3 *(the set of monotone Boolean functions is strongly $\mathcal{F}_\delta$-closed). Let $[n]$ be a partially ordered set of width at most $\frac{n}{4}$. Then, for every $\delta < \frac{n}{8}$, the set of monotone Boolean functions over $[n]$ is strongly $\mathcal{F}_\delta$-closed.*

Proof. For a sufficiently large n, let Π_n be the set of monotone Boolean functions over $[n]$, and let $\delta < \frac{n}{8}$. We will rely on the following lemma.

Lemma 4.3.1 *(Lemma 4 in [FLN+02]). For $f : [n] \to \{0,1\}$, if $\Delta(f, \Pi_n) \geq \rho$, then there exists a collection of disjoint violating pairs for f having size ρ.*

Combining Claim 4.1.3 and Lemma 4.3.1, we get the following corollary:

Corollary 4.3.2. *For a* Boolean *function $f : [n] \to \{0,1\}$, it holds that $\Delta(f, \Pi_n) \geq \rho$* if and only if *there exists a collection of disjoint violating pairs for f having size ρ.*

Now, let $f \notin \Pi_n \cup \mathcal{F}_\delta(\Pi_n)$. According to Corollary 4.3.2, there exists a collection $\mathcal{V}$ of disjoint violating pairs for f, such that $|\mathcal{V}| = \Delta(f, \Pi_n) < \delta$. According to Lemma 4.1.1, there exists a collection $\mathcal{T}$ of flat pairs for f such that $|\mathcal{T}| \geq \frac{n}{4} \geq 2\delta$. The number of pairs in $\mathcal{T}$ that share a common element with any pair in $\mathcal{V}$ is at most $2 \cdot |\mathcal{V}| < 2 \cdot \delta \leq |\mathcal{T}|$. Hence, there exists some pair $(x, y) \in \mathcal{T}$ such that $\mathcal{V} \cup \{(x, y)\}$ is a collection of disjoint pairs. By modifying the value of f on one input from (x, y), we can turn it into a violating pair. This way, we obtain a function f' such that $\Delta(f, f') = 1$, and there exists a collection of disjoint violating pairs for f' of size $|\mathcal{V}| + 1 = \Delta(f, \Pi_n) + 1$. Relying on Corollary 4.3.2 again, we get that $\Delta(f', \Pi_n) = \Delta(f, \Pi_n) + 1$. ■

5 Testing distributions that are far from a known distribution

In this section we prove Proposition 1.8, which asserts that the dual problem of distribution testing is not equivalent to the original problem, in general; and Theorem 1.9 as well as another result, which assert that for many *specific natural distribution families* the dual problem nevertheless is equivalent to the original problem.

Let us first recall the setting of distribution testing (for excellent surveys see, e.g., [Rub12, Can15]). In this context, a tester gets independent samples from an input distribution, and tries to determine whether the distribution has some property or is far from having the property. A basic problem in this field is the one of testing whether a *distribution is identical to a known distribution*. In this problem, a distribution $\mathbf{D}$ over $[n]$ is predetermined and explicitly known, and an ϵ-tester gets independent samples from a distribution $\mathbf{I}$ over $[n]$. The goal of the tester is to determine, using as few samples as possible, whether $\mathbf{I} = \mathbf{D}$ or $\mathbf{I}$ is ϵ-far from $\mathbf{D}$ in the ℓ_1 norm; that is, whether $\|\mathbf{I} - \mathbf{D}\|_1 = \sum_{i \in [n]} |\mathbf{I}(i) - \mathbf{D}(i)| \geq \epsilon$. [8]

The main question in this section is for which distribution families it holds that the dual identity testing problem is equivalent to the original problem. That is, we ask for which families of distributions $\{\mathbf{D}_n\}_{n \in \mathbb{N}}$ it holds that for every sufficiently small constant $\delta > 0$ and every sufficiently large n, the singleton $\{\mathbf{D}_n\}$ is $\mathcal{F}_\delta$-closed (cf. Definition 1.3).

5.1 The dual problem is different from the original

In $\mathbb{R}^n$ with the Euclidean metric, every singleton $\{\mathbf{D}_n\}$ is $\mathcal{F}_\delta$-closed for every $\delta > 0$. Our first observation is that the analogous fact is not true in the simplex with the ℓ_1 norm.

Proposition 5.1 *(Proposition 1.8, extended). Let $\{\mathbf{D}_n\}_{n \in \mathbb{N}}$ be a distribution family such that for every $n \in \mathbb{N}$ it holds that $\mathbf{D}_n(1) = 1 - \frac{1}{n}$ and for any $i \in [n] \setminus \{1\}$ it holds that $\mathbf{D}_n(i) = \frac{1}{n \cdot (n-1)}$. Then, for every $\delta > 0$ and sufficiently large n, it holds that $\Pi = \{\mathbf{D}_n\}$ is* not *$\mathcal{F}_\delta$-closed.*

Proof. For $\delta > 0$, let $n \in \mathbb{N}$ such that $\delta > \frac{3}{n}$. Relying on Fact 2.2, it suffices to show a distribution $\mathbf{X} \notin \{\mathbf{D}_n\} \cup \mathcal{F}_\delta(\{\mathbf{D}_n\})$ such that there does not exist $\mathbf{Z} \in \mathcal{F}_\delta(\{\mathbf{D}_n\})$ satisfying $\Delta(\mathbf{X}, \mathbf{Z}) < \delta$.

Let $\mathbf{X}$ be the distribution over $[n]$ such that $\mathbf{X}(1) = 1$ (and for every $i > 1$ it holds that $\mathbf{X}(i) = 0$). Then $0 < \Delta(\mathbf{X}, \mathbf{D}_n) = 2/n < \delta$, implying that $\mathbf{X} \notin \{\mathbf{D}_n\} \cup \mathcal{F}_\delta(\{\mathbf{D}_n\})$. Let $\mathbf{Z}$ be any distribution over $[n]$. If $\mathbf{Z}(1) > 1 - \frac{1}{n}$, then $\sum_{i=2}^{n} \mathbf{Z}(i) < \frac{1}{n}$, and hence

$$\begin{aligned}
\Delta(\mathbf{Z}, \mathbf{D}_n) &= \mathbf{Z}(1) - \mathbf{D}(1) + \sum_{i=2}^{n} |\mathbf{Z}(i) - \mathbf{D}_n(i)| \\
&\leq \frac{1}{n} + \sum_{i=2}^{n} \mathbf{Z}(i) + \sum_{i=2}^{n} \mathbf{D}_n(i) \,,
\end{aligned}$$

which is less than $3/n$, and thus $\Delta(\mathbf{Z}, \mathbf{D}_n) < \delta$, implying that $\mathbf{Z} \notin \mathcal{F}_\delta(\{\mathbf{D}_n\})$.

[8] Note that the metric space for this problem is the standard simplex in $\mathbb{R}^n$ with the ℓ_1 norm, and that the distances satisfy $\delta \in [0, 2]$. Accordingly, we slightly abuse Definition 2.1 in this section, by requiring that an ϵ-tester distinguish between Π and $\mathcal{F}_\epsilon(\Pi)$, and not between Π and $\mathcal{F}_{\epsilon \cdot n}(\Pi)$ (i.e., the proximity parameter for testing $\epsilon > 0$ is the absolute distance between "yes" instances and "no" instances, and not the relative distance).

Otherwise, $\mathbf{Z}(1) \leq 1 - \frac{1}{n}$. Note that $\mathbf{Z}(1) \leq \mathbf{D}_n(1) < \mathbf{X}(1)$, and therefore $|\mathbf{Z}(1) - \mathbf{X}(1)| - |\mathbf{Z}(1) - \mathbf{D}_n(1)| = \mathbf{X}(1) - \mathbf{D}_n(1) = \frac{1}{n}$. Hence, we get that

$$\begin{aligned}\Delta(\mathbf{Z}, \mathbf{X}) - \Delta(\mathbf{Z}, \mathbf{D}_n) &= \sum_{i=1}^{n} \Big(|\mathbf{Z}(i) - \mathbf{X}(i)| - |\mathbf{Z}(i) - \mathbf{D}_n(i)| \Big) \\ &= \frac{1}{n} + \sum_{i=2}^{n} \left(\mathbf{Z}(i) - \left| \mathbf{Z}(i) - \frac{1}{n(n-1)} \right| \right) \\ &\geq \frac{1}{n} - (n-1) \cdot \frac{1}{n(n-1)} ,\end{aligned}$$

which equals zero. (The last inequality is since for $a, b \in \mathbb{R}^+$ it holds that $b - |b - a| \geq -a$, because $|b - a| \leq |b| + |a| = b + a$.) It follows that $\mathbf{Z} \in \mathcal{F}_\delta(\{\mathbf{D}_n\})$ cannot satisfy $\Delta(\mathbf{Z}, \mathbf{X}) < \delta$ (since in such a case $\Delta(\mathbf{Z}, \mathbf{X}) - \Delta(\mathbf{Z}, \mathbf{D}_n) < 0$). ■

5.2 Distribution families for which the dual problem is equivalent to the original

The following two propositions show that for many *natural distribution families*, the singleton $\{\mathbf{D}_n\}$ is $\mathcal{F}_\delta$-closed for every sufficiently small $\delta > 0$, and thus the dual testing problem is equivalent to the original one. The first proposition refers to distribution families $\{\mathbf{D}_n\}_{n \in \mathbb{N}}$ such that $\lim_{n \to \infty} \|\mathbf{D}_n\|_\infty = 0$, whereas the second proposition refers to distribution families in which each support element has $\Omega(1)$ probability mass.

Proposition 5.2 *(distributions with low ℓ_∞ norm induce $\mathcal{F}_\delta$-closed properties). Let $\{\mathbf{D}_n\}_{n \in \mathbb{N}}$ be a family of distributions such that $\lim_{n \to \infty} \|\mathbf{D}_n\|_\infty = 0$ (where $\|\mathbf{D}_n\|_\infty = \max_{i \in [n]} \{\Pr_{\mathbf{r} \sim \mathbf{D}_n}[\mathbf{r} = i]\}$). Then, for any $\delta \in (0, \frac{1}{4})$ and a sufficiently large $n \in \mathbb{N}$, the property $\Pi = \{\mathbf{D}_n\}$ is $\mathcal{F}_\delta$-closed.*

Proof. Let $\delta \in (0, \frac{1}{4})$, and let $n \in \mathbb{N}$ be sufficiently large such that for every $i \in [n]$ it holds that $\mathbf{D}_n(i) \leq \frac{\delta}{30}$. To prove that $\Pi = \{\mathbf{D}_n\}$ is $\mathcal{F}_\delta$-closed we rely on Fact 2.2: For every $\mathbf{X} \notin \{\mathbf{D}_n\} \cup \mathcal{F}_\delta(\{\mathbf{D}_n\})$, we show that there exists $\mathbf{Z} \in \mathcal{F}_\delta(\{\mathbf{D}_n\})$ such that $\Delta(\mathbf{X}, \mathbf{Z}) < \delta$. Throughout the proof we simplify the notation by denoting $\mathbf{D} = \mathbf{D}_n$. Also, for every distribution $\mathbf{X}$, we denote the probabilistic mass of $i \in [n]$ under $\mathbf{X}$ by $\mathbf{X}_i \stackrel{\text{def}}{=} \mathbf{X}(i)$.

High-level overview. Let $\mathbf{X} \notin \{\mathbf{D}\} \cup \mathcal{F}_\delta(\{\mathbf{D}\})$, and denote $\Delta(\mathbf{X}, \mathbf{D}) = \alpha\delta$, where $\alpha \in (0, 1)$. We will show an explicit construction of a distribution $\mathbf{Z}$ that satisfies the following two requirements:

1. $\Delta(\mathbf{Z}, \mathbf{X}) < \delta$.
2. $\Delta(\mathbf{Z}, \mathbf{D}) - \Delta(\mathbf{X}, \mathbf{D}) \geq (1 - \alpha) \cdot \delta$.

Note that Requirement (2) is equivalent to the requirement that $\Delta(\mathbf{Z}, \mathbf{D}) \geq \delta$ (i.e., $\mathbf{Z} \in \mathcal{F}_\delta(\{\mathbf{D}_n\})$). For the distribution $\mathbf{Z}$ that we construct, and every $i \in [n]$, let

$$\begin{aligned}\texttt{Change}(i) &= |\mathbf{Z}_i - \mathbf{X}_i| \\ \texttt{Farther}(i) &= |\mathbf{Z}_i - \mathbf{D}_i| - |\mathbf{X}_i - \mathbf{D}_i|\end{aligned}$$

In words, $\texttt{Change}(i)$ is the magnitude of change made in the probabilistic mass of $i \in [n]$, and $\texttt{Farther}(i)$ reflects how farther $\mathbf{Z}$ is from $\mathbf{D}$, compared to the distance of $\mathbf{X}$ from $\mathbf{D}$, in $i \in [n]$. Thus, Requirement (1) is equivalent to the requirement that $\sum_i \texttt{Change}(i) < \delta$, and Requirement (2) is equivalent to the requirement that $\sum_i \texttt{Farther}(i) \geq (1-\alpha) \cdot \delta$. Intuitively, when constructing $\mathbf{Z}$, for every $i \in [n]$ we want that $\texttt{Farther}(i)$ will be as large as possible, compared to $\texttt{Change}(i)$.

For the construction itself we will rely on the following lemma, which asserts the existence of a set $\texttt{LIGHT} \subseteq [n]$ with useful properties. (The name $\texttt{LIGHT}$ is since the elements in this set have upper bounded probabilistic mass; see the exact definition in the actual proof below.)

Lemma 5.2.1. *There exists a set* $\texttt{LIGHT} \subseteq [n]$ *such that:*

1. *For every distribution* $\mathbf{Z}$ *and* $j \in \texttt{LIGHT}$, *if* $\mathbf{Z}_j \leq \min\{\mathbf{X}_j, \frac{1}{2} \cdot \mathbf{D}_j\}$, *then* $\texttt{Farther}(j) \geq \frac{1-\alpha}{1+\alpha} \cdot \texttt{Change}(j)$.

2. *The probabilistic mass of* $\texttt{LIGHT}$ *under* $\mathbf{X}$ *is substantial; in particular,* $\Pr_{j \sim \mathbf{X}}[j \in \texttt{LIGHT}] > \frac{1}{2}$.

In high level, our construction of $\mathbf{Z}$ is as follows. We first initiate $\mathbf{Z} = \mathbf{X}$, and let $\Delta < \frac{\delta}{2}$ be a parameter, which will be determined later. Since $\mathbf{Z} = \mathbf{X} \neq \mathbf{D}$, there exists $i^{\mathtt{UP}} \in [n]$ such that $\mathbf{Z}_{i^{\mathtt{UP}}} > \mathbf{D}_{i^{\mathtt{UP}}}$. We increase the probabilistic mass of $\mathbf{Z}_{i^{\mathtt{UP}}}$ by Δ, and since after the modification it holds that $\mathbf{Z}_{i^{\mathtt{UP}}} > \mathbf{X}_{i^{\mathtt{UP}}} > \mathbf{D}_{i^{\mathtt{UP}}}$, we get that $\texttt{Farther}(i^{\mathtt{UP}}) = \texttt{Change}(i^{\mathtt{UP}})$. Now, according to the aforementioned lemma, there exists a set $S \subseteq \texttt{LIGHT}$ with overall probabilistic mass of more than $\frac{\delta}{2} > \Delta$. We thus decrease the overall probabilistic mass of $\mathbf{Z}$ in S by Δ, while ensuring that for every $j \in S$ it holds that $\mathbf{Z}_j$ is sufficiently small, such that, according to the lemma, after the decrease of mass it holds that $\texttt{Farther}(j) \geq \frac{1-\alpha}{1+\alpha} \cdot \texttt{Change}(j)$.

Since we changed an overall $2 \cdot \Delta$ probabilistic mass of $\mathbf{X}$ to obtain $\mathbf{Z}$, we get that $\sum_{i \in [n]} \texttt{Change}(i) = 2 \cdot \Delta < \delta$. Also,

$$\begin{aligned}
\sum_{i \in [n]} \texttt{Farther}(i) &= \texttt{Farther}(i^{\mathtt{UP}}) + \sum_{j \in S} \texttt{Farther}(j) \\
&\geq \texttt{Change}(i^{\mathtt{UP}}) + \frac{1-\alpha}{1+\alpha} \cdot \left(\sum_{j \in S} \texttt{Change}(j) \right) \\
&= \left(1 + \frac{1-\alpha}{1+\alpha} \right) \cdot \Delta
\end{aligned}$$

and for $\Delta \geq \frac{1}{2}(1-\alpha)(1+\alpha) \cdot \delta$, this expression is at least $(1-\alpha) \cdot \delta$.

Actually, we show two different constructions for $\mathbf{Z}$, according to the distance of $\mathbf{X}$ from $\mathbf{D}$. These two different constructions are both of the form depicted above, but they differ in their choice of Δ, and in the way they decrease the probabilistic mass in the set S. Note that our analysis mandates that

$$\frac{1}{2}(1-\alpha)(1+\alpha) \cdot \delta \leq \Delta < \frac{\delta}{2} \tag{5.1}$$

If $\alpha \geq \frac{2}{3}$ (i.e., $\mathbf{X}$ is relatively far from $\mathbf{D}$), then the interval for possible values of Δ in Eq. (5.1) is quite large. In this case we can set Δ to be slightly larger than $\frac{1}{2}(1-\alpha)(1+\alpha)\cdot\delta$, and the construction of $\mathbf{Z}$ will be relatively simple. However, if $\alpha < \frac{2}{3}$, the interval for Δ in Eq. (5.1) might be arbitrarily small. Actually, in this case we set $\Delta = \frac{1}{2}(1-\alpha)(1+\alpha)\cdot\delta$, but we need to be quite careful when decreasing mass from elements in S. Details follow.

The actual proof of Proposition 5.2. We start by proving the two items of Lemma 5.2.1 and another technical fact. Let

$$\texttt{LIGHT} \stackrel{\text{def}}{=} \{j \in [n] : \mathbf{X}_j \leq (1+2\alpha\delta)\cdot\mathbf{D}_j\}$$

Claim 5.2.2 *(Item 1 in Lemma 5.2.1). For any distribution $\mathbf{Z}$ and $j \in \texttt{LIGHT}$, if $\mathbf{Z}_j \leq \min\{\mathbf{X}_j, \frac{1}{2}\cdot\mathbf{D}_j\}$, then*

$$\texttt{Farther}(j) \geq \frac{1-\alpha}{1+\alpha}\cdot\texttt{Change}(j)$$

Proof. Let $\mathbf{Z}$ and $j \in \texttt{LIGHT}$ such that $\mathbf{Z}_j \leq \min\{\mathbf{X}_j, \frac{1}{2}\cdot\mathbf{D}_j\}$. If $\mathbf{X}_j \leq \mathbf{D}_j$, then

$$\texttt{Farther}(j) = |\mathbf{Z}_j - \mathbf{D}_j| - |\mathbf{X}_j - \mathbf{D}_j| = \mathbf{X}_j - \mathbf{Z}_j = \texttt{Change}(j)$$

and we are done.

Otherwise, it holds that $\mathbf{X}_j > \mathbf{D}_j$, and since $j \in \texttt{LIGHT}$, it follows that $\mathbf{D}_j < \mathbf{X}_j \leq (1+2\alpha\delta)\cdot\mathbf{D}_j$. In particular, in this case $\mathbf{D}_j \neq 0$. Note that $\mathbf{X}_j - \mathbf{D}_j \leq 2\alpha\delta\cdot\mathbf{D}_j$, whereas since $\mathbf{Z}_j \leq \frac{1}{2}\cdot\mathbf{D}_j$, it holds that $\mathbf{D}_j - \mathbf{Z}_j \geq \frac{1}{2}\cdot\mathbf{D}_j$. Also recall that $\delta < \frac{1}{4}$. Therefore,

$$\frac{\mathbf{X}_j - \mathbf{D}_j}{\mathbf{D}_j - \mathbf{Z}_j} \leq \frac{2\alpha\delta\cdot\mathbf{D}_j}{\mathbf{D}_j/2} = 4\alpha\delta < \alpha\ . \tag{5.2}$$

Now, relying on Eq. (5.2), we deduce that

$$\mathbf{X}_j - \mathbf{Z}_j = (\mathbf{X}_j - \mathbf{D}_j) + (\mathbf{D}_j - \mathbf{Z}_j) < (1+\alpha)\cdot(\mathbf{D}_j - \mathbf{Z}_j) \tag{5.3}$$

and thus we get that

$$\begin{aligned}
\texttt{Farther}(j) &= (\mathbf{D}_j - \mathbf{Z}_j) - (\mathbf{X}_j - \mathbf{D}_j) && \text{(since } \mathbf{X}_j > \mathbf{D}_j > \mathbf{Z}_j\text{)}\\
&> (1-\alpha)\cdot(\mathbf{D}_j - \mathbf{Z}_j) && \text{(according to (5.2))}\\
&> \frac{1-\alpha}{1+\alpha}\cdot(\mathbf{X}_j - \mathbf{Z}_j) && \text{(according to (5.3))}\\
&= \frac{1-\alpha}{1+\alpha}\cdot\texttt{Change}(j)\ .
\end{aligned}$$

□

Claim 5.2.3 *(Item 2 in Lemma 5.2.1). It holds that $\sum_{j\in\texttt{LIGHT}} \mathbf{X}_j \geq \frac{1}{2}$.*

Proof. Let $\texttt{HEAVY} = [n] \setminus \texttt{LIGHT}$, and note that it suffices to prove that $\sum_{i \in \texttt{HEAVY}} \mathbf{X}_i < \frac{1}{2}$. For every $i \in \texttt{HEAVY}$, it holds that $\mathbf{X}_i - \mathbf{D}_i > 2\alpha\delta \cdot \mathbf{D}_i$ (i.e., $\mathbf{D}_i < \frac{\mathbf{X}_i - \mathbf{D}_i}{2\alpha\delta}$). Let $\Delta^+ \stackrel{\text{def}}{=} \sum_{i : \mathbf{X}_i > \mathbf{D}_i} \mathbf{X}_i - \mathbf{D}_i$, and note that $\Delta^+ = \frac{\Delta(\mathbf{X}, \mathbf{D})}{2} = \frac{\alpha\delta}{2}$. Also note that $\texttt{HEAVY} \subseteq \{i : \mathbf{X}_i > \mathbf{D}_i\}$. It follows that

$$\begin{aligned}
\sum_{i \in \texttt{HEAVY}} \mathbf{X}_i &= \sum_{i \in \texttt{HEAVY}} (\mathbf{X}_i - \mathbf{D}_i) + \sum_{i \in \texttt{HEAVY}} \mathbf{D}_i \\
&< \left(1 + \frac{1}{2\alpha\delta}\right) \cdot \sum_{i \in \texttt{HEAVY}} (\mathbf{X}_i - \mathbf{D}_i) \\
&\le \left(1 + \frac{1}{2\alpha\delta}\right) \cdot \Delta^+ .
\end{aligned}$$

Recall that $\alpha < 1$ and $\delta < \frac{1}{4}$, and thus $\left(1 + \frac{1}{2\alpha\delta}\right) \cdot \Delta^+ = \left(\frac{1}{2} + \frac{1}{4\alpha\delta}\right) \cdot \alpha\delta < \frac{1}{2}$. $\square$

Fact 5.2.4. *For every $i \in [n]$, there exists a set $S \subseteq \texttt{LIGHT} \setminus \{i\}$ such that $\frac{1}{3} \cdot \delta \le \sum_{j \in S} \mathbf{X}_j < \frac{1}{2} \cdot \delta$.*

Proof. According to Claim 5.2.3, and since every $i \in [n]$ satisfies $\mathbf{D}_i \le \frac{\delta}{30}$, it follows that $\sum_{j \in \texttt{LIGHT} \setminus \{i\}} \mathbf{X}_j > \frac{1}{2} - \frac{\delta}{30} > \frac{\delta}{3}$. Also, for every $j \in \texttt{LIGHT}$ it holds that

$$\begin{aligned}
\mathbf{X}_j &\le (1 + 2\alpha\delta) \cdot \mathbf{D}_j && \text{(since } j \in \texttt{LIGHT}\text{)} \\
&\le (1 + 2\alpha\delta) \cdot \frac{\delta}{30} && \text{(since } \mathbf{D}_j \le \tfrac{\delta}{30}\text{)} \\
&< \frac{1}{6} \cdot \delta . && \text{(since } \delta < \tfrac{1}{4}\text{)}
\end{aligned}$$

We construct S by initiating $S = \emptyset$, and adding elements from $\texttt{LIGHT} \setminus \{i\}$ to S until $\sum_{j \in S} \mathbf{X}_j \ge \frac{1}{3} \cdot \delta$. Since $\sum_{j \in \texttt{LIGHT} \setminus \{i\}} \mathbf{X}_j > \frac{\delta}{3}$, there is sufficient probabilistic mass in $\texttt{LIGHT} \setminus \{i\}$ to construct a set S with $\sum_{j \in S} \mathbf{X}_j \ge \frac{1}{3} \cdot \delta$. Also, since the mass of every element in $\texttt{LIGHT} \setminus \{i\}$ is at most $\frac{1}{6} \cdot \delta$, the construction yields a set S such that $\sum_{j \in S} \mathbf{X}_j < \frac{1}{3} \cdot \delta + \frac{1}{6} \cdot \delta = \frac{1}{2} \cdot \delta$. $\square$

We now split the rest of the proof (of Proposition 5.2) into two cases, depending on $\Delta(\mathbf{X}, \mathbf{D})$. In each case we prove the existence of a suitable $\mathbf{Z}$ using a different construction.

Case 1: Assuming $\Delta(\mathbf{X}, \mathbf{D}) \ge \frac{2}{3} \cdot \delta$. In this case $\alpha \ge \frac{2}{3}$, and we set Δ such that it might be slightly larger than the lower bound implied by Eq. (5.1). The construction of the distribution $\mathbf{Z}$ is as follows.

Construction 5.2.5. *(construction of the distribution $\mathbf{Z}$ when $\Delta(\mathbf{X}, \mathbf{D}) \ge \frac{2}{3} \cdot \delta$).*

1. Let $\mathbf{Z} = \mathbf{X}$, and let:
   (a) $i^{\text{UP}} = \text{argmax}_{i \in [n]} \{\mathbf{X}_i - \mathbf{D}_i\}$.
   (b) $S \subseteq \texttt{LIGHT} \setminus \{i^{\text{UP}}\}$ such that $\frac{1}{3} \cdot \delta \le \sum_{j \in S} \mathbf{X}_j < \frac{1}{2} \cdot \delta$.
   (c) $\Delta = \sum_{i \in S} \mathbf{X}_i$.
2. (increase Δ mass) Set $\mathbf{Z}_{i^{\text{UP}}} = \mathbf{X}_{i^{\text{UP}}} + \Delta$.

3. (decrease Δ mass) For every $j \in S$ set $\mathbf{Z}_j = 0$.

According to Fact 5.2.4, a suitable set S exists for Step (1b). Also, note that $\mathbf{Z}$ is a distribution, since we obtained it by removing a probabilistic mass of Δ from $\mathbf{X}$ at S, and adding the same magnitude of mass to i^{UP}. Since $\mathbf{X} \neq \mathbf{D}$, and $i^{\mathrm{UP}} = \mathrm{argmax}_{i\in[n]}\{\mathbf{X}_i - \mathbf{D}_i\}$, then $\mathbf{Z}_{i^{\mathrm{UP}}} > \mathbf{X}_{i^{\mathrm{UP}}} > \mathbf{D}_{i^{\mathrm{UP}}}$, implying that $\mathtt{Farther}(i^{\mathrm{UP}}) = \mathtt{Change}(i^{\mathrm{UP}}) = \Delta$. Furthermore, since for every $j \in S$ it holds that j and $\mathbf{Z}$ satisfy the conditions in Claim 5.2.2, then for every $j \in S$ it holds that $\mathtt{Farther}(j) \geq 0$. Thus,

$$\Delta(\mathbf{Z},\mathbf{D}) - \Delta(\mathbf{X},\mathbf{D}) = \mathtt{Farther}(i^{\mathrm{UP}}) + \sum_{j\in S}\mathtt{Farther}(j) \geq \mathtt{Change}(i^{\mathrm{UP}})$$

and $\mathtt{Change}(i^{\mathrm{UP}}) = \Delta \geq \frac{1}{3}\cdot\delta \geq \delta - \Delta(\mathbf{X},\mathbf{D})$. It follows that $\Delta(\mathbf{Z},\mathbf{D}) \geq \delta$, implying that $\mathbf{Z} \in \mathcal{F}_\delta(\{\mathbf{D}\})$. Since we added and removed $2\cdot\Delta$ probabilistic mass from $\mathbf{X}$ to obtain $\mathbf{Z}$, it also holds that $\Delta(\mathbf{Z},\mathbf{X}) = 2\cdot\Delta < \delta$.

Case 2: Assuming $\Delta(\mathbf{X},\mathbf{D}) < \frac{2}{3}\cdot\delta$. In this case $\alpha = \frac{\Delta(\mathbf{X},\mathbf{D})}{\delta} < \frac{2}{3}$, and $\mathbf{X}$ might be arbitrarily close to $\mathbf{D}$. In the latter case, the interval for values of Δ implied by Eq. (5.1) might be arbitrarily small. We thus set Δ to exactly match the lower bound of this interval. The construction of the distribution $\mathbf{Z}$ is as follows.

Construction 5.2.6. *(construction of the distribution $\mathbf{Z}$ when $\Delta(\mathbf{X},\mathbf{D}) < \frac{2}{3}\cdot\delta$).*

1. Let $\mathbf{Z} = \mathbf{X}$ and $\Delta = \frac{1}{2}\cdot(1-\alpha)\cdot(1+\alpha)\cdot\delta$.
2. (increase Δ mass) For $i^{\mathrm{UP}} = \mathrm{argmax}_{i\in[n]}\{\mathbf{X}_i - \mathbf{D}_i\}$ set $\mathbf{Z}_{i^{\mathrm{UP}}} = \mathbf{X}_{i^{\mathrm{UP}}} + \Delta$.
3. (decrease Δ mass)
   (a) Let $S = \emptyset$.
   (b) While $\sum_{j\in S}\mathbf{X}_j < \Delta$ do $S \leftarrow \mathrm{argmax}_{i\in\mathtt{LIGHT}\setminus(S\cup\{i^{\mathrm{UP}}\})}\{\mathbf{X}_i\}$.
   (c) For every $j \in S$ set $\mathbf{Z}_j = \frac{\sum_{j\in S}\mathbf{X}_j - \Delta}{|S|}$.

The following claim specifies conditions that Construction 5.2.6 satisfies, which we will later rely on.

Claim 5.2.7. *Construction 5.2.6 is well-defined, and it produces a distribution $\mathbf{Z}$ such that:*

1. *For $i^{\mathrm{UP}} \in [n]$ it holds that $\mathbf{Z}_{i^{\mathrm{UP}}} = \mathbf{X}_{i^{\mathrm{UP}}} + \Delta$ and $\mathbf{X}_{i^{\mathrm{UP}}} > \mathbf{D}_{i^{\mathrm{UP}}}$.*
2. *For $S \subseteq \mathtt{LIGHT}$ it holds that:*
   (a) $\sum_{j\in S}\mathbf{X}_j - \mathbf{Z}_j = \Delta$.
   (b) For every $j \in S$ it holds that $\mathbf{Z}_j \leq \min\{\mathbf{X}_j, \frac{1}{2}\cdot\mathbf{D}_j\}$.

Before proving Claim 5.2.7, let us assume for a moment that it is correct, and see how it implies that $\mathbf{Z} \in \mathcal{F}_\delta(\{\mathbf{D}\})$ and $\Delta(\mathbf{X},\mathbf{Z}) < \delta$. First, since $\Delta = \frac{1}{2}(1-\alpha)(1+\alpha)\cdot\delta < \delta/2$, it holds that $\Delta(\mathbf{Z},\mathbf{X}) = 2\cdot\Delta < \delta$. Now, since $\mathbf{Z}_{i^{\text{UP}}} > \mathbf{X}_{i^{\text{UP}}} > \mathbf{D}_{i^{\text{UP}}}$, it follows that $\texttt{Farther}(i^{\text{UP}}) = \texttt{Change}(i^{\text{UP}})$. Also, since for every $j \in S$ it holds that j and $\mathbf{Z}$ satisfy the conditions in Claim 5.2.2, it follows that $\texttt{Farther}(j) \geq \frac{1-\alpha}{1+\alpha}\cdot\texttt{Change}(j)$. Therefore,

$$\begin{aligned}
\Delta(\mathbf{Z},\mathbf{D}) - \Delta(\mathbf{X},\mathbf{D}) &= \texttt{Farther}(i^{\text{UP}}) + \sum_{j\in S}\texttt{Farther}(j) \\
&\geq \texttt{Change}(i^{\text{UP}}) + \frac{1-\alpha}{1+\alpha}\cdot\sum_{j\in S}\texttt{Change}(j) \\
&= \left(\frac{1-\alpha}{1+\alpha}+1\right)\cdot\Delta \\
&= (1-\alpha)\cdot\delta
\end{aligned}$$

which implies that $\Delta(\mathbf{Z},\mathbf{D}) \geq (1-\alpha)\cdot\delta + \Delta(\mathbf{X},\mathbf{D}) = \delta$. Hence $\mathbf{Z} \in \mathcal{F}_\delta(\{\mathbf{D}\})$ and $\Delta(\mathbf{Z},\mathbf{X}) < \delta$. To finish the proof it is thus left to prove Claim 5.2.7.

Proof of Claim 5.2.7. To see that Construction 5.2.6 is well-defined, note that according to Fact 5.2.4 there is sufficient probability mass in $\texttt{LIGHT} \setminus \{i^{\text{UP}}\}$ in order for the loop in Step (3b) of Construction 5.2.6 to complete successfully. Also, the first part of Condition (1) follows since the probabilistic mass of i^{UP} only changes in Step (2); and the second part of Condition (1) follows since $\mathbf{X} \neq \mathbf{D}$ and by the definition of i^{UP}.

Condition (2a) follows since

$$\sum_{j\in S}\mathbf{X}_j - \mathbf{Z}_j = \left(\sum_{j\in S}\mathbf{X}_j\right) - |S|\cdot\frac{\sum_{j\in S}\mathbf{X}_j - \Delta}{|S|} = \Delta\,.$$

For Condition (2b), we first need the following fact.

Fact 5.3. *For every $j \in S$ it holds that $\sum_{j'\in S}\mathbf{X}_{j'} - \Delta < \mathbf{X}_j$.*

Proof. Denote the last element that was inserted into S in Step (3b) by k, and note that $\mathbf{X}_k \leq \mathbf{X}_j$. Assume towards a contradiction that $\sum_{j'\in S}\mathbf{X}_{j'} - \Delta \geq \mathbf{X}_j$. It follows that $\sum_{j'\in S}\mathbf{X}_{j'} - \mathbf{X}_k \geq \sum_{j'\in S}\mathbf{X}_{j'} - \mathbf{X}_j \geq \Delta$. However, in this case, k would not have been added to S, since after the previous-to-last iteration of Step (3b), the overall probabilistic mass of elements in S would have already exceeded Δ. □

Now, let $j \in S$, and we show that $\mathbf{Z}_j < \min\{\mathbf{X}_j, \frac{1}{2}\cdot\mathbf{D}_j\}$.

- $\mathbf{Z}_j < \mathbf{X}_j$: Since $\mathbf{Z}_j = \frac{\sum_{j'\in S}\mathbf{X}_{j'} - \Delta}{|S|} \leq \sum_{j'\in S}\mathbf{X}_{j'} - \Delta < \mathbf{X}_j$.
- $\mathbf{Z}_j < \frac{1}{2}\cdot\mathbf{D}_j$: Recall that $\alpha < \frac{2}{3}$, and thus $\Delta > \frac{1}{2}\cdot\frac{1}{3}\cdot\delta = \delta/6$. Also, since $S \subseteq \texttt{LIGHT}$, for every $i \in S$ it holds that $\mathbf{X}_i \leq (1+2\cdot\alpha\delta)\cdot\mathbf{D}_i \leq \frac{\delta}{20}$ (where the second inequality relies on the fact that $\mathbf{D}_i \leq \frac{\delta}{30}$ for every $i \in [n]$, and on the fact that $2\cdot\alpha\delta < \frac{1}{2}$). It follows that

$$|S| \geq \frac{\Delta}{\max_{i\in S}\{\mathbf{X}_i\}} > \frac{\delta/6}{\delta/20} > 3\,.$$

Therefore,

$$\mathbf{Z}_j = \frac{\sum_{j' \in S} \mathbf{X}_{j'} - \Delta}{|S|} < \frac{\mathbf{X}_j}{3} \le \frac{1+2\alpha\delta}{3} \cdot \mathbf{D}_j$$

and note that $\frac{1+2\alpha\delta}{3} < \frac{1}{3} + \frac{1}{6} = \frac{1}{2}$.

Also, $\mathbf{Z}$ is a distribution, since by Conditions (1) and (2a) it holds that $\sum_{i\in[n]} \mathbf{Z}_i = 1$, and for every $i \in [n]$ it holds that $\mathbf{Z}_i \ge 0$. □

This completes the proof of Proposition 5.2. ■

The following proposition asserts that for distribution families $\{\mathbf{D}_n\}_{n\in\mathbb{N}}$ in which each support element has $\Omega(1)$ probability mass it holds that $\{\mathbf{D}_n\}$ is $\mathcal{F}_\delta$-closed for every sufficiently small $\delta > 0$.

Proposition 5.4 *(distributions with bounded probabilistic mass on elements in their support). For $\rho > 0$, let $\{\mathbf{D}_n\}_{n\in\mathbb{N}}$ be a distribution family such that for every $n \in \mathbb{N}$ and $i \in [n]$ it holds that either $\mathbf{D}_n(i) \ge \rho$ or $\mathbf{D}_n(i) = 0$. Then, for any $\delta \in (0, \rho)$ and every $n \in \mathbb{N}$, the property $\Pi = \{\mathbf{D}_n\}$ is $\mathcal{F}_\delta$-closed.*

Proof. Let $\delta \in (0, \rho)$ and $n \in \mathbb{N}$. We prove that $\Pi = \{\mathbf{D}_n\}$ is $\mathcal{F}_\delta$-closed, relying on Fact 2.2: For $\mathbf{X} \notin \{\mathbf{D}_n\} \cup \mathcal{F}_\delta(\{\mathbf{D}_n\})$, we show that there exists $\mathbf{Z} \in \mathcal{F}_\delta(\{\mathbf{D}_n\})$ such that $\Delta(\mathbf{X}, \mathbf{Z}) < \delta$.

Since $\mathbf{X} \ne \mathbf{D}_n$ and since $\mathbf{X}$ and $\mathbf{D}_n$ are distributions, there exist $i, j \in [n]$ such that $\mathbf{X}(i) > \mathbf{D}_n(i)$ and $\mathbf{X}(j) < \mathbf{D}_n(j)$. Since $\mathbf{X} \notin \mathcal{F}_\delta(\{\mathbf{D}_n\})$ it holds that

$$\mathbf{D}_n(j) - \mathbf{X}(j) \le \frac{\Delta(\mathbf{X}, \mathbf{D}_n)}{2} < \rho/2$$

and thus $\mathbf{X}(j) > \mathbf{D}_n(j) - \rho/2 \ge \rho/2$, where the last inequality is by the hypothesis that $\mathbf{D}_n(j) \ge \rho$. Similarly, $\mathbf{X}(i) - \mathbf{D}_n(i) < \rho/2$. Now, note that $\mathbf{D}_n(i) \le 1 - \rho$: This is the case since if $\mathbf{D}_n$ is supported on a single element $k \in [n]$ then $\mathbf{D}_n(i) = 0$, and otherwise $\mathbf{D}_n$ is supported on at least two elements each having mass at least ρ, and thus for every $k \in [n]$ it holds that $\mathbf{D}_n(k) \le 1 - \rho$. It follows that $\mathbf{X}(i) < 1 - \rho/2$.

Let $\Delta = \frac{1}{2} \cdot (\delta - \Delta(\mathbf{X}, \mathbf{D}_n))$ and note that $0 < \Delta < \rho/2$. We define $\mathbf{Z}$ as follows: $\mathbf{Z}(i) = \mathbf{X}(i) + \Delta < 1$, and $\mathbf{Z}(j) = \mathbf{X}(j) - \Delta > 0$, and for every $k \notin \{i, j\}$ it holds that $\mathbf{Z}(k) = \mathbf{X}(k)$. Note that $\mathbf{Z}$ is a distribution, since the probabilistic mass of every $i \in [n]$ is in $[0, 1]$, and $\sum_{i\in[n]} \mathbf{Z}_i = \sum_{i\in[n]} \mathbf{X}_i = 1$. Furthermore, $\Delta(\mathbf{Z}, \mathbf{X}) = 2 \cdot \Delta < \delta$, and

$$\begin{aligned}\Delta(\mathbf{Z}, \mathbf{D}_n) &= \Delta(\mathbf{X}, \mathbf{D}_n) + |\mathbf{Z}(i) - \mathbf{D}_n(i)| + |\mathbf{Z}(j) - \mathbf{D}_n(j)| \\ &= \Delta(\mathbf{X}, \mathbf{D}_n) + 2 \cdot \Delta \\ &= \delta\end{aligned}$$

which implies that $\mathbf{Z} \in \mathcal{F}_\delta(\{\mathbf{D}_n\})$, as needed. ■

The query complexity of the distribution testing problem is $\tilde{\Theta}(\sqrt{n})$ (see, e.g., [Can15]). It follows that the query complexity of the dual problem is lower bounded by $\Omega(\sqrt{n})$. Also, for every distribution family from the classes of distributions described in Propositions 5.2 and 5.4, the query complexity of the dual problem is $\tilde{O}(\sqrt{n})$.

6 Testing graphs that are far from having a property in the dense graph model

The main result in this section is Theorem 1.10, which asserts that the dual problem of testing k-colorability in the dense graph model is different than the original problem, but nevertheless the dual problem can be tested with $O(1)$ queries. We also prove Proposition 1.11, which asserts that the dual problems of testing the existence of a ρ-clique and of testing graph isomorphism are different from the original problems.

Let us first recall the setting of property testing in the dense graph model. The metric space in this model consists of simple, undirected graphs, and the absolute distance between two graphs on v vertices is the size of the symmetric difference between their edge sets. A property of graphs is a set of graphs closed under taking isomorphisms of the graphs. A graph on v vertices is represented by a corresponding string $x \in \{0,1\}^n$, where $n = \binom{v}{2}$, such that the i^{th} edge is included in the graph if and only if $x_i = 1$. A property of graphs is accordingly denoted by $\Pi = \{\Pi_n\}_{n \in \mathcal{N}}$, where $\mathcal{N} = \left\{\binom{v}{2} : v \in \mathbb{N}\right\}$. The testing problem is as follows: An ϵ-tester gets oracle access to $x \in \binom{v}{2}$, corresponding to an input graph over v vertices, and needs to decide whether the graph has the property, or whether it is $\epsilon \cdot \binom{v}{2}$-far from any graph having the property.

6.1 Testing the property of being far from k-colorable

In this section we study the dual problem of k-colorability: For every $\epsilon > 0$, we are interested in the problem of ϵ-testing the set of graphs that are $\left(\epsilon \cdot \binom{v}{2}\right)$*-far from being k-colorable*, where v is the number of vertices in the graph. We first show that this problem is different from its original problem, and then show that its query complexity is $O(1)$.

Proposition 6.1 *(the set of k-colorable graphs is not $\mathcal{F}_\delta$-closed). For any $k \geq 2$ and $v \geq k+1$, let $n = \binom{v}{2}$ and $\delta \geq 2$. Then, the set of graphs over v vertices that are k-colorable, denoted by $\Pi_n \subseteq \{0,1\}^n$, is not $\mathcal{F}_\delta$-closed.*

Proof. We show a graph G such that $\Delta(G, \Pi_n) = 1$, and all neighbors of G are either in Π_n or adjacent to Π_n. It follows that for every $\delta \geq 2$, there does not exist a path (i.e., a sequence of graphs such that their bit-string representations induce a path in $\{0,1\}^n$) from G to $\mathcal{F}_\delta(\Pi_n)$ such that every graph subsequent to G on the path is neither in Π_n nor adjacent to Π_n. Relying on Proposition 2.3, this implies that Π_n is not $\mathcal{F}_\delta$-closed.

Let G be a graph that contains a single clique on $k+1$ vertices, and no other edges. Note that G is not k-colorable, but that removing any edge from G turns G into a k-colorable graph; thus, $\Delta(G, \Pi_n) = 1$. Now, let G' be a graph that disagrees with G on a single edge (i.e., $\Delta(G, G') = 1$); we want to prove that $\Delta(G', \Pi_n) \leq 1$. As mentioned, removing any edge from G turns it into a k-colorable graph; thus, it suffices to show that any graph G' obtained by adding an edge to G satisfies $\Delta(G', \Pi_n) \leq 1$. To see this, note that any such graph is comprised of a $(k+1)$-clique (the same one that existed in G) and an additional edge; by removing any edge from the clique, we obtain a k-colorable graph. ■

We now show that *the query complexity of the dual problem of k-colorability is* $O(1)$. Loosely speaking, our first step towards this result is to prove that graphs that are far-from-far from being k-colorable are relatively close to being k-colorable. Specifically, for every sufficiently small $\epsilon > 0$ we show that there exists $\alpha \in (0,1)$ such that for every sufficiently large $n \in \mathcal{N}$ it holds that $\mathcal{F}_{\epsilon \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n)) \subseteq \{G : \Delta(G, \Pi_n) \leq (\alpha \cdot \epsilon) \cdot n\}$. The meaning of this result is that the dual problem of k-colorability reduces to the corresponding tolerant testing problem.

Proposition 6.2 *(graphs that are far-from-far from being k-colorable are relatively close to being k-colorable). Let* $\Pi = \{\Pi_n\}_{n \in \mathcal{N}}$ *be the property of k-colorable graphs, where* $\Pi_n \subseteq \{0,1\}^n$ *consists of graphs over v vertices such that* $n = \binom{v}{2}$*. Then, there exists* $\alpha \in (0,1)$ *such that for every sufficiently small* $\epsilon > 0$ *and sufficiently large* $n \in \mathcal{N}$ *it holds that* $\mathcal{F}_{\epsilon \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n)) \subseteq \{G : \Delta(G, \Pi_n) \leq (\alpha \cdot \epsilon) \cdot n\}$.

Proof. We rely on Proposition 2.5, which implies the following: If there exists $m = O(1)$ such that every graph G satisfying $\Delta(G, \Pi_n) < \delta$ can be modified into a graph G' that is farther away from Π_n, compared to G, by adding and/or removing at most m edges from G, then the distance of any graph in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$ from Π_n is at most $\left(1 - \frac{1}{m}\right) \cdot \delta$. It thus suffices to show a way to modify every graph G that is not δ-far from being k-colorable into a graph G' that is farther away from being k-colorable, with only $O(1)$ changes.

Throughout the proof, it will be convenient to think of the number of vertices, denoted by v, as the primary asymptotic parameter (recall that $n = \binom{v}{2}$). Let $\alpha = \left(1 - \frac{1}{\binom{k+1}{2}}\right)$ and $\epsilon < \frac{1}{8 \cdot k^2 \cdot (k+1)}$. For a sufficiently large $v \in \mathbb{N}$, let $n = \binom{v}{2}$ and $\delta = \epsilon \cdot n$. According to Proposition 2.5, it suffices to construct, for any graph G with v vertices satisfying $\Delta(G, \Pi_n) < \delta$, a corresponding graph G' such that $\Delta(G, G') < \binom{k+1}{2}$ and $\Delta(G', \Pi_n) \geq \Delta(G, \Pi_n) + 1$.

In the following arguments, given a graph G and a k-partition of its vertices, we say that an edge (u, u') is a *violating edge* if u and u' are in the same cell of the partition. The distance of G from being k-colorable is the minimum, over all k-partitions P of the vertices of G, of the number of violating edges for P. We first prove that G has an independent set of size $k+1$.

Lemma 6.2.1. *Let G be a graph on v vertices satisfying* $\Delta(G, \Pi_n) < \delta$*. Then, there exists an independent set of size* $k + 1$ *in G.*

Proof. Since the distance of G from being k-colorable is less than δ, there exists a k-partition of the vertices of G with less than δ violating edges. Let U be the subgraph of G corresponding to the biggest cell in this k-partition, and note that $|U| \geq v/k$. The average degree of the vertices in U is less than $2\delta/|U| \leq \frac{2 \cdot \epsilon \cdot n}{v/k} < \epsilon \cdot k \cdot v < \frac{|U|}{8 \cdot (k+1)}$. Hence, at most half of the vertices in U have degree more than $\frac{|U|}{4 \cdot (k+1)}$; by dropping these vertices, we obtain a subgraph U' such that $|U'| \geq \frac{v}{2 \cdot k}$, and every vertex in U' has degree at most $\frac{|U|}{4 \cdot (k+1)} \leq \frac{|U'|}{2 \cdot (k+1)}$.

Now, for $i = 1, ..., k+1$, we choose a vertex in U', and remove all of its neighbors from U'. This process can indeed continue for $k+1$ iterations, because after the i^{th} iteration, the number of vertices in the resulting subgraph is at least $|U'| - i \cdot \frac{|U'|}{2 \cdot (k+1)}$. To conclude, observe that the chosen $k+1$ vertices form an independent set in G. □

Let G be a graph with v vertices such that $\Delta(G, \Pi_n) < \delta$, and let I be an independent set of size $k+1$ in G. We modify G into G' by adding $\binom{k+1}{2}$ edges between all pairs of vertices in I. To see that G' is farther away from being k-colorable, compared to G, note that for any k-partition P of the vertices of G, the number of violating edges for P in G' is larger than the number of violating edges for P in G. This is the case since at least two vertices in I are in the same cell of P (because $|I| = k+1$), forming a violating edge for P in G', whereas no edges were removed when modifying G to G' (and thus all violating edges for P in G are also violating edges for P in G'). ■

We now rely on a general result of Fischer and Newman [FN07], which asserts that in the dense graph model, if a property Π is testable with $O(1)$ queries, then the corresponding tolerant testing problem can also be solved with $O(1)$ queries. Since the original problem of k-colorability can indeed be tested with $O(1)$ queries [GGR98], the tolerant testing problem can also be solved with $O(1)$ queries, which implies (using Proposition 6.2) that the dual problem can also be solved with $O(1)$ queries.

Specifically, recall the following definition of tolerant testers and result from [FN07]. (The definition replaces the standard pair of relative distances (ϵ, ϵ') with one relative distance ϵ and one multiplcative factor $\alpha < 1$ such that $\epsilon' = \alpha \cdot \epsilon$.)

Definition 6.3 *((α, ϵ)-estimation tester; see [FN07, Def. 2]). For a set Σ, and a property $\Pi = \{\Pi_n\}_{n \in \mathbb{N}}$ such that $\Pi_n \subseteq \Sigma^n$, and $\epsilon > 0$, and $\alpha \in (0,1)$, an (α, ϵ)-*estimation tester* for Π is a probabilistic algorithm T that for every $n \in \mathbb{N}$ and $x \in \Sigma^n$ satisfies the following two conditions:*

1. *If $\Delta(x, \Pi_n) \leq \alpha \cdot \epsilon \cdot n$, then $Pr[T^x(1^n) = 1] \geq \frac{2}{3}$.*
2. *If $\Delta(x, \Pi_n) \geq \epsilon \cdot n$, then $Pr[T^x(1^n) = 0] \geq \frac{2}{3}$.*

The query complexity of (α, ϵ)-estimation testers is defined analogously to Definition 2.1.

Theorem 6.4 *(testing implies estimation in the dense graph model). Let Π be a property of graphs in the dense graph model with query complexity $O(1)$. Then, for every $\epsilon > 0$ and $\alpha \in (0,1)$, there exists an (α, ϵ)-estimation tester for Π with query complexity $O(1)$.*

As mentioned above, Theorem 1.10 follows as a corollary of Proposition 6.2 and of Theorem 6.4.

6.2 Testing the property of being far from having a large clique

In this section we show that the dual problem of testing ρ-clique is different from its original problem. The dual problem is the following: For $\rho \in (0,1)$ and $\epsilon > 0$, we are interested in the problem of ϵ-testing the set of graphs that are $\left(\epsilon \cdot \binom{v}{2}\right)$*-far from having a clique of size $\rho \cdot v$*, where v is the number of vertices in the graph.

Proposition 6.5 *(the set of graphs with a clique of size $\rho \cdot v$ is not $\mathcal{F}_\delta$-closed). For any $\rho \in \left(0, \frac{1}{2}\right]$, and $\delta \geq 2$, and even $v \geq 4$, the property of graphs on v vertices containing a clique of size $\rho \cdot v$ is not $\mathcal{F}_\delta$-closed.*

Proof. For $\rho \in \left(0, \frac{1}{2}\right]$, and $\delta \geq 2$, and an even $v \geq 4$, and $n = \binom{v}{2}$, let $\Pi \subseteq \{0,1\}^n$ be the set of graphs containing a clique of size $\rho \cdot v$. Similarly to the proof of Proposition 6.1, it suffices to show a graph G such that $\Delta(G, \Pi) = 1$, and all neighbors of G are either in Π or adjacent to Π (since we can then rely on Proposition 2.3 to deduce that Π is not $\mathcal{F}_\delta$-closed).

Let $G = (V, E)$ be as follows. We bisect $V = V_1 \cup V_2$, and since $\rho \leq \frac{1}{2}$ and $v = |V|$ is even, it holds that $|V_1| = |V_2| \geq \lceil \rho \cdot v \rceil$. We define G such that it contains two vertex-disjoint "almost cliques" of size $\lceil \rho \cdot v \rceil$, one in V_1 and the other in V_2, where an "almost clique" is a clique from which one edge is omitted. Other than the two "almost cliques", G contains no additional edges. Since G contains no clique of size $\rho \cdot v$, it follows that $G \notin \Pi$. Also, since we can create such a clique in G by adding a single edge, it follows that $\Delta(G, \Pi) = 1$. Now, let G' be neighbor of G. We wish to prove that $\Delta(G', \Pi) \leq 1$.

- If G' was obtained by adding an edge to G, then either $G' \in \Pi$ (if the edge completed one of the two "almost cliques" to a clique), or, otherwise, we can add an edge to G' that completes one of the "almost cliques" to a clique, in which case $\Delta(G', \Pi) = 1$. Either way, $\Delta(G', \Pi) \leq 1$.

- Otherwise, G' was obtained by removing an edge from one of the "almost cliques". However, in this case we can still add an edge to the other "almost clique", turning it to a clique of size $\lceil \rho \cdot v \rceil$. Thus $\Delta(G', \Pi) = 1$. ■

Since for every $v \in \mathbb{N}$ and $n = \binom{v}{2}$ there exist graphs with v vertices that are $\Omega(n)$-far from having clique of size $\rho \cdot v$ (e.g., the graph with no edges), testing the dual problem of k-colorability with one-sided error requires $\Omega(n)$ queries.

6.3 Testing the property of being far from isomorphic to a graph

In this section we show that the dual problem of testing graph isomorphism is different from its original problem. For a graph G on v vertices that is predetermined and explicitly known in advance, the dual problem consists of ϵ-testing the set of graphs that are $\left(\epsilon \cdot \binom{v}{2}\right)$*-far from being isomorphic to* G.

Proposition 6.6 *(graph families that induce properties that are not $\mathcal{F}_\delta$-closed). There exists a graph family $\{G_n\}_{n \in \mathcal{N}}$ such that for every $\delta \geq 2$ and $n \in \mathcal{N}$, the property of graphs that are isomorphic to G_n is not $\mathcal{F}_\delta$-closed.*

Proof. For $v \in \mathbb{N}$ and $n = \binom{v}{2}$, let G_n be a graph with v vertices and a single edge. We will show that for every $\delta \geq 2$, the set $\Pi_n \subseteq \{0,1\}^n$ of graphs that are isomorphic to G_n is not $\mathcal{F}_\delta$-closed. First note that Π_n is exactly the set of vectors with Hamming weight 1 (i.e., $\Pi_n = B[\emptyset, 1] \setminus \{\emptyset\}$); this is the case because each vector with Hamming weight 1 represents a graph that is isomorphic to G_n, and all vectors representing graphs that are isomorphic to a given graph have the same Hamming weight. To see that Π_n is not $\mathcal{F}_\delta$-closed, note that any path from $\emptyset \notin \Pi_n \cup \mathcal{F}_\delta(\Pi_n)$ must pass through Π_n (i.e., through a vector with Hamming weight 1). Relying on Proposition 2.3, Π_n is not $\mathcal{F}_\delta$-closed. ■

Fischer and Matsliah proved [FM08] that the query complexity of this version of the graph isomorphism is $\tilde{\Theta}(\sqrt{v})$. Thus, the query complexity of the dual problem is lower bounded by $\Omega(\sqrt{v})$. Also, according to Theorem 1.5, and since the testing problem is not trivial, testing the dual problem with one-sided error requires $\Omega(n)$ queries.

7 Testing graphs that are far from having a property in the bounded-degree model

The main results in this section are Theorems 1.12 and 1.13, which assert that the dual problems of testing connectivity and of testing cycle-free graphs in the bounded-degree model are both different from the original problems, but that both dual problems can nevertheless be tested with $O(1)$ queries. We also prove Proposition 1.14, which asserts that the dual problem of testing bipartiteness in this model is not equivalent to the original problem.

Let us first recall the setting of property testing in the bounded-degree model. In this model, we fix some function $d : \mathbb{N} \to \mathbb{N}$, and the underlying metric space consists of graphs over the vertex-set $[n]$ such that the degree of every vertex in the graph is at most $d(n)$. Typically, we are interested in $d = O(1)$. The absolute distance between a pair of graphs in this model is the same as in the metric space of the dense graph model: The size of the symmetric difference of their edge-sets. [9]

A property of graphs in this model is a set of of graphs closed under taking isomorphisms of the graphs, and is denoted by $\Pi = \{\Pi_n\}_{n\in\mathbb{N}}$ such that Π_n consists of graphs over the vertex-set $[n]$. A testing scenario for a property is as follows: Given an input graph over $[n]$ with degree bound d, we fix in advance an arbitrary ordering of the neighbors of each vertex in the graph. Then, an ϵ-tester may issue queries of the form "who is the i^{th} neighbor of $u \in [n]$?", to be answered either by the name of the neighbor (if such exists), or by an indication that u has less than i neighbors. The tester needs to determine whether the graph has the property or is $(\epsilon \cdot d \cdot n)$-far from any graph having the property.

7.1 Testing the property of being far from connected

In this section we study the dual problem of connectivity: For every $\epsilon > 0$, we are interested in the problem of ϵ-testing the set of graphs that are $(\epsilon \cdot d \cdot n)$-*far from being connected*. We show that this problem is different from its original problem, but that the query complexity of the dual problem is nevertheless $\text{poly}(1/\epsilon)$, as is the case for the original problem.

Preliminaries. For $d \geq 2$ and $n \in \mathbb{N}$, we will be concerned with graphs with maximal degree d over the vertex-set $[n]$. Similar to many texts discussing the bounded-degree model (see, e.g., [GR02, Sec. 2] and [BOT02, Sec. 3]), we allow multiple edges and self-loops, and define that adding a self-loop to a vertex increases its degree by 2. The set of connected graphs in this space is denoted by Π_n. For $\epsilon > 0$ and $\delta = \epsilon \cdot d \cdot n$, the standard problem of

[9] In some sources, each edge is counted twice towards the distance. For simplicity, we avoid doing so.

testing Π_n consists of distinguishing between Π_n and $\mathcal{F}_\delta(\Pi_n)$, and the dual problem consists of distinguishing between $\mathcal{F}_\delta(\Pi_n)$ and $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$.

High-level overview. Our starting point is a structural result, expressing the distance of a graph from being connected in this space by a formula that consists of a weighted count of the connected components of the graph and of the degrees of its vertices. This formula, which is presented in Section 7.1.1, might be of independent interest. Then, in Section 7.1.2, we use this formula to study the distance of graphs in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$ from Π_n. First, we show that $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$ contains graphs that are not connected, and even graphs that are $\Omega(n)$-far from being connected. Nevertheless, the main point of Section 7.1.2 is that the distance of graphs in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$ from being connected is at most $(1 - 1/4d) \cdot \delta$. The latter fact implies that graphs in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$ are significantly closer to being connected, compared to graphs in $\mathcal{F}_\delta(\Pi_n)$; specifically, the distance gap is at least $\delta/4d = \Omega(n)$.

It follows that in order to distinguish between graphs in $\mathcal{F}_\delta(\Pi_n)$ and graphs in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$ it suffices to estimate the distance of the graph from being connected in this space, up to an additive error of $\Omega(n)$. In Section 7.1.3 we show that the latter task can be done, using only $O(1)$ queries. [10] This fact relies again on the combinatorial formula from Section 7.1.1; in particular, the formula only contains (weighted) counts of connected components and of vertex degrees, and we show that such counts can be efficiently estimated, using variations of known sampling algorithms.

Notation. For a graph G over $[n]$ and $i \in [n]$, we define the number of *free degrees* of i in G to be $\mathtt{fd}(i) = d - \deg(i)$. The number of free degrees of a connected component c in G is the sum of the free degrees of its vertices (i.e., $\mathtt{fd}(c) = \sum_{i \in c} \mathtt{fd}(i)$), and the number of free degrees of G is $\sum_{i \in [n]} \mathtt{fd}(i)$. Also, for any $k \in \mathbb{N}$, let $C^k(G)$ be the set of connected components in G with k free degrees; that is, $C^k(G) = \{c : \mathtt{fd}(c) = k\}$. Also let $C^{k+}(G) = \{c : \mathtt{fd}(c) \geq k\}$, and let $C(G) = C^{0+}(G)$ be the set of all connected components in G. When G is clear from context, we will usually use a short-hand notation, and denote $C^k = C^k(G)$.

7.1.1 The distance of a graph from being connected in the bounded-degree model

The distance of a graph from being connected can be expressed using a formula that is based on the number of connected components of various types (e.g., C^0 and C^{2+}) in the graph. We first present this formula in the case when the degree bound d is even. In this special case the formula simplifies to a nicer form. After that, we generalize the formula for any $d \geq 2$.

[10] Marko and Ron [MR06] also considered the problem of estimating the distance of a graph from being connected. However, they were interested in distances in the general sparse graphs model, whereas we are concerned with distances in the bounded-degree model. The distance of a graph from being connected in both models can be significantly different (see Lemma 7.4 and [MR06, Sec. 2.1]).

Warm-up: Even degree bound d. For a graph G with maximal degree d, where d is even, let

$$\mathtt{wc}(G) \stackrel{\text{def}}{=} 2 \cdot \left|C^0(G)\right| + \left|C^{2+}(G)\right| - 1 \tag{7.1}$$

be the *weighted count* of connected components in G. We will see (in Lemma 7.3) that the weighted count of components in a graph equals the distance of the graph from being connected. But let us first explain the intuition behind the formula.

Given a graph G that is not connected, how can we modify it into a connected graph using the least number of edge modifications? If every component in the graph had at least two free degrees, then we could connect all r components, by adding $r-1$ edges (e.g., by considering an ordered sequence of the r components, and connecting vertices from each pair of subsequent components in the sequence). However, components in C^0 are "saturated" with edges – we cannot add any more edges to vertices in them without violating the degree bound d. Thus, to connect any such component to the rest of the graph, we must first remove an edge from the component. The intuition for the formula in Eq. (7.1) is that it expresses the number of edge changes to the components in $C^0 \cup C^{2+}$ in the aforementioned modification procedure (i.e., $\left|C^0\right| + (\left|C^0\right| + \left|C^{2+}\right| - 1)$).

Indeed, we did not account at all for components in C^1. However, when d is even, it holds that $\left|C^1\right| = 0$. This is the case since in a connected component c, the sum of vertex degrees cannot be $d \cdot |c| - 1$, which (given that d is even) is an odd number. The treatment of connected components in C^1 is what will create complications later, in the case of a general d.

Before formally proving that $\Delta(G, \Pi_n) = \mathtt{wc}(G)$, we first state and prove two auxiliary claims, which will be of use also in the general case.

Claim 7.1. *Let G be a graph with $r > 1$ connected components, and $G' \in \Pi_n$ be a connected graph. Then, there are at least $r-1$ edges in G' that do not exist in G.*

Proof. Fix some connected component c_1 in G. Since G' is connected, there is at least one edge in G' between a vertex in c_1 and a vertex in $[n] \setminus c_1$, and this edge is missing in G. Denote by c_2 the connected component (in G) of the end-point of the said edge in $[n] \setminus c_1$. Then, there must be at least one edge in G' connecting $c_1 \cup c_2$ to $[n] \setminus (c_1 \cup c_2)$, and this edge is missing in G. By iteratively applying this argument $r-1$ times (such that for the t^{th} iteration, we argue that the vertices in $\bigcup_{j \in [t]} c_j$ must be connected to $[n] \setminus \bigcup_{j \in [t]} c_j$ in G'), we get that $r-1$ edges in G' are missing in G. $\square$

Claim 7.2. *For $d \geq 2$, let G be a graph with maximal degree d over $[n]$, and let $c \in C^0(G)$. Then, there exists an edge in c such that removing it does not disconnect c.*

Proof. Let $mst(c)$ be an arbitrary minimum spanning tree of c. The number of edges in $mst(c)$ is $|c| - 1$. Since $\mathtt{fd}(c) = 0$ and $d \geq 2$, the number of edges in c is $\frac{1}{2} \cdot d \cdot |c| \geq |c|$. Thus, there exists an edge in c that is not in $mst(c)$, and removing it does not disconnect c. $\square$

We now prove that in the special case where d is even, the combinatorial formula in Eq. (7.1) indeed expresses the distance of a graph from being connected.

Lemma 7.3. *For an* even $d \geq 2$ *and a sufficiently large* n*, every graph* G *with maximal degree* d *over* $[n]$ *that is not connected satisfies* $\Delta(G, \Pi_n) = \mathtt{wc}(G)$*.*

Proof. Let G be a graph with maximal degree d over $[n]$. We first show that $\Delta(G, \Pi_n) \leq \mathtt{wc}(G)$: We modify G to a connected graph, by adding and removing at most $\mathtt{wc}(G)$ edges. For the modification, we first remove an edge from each connected component $c \in C^0$; according to Claim 7.2, this modification can be done without disconnecting any component in C^0. As explained above, since $C^1 = \emptyset$, at this point all connected components have at least two free degrees. Then, we add edges between the connected components in the graph; specifically, fixing some arbitrary order of the components $c_0, c_1, ..., c_r$, where $r = |C^0| + |C^{2+}|$, we add an edge between a vertex in c_i that has free degrees and a vertex in c_{i+1} that has free degrees, for every $i \in [r]$. The first step amounts to $|C^0|$ edge removals, and the second step amounts to $|C^0| + |C^{2+}| - 1$ edge additions. Overall, we modified $2 \cdot |C^0| + |C^{2+}| - 1 = \mathtt{wc}(G)$ edges in G to obtain a connected graph.

To show that $\Delta(G, \Pi_n) \geq \mathtt{wc}(G)$, we fix an arbitrary connected graph $G' \in \Pi_n$, and show that $\Delta(G, G') \geq \mathtt{wc}(G)$. Relying on Claim 7.1, we deduce that there are $|C^0(G)| + |C^{2+}(G)| - 1$ edges in G' that do not exist in G. Now, for every $c \in C^0(G)$, there must be an edge between its vertices (in G) that does not exist in G' – otherwise, the component cannot be connected to the rest of the graph in G'. Thus, the number of edges in G that do not exist in G' is at least $|C^0(G)|$. Overall, the symmetric difference between the edge-sets of G and G' is of size at least $2 \cdot |C^0(G)| + |C^{2+}(G)| - 1 = \mathtt{wc}(G)$. Thus, for every $G' \in \Pi_n$ it holds that $\Delta(G, G') \geq \mathtt{wc}(G)$, which implies that $\Delta(G, \Pi_n) \geq \mathtt{wc}(G)$. ■

The case of a general degree bound d. As mentioned before, in the case of a general d it does not necessarily hold that $|C^1| = 0$, and this fact complicates things. In the general case, the *weighted count* of connected components in a graph G is defined as follows:

$$\mathtt{wc}(G) \stackrel{\text{def}}{=} 2 \cdot |C^0| + |C^{1+}| - 1 + \max\left\{0, |C^{1+}| - 1 - \left\lceil \frac{\mathtt{fd}(G)}{2} \right\rceil \right\} . \tag{7.2}$$

First observe that when $|C^1| = 0$ (and in particular, when d is even), the formula in Eq. (7.2) agrees with the formula in Eq. (7.1). This is true because in this case $|C^{1+}| = |C^{2+}|$ and the value of the right-most expression in Eq. (7.2) is zero (because $\mathtt{fd}(G) \geq 2 \cdot |C^{2+}|$, which implies that $|C^{2+}| - \left\lceil \frac{\mathtt{fd}(G)}{2} \right\rceil \leq 0$). The following lemma, which is the main result in this section, asserts that $\mathtt{wc}(G)$ equals the distance of G from being connected also in the general case.

Lemma 7.4. *For any* $d \geq 2$ *and* $n \in \mathbb{N}$*, every graph* G *with maximal degree* d *over* $[n]$ *that is not connected satisfies* $\Delta(G, \Pi_n) = \mathtt{wc}(G)$*.*

The proof of Lemma 7.4 relies mostly on ideas similar to the ideas in the proof of Lemma 7.3, but it is significantly more tedious (reflecting the more complex expression for $\mathtt{wc}(G)$). Readers that are not interested in the technical details can safely skip the proof, and continue reading from Section 7.1.2.

Proof of Lemma 7.4. Let us begin with a short overview of the proof. Given a graph $G \notin \Pi_n$, we wish to show that $\Delta(G, \Pi_n) = \mathtt{wc}(G)$. To show that $\Delta(G, \Pi_n) \leq \mathtt{wc}(G)$, we will present an algorithm that modifies G to a connected graph by at most $\mathtt{wc}(G)$ edge removals and additions. This algorithm will be a natural one, extending the basic algorithm (for the case of an even d) described in the proof of Lemma 7.3. The analysis of the algorithm will be relatively straightforward, but will involve some tedious calculations.

To show that $\Delta(G, \Pi_n) \geq \mathtt{wc}(G)$, we will show that the symmetric difference of the edge-set of G and of any $G' \in \Pi_n$ is of size at least $\mathtt{wc}(G)$. This will be done relying on two simple observations. The first, similar to the proof of Lemma 7.3, is that an edge must be removed from any connected component in $C^0(G)$ in order to obtain a connected graph. The second observation is that the number of free degrees in a graph must be non-negative, otherwise it means that a vertex in the graph has violated the degree bound d. Thus, if adding to G edges that are missing in order to make it connected causes its number of free degrees in the graph to become negative, it follows that edges need to also be removed from G in order to obtain a graph that does not violate the degree bound. For details, see Claim 7.4.2.

The actual proof. Let $G \notin \Pi_n$ be a graph with maximal degree d over $[n]$. For technical reasons, it will be useful to work with an equivalent definition for $\mathtt{wc}(G)$, as follows. Let $\mathtt{aux}(G) \stackrel{\text{def}}{=} |C^{1+}| - 1 - \left\lceil \frac{\mathtt{fd}(G)}{2} \right\rceil$ be an auxiliary term; then, Eq. (7.2) is equivalent to the following definition:

$$\mathtt{wc}(G) \stackrel{\text{def}}{=} \begin{cases} 2 \cdot |C^0| + |C^{1+}| - 1 & \mathtt{aux}(G) \leq 0 \\ 2 \cdot (|C| - 1) - \left\lceil \frac{\mathtt{fd}(G)}{2} \right\rceil & \mathtt{aux}(G) > 0 \end{cases}. \tag{7.3}$$

We first show that $\Delta(G, \Pi_n) \leq \mathtt{wc}(G)$. In particular, we show that the following algorithm modifies G to a connected graph, by adding and removing at most $\mathtt{wc}(G)$ edges.

Algorithm 1. On an input graph $G \notin \Pi_n$, do the following:

1. Remove an edge from every connected component in C^0, without disconnecting any of the components. (Recall that this is possible according to Claim 7.2.)
2. Connect the components that now have 2 or more free degrees (i.e., all components that were originally in $C^0 \cup C^{2+}$). Specifically, fix an arbitrary order of the components, $c_1, c_2, ..., c_r$, and add an edge between c_i and c_{i+1} for every $i \in [r-1]$. This does not violate the degree bound d, since after Step (1) all these components have at least 2 free degrees.
3. At this point, the graph consists of a connected component that contains all vertices that were originally in $C^0 \cup C^{2+}$, which we call the *main* connected component and denote by c_0; and an additional collection of components, that is C^1. Execute the following loop: While $\mathtt{fd}(c_0) > 0$ and the graph is not connected, take an arbitrary vertex $i \in c_0$ such that $\mathtt{fd}(i) > 0$, and connect i to a suitable vertex in a connected component $c \neq c_0$ such that $\mathtt{fd}(c) = 1$.

4. If the previous step resulted in a connected graph, then we are done. Otherwise, at this point the graph consists of the (extended) main component c_0, which now has no free degrees (i.e., $\mathtt{fd}(c_0) = 0$), and an additional collection $S \subseteq C^1$ of connected components. Split S into pairs of components, and for each pair of components, do the following step: Remove an edge from c_0, thereby freeing two free degrees in c_0 without disconnecting it (this is possible according to Claim 7.2, and since $\mathtt{fd}(c_0) = 0$ at this point); and connect each of the pair of components to a vertex in c_0 that now has a free degree (thereby reducing the free degrees in c_0 to zero again). If after finishing the pairs in S there is a remainder of a single (unpaired) component, remove another edge from c_0 and connect the last component to c_0.

When Algorithm 1 finishes its execution, the resulting graph is a connected graph that does not violate the degree bound d. It is thus left to show that the number of edge modifications that Algorithm 1 makes is at most $\mathtt{wc}(G)$.

Claim 7.4.1. *On any input graph $G \notin \Pi_n$, Algorithm 1 makes $\mathtt{wc}(G)$ edge modifications to G.*

Proof. First note that in Step (1) we remove $\left|C^0\right|$ edges, whereas in Step (2) we add $\left|C^0\right| + \left|C^{2+}\right| - 1$ edges. In order to account for the number of modifications in Steps (3) and (4) we need to make some preliminary calculations about the state of the graph when these steps of the algorithm are executed. The actual count of the number of modifications in these steps will be based on a case-analysis, depending on the said calculations.

In the description of Step (3), we defined a *main component* c_0 that consists of all vertices that originally resided in $C^0 \cup C^{2+}$. We start by calculating the number of free degrees in c_0 in the beginning of Step (3), which we denote by $\mathtt{fd}^{(St3)}(c_0)$. In the beginning of Step (2), the vertices in c_0 had $\sum_{c \in C^{2+}} \mathtt{fd}(c) + 2 \cdot \left|C^0\right|$ free degrees; and during Step (2) we added $\left|C^0\right| + \left|C^{2+}\right| - 1$ edges between the vertices of c_0, lowering the free degrees of c_0 by twice this much. Therefore, in the beginning of Step (3) it holds that

$$\begin{aligned} \mathtt{fd}^{(St3)}(c_0) &= \sum_{c \in C^{2+}} \mathtt{fd}(c) + 2 \cdot \left|C^0\right| - 2 \cdot \left(\left|C^0\right| + \left|C^{2+}\right| - 1\right) \\ &= \mathtt{fd}(G) - \left|C^1\right| - 2 \cdot \left|C^{2+}\right| + 2\,. \end{aligned} \tag{7.4}$$

If $\mathtt{fd}^{(St3)}(c_0) \geq \left|C^1\right|$, then the loop in Step (3) will end when the graph is connected; and otherwise, the loop will end after $\mathtt{fd}^{(St3)}(c_0)$ iterations, and we will continue to Step (4). In the latter case, the number of additional components with a single free degree that remain in the beginning of Step (4) is $|S| = \left|C^1\right| - \mathtt{fd}^{(St3)}(c_0)$. Relying on Eq. (7.4), it follows that:

$$\begin{aligned}
\left\lceil \frac{|S|}{2} \right\rceil &= \left\lceil \frac{\left|C^1\right| - \mathtt{fd}^{(St3)}(c_0)}{2} \right\rceil \\
&= \left\lceil \frac{\left|C^1\right| - \left(\mathtt{fd}(G) - \left|C^1\right| - 2\cdot\left|C^{2+}\right| + 2\right)}{2} \right\rceil \\
&= \left|C^{1+}\right| - 1 - \left\lceil \frac{\mathtt{fd}(G)}{2} \right\rceil && (7.5) \\
&= \mathtt{aux}(G)\ . && (7.6)
\end{aligned}$$

We now count the number of modifications in Steps (3) and (4), based on a case analysis, depending on whether $\mathtt{fd}^{(St3)}(c_0) \geq \left|C^1\right|$ (i.e., the algorithm has not executed Step (4)).

- **Case 1:** $\mathtt{fd}^{(St3)}(c_0) \geq \left|C^1\right|$. In this case, the loop in Step (3) ends after $\left|C^1\right|$ iterations, with all components in C^1 being connected to the main component. The overall number of modifications in this case equals $2\cdot\left|C^0\right| + \left|C^{2+}\right| - 1 + \left|C^1\right| = 2\cdot\left|C^0\right| + \left|C^{1+}\right| - 1$. Also, relying on the fact that $\left\lceil \frac{\left|C^1\right| - \mathtt{fd}^{(St3)}(c_0)}{2} \right\rceil = \mathtt{aux}(G)$ (by Eq. (7.6)) and on the fact that $\mathtt{fd}^{(St3)}(c_0) \geq \left|C^1\right|$, it follows that $\mathtt{aux}(G) \leq 0$. According to Eq. (7.3), this implies that $\mathtt{wc}(G) = 2\cdot\left|C^0\right| + \left|C^{1+}\right| - 1$. Thus, in this case, Algorithm 1 performed $\mathtt{wc}(G)$ modifications to G.

- **Case 2:** $\mathtt{fd}^{(St3)}(c_0) < \left|C^1\right|$. In this case, the loop in Step (3) ends after $\mathtt{fd}^{(St3)}(c_0)$ iterations, when $\mathtt{fd}(c_0) = 0$, and we continue to Step (4). In Step (4), we are left with $|S| = \left|C^1\right| - \mathtt{fd}^{(St3)}(c_0) > 0$ components with a single free degree, alongside the extended main component c_0. For every pair of components in S, we remove one edge and add two, and for a possible last remainder component, we remove an edge and add an edge; this amounts to $\frac{3}{2}\cdot\left\lfloor\frac{|S|}{2}\right\rfloor + 2\cdot\left(\left\lceil\frac{|S|}{2}\right\rceil - \left\lfloor\frac{|S|}{2}\right\rfloor\right) = |S| + \left\lceil\frac{|S|}{2}\right\rceil$ edges. Overall, the number of modifications in this case is

$$\begin{aligned}
&2\cdot\left|C^0\right| + \left|C^{2+}\right| - 1 + \mathtt{fd}^{(St3)}(c_0) + |S| + \left\lceil\frac{|S|}{2}\right\rceil \\
&= 2\cdot\left|C^0\right| + \left|C^1\right| + \left|C^{2+}\right| - 1 + \left\lceil\frac{|S|}{2}\right\rceil && \left(|S| = \left|C^1\right| - \mathtt{fd}^{(St3)}(c_0)\right) \\
&= 2\cdot\left|C^0\right| + \left|C^{1+}\right| - 1 + \left(\left|C^{1+}\right| - 1 - \left\lceil\frac{\mathtt{fd}(G)}{2}\right\rceil\right) && \text{(by Eq. (7.5))} \\
&= 2\cdot(|C| - 1) - \left\lceil\frac{\mathtt{fd}(G)}{2}\right\rceil\ .
\end{aligned}$$

  Now, since $|S| > 0$, according to Eq. (7.6) it follows that $\mathtt{aux}(G) > 0$, which (according to Eq. (7.3)) implies that $\mathtt{wc}(G) = 2\cdot(|C| - 1) - \left\lceil\frac{\mathtt{fd}(G)}{2}\right\rceil$. Thus, in this case it also holds that Algorithm 1 performed $\mathtt{wc}(G)$ modifications to G.

This completes the proof of Claim 7.4.1. $\square$

For the other direction, we prove that for any graph G that is not connected it holds that $\Delta(G, \Pi_n) \geq \mathtt{wc}(G)$.

Claim 7.4.2. *Let $G \notin \Pi_n$. Then, for every connected graph $G' \in \Pi_n$ it holds that $\Delta(G, G') \geq \mathtt{wc}(G)$.*

Proof. Our proof relies on a case analysis, according to the value of $\mathtt{aux}(G)$.

Case 1: $\mathtt{aux}(G) \leq 0$. According to Eq. (7.3), we have $\mathtt{wc}(G) = 2 \cdot \left|C^0\right| + \left|C^{1+}\right| - 1$. Relying on Claim 7.1, there exist $|C(G)| - 1 = \left|C^0(G)\right| + \left|C^{1+}(G)\right| - 1$ edges in G' that do not exist in G. Also, for every component $c \in C^0(G)$, there must exist an edge between its vertices (in G) that does not exist in G' – otherwise, the component cannot be connected to the rest of the graph in G'. Thus, the number of edges in components in $C^0(G)$ that do not exist in G' is at least $\left|C^0(G)\right|$. Therefore, the symmetric difference between the edge-sets of G and of G' is of size at least $2 \cdot \left|C^0\right| + \left|C^{1+}\right| - 1 = \mathtt{wc}(G)$, which finishes the first case.

Case 2: $\mathtt{aux}(G) > 0$. According to Eq. (7.3), we have $\mathtt{wc}(G) = 2 \cdot (|C(G)| - 1) - \left\lceil \frac{\mathtt{fd}(G)}{2} \right\rceil$. Relying on Claim 7.1, there exist $|C(G)| - 1$ edges in G' that do not exist in G. We now show that there also exist many edges in G that do not exist in G', relying on a count of free degrees in G.

Consider the graph G'', obtained by adding to G the said $|C(G)| - 1$ edges in G' that do not exist in G, disregarding the degree bound d. The number of free degrees in G'' is:

$$\mathtt{fd}(G'') = \mathtt{fd}(G) - 2 \cdot (|C(G)| - 1) \ . \tag{7.7}$$

Combining Eq. (7.7) with the assumption that $\mathtt{aux}(G) > 0$, we get that $\mathtt{fd}(G'') < 0$:

$$\begin{aligned} 0 < \mathtt{aux}(G) &= \left|C^{1+}(G)\right| - 1 - \left\lceil \frac{\mathtt{fd}(G)}{2} \right\rceil \\ &\leq |C(G)| - 1 - \frac{\mathtt{fd}(G)}{2} \\ &= -\frac{1}{2} \cdot \mathtt{fd}(G'') \ . \end{aligned}$$

The fact that $\mathtt{fd}(G'') < 0$ implies that there exist vertices in G'' that violate the degree bound d. Since removing a single edge from G'' creates two additional free degrees in the graph, it follows that we need to remove at least $\left\lceil \frac{|\mathtt{fd}(G'')|}{2} \right\rceil$ edges from G'' in order to obtain a graph in which the degree bound is not violated, and in particular in order to obtain the graph G'. Thus, using Eq. (7.7), the number of edges in G'' that do not exist in G' is at least

$$\left\lceil \frac{-\mathtt{fd}(G'')}{2} \right\rceil = (|C(G)| - 1) - \left\lceil \frac{\mathtt{fd}(G)}{2} \right\rceil \ .$$

Note that the aforementioned edges (that exist in G'' but not in G') are also edges in G that do not exist in G'. This is the case since the only edges that exist in G'' but not in G are the ones that we added, which also exist in G'. Hence, overall, the size of the symmetric difference between the edge-sets of G and of G' is of size at least:

$$|C(G)| - 1 + \left\lceil \frac{-\mathtt{fd}(G'')}{2} \right\rceil = \mathtt{wc}(G)$$

which implies that $\Delta(G, G') \geq \mathtt{wc}(G)$, and finishes the second case. Hence, for every $G' \in \Pi_n$ it holds that $\Delta(G, G') \geq \mathtt{wc}(G)$. □

Claim 7.4.2 implies that for every $G \notin \Pi_n$ it holds that $\Delta(G, \Pi_n) \geq \mathtt{wc}(G)$. This completes the proof of Lemma 7.4. ■

7.1.2 Graphs that are far-from-far from being connected.

In this section we prove that $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$ contains graphs that are not connected, and even graphs that are $\Omega(n)$-far from being connected. On the other hand, we show that the distance of any such graph from being connected is at most $\left(1 - \frac{1}{4d}\right) \cdot \delta$.

Proposition 7.5 *(the set of connected graphs is not $\mathcal{F}_\delta$-closed). For any $d \geq 3$, and $\delta \geq 2$, and sufficiently large n, the set of connected graphs Π_n is not $\mathcal{F}_\delta$-closed. Moreover, for any $d \geq 6$, and $\epsilon > 0$, and sufficiently large n, and $\delta = \epsilon \cdot d \cdot n$, it holds that $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$ contains graphs that are $\Omega(n)$-far from Π_n.*

Proof. For the first part of the statement, similarly to the proofs of Propositions 6.1 and 6.5, it suffices to show a graph G such that $\Delta(G, \Pi_n) = 1$, and all neighbors of G (i.e., graphs that disagree with G on one edge) are either in Π_n or adjacent to Π_n. The graph G consists of two disjoint cycles. Observe that G is not connected, but one can connect the two cycles by adding an edge (since $d \geq 3$); thus, $\Delta(G, \Pi_n) = 1$. However, after adding any edge to G, or removing any edge from it, the resulting graph G' still satisfies $\Delta(G', \Pi_n) \leq 1$: This is since the addition of an edge or removal of an edge does not disconnect either of the two cycles, and thus we can still connect the cycles by adding an edge between them. Relying on Proposition 2.3, we deduce that Π_n is not $\mathcal{F}_\delta$-closed.

For the "moreover" part, we need the following definition. For $m \in \mathbb{N}$, a connected graph H is *m-resilient* if for any $r \in \mathbb{N}$, splitting H into $1 + r$ connected components cannot be done with less than $m \cdot r$ edge removals from H. The intuitive meaning of this definition is that in order to split an m-resilient graph to two components, we need to remove m edges from the graph, and to split either of these two components, we must remove an additional m edges from that component (and so forth); that is, intuitively, whenever splitting an m-resilient graph to connected components, each of the components is also m-resilient. Note that the notion of m-resiliency extends the notion of m-edge-connectivity: The latter means that the graph cannot be disconnected by removing less than m edges, whereas to achieve the former, we wish that after disconnecting the graph, this feature will also be *preserved* in each resulting connected component. An example for an m-resilient graph is a "multi-path", that is a path in which any two adjacent vertices are connected by m parallel edges.

Let $d \geq 6$, let $\epsilon > 0$, let n be sufficiently large, and let $\delta = \epsilon \cdot d \cdot n$. Our construction of a graph $G \in \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$ such that $\Delta(G, \Pi_n) = \Omega(n)$ is as follows. The graph G consists of $\frac{\delta}{6}$ connected components that are each $\lfloor d/2 \rfloor$-resilient and have maximal degree at most d (e.g., each component is a "multi-path" as above). According to Claim 7.1, the distance of G from being connected is at least $\frac{\delta}{6} - 1 = \Omega(n)$.

Now, let $H \in \mathcal{F}_\delta(\Pi_n)$. Relying on Lemma 7.4 and on Eq. (7.2), we have

$$\delta \leq \Delta(H, \Pi_n) = \mathtt{wc}(H) \leq 2 \cdot |C^{\cdot}(H)|$$

and hence the number of connected components in H is at least $\frac{\delta}{2}$. Since G consists of connected components that are $\lfloor d/2 \rfloor$-resilient, creating additional $\frac{\delta}{2} - \frac{\delta}{6} = \frac{\delta}{3}$ connected components in G requires the removal of at least $\left\lfloor \frac{d}{2} \right\rfloor \cdot \frac{\delta}{3} \geq \delta$ edges from G (where the inequality is since $d \geq 6$). Thus, the symmetric difference between the edge-sets of H and G is of size at least δ, which implies that $\Delta(G, H) \geq \delta$. It follows that G is δ-far from any $H \in \mathcal{F}_\delta(\Pi_n)$, which implies that $G \in \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$. ■

Nevertheless, building on Lemma 7.4, we now show that the distance of graphs in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$ from Π_n is $(1 - \Omega(1)) \cdot \delta$.

Proposition 7.6 *(graphs that are far-from-far from being connected are relatively close to being connected). Let $d \geq 2$, let $\epsilon < \frac{1}{2 \cdot d}$, let n be a sufficiently large integer and let $\delta = \epsilon \cdot d \cdot n$. Then, for every graph $G \in \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$, it holds that $\Delta(G, \Pi_n) < \left(1 - \frac{1}{4d}\right) \cdot \delta$.*

Proof. Relying on Proposition 2.5, it suffices to show a way to modify every graph G that is not δ-far from being connected into a graph G' that is farther away from being connected, with only $4 \cdot d$ changes. The intuition for the modification procedure (of G to G') is as follows. Recall that, according to Lemma 7.4 and Eq. (7.2), the distance of a graph from being connected is proportional to the number of its connected components, and (in some cases) inversely proportional to the number of free degrees in the graph. Accordingly, to modify a graph G to a graph that is farther away from being connected, we remove edges from G in order to isolate a small connected component, and then, in order to decrease the number of free degrees to its original value, we add edges within the new component as well as between vertices of the original connected component (from which the new component was detached). This modification procedure is depicted in the proof of the following claim.

Claim 7.6.1. *For every $G \notin \mathcal{F}_\delta(\Pi_n)$ there exists G' such that $\Delta(G, G') \leq 4 \cdot d$ and $\Delta(G', \Pi_n) \geq \Delta(G, \Pi_n) + 1$.*

Proof. First note that there exists a connected component c in G with at least 3 vertices. This is the case since otherwise, the number of connected components in G is at least $n/2 > \epsilon \cdot d \cdot n$ (because $\epsilon < \frac{1}{2 \cdot d}$), and relying on Claim 7.1, it follows that G is δ-far from being connected, which contradicts the hypothesis.

Warm-up: When d is even. Let us first consider the case in which the degree bound d is even; this case uses ideas similar to the ideas in the proof for the case of a general degree bound d, but avoids many tedious technicalities. Recall that, according to Lemma 7.3, in this case the distance of a graph H from being connected is $\Delta(H, \Pi_n) = 2 \cdot |C^0(H)| + |C^{2+}(H)| - 1$. To modify G into G', we isolate two vertices $i, j \in [n]$ from the aforementioned connected component c, by removing all edges incident to them; and then we add d multiple edges between these two vertices. Overall, we removed at most $2 \cdot d$ edges, and added d edges, and so $\Delta(G, G') \leq 3 \cdot d$.

Compared to G, the modified graph G' has an additional component with no free degrees (the component $\{i, j\}$), and the vertices in $c \setminus \{i, j\}$ have more free degrees. Thus, two cases are possible: Either it is that c originally had free degrees in G (i.e., $c \in C^{2+}(G)$), in which case $|C^0(G')| = |C^0(G)| + 1$ and $|C^{2+}(G')| \geq |C^{2+}(G)|$; or it is the case that c originally had no free degrees in G, in which case $|C^0(G')| = |C^0(G)|$ and $|C^{2+}(G')| \geq |C^{2+}(G)| + 1$. In both cases it holds that $\Delta(G', \Pi_n) \geq \Delta(G, \Pi_n) + 1$.

The general case. For the case of a general degree bound d, we construct the graph G' is as follows. Fix a connected component c with three or more vertices, and two vertices $i, j \in c$. Remove all edges incident to i and to j from the graph, and add d multiple edges between i and j. Thus, the component c has split to two non-empty sets: $c_0 = c \setminus \{i, j\}$ and $c_1 = \{i, j\}$. Now, note that the first removal step has increased the number of free degrees of vertices in c_0, by an amount denoted by $m \leq 2 \cdot d$ (i.e., m is the number of edges in G that connected i and j to vertices in $c \setminus \{i, j\}$). Consequently, at this point we can add $\lfloor m/2 \rfloor$ edges between vertices in c_0 (some of these edges might be multiple edges and/or self-loops). This completes the modification of G to G'.

Overall, we removed at most $2 \cdot d$ edges from G, and added at most $2 \cdot d$ edges to it, to obtain the graph G'; thus, $\Delta(G, G') \leq 4 \cdot d$. Therefore we only need to prove that $\Delta(G', \Pi_n) \geq \Delta(G, \Pi_n) + 1$. To do this, we will track the changes made to the graph, and in particular the changes to its number of free degrees and the changes to its connected components.

Fact 7.6.1.1. *After the modification of G to G', the number of free degrees in the graph has not increased; that is,* $\texttt{fd}(G') \leq \texttt{fd}(G)$.

Proof. Denote by $deg_G(i)$ and $deg_G(j)$ the degrees of i and of j, respectively, in G (i.e., before the modification), and note that

$$\texttt{fd}(G') - \texttt{fd}(G) = m + \deg_G(i) + \deg_G(j) - 2 \cdot \lfloor m/2 \rfloor - 2 \cdot d\,. \tag{7.8}$$

If $\deg_G(i) + \deg_G(j) < 2d$, then the expression in Eq. (7.8) is at most zero. Otherwise, if $\deg_G(i) = \deg_G(j) = d$, then m must be an even number. This is the case since, denoting the number of edges (in G) between i and j by f, then $2d = \deg(i) + \deg(j) = m + 2f$, which implies that $m = 2 \cdot (d - f)$. Thus, in this case, $2 \cdot \lfloor m/2 \rfloor = m$, which implies that the expression in Eq. (7.8) equals zero. □

Let us see what happened to the connected components of G when modified to G'. The only connected component in G that was changed is c, which was split into at least two connected components: The component $c_1 = \{i, j\}$, which has no free degrees in G', and the component or components containing the vertices in $c_0 = c \setminus \{i, j\}$. Thus, there are more connected components in G', and at least one of them (i.e., c_1) is without free degrees. Combined with Fact 7.6.1.1, this will now allow us to prove that $\Delta(G', \Pi_n) \geq \Delta(G, \Pi_n) + 1$.

For any graph H, denote $\varphi_1(H) = 2 \cdot |C^0(H)| + |C^{1+}(H)| - 1$ and $\varphi_2(H) \stackrel{\text{def}}{=} |C^{1+}(H)| - 1 - \left\lceil \frac{\mathtt{fd}(H)}{2} \right\rceil$. Then, according to Lemma 7.4, it holds that:

$$\Delta(H, \Pi_n) = \varphi_1(H) + \max\{0, \varphi_2(H)\} \ . \tag{7.9}$$

We prove that $\Delta(G', \Pi_n) \geq \Delta(G, \Pi_n) + 1$ by relying on Eq. (7.9), and considering three separate cases.

Case 1: $|C^0(G')| \geq |C^0(G)| + 2$. First note that $|C^{1+}(G')| \geq |C^{1+}(G)| - 1$, since the modification of G to G' is equivalent to the removal of one connected component (i.e., of c) and the addition of two or more connected components (i.e., of c_1 and of the components containing the vertices in $c \setminus \{i, j\}$). Relying on this fact, and on Fact 7.6.1.1, it follows that $\varphi_2(G') \geq \varphi_2(G) - 1$. However, since $|C^0(G')| \geq |C^0(G)| + 2$, and relying again on the fact that $|C^{1+}(G')| \geq |C^{1+}(G)| - 1$, it follows that $\varphi_1(G') \geq \varphi_1(G) + 3$. Thus, by Eq. (7.9), it holds that $\Delta(G', \Pi_n) - \Delta(G, \Pi_n) \geq 2$.

Case 2: $|C^0(G')| = |C^0(G)|$. Since we know that an additional connected component with no free degrees was created in G' (i.e., the component c_1), this case is possible only if the component c was originally (i.e., in G) a component without free degrees, and after the modification, the connected components that consist of vertices in $c_0 = c \setminus \{i, j\}$ all have free degrees. Thus, in this case, it holds that $|C^{1+}(G')| \geq |C^{1+}(G)| + 1$. It follows that $\varphi_1(G') \geq \varphi_1(G) + 1$, and, relying on Fact 7.6.1.1, that $\varphi_2(G') > \varphi_2(G)$. Overall, we get that $\Delta(G', \Pi_n) - \Delta(G, \Pi_n) \geq \varphi_1(G') - \varphi_1(G) \geq 1$.

Case 3: $|C^0(G')| = |C^0(G)| + 1$. In this case it necessarily holds that $|C^{1+}(G')| \geq |C^{1+}(G)|$. To see that this is true, assume otherwise; it follows that c was a component with free degrees in G, but that no component that consists of vertices in c_0 has free degrees in G'. However, this implies that there are at least two additional components without free degrees in G', compared to G (the component c_1, and a component containing vertices in c_0), which contradicts the hypothesis of the current case. Therefore, it follows that $\varphi_1(G') \geq \varphi_1(G) + 2$, and (relying on Fact 7.6.1.1) that $\varphi_2(G') \geq \varphi_2(G)$. Overall, we get that $\Delta(G', \Pi_n) \geq \Delta(G, \Pi_n) + 2$. □

This completes the proof of Proposition 7.6. ■

A comment about non-simple graphs. Recall that in the preliminary definitions of the current section (i.e., Section 7.1), we assumed that the space of graphs we are dealing with also contains graphs with multiple edges and self-loops. Throughout Section 7.1.2, we relied on the assumption that such non-simple graphs exist in our metric space. Most notably, we relied on this assumption in Claim 7.6.1, which was the main step in proving Proposition 7.6. We believe that it is possible to prove a claim similar to Claim 7.6.1, and thus also obtain a result similar to Proposition 7.6, without relying on the existence of non-simple graphs, but it was not our focus in this work.

7.1.3 The dual problem of connectivity in the bounded-degree model

Proposition 7.5 implies that the dual problem of connectivity in the bounded-degree model is "very different" from its original problem, in the sense that $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$ contains graphs that are $\Omega(n)$-far from Π_n. However, Proposition 7.6 implies that there is a gap of $\frac{1}{4d} \cdot \delta = \Omega(n)$ between the distance of graphs in $\mathcal{F}_\delta(\Pi_n)$ from Π_n and the distance of graphs in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$ from Π_n. Thus, to show a tester for the dual problem, it suffices to show that the distance of a graph from Π_n can be estimated using a small number of queries.

Relying on Lemma 7.4, for a given graph G, this is equivalent to estimating the following quantity:

$$2 \cdot \left|C^0\right| + \left|C^{1+}\right| - 1 + \max\left\{0, \left|C^{1+}\right| - 1 - \left\lceil \frac{\mathtt{fd}(G)}{2} \right\rceil \right\} . \tag{7.10}$$

We will see that *each* of the terms in Eq. (7.10) can be estimated up to an additive error of $\gamma \cdot n$, for any $\gamma > 0$, using only $\text{poly}(1/\gamma)$ queries.

A preliminary discussion of the estimation algorithm. First note that the term $\mathtt{fd}(G)$ can be estimated by straightforward sampling. This is the case since $\mathtt{fd}(G) = d \cdot n - \sum_{i \in [n]} \deg(i)$, and the average degree of a vertex in the graph can be estimated, with high probability, by outputting the average degree in a sample of uniformly chosen vertices.

It is thus left to handle the terms $\left|C^0\right|$ and $\left|C^{1+}\right|$; for simplicity, we focus on the term $\left|C^0\right|$ (the term $\left|C^{1+}\right|$ can be handled very similarly). The estimation algorithm for $\left|C^0\right|$ is based on the algorithm of Chazelle, Rubinfeld, and Trevisan [CRT05] for estimating the number of connected components in a graph. In particular, for every vertex $i \in [n]$, let $c(i)$ be the connected component in which i resides, and let

$$s(i) = \begin{cases} \frac{1}{|c(i)|} & c(i) \in C^0 \\ 0 & c(i) \notin C^0 \end{cases} .$$

For a fixed component $c \in C^0$, we have $\sum_{i:c(i)=c} s(i) = 1$. Therefore, we get that $\sum_{i \in [n]} s(i) = \left|C^0\right|$. Hence, to estimate $\left|C^0\right|$, it suffices to estimate the average value of $s(i)$, over all $i \in [n]$. Given a fixed $i \in [n]$, we can compute $s(i)$ using $|c(i)| \cdot d$ queries, by running a BFS from i, and counting the number of free degrees in its connected component. When $|c(i)| = O(1)$, this requires only $O(1)$ queries; but when $|c(i)|$ is large, the BFS requires too much queries. However, in the latter case, $s(i)$ is very small; in this case, we can obtain a rough estimate of $s(i)$ by choosing a sufficiently small fixed value (actually, we just take the value zero). More specifically, given an estimation parameter $\gamma > 0$, for any vertex $i \in [n]$, let

$$\tilde{s}(i) = \begin{cases} s(i) & \text{if } |c(i)| \leq 1/\gamma \\ 0 & \text{o.w.} \end{cases} .$$

Note that given a vertex $i \in [n]$, we can exactly compute $\tilde{s}(i)$ using $\frac{d}{\gamma}$ queries. This is done by performing a BFS, starting from i, and halting if we encountered more than $\frac{1}{\gamma}$ vertices in the connected component of i (in which case it holds that $\tilde{s}(i) = 0$). Also note that for every $i \in [n]$ it holds that $|\tilde{s}(i) - s(i)| < \gamma$. Therefore,

$$\left| \sum_{i \in [n]} \tilde{s}(i) - \left|C^0\right| \right| \leq \sum_{i \in [n]} |\tilde{s}(i) - s(i)| < \gamma \cdot n .$$

Thus, to estimate $\left|C^0\right|$ up to an additive error of $2\gamma \cdot n$, with high probability, it suffices to estimate the average value of $\tilde{s}(i)$ over the vertices in the graph up to an additive error of γ, with high probability. Relying on Chernoff's inequality, the latter can be done by uniformly sampling $O(\gamma^{-2})$ vertices, computing the $\tilde{s}$ value of each vertex (using $\frac{d}{\gamma}$ queries), and outputting the average $\tilde{s}$ value of vertices in the sample. The query complexity of this estimation procedure is $O(\gamma^{-2} \cdot \frac{d}{\gamma}) = O(\gamma^{-3} \cdot d)$. The same holds for $\left|C^{1+}\right|$.

The tester itself. Let us spell out the tester for the dual problem of connectivity that is obtained by combining the above estimation algorithms.

Theorem 7.7 *(a tester for the dual problem of connectivity). Let $d \geq 2$, let $\epsilon < \frac{1}{2 \cdot d}$, let n be a sufficiently large integer and let $\delta = \epsilon \cdot d \cdot n$. Then, there exists an algorithm with query complexity $O\left(\epsilon^{-3} \cdot d\right)$ that accepts, with high probability, every graph in $\mathcal{F}_\delta(\Pi_n)$, and rejects, with high probability, every graph in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$.*

Proof. Given an input graph G over the vertex-set $[n]$, the algorithm estimates $\Delta(G, \Pi_n)$, such that with high (constant) probability, the estimated value is correct up to an additive error of $\frac{\delta}{8 \cdot d} = \frac{\epsilon}{8} \cdot n$. It then accepts G if and only if the estimated value is at least $\left(1 - \frac{1}{8 \cdot d}\right) \cdot \delta$. The correctness of the algorithm follows from Proposition 7.6. The query complexity of the algorithm is simply the query complexity of the estimation procedure: To estimate the average degree of a vertex in the graph up to an error of $O(\epsilon \cdot n)$, we perform $O(\epsilon^{-2} \cdot d)$ queries; and to estimate each of the two terms $\left|C^0(G)\right|$ and $\left|C^{1+}(G)\right|$ up to an error of $O(\epsilon \cdot n)$, we perform $O(\epsilon^{-3} \cdot d)$ queries. ■

7.2 Testing the property of being far from cycle-free

In this section we study the dual problem of testing cycle-free graphs: For every $\epsilon > 0$, we are interested in the problem of ϵ-testing the set of graphs that are $(\epsilon \cdot d \cdot n)$-*far from being cycle-free*. We show that this problem is different from its original problem, but that the query complexity of the dual problem is nevertheless $\text{poly}(1/\epsilon)$, as is the case for the original problem.

Preliminaries. For $d \geq 2$ and $n \in \mathbb{N}$, we will be concerned with graphs with maximal degree d over the vertex-set $[n]$. For a graph G over $[n]$, let $E(G)$ be the edge-set of G, and let $C(G)$ be the set of connected components in G. Similar to other texts discussing the problem of testing cycle-free graphs in this model (see, e.g. [GR02, Sec. 4] and [MR06, Sec. 5]), we consider *only simple graphs*. The set of cycle-free graphs in this space is denoted by Π_n. For $\epsilon > 0$ and $\delta = \epsilon \cdot d \cdot n$, the standard problem of testing Π_n consists of distinguishing between Π_n and $\mathcal{F}_\delta(\Pi_n)$, and the dual problem consists of distinguishing between $\mathcal{F}_\delta(\Pi_n)$ and $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$.

High-level overview. Our starting point is two results of Marko and Ron [MR06, Sec. 5] about cycle-free graphs in the bounded-degree model. Specifically, they observed that the distance of a graph from being cycle-free in this model is $\Delta(G, \Pi_n) = |E(G)| + |C(G)| - n$, and proved that given an input graph G, this quantity can be estimated, up to an $\Omega(n)$ additive error, using only $O(1)$ queries.

Our contribution primarily consists of the analysis of the distance of graphs in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$ from being cycle-free. Specifically, we show that there exist graphs in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$ that are not cycle-free (i.e., that Π_n is not $\mathcal{F}_\delta$-closed), but on the other hand, the distance of graphs in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$ from being cycle-free is at most $\frac{2}{3} \cdot \delta$. The latter fact implies that graphs in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$ are significantly closer to being cycle-free, compared to graphs in $\mathcal{F}_\delta(\Pi_n)$; in particular, the distance gap is at least $\delta/3 = \Omega(n)$. It follows that the dual problem can be solved using the algorithm of [MR06] to estimate the distance of a graph from the property.

Proposition 7.8 *(the set of cycle-free graphs is not $\mathcal{F}_\delta$-closed). For any $d \geq 2$, and $\delta \geq 2$, and sufficiently large n, the set of cycle-free graphs Π_n is not $\mathcal{F}_\delta$-closed.*

Proof. Similarly to the proofs of Propositions 6.1, 6.5, and 7.5, it suffices to show a graph G such that $\Delta(G, \Pi_n) = 1$, and all neighbors of G (i.e., graphs that disagree with G on one edge) are either in Π_n or adjacent to Π_n (since we can then rely on Proposition 2.3 to deduce that Π_n is not $\mathcal{F}_\delta$-closed for every $\delta \geq 2$). The graph G over $[n]$ consists of a single triangle and of additional $n - 3$ isolated vertices. The graph is not cycle-free, but can be made cycle-free by removing a single edge from the triangle; thus, $\Delta(G, \Pi_n) = 1$. However, note that adding any edge to G yields a graph G' such that $\Delta(G', \Pi_n) \leq 1$: This is the case since any additional edge either connects an additional vertex to the triangle, or connects two isolated vertices (recall that the metric space is comprised only of simple graphs); in both cases, removing an edge from the original triangle turns G' into a cycle-free graph. ■

We now show that the distance of graphs in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$ from being cycle-free is nevertheless at most $\frac{2}{3} \cdot \delta$.

Proposition 7.9 *(graphs that are far-from-far from being cycle-free are relatively close to being cycle-free). For $d \geq 3$, let $\epsilon < \frac{1}{12 \cdot d^2}$, let n be a sufficiently large integer, and let $\delta = \epsilon \cdot d \cdot n$. Then, for every $G \in \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$ it holds that $\Delta(G, \Pi_n) \leq \frac{2}{3} \cdot \delta$.*

Proof. Relying on Proposition 2.5, it suffices to show how to modify every graph that is not δ-far from being cycle-free into a graph that is farther away from being cycle-free, by adding at most *three edges* to the original graph. This modification procedure is depicted in the proof of the following claim.

Claim 7.9.1. *For every $G \notin \mathcal{F}_\delta(\Pi_n)$ there exists G' such that $\Delta(G, G') \le 3$ and $\Delta(G', \Pi_n) = \Delta(G, \Pi_n) + 1$.*

Proof. Our proof is based on a case analysis, depending on the number of connected components in G. Specifically, if $|C(G)|$ is not too large (i.e., $|C(G)| \le \frac{n}{6 \cdot d}$), we will show that there exist two non-adjacent vertices with degree at most $d-1$ in the same connected component in the graph. Connecting the two vertices by an edge yields G' as required. Otherwise, if $|C(G)|$ is large (i.e., $|C(G)| > \frac{n}{6 \cdot d}$), we will show that there exist three non-adjacent vertices with degree at most one in the graph. Adding edges between three such vertices, creating a new triangle in the graph, yields G' as required.

For the proof itself, first note that, since $\Delta(G, \Pi_n) = |E(G)| + |C(G)| - n$, and since $\Delta(G, \Pi_n) \le \delta = \epsilon \cdot d \cdot n$, we get that

$$|E(G)| = \Delta(G, \Pi_n) + n - |C(G)| < (1 + \epsilon \cdot d) \cdot n - |C(G)| \,. \tag{7.11}$$

Then, the two cases of the proof are as follows.

Case 1: $|C(G)| \le \frac{n}{6 \cdot d}$**.** Denote the number of vertices with degree d in G by m. Then, relying on Eq. (7.11), we get that

$$m \cdot d \le \sum_{i \in [n]} \deg(i) = 2 \cdot |E(G)| < (2 + 2 \cdot \epsilon \cdot d) \cdot n \,.$$

It follows that $m < \left(\frac{2}{d} + 2 \cdot \epsilon\right) \cdot n$, and since $d \ge 3$ and $\epsilon < \frac{1}{12 \cdot d^2} < \frac{1}{6}$, we get that $m < \frac{5}{6} \cdot n$. Therefore, there exist more than $n/6$ vertices with degree at most $d-1$ in the graph. Hence, the expected number of vertices with degree at most $d-1$ in a uniformly chosen connected component in the graph is $\frac{n-m}{|C(G)|} > \frac{n/6}{n/6d} = d$. Since the inequality is strict, it follows that there exists a connected component in which there are at least $d+1$ vertices that each have degree at most $d-1$. At least two of these vertices are not adjacent; connecting them by an edge yields a graph G' such that $|C(G')| = |C(G)|$ and $|E(G')| = |E(G)| + 1$. It follows that $\Delta(G', \Pi_n) = \Delta(G, \Pi_n) + 1$.

Case 2: $|C(G)| > \frac{n}{6 \cdot d}$**.** Relying on the hypothesis of the case and on Eq. (7.11), we get that

$$|E(G)| < \left(1 + \epsilon \cdot d - \frac{1}{6 \cdot d}\right) \cdot n \,.$$

Now, since $\epsilon < \frac{1}{12 \cdot d^2}$, it follows that $|E(G)| < \left(1 - \frac{1}{12 \cdot d}\right) \cdot n$, which implies that there exist $\Omega(n)$ vertices with degree at most one in the graph. For a sufficiently large n, it follows that there exist at least three non-adjacent vertices in the graph with degree at most one. To construct G', we add edges between these three vertices (i.e., we add a triangle on these vertices). This yields a graph that does not violate the degree bound (since $d \ge 3$) and that satisfies $|C(G')| \ge |C(G)| - 2$ and $|E(G')| = |E(G)| + 3$. It follows that $\Delta(G', \Pi_n) \ge \Delta(G, \Pi_n) + 1$. $\square$

This completes the proof of Proposition 7.9. ■

Proposition 7.9 implies that there is a gap of $\delta/3 = \Omega(n)$ between the distance of graphs in $\mathcal{F}_\delta(\Pi_n)$ from Π_n and the distance of graphs in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$ from Π_n. Thus, to distinguish between graphs in $\mathcal{F}_\delta(\Pi_n)$ and graphs in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$, it suffices to estimate the distance of an input graph from Π_n, up to an additive error of $\frac{1}{6} \cdot \delta = \frac{\epsilon \cdot d}{6} \cdot n$. Using the algorithm of Marko and Ron [MR06, Sec. 5], this can be done using $O(\epsilon^{-3} \cdot d^{-3})$ queries. Thus, we have the following result:

Theorem 7.10 *(a tester for the dual problem of testing cycle-free graphs). Let $d \geq 3$, let $\epsilon < \frac{1}{12 \cdot d^2}$, let n be a sufficiently large integer and let $\delta = \epsilon \cdot d \cdot n$. Then, there exists an algorithm with query complexity $O\left(\epsilon^{-3} \cdot d^{-3}\right)$ that accepts, with high probability, every graph in $\mathcal{F}_\delta(\Pi_n)$, and rejects, with high probability, every graph in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$.*

7.3 Testing the property of being far from bipartite

In this section we study the dual problem of bipartiteness, and, more generally, of testing k-colorability: For $k \geq 2$ and every $\epsilon > 0$, we are interested in the problem of ϵ-testing the set of graphs that are $(\epsilon \cdot d \cdot n)$*-far from being k-colorable*. We show that this problem is different from its original problem. Similar to the problem of testing cycle-free graphs (i.e., to Section 7.2), in the current section we also consider *only simple graphs*.

Proposition 7.11 *(the set of k-colorable graphs with degree bound d is not $\mathcal{F}_\delta$-closed). For any $k \geq 2$, and $d \geq k+1$, and sufficiently large $n \in \mathbb{N}$, and $\delta \geq 2$, the set of k-colorable graphs over $[n]$ with degree bound d, denoted by Π_n, is not $\mathcal{F}_\delta$-closed.*

Proof. Similarly to the proofs of Propositions 6.1, 6.5, 7.5, and 7.8, it suffices to show a graph G such that $\Delta(G, \Pi_n) = 1$, and all neighbors of G are either in Π_n or adjacent to Π_n (since we can then rely on Proposition 2.3 to deduce that Π_n is not $\mathcal{F}_\delta$-closed for every $\delta \geq 2$). The construction of G is identical to the one in the proof of Proposition 6.1: The graph G contains a single $(k+1)$ clique alongside $n-(k+1)$ isolated vertices. In the proof of Proposition 6.1 we showed that adding or removing an edge from G yields a graph that is either k-colorable, or adjacent to the set of k-colorable graphs. To conclude the proof, we observe that all graphs involved in the proof do not violate the degree degree bound d. ■

For $k = 2$ (i.e., testing bipartiteness), the query complexity of the original k-colorability problem is $\tilde{\Theta}(\sqrt{n})$: The lower bound was shown in [GR02] and the upper bound in [GR99]. Therefore, the query complexity of the dual problem is lower bounded by $\Omega(\sqrt{n})$. For $k = 3$, the original problem requires $\Omega(n)$ queries [BOT02], and thus so does the dual problem.

8 A generalization: On being δ'-far from δ-far

In this section we study a more general notion of dual testing problems. Given a property Π, we consider *two* proximity parameters, $\epsilon > 0$ and $\epsilon' > 0$, such that $\epsilon > 0$ determines the "yes" inputs for testing, and $\epsilon' > 0$ is the proximity parameter that determines the distance of the "no" inputs from the "yes" inputs. That is, given $\epsilon, \epsilon' > 0$, the generalized dual problem consists of distinguishing between $\mathcal{F}_{\epsilon \cdot n}(\Pi_n)$ and $\mathcal{F}_{\epsilon' \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n))$. The main result in this section is Theorem 1.15, which asserts the existence of testers (with query complexity that only depends on ϵ and on ϵ') for three generalized dual problems.

Our formal definition of generalized dual problems, which is presented below, coincides with the *standard notion of property testing*, and is thus more natural in that context than the non-generalized notion of dual problems (which we have used so far). Specifically, given a set $\Pi = \{\Pi_n\}_{n \in \mathbb{N}}$ and a parameter $\epsilon > 0$, we will consider the *generalized ϵ-dual problem of* Π, which is just the problem of testing the fixed property $\{\mathcal{F}_{\epsilon \cdot n}(\Pi_n)\}_{n \in \mathbb{N}}$ with an arbitrarily small proximity parameter, denoted by $\epsilon' > 0$. More formally:

Definition 8.1 *(generalized dual problems). For a set Σ, and $\Pi = \{\Pi_n\}_{n \in \mathbb{N}}$ such that $\Pi_n \subseteq \Sigma^n$, and two parameters $\epsilon, \epsilon' > 0$, an ϵ'*-tester for the generalized ϵ-dual problem of Π *is a probabilistic algorithm T that gets oracle access to $x \in \Sigma^n$ and satisfies the following two conditions:*

1. *If $x \in \mathcal{F}_{\epsilon \cdot n}(\Pi_n)$ then $\Pr[T^x(1^n) = 1] \geq \frac{2}{3}$.*
2. *If $x \in \mathcal{F}_{\epsilon' \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n))$ then $\Pr[T^x(1^n) = 0] \geq \frac{2}{3}$.*

The query complexity of a generalized dual problem is defined in the natural way, and is a function of ϵ, ϵ', and n.

When $\epsilon' > \epsilon$, it is possible that $\mathcal{F}_{\epsilon' \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n))$ does not contain Π_n, and it might even be that $\mathcal{F}_{\epsilon' \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n)) = \emptyset$ (e.g., consider a "pathological" example in which $\Pi_n \subseteq \{0,1\}^n$, and $|\mathcal{F}_{\epsilon \cdot n}(\Pi_n)| > 1$, and $\epsilon' = 1$). Even if that does not happen, the following observation asserts that when $\epsilon' > \epsilon$, the problem of distinguishing $\mathcal{F}_{\epsilon \cdot n}(\Pi_n)$ from $\mathcal{F}_{\epsilon' \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n))$ reduces to our standard notion of dual testing problems.

Observation 8.2 *($\epsilon' > \epsilon$ reduces to $\epsilon' = \epsilon$). Let $\Pi = \{\Pi_n\}_{n \in \mathbb{N}}$ such that the query complexity of the dual problem of Π is $q(n, \epsilon)$ (i.e., for every $\epsilon > 0$ and $n \in \mathbb{N}$, a tester can distinguish between $\mathcal{F}_{\epsilon \cdot n}(\Pi_n)$ and $\mathcal{F}_{\epsilon \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n))$ using $q(n, \epsilon)$ queries). Then, for every $\epsilon' > \epsilon > 0$, the problem of distinguishing $\mathcal{F}_{\epsilon \cdot n}(\Pi_n)$ from $\mathcal{F}_{\epsilon' \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n))$ can be solved using $q(n, \epsilon)$ queries.*

Proof. For every $n \in \mathbb{N}$ and $\epsilon' > \epsilon$, observe that $\mathcal{F}_{\epsilon' \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n)) \subseteq \mathcal{F}_{\epsilon \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n))$, since every input that is $(\epsilon' \cdot n)$-far from $\mathcal{F}_{\epsilon \cdot n}(\Pi_n)$ is also $(\epsilon \cdot n)$-far from Π_n. Thus, an algorithm that distinguishes $\mathcal{F}_{\epsilon \cdot n}(\Pi_n)$ from $\mathcal{F}_{\epsilon \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n))$ also distinguishes $\mathcal{F}_{\epsilon \cdot n}(\Pi_n)$ from $\mathcal{F}_{\epsilon' \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n))$. ■

We thus focus on the case of $\epsilon' \leq \epsilon$. Our main results in this section are obtained by *reducing generalized dual problems to tolerant testing problems*. We say that a generalized dual problem reduces to the corresponding tolerant testing problem if for every $\epsilon' \leq \epsilon$, the distance of inputs in $\mathcal{F}_{\epsilon' \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n))$ from Π_n is bounded away from $\epsilon \cdot n$; specifically, if for every sufficiently small $\epsilon > 0$ and every $\epsilon' \leq \epsilon$ there exists $\alpha \in (0,1)$, which may depend on ϵ and on ϵ', such that for every sufficiently large n it holds that $\mathcal{F}_{\epsilon' \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n)) \subseteq \{y : \Delta(y, \Pi_n) \leq \alpha \cdot \epsilon \cdot n\}$.

Proposition 8.3 *(generalized dual problems that reduce to tolerant testing). The following problems reduce to their corresponding tolerant testing problems:*

1. *The generalized dual problem of any error-correcting code with constant relative distance.*
2. *The generalized dual problem of monotone Boolean functions over the Boolean hypercube.*
3. *The generalized dual problem of k-colorable graphs in the dense graph model.*
4. *The generalized dual problem of connected graphs in the bounded-degree graphs model.*
5. *The generalized dual problem of cycle-free graphs in the bounded-degree graphs model.*

Proof. All the metric spaces corresponding to properties in Proposition 8.3 are graphical, and thus we restrict our discussion to graphical metric spaces. Fix a property Π_n, and let $\epsilon, \epsilon' > 0$. Relying on Proposition 2.5, if there exists $m = O(1)$ such that for every input x that is not $(\epsilon \cdot n)$-far from Π_n there exists an input x' that is farther from Π_n such that $\Delta(x, x') \leq m$, then the distance of inputs in $\mathcal{F}_{\epsilon' \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n))$ from Π_n is at most $\epsilon \cdot n - \frac{\epsilon' \cdot n}{m} = \left(1 - \frac{\epsilon'}{\epsilon \cdot m}\right) \cdot \epsilon \cdot n$. Hence, in order to reduce a generalized ϵ-dual problem to the corresponding tolerant testing problem, it suffices to show the above (and set $\alpha = 1 - \frac{\epsilon'}{\epsilon \cdot m}$).

For Items (1) and (2), we use the fact that the corresponding properties are strongly $\mathcal{F}_{\epsilon \cdot n}$-closed for a sufficiently small $\epsilon > 0$ (see Propositions 3.4 and 4.3, respectively), which corresponds to the case of $m = 1$. Item (3) is proved by observing that (in the proof of Proposition 6.2) we showed that for every graph G that is not $(\epsilon \cdot n)$-far from being k-colorable, there exists a graph G' that is farther from being k-colorable, compared to G, such that $\Delta(G, G') = O(1)$. Items (4) and (5) follow from Claim 7.6.1 and Claim 7.9.1, respectively. ■

Relying on Proposition 8.3 and on several previous results regarding standard dual problems, we obtain the following upper bounds on the query complexity of generalized dual problems.

Corollary 8.4 *(testers for generalized dual problems; Theorem 1.15, restated).*

1. *The query complexity of the generalized dual problem of k-colorable graphs in the dense graphs model is $F(\epsilon, \epsilon')$, for some function F that does not depend on n.*
2. *The query complexity of the generalized dual problem of connected graphs in the bounded-degree graphs model is* $\text{poly}(1/\min\{\epsilon', \epsilon\})$.
3. *The query complexity of the generalized dual problem of cycle-free graphs in the bounded-degree graphs model is* $\text{poly}(1/\min\{\epsilon', \epsilon\})$.

Proof. Relying on Proposition 8.3, the query complexity of all three problems is upper bounded by the query complexity of their corresponding tolerant testing problems. Specifically, in the proof of Proposition 8.3 we showed that the query complexity of the dual problems is upper bounded by the query complexity of distinguishing between $\mathcal{F}_{\epsilon \cdot n}(\Pi_n)$ and $\{y : \Delta(y, \Pi_n) \leq \alpha \cdot \epsilon \cdot n\}$, where $\alpha = 1 - \frac{\epsilon'}{\epsilon \cdot O(1)} = 1 - \Omega(\epsilon'/\epsilon)$.

The upper bound in Item (1) follows due to the tolerant tester by Fischer and Newman [FN07]; for a discussion of its query complexity, see [FN07, Sec. 7]. For Item (2) we can use the estimation algorithm presented in Section 7.1.3; and similarly, for Item (3) we can use the estimation algorithm by Marko and Ron [MR06, Sec. 5]. The query complexity of both estimation algorithms is $\mathrm{poly}(1/\rho)$, where ρ is the algorithm's additive error in estimating the relative distance from Π_n. In both cases (corresponding to Items (2) and (3)), the difference between the relative distance of "yes" instances from Π_n and the relative distance of "no" instances from Π_n is $\epsilon - \alpha \cdot \epsilon = \Omega(\epsilon')$. Hence, setting $\rho = O(\epsilon')$, we obtain testers for the corresponding generalized dual problems with query complexity $\mathrm{poly}(1/\epsilon')$. ■

9 Open questions

In the current work we were able to prove one general lower bound on dual testing problems, and several specific upper bounds. However, many interesting and natural *general questions* that concern dual testing problems are left without answer. In this section we suggest a few of these questions, which we suspect might lead towards better understanding of dual testing problems and of "far-from-far" sets.

9.1 Can the query complexity of a dual problem be significantly higher than that of the original problem?

Recall that the (two-sided error) query complexity of a dual problem is lower bounded by the query complexity of the original problem. A natural question is thus:

> **Question 1.** *Does there exist a property such that the query complexity of its dual problem is significantly higher than that of the original problem?*

Note that one of the upper bounds for a dual problem given in this work (i.e., the dual problem of k-colorability in the dense graph model) is significantly higher than the known upper bound for the corresponding original problem. [11]

9.2 Upper bounds for dual problems without reductions to tolerant testing

All the testers we presented for dual problems that are different than the original problems relied on reductions to tolerant testing. Thus, these testers do not fully exploit the structure of "far-from-far" sets, but rather only use the fact that "far-from-far" inputs are sufficiently close to the property. Hence, we ask:

> **Question 2.** *Does there exist a tester for a natural dual problem (that is different than the original problem) that uses significantly fewer queries than the corresponding tolerant tester?*

[11] The original problem is testable using $\mathrm{poly}(1/\epsilon)$ queries [GGR98], whereas the upper bound for the dual problem is a function that has a tower-type dependency on ϵ. The latter is the complexity of the tolerant tester by Fischer and Newman [FN07], which relies on Szemerédi's regularity lemma.

Note that when the dual problem is equivalent to the original problem, the dual problem might indeed be easier to test than the corresponding tolerant testing problem (e.g., in the case of testing whether a distribution is uniform; see Footnote 5).

9.3 Do all graph partition problems reduce to tolerant testing?

Proposition 6.2 asserts that the dual problem of k-colorability in the dense graph model reduces to the corresponding tolerant testing problem. The property of k-colorability is a special case of the *general graph partition problem* (see, e.g., [GGR98, Sec. 1.2.3.1]), but we were unable to prove an analogous result for the general graph partition problem.

Question 3. *Does the general graph partition problem in the dense graph model reduce to tolerant testing?*

Recall that (according to Proposition C.1 and Remark C.2), not all dual problems in the dense graph model reduce to tolerant testing. An initial step towards answering Question 3 might be answering it for the special case of the property of graphs having a *large clique* (see Proposition 1.11).

Acknowledgments

The author thanks his advisor, Oded Goldreich, for suggesting the core questions and observations leading to this work, and for his guidance and support during the research and writing process. The author is grateful to Neta Atzmon for several helpful observations and discussions regarding "far-from-far" sets and dual property testing problems, and for her valuable comments on a draft of this paper. The author thanks Gil Cohen for a helpful discussion regarding decomposition of posets to monotone chains, and Clément Canonne for useful discussions and suggestions. The author also thanks anonymous reviewers for helpful suggestions. This research was partially supported by the Israel Science Foundation (grant No. 671/13).

References

[BFR+13] Tuğkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren D. Smith, and Patrick White. Testing closeness of discrete distributions. *Journal of the ACM*, 60(1):4:1–4:25, 2013.

[BLR90] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47:549–595, 1990.

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

[Can15] Clément Canonne. Your data is big. But is it blue? *Electronic Colloquium on Computational Complexity: ECCC*, 22:63, 2015.

[CRT05] Bernard Chazelle, Ronitt Rubinfeld, and Luca Trevisan. Approximating the minimum spanning tree weight in sublinear time. *SIAM Journal of Computing*, 34(6):1370–1379, 2005.

[Dil50] Robert P. Dilworth. A decomposition theorem for partially ordered sets. *Annals of Mathematics (2)*, 51:161–166, 1950.

[Fis05] Eldar Fischer. The difficulty of testing for isomorphism against a graph that is given in advance. *SIAM Journal of Computing*, 34(5):1147–1158, 2005.

[FLN$^+$02] Eldar Fischer, Eric Lehman, Ilan Newman, Sofya Raskhodnikova, Ronitt Rubinfeld, and Alex Samorodnitsky. Monotonicity testing over general poset domains. In *Proc. 34th Annual ACM Symposium on Theory of Computing (STOC)*, pages 474–483, 2002.

[FM08] Eldar Fischer and Arie Matsliah. Testing graph isomorphism. *SIAM Journal of Computing*, 38(1):207–225, 2008.

[FN07] Eldar Fischer and Ilan Newman. Testing versus estimation of graph properties. *SIAM Journal of Computing*, 37(2):482–501, 2007.

[GGL$^+$00] Oded Goldreich, Shafi Goldwasser, Eric Lehman, Dana Ron, and Alex Samorodnitsky. Testing monotonicity. *Combinatorica*, 20(3):301–337, 2000.

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

[GKNR12] Oded Goldreich, Michael Krivelevich, Ilan Newman, and Eyal Rozenberg. Hierarchy theorems for property testing. *Computational Complexity*, 21(1):129–192, 2012.

[GM12] Gary Gordon and Jennifer McNulty. *Matroids: A Geometric Introduction*. Cambridge University Press, 2012.

[Gol10] Oded Goldreich, editor. *Property Testing - Current Research and Surveys*, volume 6390 of *Lecture Notes in Computer Science*. Springer, 2010.

[GR99] Oded Goldreich and Dana Ron. A sublinear bipartiteness tester for bounded degree graphs. *Combinatorica*, 19(3):335–373, 1999.

[GR02] Oded Goldreich and Dana Ron. Property testing in bounded degree graphs. *Algorithmica*, 32(2):302–343, 2002.

[KD06] Jörg Koppitz and Klaus Denecke. *M-Solid Varieties of Algebras (Advances in Mathematics)*. Springer-Verlag New York, Inc., 2006.

[MR06] Sharon Marko and Dana Ron. Distance approximation in bounded-degree and general sparse graphs. In *Proc. 10th International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*, pages 475–486, 2006.

[PRR06] Michal Parnas, Dana Ron, and Ronitt Rubinfeld. Tolerant property testing and distance approximation. *Journal of Computer and System Sciences*, 72(6):1012–1042, 2006.

[RS96] Ronitt Rubinfeld and Madhu Sudan. Robust characterizations of polynomials with applications to program testing. *SIAM Journal of Computing*, 25(2):252–271, 1996.

[Rub12] Ronitt Rubinfeld. Taming big probability distributions. *ACM Crossroads*, 19(1):24–28, 2012.

[Tel14] Roei Tell. Deconstructions of reductions from communication complexity to property testing using generalized parity decision trees. *Electronic Colloquium on Computational Complexity: ECCC*, 21:115, 2014.

[vdV93] Marcel L.J. van de Vel. *Theory of Convex Structures*. North-Holland Mathematical Library. Elsevier Science, 1993.

[VV11] Gregory Valiant and Paul Valiant. Estimating the unseen: an n/log(n)-sample estimator for entropy and support size, shown optimal via new clts. In *Proc. 43rd Annual ACM Symposium on Theory of Computing (STOC)*, pages 685–694, 2011.

Appendix A Conditions for a set to be $\mathcal{F}_\delta$-closed

In this appendix we characterize the sets that are $\mathcal{F}_\delta$-closed in any metric space, and present sufficient and necessary conditions for a set to be $\mathcal{F}_\delta$-closed in graphical metric spaces

A.1 Characterizations of $\mathcal{F}_\delta$-closed sets in general metric spaces

Intuitively, we expect that any set will be far from being far from itself; that is, we expect every set Π to satisfy $\Pi \subseteq \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$. This is indeed the case:

Fact A.1 *(a set is always far from being far from itself). For any space Ω, $\delta > 0$, and $\Pi \subseteq \Omega$, it holds that $\Pi \subseteq \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$.*

Proof. Assume towards a contradiction that there exists $x \in \Pi \setminus \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$. Since $x \notin \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$, there exists $z \in \mathcal{F}_\delta(\Pi)$ such that $\Delta(x,z) < \delta$. However, since $x \in \Pi$, then $\Delta(z,\Pi) \le \Delta(z,x) < \delta$, which contradicts $z \in \mathcal{F}_\delta(\Pi)$. ■

However, not every set Π satisfies $\Pi = \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ (i.e., not every set is $\mathcal{F}_\delta$-closed). The following theorem presents several equivalent characterizations of the $\mathcal{F}_\delta$-closed sets for any fixed Ω and δ. After the proof we discuss the meaning of some of these characterizations.

Theorem A.2 *(characterizations of $\mathcal{F}_\delta$-closed sets). For any Ω, $\delta > 0$, and $\Pi \subseteq \Omega$, the following statements are equivalent:*

1. *Π is $\mathcal{F}_\delta$-closed (i.e., $\Pi = \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$).*
2. *For every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ there exists $z \in \mathcal{F}_\delta(\Pi)$ such that $\Delta(z, x) < \delta$.*
3. *There exists $\Pi' \subseteq \Omega$ such that $\Pi = \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi'))$.*
4. *There exists $\Pi'' \subseteq \Omega$ such that $\Pi = \mathcal{F}_\delta(\Pi'')$.*
5. *There exists $\Pi'' \subseteq \Omega$ such that $\Pi = \bigcap_{x\in\Pi''} \mathcal{F}_\delta(\{x\})$.*
6. *There exists $\Pi'' \subseteq \Omega$ such that $\Pi = \Omega \setminus \bigcup_{x\in\Pi''} B[x, \delta)$.*

Proof. For the proof we will need the following two facts:

Fact A.2.1 *(far-sets are intersections of sets that are far from singletons). For any Ω, $\delta > 0$ and $\Pi \subseteq \Omega$ it holds that $\mathcal{F}_\delta(\Pi) = \bigcap_{x\in\Pi} \mathcal{F}_\delta(\{x\})$.*

Proof. For any $z \in \Omega$ it holds that $z \in \mathcal{F}_\delta(\Pi)$ if and only if z is δ-far from every $x \in \Pi$, which holds if and only if $z \in \mathcal{F}_\delta(\{x\})$ for every $x \in \Pi$. □

Fact A.2.2 *(downwards monotonicity of $\mathcal{F}_\delta$). For any Ω, $\delta > 0$ and $A, B \subseteq \Omega$, if $A \subseteq B$, then $\mathcal{F}_\delta(A) \supseteq \mathcal{F}_\delta(B)$.*

Proof. Relying on Fact A.2.1,

$$\mathcal{F}_\delta(A) = \bigcap_{a\in A} \mathcal{F}_\delta(\{a\}) \supseteq \bigcap_{b\in B} \mathcal{F}_\delta(\{b\}) = \mathcal{F}_\delta(B)$$

□

We now prove the equivalences of Conditions (1)–(4).

(1) $\implies$ (2) Since Π is $\mathcal{F}_\delta$-closed, every $x \notin \Pi$ satisfies $x \notin \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$. Equivalently, every $x \notin \Pi$ satisfies $\Delta(x, \mathcal{F}_\delta(\Pi)) < \delta$. Thus, for every $x \notin \Pi$, there exists $z \in \mathcal{F}_\delta(\Pi)$ such that $\Delta(x, z) < \delta$. In particular, this holds for every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$.

(2) $\implies$ (1) For any $x \in \Omega$, if there exists $z \in \mathcal{F}_\delta(\Pi)$ such that $\Delta(x, z) < \delta$, then $x \notin \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$. Combining this fact with the hypothesis, we deduce that $\overline{\Pi \cup \mathcal{F}_\delta(\Pi)} \cap \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) = \emptyset$. Also, since $\delta > 0$ it holds that $\mathcal{F}_\delta(\Pi) \cap \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) = \emptyset$.

Now observe that $\Omega = \Pi \cup \mathcal{F}_\delta(\Pi) \cup \overline{\Pi \cup \mathcal{F}_\delta(\Pi)}$. Since we showed that $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) \cap \mathcal{F}_\delta(\Pi) = \emptyset$ and $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) \cap \overline{\Pi \cup \mathcal{F}_\delta(\Pi)} = \emptyset$ it follows that $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) \subseteq \Pi$. By Fact A.1 it holds that $\Pi \subseteq \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$, and therefore $\Pi = \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$.

(1) $\implies$ (3) Follows by setting $\Pi' = \Pi$, since $\Pi = \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$.

(3) $\implies$ (4) Follows by setting $\Pi'' = \mathcal{F}_\delta(\Pi')$, since $\Pi = \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi')) = \mathcal{F}_\delta(\Pi'')$.

(4) $\Longrightarrow$ (1) Let $\Pi = \mathcal{F}_\delta(\Pi'')$ for some $\Pi'' \subseteq \Omega$. By Fact A.1 it holds that $\Pi'' \subseteq \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi''))$, whereas by Fact A.2.2, we get that $\Pi = \mathcal{F}_\delta(\Pi'') \supseteq \mathcal{F}_\delta(\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi''))) = \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$. Using Fact A.1 again, we know that $\Pi \subseteq \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$, and thus $\Pi = \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$.

(4) $\Longleftrightarrow$ (5) By Fact A.2.1.

(5) $\Longleftrightarrow$ (6) Follows since for any $x \in \Omega$ it holds that $\mathcal{F}_\delta(\{x\}) = \Omega \setminus B[x, \delta)$, and by De-Morgan's laws. ■

Condition (2) in Theorem A.2 is the basic technical tool that we use in the paper to evaluate whether sets are $\mathcal{F}_\delta$-closed. Note that Condition (2) is in fact a collection of local conditions, where by "local" we mean that each condition depends only on a ball of radius 2δ in Ω. Note that Condition (5) implies that any intersection of $\mathcal{F}_\delta$-closed sets is $\mathcal{F}_\delta$-closed. In addition, Condition (6) provides another appealing interpretation for $\mathcal{F}_\delta$-closed sets: The $\mathcal{F}_\delta$-closed sets are exactly the sets obtained by starting from the entire space Ω and removing any union of balls from the potentially small collection $\{B[x, \delta)\}_{x \in \Omega}$.

The equivalence of Conditions (3) and (4) implies that $\{\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))\}_{\Pi \subseteq \Omega} = \{\mathcal{F}_\delta(\Pi)\}_{\Pi \subseteq \Omega}$, and that the operator $\mathcal{F}_\delta$ is a bijection between these two collections. (This is since the collection $\{\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))\}_{\Pi \subseteq \Omega}$ is the image of $\{\mathcal{F}_\delta(\Pi)\}_{\Pi \subseteq \Omega}$ under $\mathcal{F}_\delta$; and by Condition (4), every set of the form $\mathcal{F}_\delta(\Pi)$ is $\mathcal{F}_\delta$-closed, which implies that the collection $\{\mathcal{F}_\delta(\Pi)\}_{\Pi \subseteq \Omega}$ is also the image of $\{\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))\}_{\Pi \subseteq \Omega}$ under $\mathcal{F}_\delta$.)

A.2 Conditions for a set to be $\mathcal{F}_\delta$-closed in graphical metric spaces

We now focus only on graphical metric spaces, which are connected undirected graphs, equipped with the shortest path metric. Since the distances in such spaces are integer-valued, we assume throughout the section that $\delta \in \mathbb{N}$. As an initial observation, note that for any graphical Ω it holds that $\min_{x \neq y \in \Omega}\{\Delta(x, y)\} = 1$. Note that in any space Ω, if $\delta \leq \min_{x \neq y \in \Omega}\{\Delta(x, y)\}$, then all sets in Ω are $\mathcal{F}_\delta$-closed (since $\mathcal{F}_\delta(\Pi) = \overline{\Pi}$ for any set Π). Thus, in every graphical space, all sets are $\mathcal{F}_1$-closed. Accordingly, in this section we are mainly interested in integer values of $\delta \geq 2$.

Loosely speaking, a necessary condition for a set Π in a graphical space to be $\mathcal{F}_\delta$-closed is that it does not "enclose" some vertex $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ from "all sides". In particular, the following proposition shows that if Π is $\mathcal{F}_\delta$-closed, then every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ is connected to $\mathcal{F}_\delta(\Pi)$ via a path that does not intersect Π (nor any vertex that is adjacent to Π).

Proposition A.3 *(sets that "enclose" some vertex are not $\mathcal{F}_\delta$-closed). For a graphical Ω and $\delta \geq 2$, let $\Pi \subseteq \Omega$ be an $\mathcal{F}_\delta$-*closed set*. Then, for every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$, there exists a path $x = v_0, v_1, ..., v_l = z$ such that $z \in \mathcal{F}_\delta(\Pi)$, and for every $i \in [l]$ it holds that $\Delta(v_i, \Pi) \geq 2$.*

Note that $x = v_0$ itself may be adjacent to Π, and the requirement is that the vertices subsequent to x in the path to $\mathcal{F}_\delta(\Pi)$ will neither be in Π nor adjacent to Π.

Proof of Proposition A.3. Let Ω and $\delta \geq 2$. The key observation is that, for *every* set Π (not necessarily an $\mathcal{F}_\delta$-closed set) and every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$, a shortest path from x to Π does not intersect $\mathcal{F}_\delta(\Pi)$ nor any vertex adjacent to $\mathcal{F}_\delta(\Pi)$.

Fact A.3.1. *For a graphical Ω, and $\delta \geq 2$, let $\Pi \subseteq \Omega$ be a set (not necessarily an $\mathcal{F}_\delta$-closed set). Then, for every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ and a shortest path from x to Π, every vertex v subsequent to x on the path satisfies $\Delta(v, \mathcal{F}_\delta(\Pi)) \geq 2$.*

Proof. Let $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$, and let $p \in \Pi$ such that $\Delta(x, \Pi) = \Delta(x, p)$. Let P be a shortest path from x to p. Since P is a shortest path, for every vertex v subsequent to x on the path it holds that v is closer to p than x; since $x \notin \mathcal{F}_\delta(\Pi)$, we get that, $\Delta(v, p) \leq \Delta(x, p) - 1 \leq \delta - 2$. Thus, every neighbor v' of v satisfies $\Delta(v', \Pi) \leq \Delta(v, \Pi) + 1 \leq \delta - 1$, which implies that $v' \notin \mathcal{F}_\delta(\Pi)$. It follows that $\Delta(v, \mathcal{F}_\delta(\Pi)) \geq 2$. □

Now, let Π be an $\mathcal{F}_\delta$-closed set, and let $\Pi' = \mathcal{F}_\delta(\Pi)$. Then, $\Pi = \mathcal{F}_\delta(\Pi')$, which implies that $\Pi' \cup \mathcal{F}_\delta(\Pi') = \Pi \cup \mathcal{F}_\delta(\Pi)$. According to Fact A.3.1, for every $x \notin \Pi' \cup \mathcal{F}_\delta(\Pi') = \Pi \cup \mathcal{F}_\delta(\Pi)$, a shortest path from x to $\Pi' = \mathcal{F}_\delta(\Pi)$ does not intersect $\mathcal{F}_\delta(\Pi') = \Pi$ nor any vertex adjacent to Π. ■

The condition in Proposition A.3 is not sufficient to deduce that a set is $\mathcal{F}_\delta$-closed. To see this, consider the graph depicted in Figure 1 and $\delta = 3$. Let $\Pi = \{p\}$, and note that $\mathcal{F}_3(\{p\}) = \{z\}$. Each vertex $v_1, ..., v_4 \notin \{p\} \cup \mathcal{F}_3(\{p\})$ has a path starting from itself and reaching z such that the path does not intersect p or any of its neighbors. Thus, $\{p\}$ meets the necessary condition implied by Proposition A.3. However, since $\mathcal{F}_3(\mathcal{F}_3(\{p\})) = \{p, v_1\}$, it follows that $\{p\}$ is not $\mathcal{F}_3$-closed.

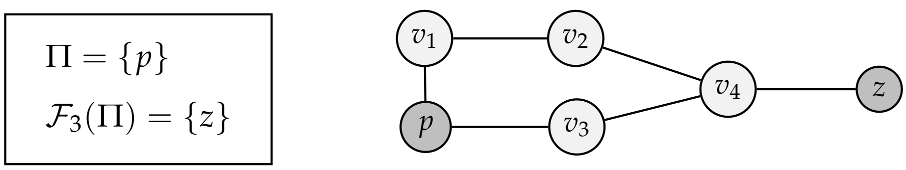

Figure 1: The singleton $\{p\}$ is not $\mathcal{F}_3$-closed, although the necessary condition stated in Proposition A.3 is satisfied.

We now demonstrate that the condition implied by Proposition A.3 is not sufficient for a set to be $\mathcal{F}_\delta$-closed even when the metric space is the Boolean hypercube.

Proposition A.4 *(the condition in Proposition A.3 is not sufficient to be $\mathcal{F}_\delta$-closed in the hypercube). For $n \geq 3$, let H_n be the n-dimensional Boolean hypercube. Then, there exists a set $\Pi \subseteq H_n$ such that for every $4 \leq \delta \leq n - 1$:*

1. *For every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ there exists a path $p = v_0, v_1, ...x = v_r, ..., v_l = z$ such that for every $i \in [l]$ it holds that $\Delta(v_i, \Pi) \geq 2$.*
2. *Π is not $\mathcal{F}_\delta$-closed.*

Proof. For the proof it will be convenient to identify every vertex $v \in \{0,1\}^n$ of H_n with the corresponding subset of $[n]$; that is, the subset $\{i \in [n] : v_i = 1\}$. Let

$$\Pi = \{\{1\}, \{2\}, ..., \{n-2\}\}$$

and let $4 \leq \delta \leq n-1$.

To prove the first statement, for any $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$, we show a path satisfying the requirements. First note that since $\Pi \subseteq \{v : |v| = 1\}$, for any w such that $|w| \geq 2$ it holds that $\Delta(w, \Pi) \geq |w| - 1$, since we need to remove at least $|w| - 1$ elements from w to reach Π. In particular, this implies that:

- For every w such that $|w| \geq 3$ it holds that $\Delta(w, \Pi) \geq 2$.
- $\Delta([n], \Pi) \geq n-1$, and since $\delta \leq n-1$ we get that $[n] \in \mathcal{F}_\delta(\Pi)$.

Combining these two facts, we deduce that if $|x| \geq 2$, then there exists a path from x to $[n] \in \mathcal{F}_\delta(\Pi)$ such that every vertex v subsequent to x in the path satisfies $\Delta(v, \Pi) \geq 2$: This path is obtained by just adding elements to x (in arbitrary order). It is thus left to show that for every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ such that $|x| \leq 1$ there exists a path from x to $\mathcal{F}_\delta(\Pi)$ that does not intersect Π nor vertices adjacent to Π. Note that it suffices to show such a path from x to x' such that $|x'| = 2$.

Now, the only vertices that satisfy both $|x| \leq 1$ and $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ are $\emptyset$, $\{n-1\}$, and $\{n\}$. For $\emptyset$, we take the path $\emptyset, \{n\}, \{n-1, n\}$, and indeed $\{n\}$ and $\{n-1, n\}$ are neither in Π nor adjacent to Π. Similarly, for $\{n\}$ we take the path $\{n\}, \{n-1, n\}$, whereas for $\{n-1\}$ we take the path $\{n-1\}, \{n-1, n\}$. This completes the proof of Item (1).

To show that Π is not $\mathcal{F}_\delta$-closed, we rely on Condition (2) of Theorem A.2. Note that $\Delta(\emptyset, \Pi) = 1$, and hence $\emptyset \notin \Pi \cup \mathcal{F}_\delta(\Pi)$. We will show that for every $z \in \mathcal{F}_\delta(\Pi)$ it holds that $\Delta(z, \emptyset) \geq \delta$. Assume towards a contradiction that there exists $z \in \mathcal{F}_\delta(\Pi)$ such that $\Delta(z, \emptyset) \leq \delta - 1$, which implies that $|z| \leq \delta - 1$.

- If $|z| \leq \delta - 2$, then we can remove all elements from z, and add the element 1, to obtain $\{1\} \in \Pi$. Therefore $\Delta(z, \Pi) \leq \Delta(z, \{1\}) \leq |z| + 1 \leq \delta - 1$, which contradicts $z \in \mathcal{F}_\delta(\Pi)$.
- If $|z| = \delta - 1 \geq 3$, since $\bigcup_{p \in \Pi} p = [n] \setminus \{n, n-1\}$, it follows that z intersects the set $\bigcup_{p \in \Pi} p$. Thus, for some $p \in \Pi$, it holds that $z \cap p \neq \emptyset$, and since Π only contains singletons, it follows that $z \cap p = p$. By removing the $\delta - 2$ elements that are not in $z \cap p$ from z, we obtain $p \in \Pi$, meaning that $\Delta(z, \Pi) \leq \Delta(z, p) \leq \delta - 2$, which contradicts $z \in \mathcal{F}_\delta(\Pi)$.

Having shown that Π is not $\mathcal{F}_\delta$-closed, the proposition follows. ■

We now present a *sufficient* condition for a set in a graphical metric space to be $\mathcal{F}_\delta$-closed. Recall that a set is *strongly $\mathcal{F}_\delta$-closed* if for any $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ there exist a neighbor x' of x that is farther from Π than x itself.

Definition A.5 *(strongly $\mathcal{F}_\delta$-closed sets; Definition 2.4, restated). For a graphical Ω and $\delta > 0$, a set $\Pi \subseteq \Omega$ is* strongly $\mathcal{F}_\delta$-closed *if and only if for every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ there exists a neighbor x' of x such that $\Delta(x', \Pi) = \Delta(x, \Pi) + 1$.*

Proposition A.6 *(strongly $\mathcal{F}_\delta$-closed sets are $\mathcal{F}_\delta$-closed). Let Ω be a graphical space, let $\delta > 0$, and let $\Pi \subseteq \Omega$ be a strongly $\mathcal{F}_\delta$-closed set. Then, Π is $\mathcal{F}_\delta$-closed.*

Proof. We will show that for every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ there exists $z \in \mathcal{F}_\delta(\Pi)$ such that $\Delta(x, z) = \delta - \Delta(x, \Pi) < \delta$, and rely on Item (2) of Theorem A.2 to deduce that Π is $\mathcal{F}_\delta$-closed. Let $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ and denote $x_0 = x$. By the hypothesis, x_0 has a neighbor x_1 such that $\Delta(\Pi, x_1) = \Delta(\Pi, x_0) + 1$. If $\Delta(x_1, \Pi) = \delta$ we are done, since this implies that $\Delta(x, \Pi) = \delta - 1$ and hence $\Delta(x, x_1) = 1 = \delta - \Delta(x, \Pi)$. Otherwise, note that $x_1 \notin \Pi \cup \mathcal{F}_\delta(\Pi)$, since $\Delta(x_1, \Pi) > \Delta(x_0, \Pi) > 0$, and hence we can apply the hypothesis again to obtain a neighbor x_2 of x_1 such that $\Delta(x_2, \Pi) = \Delta(x_1, \Pi) + 1$. Repeatedly applying this step, in the i^{th} application we have that $\Delta(x_i, \Pi) = \Delta(x, \Pi) + i$ and $\Delta(x_i, x) = i$. As long as $i < \delta - \Delta(x, \Pi)$ we can continue applying the step, since $\Delta(x_i, \Pi) = \Delta(x, \Pi) + i < \delta$, and hence $x_i \notin \Pi \cup \mathcal{F}_\delta(\Pi)$, and so we rely on the hypothesis to obtain x_{i+1}. When $i = \delta - \Delta(x, \Pi)$ we get that $\Delta(x_{\delta - \Delta(x,\Pi)}, \Pi) = \delta$ and $\Delta(x_{\delta - \Delta(x,\Pi)}, x) = \delta - \Delta(x, \Pi)$, which is what we wanted. ∎

The condition in Definition 2.4 is more convenient to evaluate in some cases than the conditions in Theorem A.2: When one seeks to prove that a set is strongly $\mathcal{F}_\delta$-closed, and given a vertex $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$, one does not need to reason about $\mathcal{F}_\delta(\Pi)$, but only to find an immediate neighbor of x that is farther away from Π than x. Unfortunately, being strongly $\mathcal{F}_\delta$-closed is not a necessary condition for being $\mathcal{F}_\delta$-closed.

To see this, consider the graph depicted in Figure 2, with $\delta = 3$. Let $\Pi = \{p\}$, and note that $\mathcal{F}_\delta(\{p\}) = \{z\}$, and $\mathcal{F}_\delta(\mathcal{F}_\delta(\{p\})) = \mathcal{F}_\delta(\{z\}) = \{p\}$. Hence $\{p\}$ is $\mathcal{F}_\delta$-closed. However, the vertex b does not lie on a shortest path between $\{p\}$ and $\{z\}$, and thus $\{p\}$ is not strongly $\mathcal{F}_\delta$-closed.

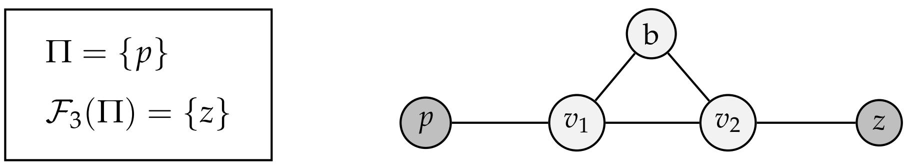

Figure 2: The singleton $\{p\}$ is $\mathcal{F}_3$-closed but not strongly $\mathcal{F}_3$-closed.

Moreover, being strongly $\mathcal{F}_\delta$-closed is not a necessary condition for being $\mathcal{F}_\delta$-closed even in the special case where the graph is the Boolean hypercube.

Proposition A.7 *($\mathcal{F}_\delta$-closed sets that are not strongly $\mathcal{F}_\delta$-closed). For $n \geq 9$ and $4 \leq \delta \leq \frac{n}{2}$ such that $\delta - 1$ divides n, there exist sets in the Boolean hypercube that are $\mathcal{F}_\delta$-closed but are not strongly $\mathcal{F}_\delta$-closed.*

Proof. Similar to the proof of Proposition A.4, in the current proof it will be convenient to identify every vertex $v \in \{0,1\}^n$ with the corresponding subset of $[n]$ that v indicates (i.e., the set $\{i : v_i = 1\}$). Also recall that for $x, y \in \{0,1\}^n$ we denote by $\mathtt{sd}(x, y)$ the symmetric difference between x and y, and that $\Delta(x, y) = |\mathtt{sd}(x, y)|$.

Let $n \in \mathbb{N}$ and δ be as in the hypothesis. The set Π is an equipartition of $[n]$ to $n/(\delta - 1)$ sets, each of cardinality $\delta - 1$; specifically,

$$\Pi = \{\{1, ..., \delta - 1\}, \{\delta, ..., 2 \cdot \delta - 2\}, ..., \{n - \delta + 2, ..., n\}\} \ .$$

We will first show that Π is not strongly $\mathcal{F}_\delta$-closed, and then show that Π is $\mathcal{F}_\delta$-closed.

Claim A.7.1. *Π is not strongly $\mathcal{F}_\delta$-closed.*

Proof. Note that $\Delta(\emptyset, \Pi) = \delta - 1 \in (0, \delta)$, and hence $\emptyset \notin \Pi \cup \mathcal{F}_\delta(\Pi)$; we show that $\emptyset$ has no neighbor that is farther from Π than $\emptyset$ itself. Note that the neighbors of $\emptyset$ are singletons. Since $\bigcup_{p \in \Pi} p = [n]$, for every singleton x' there exists $p \in \Pi$ such that $p \cap x' \neq \emptyset$, which implies that $\Delta(x', \Pi) \leq \Delta(x', p) \leq \delta - 2$. It follows that $\Delta(x', \Pi) < \Delta(\emptyset, \Pi)$. Thus, Π is not strongly $\mathcal{F}_\delta$-closed. □

To prove that Π is $\mathcal{F}_\delta$-closed we will need the following two facts:

Fact A.7.2 *(all sets of size at least $2 \cdot \delta - 1$ are in $\mathcal{F}_\delta(\Pi)$). There exists $z \subseteq [n]$ satisfying $|z| \geq 2 \cdot \delta - 1$, and for any such z it holds that $z \in \mathcal{F}_\delta(\Pi)$.*

Proof. Since $2 \cdot \delta - 1 \leq n$ there exist sets of cardinality $2 \cdot \delta - 1$. Every such set z satisfies $z \in \mathcal{F}_\delta(\Pi)$, since $\Pi \subseteq \{v : |v| = \delta - 1\}$, and since we need to remove at least δ elements from z to obtain a set of cardinality $\delta - 1$. □

Fact A.7.3 *(there exist sets of size 3 that are in $\mathcal{F}_\delta(\Pi)$). There exists $z \subseteq [n]$ such that $|z| = 3$ and for every $p \in \Pi$ it holds that $|z \cap p| \leq 1$. For any such z it holds $z \in \mathcal{F}_\delta(\Pi)$.*

Proof. To see that z as in the statement exists, note that $\frac{n}{\delta - 1} > 2$, and hence there exist at least three distinct subsets in Π. A suitable z is comprised of three elements, each from one of those three distinct subsets in Π. For such a set z it holds that

$$\begin{aligned}
|\mathtt{sd}(z, p)| &= |(z \cup p) \setminus (z \cap p)| \\
&= |z| + |p| - 2 \cdot |z \cap p| \\
&\geq 3 + (\delta - 1) - 2 \cdot 1 \\
&= \delta
\end{aligned}$$

and thus $\Delta(z, \Pi) \geq \delta$. □

It is thus left to show that Π is $\mathcal{F}_\delta$-closed. To do this we rely on Condition (2) from Theorem A.2: For $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ we show that there exists $z \in \mathcal{F}_\delta(\Pi)$ such that $\Delta(x, z) \leq \delta - 1$.

Let $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$. First, relying on Fact A.7.2 and on the hypothesis that $x \notin \mathcal{F}_\delta(\Pi)$, it follows that $|x| < 2 \cdot \delta - 1$. Now, if $|x| \in [\delta, 2 \cdot \delta - 1)$, then we can add $(2 \cdot \delta - 1) - |x|$ elements from $[n] \setminus x$ to x, thereby obtaining a subset z of cardinality $|z| = 2 \cdot \delta - 1$ satisfying

$\Delta(x,z) = (2 \cdot \delta - 1) - |x| \leq \delta - 1$. Relying on Fact A.7.2, again, it holds that $z \in \mathcal{F}_\delta(\Pi)$. Hence the condition holds.

We are left with the case of $|x| \leq \delta - 1$. In this case we show that it is possible to modify x to a subset as in Fact A.7.3 (i.e., a subset z such that $|z| = 3$ and $|z \cap p| \leq 1$ for every $p \in \Pi$), by at most $\delta - 1$ actions of adding elements to x or removing elements from it. Since such z is in $\mathcal{F}_\delta(\Pi)$, once we show this it will follow that there exists $z \in \mathcal{F}_\delta(\Pi)$ such that $\Delta(x,z) \leq \delta - 1$.

Recall that for $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ such that $|x| \leq \delta - 1$, we wish to present a set z such that $\Delta(x,z) \leq \delta - 1$, and $|z| = 3$, and for every $p \in \Pi$ it holds that $|z \cap p| \leq 1$. Also recall that, as mentioned in the proof of Fact A.7.3, since $\frac{n}{\delta-1} > 2$, there exist at least three distinct subsets in Π. We proceed by a case analysis:

- If $x = \emptyset$, then we can reach a suitable z with three actions (which is less than $\delta \geq 4$) by adding one element from each of three distinct subsets in Π.
- If x intersects with a single subset $p \in \Pi$, then it holds that $|x| = |x \cap p| \leq \delta - 2$, otherwise $x = p \in \Pi$, contradicts $x \notin \Pi$. Therefore we can remove $|x| - 1 \leq \delta - 3$ arbitrary elements from x, and then add to x two elements from two distinct subsets $p_1, p_2 \neq p$ from Π, thereby reaching a suitable z with at most $\delta - 1$ actions.
- If x intersects with $k \geq 2$ subsets of Π, denote these subsets by $\{p_1, ..., p_k\}$. We start by removing all elements from x, except for a single element from p_1 and a single element from p_2. Since $|x| \leq \delta - 1$ we performed at most $\delta - 3$ actions so far. We now add to x an element from a subset $p_3 \in \Pi$ such that $p_3 \neq p_1, p_2$, thereby reaching a suitable z with at most $\delta - 2$ actions. ■

Appendix B Existence and prevalence of sets that are not $\mathcal{F}_\delta$-closed

The focus of this appendix is on proving the existence, and in some sense the abundance, of sets that are not $\mathcal{F}_\delta$-closed. First, we will demonstrate the *generality* of the phenomenon of sets that are not $\mathcal{F}_\delta$-closed, by showing that for *any metric space* in which not all points are equidistant and any δ that is not "too extreme" there exist non-trivial sets that are $\mathcal{F}_\delta$-closed and non-trivial sets that are not $\mathcal{F}_\delta$-closed. Next, we will *lower bound the number* of sets that are not $\mathcal{F}_\delta$-closed in two special cases: One is when we assume some conditions on the structure of the metric space and the other is in the Boolean hypercube.

Let us now consider a fixed Ω, and delineate two "extreme" settings for δ that collapse $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ to a trivial operator. On the one hand, if $\delta > \sup_{x,y \in \Omega}\{\Delta(x,y)\}$, then $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) = \Omega$ for any non-empty set Π (since $\mathcal{F}_\delta(\Pi) = \emptyset$), in which case all non-trivial sets are not $\mathcal{F}_\delta$-closed. On the other hand, if $\delta \leq \inf_{x \neq y \in \Omega}\{\Delta(x,y)\}$, then $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) = \Pi$ for any set Π (since $\mathcal{F}_\delta(\Pi) = \overline{\Pi}$), in which case all sets are $\mathcal{F}_\delta$-closed. Thus, disregarding for a moment the "boundary case" in which $\delta = \sup_{x \neq y}\{\Delta(x,y)\}$, we restrict our investigation to settings of δ such that

$$\delta \in \left(\inf_{x \neq y \in \Omega} \{\Delta(x,y)\}, \sup_{x,y \in \Omega} \{\Delta(x,y)\} \right) . \tag{B.1}$$

Needless to say, if neither the supremum nor the infimum exist, then Eq. (B.1) poses no restriction at all on δ. It turns out that for *every* δ that satisfies Eq. (B.1), there exists a non-trivial $\Pi \subseteq \Omega$ that is $\mathcal{F}_\delta$-closed and a non-trivial $\Pi' \subseteq \Omega$ that is not $\mathcal{F}_\delta$-closed.

Theorem B.1 *(Theorem 1.2, restated). For any Ω, if $\delta > 0$ satisfies Eq. (B.1), then there exists a non-trivial $\Pi \subseteq \Omega$ that is $\mathcal{F}_\delta$-closed and a non-trivial $\Pi' \subseteq \Omega$ that is not $\mathcal{F}_\delta$-closed.*

Proof. Since $\delta < \sup_{x,y \in \Omega}\{\Delta(x,y)\}$ there exist $x, y \in \Omega$ such that $\Delta(x,y) \geq \delta$. Let $\Pi = \mathcal{F}_\delta(\{x\})$, and note that $\Pi \notin \{\emptyset, \Omega\}$ since $x \notin \Pi$ and $y \in \Pi$. By Condition (4) of Theorem A.2 it holds that Π is $\mathcal{F}_\delta$-closed.

Now, since $\delta > \inf_{x \neq y \in \Omega}\{\Delta(x,y)\}$ there exist $x', y' \in \Omega$ such that $\Delta(x',y') < \delta$. Let $\Pi' = \Omega \setminus \{x'\}$, and note that $\Pi' \notin \{\emptyset, \Omega\}$ since $x' \notin \Pi'$ and $y' \in \Pi'$. Since $\Delta(x', \Pi') \leq \Delta(x',y') < \delta$ it follows that $x' \notin \mathcal{F}_\delta(\Pi')$, and thus $\mathcal{F}_\delta(\Pi') = \emptyset$ and $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi')) = \Omega \neq \Pi'$. Therefore Π' is not $\mathcal{F}_\delta$-closed. ■

For spaces in which the supremum in Eq. (B.1) is attained (e.g., finite metric spaces) such non-trivial sets exist *if and only if* $\delta \in \left(\inf_{x \neq y \in \Omega}\{\Delta(x,y)\}, \max_{x,y \in \Omega}\{\Delta(x,y)\}\right]$. (Note that now the right boundary of the interval is closed.) This is the case because when $\delta = \max_{x,y \in \Omega}\{\Delta(x,y)\}$, where the maximal distance is attained between $u, v \in \Omega$, we have the $\mathcal{F}_\delta$-closed set $\Pi = \mathcal{F}_\delta(\{u\}) \neq \emptyset$ (similar to the proof of Theorem B.1).

Theorem B.1 implies that for any Ω and $\delta > 0$ that satisfies Eq. (B.1) there *exist* non-trivial $\mathcal{F}_\delta$-closed sets and non-trivial sets that are not $\mathcal{F}_\delta$-closed. The following proposition assumes slightly stricter conditions on the structure of Ω with respect to a parameter δ, and under these conditions yields a lower bound on the *number* of sets that are not $\mathcal{F}_\delta$-closed.

Proposition B.2 *(lower bound on the number of sets that are not $\mathcal{F}_\delta$-closed). Let Ω be a metric space and $\delta > 0$. Assume that for $n \in \mathbb{N}$ and $m \geq 2$ there exist $x_1, ..., x_n \in \Omega$ such that for every $i \neq j \in [n]$ it holds that $\Delta(x_i, x_j) \geq 2\delta$ and $2 \leq |B[x_i, \delta)| \leq m$. Then, the probability that a uniformly chosen random set is $\mathcal{F}_\delta$-closed is at most $(1 - 2^{-m})^n$.*

Proof. By the hypothesis, for any $i \in [n]$ it holds that $|B[x_i, \delta)| \geq 2$. Therefore, if we choose Π such that $\Pi \cap B[x_i, \delta) = B[x_i, \delta) \setminus \{x_i\}$, we get a set such that $x_i \notin \Pi$ and $B[x_i, \delta) \cap \Pi \neq \emptyset$ and $B[x_i, \delta) \cap \mathcal{F}_\delta(\Pi) = \emptyset$. According to Item (2) of Theorem A.2, such a set is not $\mathcal{F}_\delta$-closed, regardless of the way the set is defined in the rest of Ω. Therefore it suffices to lower bound the probability that a random set will be of this form in any of the n balls of radius δ whose existence is guaranteed by the hypothesis.

For any fixed $i \in [n]$, the probability that a uniformly chosen Π satisfies $\Pi \cap B[x_i, \delta) = B[x_i, \delta) \setminus \{x_i\}$ is $2^{-|B[x_i, \delta)|}$. Since, by the hypothesis, it holds that $|B[x_i, \delta)| \leq m$, then this probability is lower bounded by 2^{-m}. Thus, the probability that $\Pi \cap B[x_i, \delta) \neq B[x_i, \delta) \setminus \{x_i\}$ is at most $1 - 2^{-m}$. Also note that by the hypothesis, for any $i \neq j \in [n]$ it holds that

$\Delta(x_i, x_j) \geq 2\delta$, and hence $B[x_i, \delta) \cap B[x_j, \delta)$ are disjoint, implying that the events $\Pi \cap B[x_i, \delta) \neq B[x_i, \delta) \setminus \{x_i\}$ for all $i \in [n]$ are independent. Therefore, the probability that for every $i \in [n]$ it holds that $\Pi \cap B[x_i, \delta) \neq B[x_i, \delta) \setminus \{x_i\}$ is upper bounded by $(1 - 2^{-m})^n$. It follows that probability that the set is $\mathcal{F}_\delta$-closed is at most $(1 - 2^{-m})^n$. ■

If the collection of balls in Proposition B.2 satisfies $n \geq 2^m$, then we get that the majority of sets in Ω are not $\mathcal{F}_\delta$-closed. However, the lower bound in Proposition B.2 is far from tight for some spaces. In particular, in the special case of the Boolean hypercube, Proposition B.3 presents a tighter lower bound, relying on a simple argument tailored to this specific case.

Proposition B.3 *(most sets in the Boolean hypercube are not $\mathcal{F}_\delta$-closed). For the n-dimensional Boolean hypercube H_n and $\delta \geq 3$, the probability that a uniformly chosen $\Pi \subseteq H_n$ is $\mathcal{F}_\delta$-closed is at most $2^{-\Omega(n^2)}$.*

Proof. First observe that any Π that satisfies $\Pi \neq H_n$ and $\mathcal{F}_\delta(\Pi) = \emptyset$ is not $\mathcal{F}_\delta$-closed. We show that a uniformly chosen random Π satisfies both conditions with very high probability.

For any $z \in H_n$ it holds that $z \in \mathcal{F}_\delta(\Pi)$ if and only if $B[z, \delta - 1] \cap \Pi = \emptyset$. For a fixed $z \in H_n$ this happens with probability $2^{-|B[z,\delta-1]|}$, and since since $\delta \geq 3$ this expression is upper bounded by $2^{-\left(1+n+\binom{n}{2}\right)} = 2^{-\Omega(n^2)}$. By union-bounding over all $z \in H_n$, the probability that there exists some $z \in \mathcal{F}_\delta(\Pi)$ is at most $2^{n-\Omega(n^2)}$. Also, the probability that $\Pi = H_n$ is 2^{-2^n}. Thus the probability that a random set is $\mathcal{F}_\delta$-closed is at most

$$2^{n-\Omega(n^2)} + 2^{-2^n} = 2^{-\Omega(n^2)} . \quad ■$$

Appendix C On the distance of points in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ from Π

One might mistakenly think that even in cases where $\Pi \neq \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ (i.e., Π is not $\mathcal{F}_\delta$-closed), all points in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ are, in some sense, close to Π. Indeed, since for any $\delta > 0$ it holds that $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) \cap \mathcal{F}_\delta(\Pi) = \emptyset$, the points in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ cannot be δ-far from Π. In the current appendix we ask *how far from Π can points in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ be*, in general metric spaces as well as in graphical metric spaces. In particular, we show that points in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ can be almost δ-far from Π; an asymptotic version of the foregoing statement (see next) implies, in particular, that not all dual problems reduce to tolerant testing.

In the context of property testing, a typical setting involves a sequence $\Pi = \{\Pi_n\}_{n\in\mathbb{N}}$ such that $\Pi_n \subseteq \{0,1\}^n$, and the distance of interest is $\delta = \epsilon \cdot n$, for a small constant $\epsilon > 0$. Recall that, as mentioned in Section 1.4, the dual testing problem of $\Pi = \{\Pi_n\}_{n\in\mathbb{N}}$ reduces to the problem of tolerant testing if for every *sufficiently small* $\epsilon > 0$ it holds that the distance of points in $\mathcal{F}_{\epsilon\cdot n}(\mathcal{F}_{\epsilon\cdot n}(\Pi_n))$ is bounded away from $\epsilon \cdot n$. We now show a property Π for which this does not happen. That is, we show a *single fixed set* $\Pi = \{\Pi_n\}_{n\in\mathbb{N}}$ such that for *every* small ϵ it holds that points in $\mathcal{F}_{\epsilon\cdot n}(\mathcal{F}_{\epsilon\cdot n}(\Pi_n))$ are not close to Π_n (actually, we show that the latter holds for an infinite sequence of ϵ's that tends to zero).

Proposition C.1 *(a dual problem that does not reduce to tolerant testing). There exists $\Pi = \{\Pi_n\}_{n\in\mathbb{N}}$, where $\Pi_n \subseteq \{0,1\}^n$, that satisfies the following. For every $\epsilon = 2^{-k}$ such that $k \geq 2$, and sufficiently large n, there exists $x \in \mathcal{F}_{\epsilon\cdot n}(\mathcal{F}_{\epsilon\cdot n}(\Pi_n))$ such that $\Delta(x, \Pi_n) \geq \epsilon \cdot n - 2$.*

Proof. Let us start by describing the intuition behind our construction. Fix any $\epsilon > 0$. The basic observation is that if Π_n contains all the strings of Hamming weight ℓ, for some $\ell \in [n]$, and all the strings of Hamming weight roughly $h = \ell + 2 \cdot (\epsilon \cdot n - 1)$, and Π_n does *not* contain *any* string of Hamming weight in between these two values, then $\mathcal{F}_{\epsilon \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n))$ contains a string whose distance from Π_n is at least $\epsilon \cdot n - 2$ (see Lemma C.1.2). Intuitively, this is the case since a string x of Hamming weight $\frac{h+\ell}{2}$ is, on the one hand, far from being far from Π_n (since all strings of Hamming weight in between ℓ and h are not $(\epsilon \cdot n)$-far from Π_n, which implies that any shortest path from x to $\mathcal{F}_{\epsilon \cdot n}(\Pi_n)$ passes through Π_n); but x is, on the other hand, not close to Π_n (since $\Delta(x, \Pi_n) \approx \epsilon \cdot n - 1$). See Figure 3 for a graphical illustration.

For any $\epsilon > 0$, if Π_n is as above (i.e., Π_n contains all strings with Hamming weight ℓ or h and no string with Hamming weight in between these values), then we say that Π_n *exhibits a disruptive pattern for* ϵ. Thus, it suffices to construct $\Pi = \{\Pi_n\}$ such that for every $\epsilon = 2^{-k}$ and sufficiently large n it holds that Π_n exhibits a disruptive pattern for ϵ.

Accordingly, for any n, we construct Π_n such that it exhibits many disruptive patterns at once, for many values of ϵ. Fixing any n, we first define roughly $\log(n)$ pairwise-disjoint subsets of $\{0,1\}^n$, which we call *regions*. Specifically, for any $1 < k < \log(n)$, the region that corresponds to k includes all strings with relative Hamming weight more than $1 - 2^{-(k-2)}$ and at most $1 - 2^{-(k-1)}$. Since the region that corresponds to k contains strings with about $2^{-(k-1)}$ different weights, we can define Π_n in this region such that it exhibits a disruptive pattern for $\epsilon = 2^{-k}$. Hence, for any fixed $\epsilon = 2^{-k}$ and $n \geq \exp(k)$ it holds that Π_n contains a disruptive pattern for ϵ. Details follow.

The actual proof. We start by defining a disruptive pattern for $\epsilon > 0$, and proving that it indeed yields a string in $\mathcal{F}_{\epsilon \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n))$ whose distance from Π_n is $\epsilon \cdot n - O(1)$.

Definition C.1.1 *(disruptive pattern for ϵ). Let $\epsilon > 0$ and $n \in \mathbb{N}$ such that $\epsilon \cdot n \geq 2$. We say that $\Pi_n \subseteq \{0,1\}^n$* exhibits a disruptive pattern for ϵ *if there exist $\ell, h \in [n]$ such that $h - \ell = 2 \cdot (\lfloor \epsilon \cdot n \rfloor - 1)$, and the following hold:*

1. *Π_n contains all strings with Hamming weight ℓ, and all strings with Hamming weight h.*

2. *Π_n does not contain any string with Hamming weight w such that $\ell < w < h$.*

Lemma C.1.2. *Let $\epsilon > 0$ and $n \in \mathbb{N}$ such that $\epsilon \cdot n \geq 2$. If $\Pi_n \subseteq \{0,1\}^n$ exhibits a disruptive pattern for ϵ, then $\mathcal{F}_{\epsilon \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n))$ contains a string x such that $\Delta(x, \Pi_n) \geq \epsilon \cdot n - 2$.*

Proof. Let $\ell, h \in [n]$ as in Definition C.1.1, and denote by B the set of strings with Hamming weight w such that $\ell < w < h$. We will show that any string $x \in B$ with Hamming weight $\frac{h+\ell}{2}$ is, on one hand, in $\mathcal{F}_{\epsilon \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n))$, and is, on the other hand, $(\epsilon \cdot n - 2)$-far from Π_n.

First note that there *exists* a string with Hamming weight $\frac{\ell+h}{2}$. This is the case because $h - \ell = 2 \cdot (\lfloor \epsilon \cdot n \rfloor - 1) \geq 2$ (by our hypothesis that $\epsilon \cdot n \geq 2$). Let x be such a string, and observe that $x \in B$. The lemma follows from the following two facts:

1. **The distance of x from Π_n is at least $\epsilon \cdot n - 2$.** This is the case because $B \cap \Pi_n = \emptyset$ (by our hypothesis), and thus any path from x to Π_n has to pass through some $u \notin B$ such

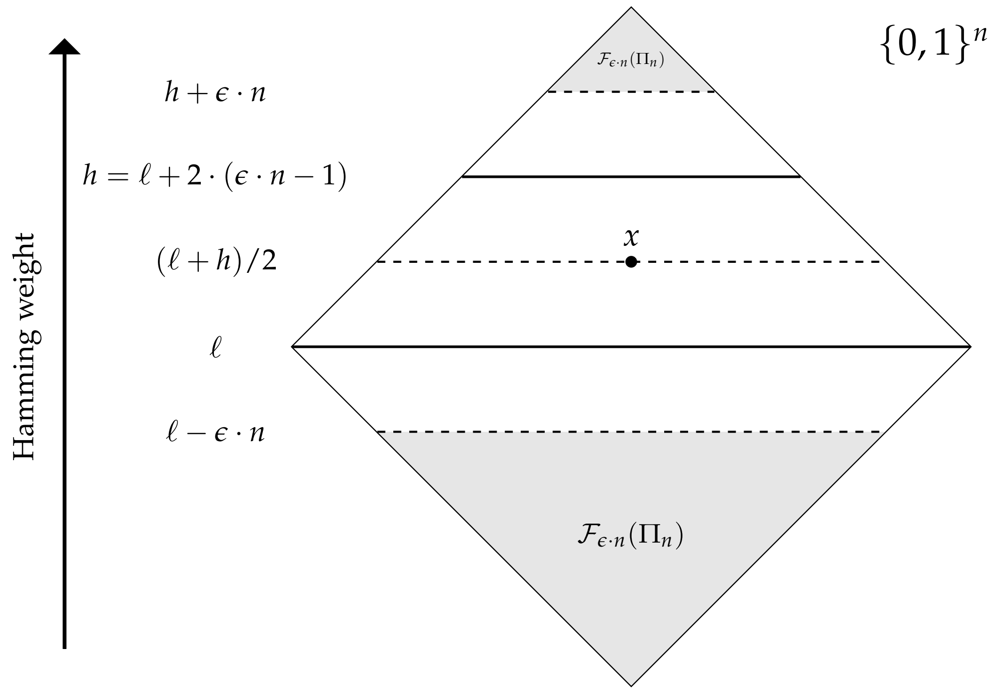

Figure 3: Graphical illustration of a disruptive pattern for $\epsilon = 1/8$. The rhombus represents $\{0,1\}^n$ such that lower points inside the rhombus are strings with low Hamming weight, and higher points are strings with high Hamming weight. The set Π_n consists of all strings with Hamming weight $\ell = n/2$ or $h = 3/4n - 2$, and thus $\mathcal{F}_{\epsilon\cdot n}(\Pi_n)$ is contained in the gray areas. The string x satisfies $x \in \mathcal{F}_{\epsilon\cdot n}(\mathcal{F}_{\epsilon\cdot n}(\Pi_n))$ as well as $\Delta(x, \Pi_n) = \frac{h-\ell}{2} = \epsilon \cdot n - 1$.

that u has a neighbor in B. However, any such u has Hamming weight either ℓ or h, and thus $u \in \Pi_n$. It follows that $\Delta(x, \Pi_n) = \frac{h-\ell}{2} = \lfloor \epsilon \cdot n \rfloor - 1 \geq \epsilon \cdot n - 2$.

2. **The distance of x from $\mathcal{F}_{\epsilon\cdot n}(\Pi_n)$ is more than $\epsilon \cdot n$.** To see that this holds, first note that $B \cap \mathcal{F}_{\epsilon\cdot n}(\Pi_n) = \emptyset$ (because for every $y \in B$, we can flip at most $\frac{h-\ell}{2} < \epsilon \cdot n$ bits in y, to obtain a string with Hamming weight ℓ or h). Thus, every path from x to $\mathcal{F}_{\epsilon\cdot n}(\Pi_n)$ has to pass through some $u \notin B$ such that u has a neighbor in B; but any such u has Hamming weight ℓ or h, and thus $u \in \Pi_n$. The length of any such path is $\Delta(x, u) + \Delta(u, \mathcal{F}_{\epsilon\cdot n}(\Pi_n)) \geq 1 + \epsilon \cdot n$, and hence $\Delta(x, \mathcal{F}_{\epsilon\cdot n}(\Pi_n)) \geq 1 + \epsilon \cdot n$. □

For every $n \in \mathbb{N}$, in order to define $\Pi_n \subseteq \{0,1\}^n$, we first define the pairwise-disjoint *regions* in $\{0,1\}^n$. For each $k \in \{2, ..., \lfloor \log(n) \rfloor - 1\}$ we will have a corresponding region in $\{0,1\}^n$, which is denoted by $\mathcal{R}_k^n$ and defined as follows: The region $\mathcal{R}_k^n$ is the set of all strings with relative Hamming weight more than $1 - 2^{-(k-2)}$ and at most $1 - 2^{-(k-1)}$; that is,

$$\mathcal{R}_k^n = \left\{ x \in \{0,1\}^n : 1 - 2^{-(k-2)} < \frac{\|x\|_1}{n} \leq 1 - 2^{-(k-1)} \right\} ,$$

where $\|x\|_1$ is the Hamming weight of x. First, observe that for any n, the regions in $\{0,1\}^n$ are indeed pairwise-disjoint.

Fact C.1.3 *(the regions are pairwise-disjoint). For every $n \in \mathbb{N}$, the regions in $\{0,1\}^n$, corresponding to different values of $k \in \{2, ..., \lfloor \log(n) \rfloor - 1\}$, are pairwise-disjoint.*

Proof. For any $k, k' \in \{2, ..., \lfloor \log(n) \rfloor - 1\}$ such that $k' > k$, the maximal Hamming weight of a string in $\mathcal{R}_k^n$ is $\left\lfloor \left(1 - 2^{-(k-1)}\right) \cdot n \right\rfloor$, whereas the minimal Hamming weight of a string in $\mathcal{R}_{k'}^n$ is $\left\lfloor \left(1 - 2^{-(k'-2)}\right) \cdot n \right\rfloor + 1 \geq \left\lfloor \left(1 - 2^{-(k-1)}\right) \cdot n \right\rfloor + 1$. □

For any $n \in \mathbb{N}$, we define Π_n as follows. First, Π_n includes all strings that are outside the regions in $\{0,1\}^n$ (i.e., the all-zero string is included in Π_n, and so are all strings with Hamming weight more than $\left(1 - 2^{-(\lfloor \log(n) \rfloor - 2)}\right) \cdot n$).[12] Now, relying on Fact C.1.3, we can define Π_n independently in each of the pairwise-disjoint regions. For $k \in \{2, ..., \lfloor \log(n) \rfloor - 1\}$, we will define Π_n in $\mathcal{R}_k^n$ such that it exhibits a disruptive pattern for $\epsilon = 2^{-k}$. Specifically, we define Π_n such that it contains all the strings with *minimal* Hamming weight in $\mathcal{R}_k^n$, where this minimal weight is denoted by ℓ_k^n, and Π_n contains all strings in $\mathcal{R}_k^n$ with Hamming weight at least $\ell_k^n + 2 \cdot (\lfloor 2^{-k} \cdot n \rfloor - 1)$, and Π_n does not contain any string with Hamming weight in between these two values. Let us now verify that Π_n indeed exhibits a disruptive pattern for $\epsilon = 2^{-k}$ in $\mathcal{R}_k^n$.

Claim C.1.4 *(Π_n exhibits a disruptive patterns). For any $2 \leq k \leq \lfloor \log(n) \rfloor - 1$ it holds that Π_n exhibits a disruptive pattern for $\epsilon = 2^{-k}$.*

Proof. Let k be as in the hypothesis, and let $\epsilon_k = 2^{-k}$. First observe that $\epsilon_k \cdot n \geq 2$ (because $k \leq \lfloor \log(n) \rfloor - 1$). Next, denote by ℓ_k^n the minimal Hamming weight of a string in $\mathcal{R}_k^n$, and by h_k^n the maximal Hamming weight of a string in $\mathcal{R}_k^n$; then, it holds that $h_k^n - \ell_k^n \geq 2 \cdot (\epsilon_k \cdot n - 1)$, because:

$$
\begin{aligned}
h_k^n - \ell_k^n &= \left\lfloor \left(1 - 2^{-(k-1)}\right) \cdot n \right\rfloor - \left(\left\lfloor \left(1 - 2^{-(k-2)}\right) \cdot n \right\rfloor + 1 \right) \\
&\geq \left(1 - 2^{-(k-1)}\right) \cdot n - 1 - \left(\left(1 - 2^{-(k-2)}\right) \cdot n + 1 \right) \\
&= 2^{-(k-1)} \cdot n - 2 \\
&= 2 \cdot (\epsilon_k \cdot n - 1) \ .
\end{aligned}
$$

In particular, for $\ell = \ell_k^n$ there exists $h \leq h_k^n$ such that $h - \ell = 2 \cdot (\lfloor \epsilon_k \cdot n \rfloor - 1)$, and $\mathcal{R}_k^n$ contains all strings with Hamming weight $w \in [\ell, h]$. By the definition of Π_n, it holds that Π_n contains all strings with Hamming weight ℓ, and all strings with Hamming weight h, and no string with Hamming weight in between these two values. According to Definition C.1.1, it holds that Π_n contains a disruptive pattern for $\epsilon_k = 2^{-k}$. □

Let $\Pi = \cup_{n \in \mathbb{N}} \Pi_n$. For any fixed $\epsilon = 2^{-k}$, where $k \geq 2$, and any $n \geq 2^{k+1}$, by combining Claim C.1.4 and Lemma C.1.2, we deduce that there exists a string in $\mathcal{F}_{\epsilon \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n))$ whose distance from Π_n is at least $\epsilon \cdot n - 2$. Proposition C.1 follows. ■

[12] The choice to include these strings in Π_n is arbitrary, and does not affect the proof.

Remark C.2. The property constructed in the proof of Proposition C.1 can be presented as a property of graphs in the dense graph model (see Section 1.2.4 for a definition of the latter). For any $n \in \mathbb{N}$, the set Π_n constructed in the proof is of the form "all strings with Hamming weight in W_n", where W_n is a set of weights. Relying on a natural bijection between strings and graphs, Π_n corresponds to the set of all graphs over v vertices (such that $n = \binom{v}{2}$) with number of edges in W_n. The point is that Π_n is closed under taking graph isomorphisms, and is thus a graph property.

Another mistaken intuition is that even when $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ contains points that are far from Π, not all points in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ are so (i.e., $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ also contains points that are closer to Π). The following proposition demonstrates that this is not the case: There exist spaces and sets in which *all* points in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ are either in Π or almost δ-far from Π.

Proposition C.3 *(all points in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) \setminus \Pi$ might be almost δ-far from Π). For every odd integer $\delta \geq 3$, there exist Ω and $\Pi \subseteq \Omega$ such that Π is not $\mathcal{F}_\delta$-closed, and every $x \in \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) \setminus \Pi$ satisfies $\Delta(x, \Pi) = \delta - 1$.*

Proof. For an odd integer $\delta \geq 3$, let Ω be a graph that is a simple path of length $\delta - 1$. We call this path the *base path*, and denote its vertices by $v_0, v_1, ..., v_{\delta-1}$. Now add to Ω another simple path, this time of length $(\delta-1)/2 + 1$, starting from $v_{(\delta-1)/2}$. We call this path the *additional path*, and denote its vertices by $v_{(\delta-1)/2} = z_0, z_1, ..., z_{(\delta-1)/2+1}$. The only vertex belonging to both the base path and the additional path is $v_{(\delta-1)/2} = z_0$, and the two paths are edge-disjoint.

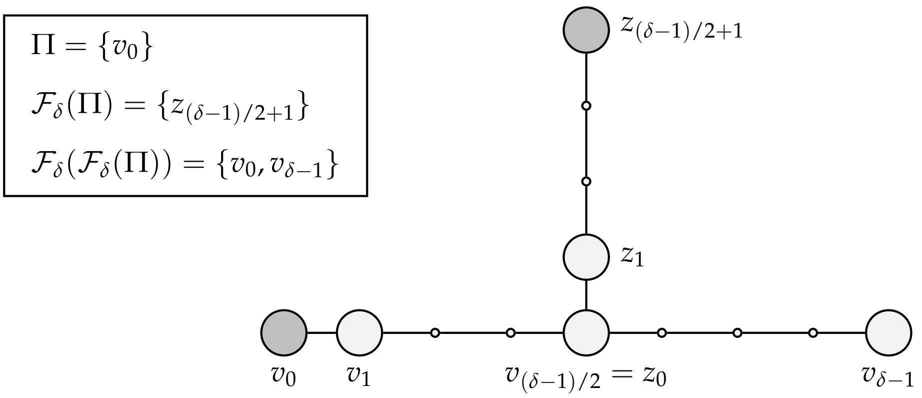

Figure 4: The space Ω.

Let $\Pi = \{v_0\}$. For every vertex v_i on the base path, it holds that $\Delta(v_i, \Pi) = i < \delta$. Also, for every vertex z_i on the additional path it holds that $\Delta(z_i, \Pi) = \Delta(z_i, z_0) + \Delta(z_0, \Pi) = i + (\delta-1)/2$. Thus, the only vertex that is δ-far from Π is $z_{(\delta-1)/2+1}$, implying that $\mathcal{F}_\delta(\Pi) = \{z_{(\delta-1)/2+1}\}$.

Now, note that for every vertex z_i on the additional path it holds that $\Delta(z_i, \mathcal{F}_\delta(\Pi)) =$

$(\delta-1)/2+1-i<\delta$. Also, for every vertex v_i on the original path it holds that

$$\Delta(v_i,\mathcal{F}_\delta(\Pi)) = \Delta(v_i,v_{(\delta-1)/2}) + \Delta(z_0,z_{(\delta-1)/2+1}) = \left|i-\frac{\delta-1}{2}\right| + \left(\frac{\delta-1}{2}+1\right)$$

and thus $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) = \{v_0, v_{\delta-1}\}$. Therefore, only $v_{\delta-1}$ satisfies $v_{\delta-1} \in \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) \setminus \Pi$, and it holds that $\Delta(v_{\delta-1},\Pi) = \delta-1$. ■

We now turn our attention to *graphical metric spaces*, and show a sufficient condition for deducing that the distance of vertices in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ is bounded away from δ.

Proposition C.4. *Let Ω be a graphical space, let $\Pi \subseteq \Omega$, let $\delta \geq 2$, and let $\delta' \leq \delta$. If there exists an integer m such that for every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ there exists x' satisfying $\Delta(x,x') \leq m$ and $\Delta(x',\Pi) > \Delta(x,\Pi)$, then $\mathcal{F}_{\delta'}(\mathcal{F}_\delta(\Pi)) \subseteq \left\{y : \Delta(y,\Pi) \leq \delta - \frac{\delta'}{m}\right\}$.*

Proof. We will show that if $\Delta(y,\Pi) > \delta - \frac{\delta'}{m}$, then $y \notin \mathcal{F}_{\delta'}(\mathcal{F}_\delta(\Pi))$. This will be done by showing a path $y_0,...,y_t$ of length less than δ' such that $y = y_0$ and $y_t \in \mathcal{F}_\delta(\Pi)$. Specifically, we start with $y = y_0$, and proceed in iterations, where in iteration i we begin at vertex y_i and rely on the hypothesis to obtain y_{i+1} that is farther away from Π, compared to y_i, such that the distance between y_i and y_{i+1} is at most m. After at most $t = \delta - \Delta(y,\Pi)$ iterations, we obtain y_t such that $\Delta(y_t,\Pi) \geq t + \Delta(y_0,\Pi) = \delta$ and $\Delta(y,y_t) \leq m \cdot t = m \cdot (\delta - \Delta(y,\Pi)) < \delta'$ (where the last inequality holds because $\Delta(y,\Pi) > \delta - \frac{\delta'}{m}$). ■

A special case of Proposition C.4 is when a set is strongly $\mathcal{F}_\delta$-closed. Specifically, for a graphical space Ω and $\delta \geq 2$, if $\Pi \subseteq \Omega$ is strongly $\mathcal{F}_\delta$-closed, then for every $\delta' \leq \delta$ the condition in Proposition C.4 is met with $m = 1$, which implies that $\mathcal{F}_{\delta'}(\mathcal{F}_\delta(\Pi)) \subseteq \{y : \Delta(y,\Pi) \leq \delta - \delta'\}$.

Appendix D The mapping $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ is a closure operator in $\mathcal{P}(\Omega)$

The notion of closure operators (or *hull operators*; see, e.g., [KD06, Chp. 2] or [vdV93, Chp. 1]) is prevalent in many mathematical fields, including algebra, topology, matroid theory, and computational geometry. We show that the operator $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ is a closure operator on Ω, a statement that gives some structure to the relationship between Π and $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$.

Definition D.1 *(closure operators). A* closure operator *on a set Ω is an operator $cl : \mathcal{P}(\Omega) \to \mathcal{P}(\Omega)$ such that for any $\Pi, \Pi' \subseteq \Omega$ it holds that*

1. *(extensive) $\Pi \subseteq cl(\Pi)$.*
2. *(upwards monotone) $\Pi \subseteq \Pi' \implies cl(\Pi) \subseteq cl(\Pi')$.*
3. *(idempotent) $cl(cl(\Pi)) = cl(\Pi)$.*

Proposition D.2 *($\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ is a closure operator). For any Ω and $\delta > 0$ it holds that $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ is a closure operator on Ω.*

Proof. Axiom (1) follows from Fact A.1. Axiom (2) follows by applying Fact A.2.2 twice to the expression $\Pi \subseteq \Pi'$. Axiom (3) is essentially the requirement that for any set Π it holds that $\mathcal{F}_\delta^{(4)}(\Pi) = \mathcal{F}_\delta^{(2)}(\Pi)$ (i.e., four applications of $\mathcal{F}_\delta$ on Π are equivalent to two applications); or, equivalently, that any set of the form $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ is $\mathcal{F}_\delta$-closed. The latter statement follows from Condition (3) in Theorem A.2. ∎

A closure operator is characterized by the collection of closed sets $\{cl(\Pi)\}_{\Pi\subseteq\Omega}$. In particular, the collection of closed sets under the operator $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ is $\{\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))\}_{\Pi\subseteq\Omega}$, which according to Theorem A.2 is exactly the collection of $\mathcal{F}_\delta$-closed sets. In general, any closure operator maps any set Π to its closure, which is the unique smallest closed set containing Π. The following proposition substantiates that this is indeed the case in the special case of the operator $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$: The proposition states that $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ is the intersection of all $\mathcal{F}_\delta$-closed sets containing Π. Since $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ is itself an $\mathcal{F}_\delta$-closed set, this implies that $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ it the unique $\mathcal{F}_\delta$-closed set that contains Π, and that this set is minimal (i.e., does not contain any other $\mathcal{F}_\delta$-closed set containing Π).

Proposition D.3 *($\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ is the unique minimal $\mathcal{F}_\delta$-closed set containing Π). For any Ω, $\delta > 0$ and $\Pi \subseteq \Omega$ it holds that*

$$\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) = \bigcap_{\Pi':\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi'))\supseteq\Pi} \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi'))$$

Proof. We follow the standard proof that for any closure operator cl it holds that $cl(\Pi) = \bigcap_{\Pi':cl(\Pi')\supseteq\Pi} cl(\Pi')$. This standard proof relies on the fact that for general closure operators, the intersection of closed sets is closed; in the specific case of $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$, this fact follows immediately from Condition (5) in Theorem A.2, and was mentioned in the discussion after the proof of Theorem A.2.

Let $\mathcal{I} = \{\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi')) : \Pi' \subseteq \Omega \wedge \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi')) \supseteq \Pi\}$. We seek to prove that

$$\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) = \bigcap_{\Phi\in\mathcal{I}} \Phi$$

To see that $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) \supseteq \bigcap_{\Phi\in\mathcal{I}} \Phi$, note that by Condition (1) of Definition D.1 it holds that $\Pi \subseteq \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$, and thus $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) \in \mathcal{I}$. For the other direction, to see that $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) \subseteq \bigcap_{\Phi\in\mathcal{I}} \Phi$, note that any $\Phi \in \mathcal{I}$ satisfies $\Pi \subseteq \Phi$; and thus

$$\Pi \subseteq \bigcap_{\Phi\in\mathcal{I}} \Phi \tag{A.1}$$

Relying on Condition (2) of Definition D.1 and on Eq. (A.1), we get that

$$\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) \subseteq \mathcal{F}_\delta\left(\mathcal{F}_\delta\left(\bigcap_{\Phi\in\mathcal{I}} \Phi\right)\right) \tag{A.2}$$

Since every $\Phi \in \mathcal{I}$ is of the form $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi'))$ for some $\Pi' \subseteq \Omega$, it holds that every $\Phi \in \mathcal{I}$ is $\mathcal{F}_\delta$-closed. Relying on the fact that the intersection of $\mathcal{F}_\delta$-closed sets is $\mathcal{F}_\delta$-closed, we get that $\bigcap_{\Phi\in\mathcal{I}} \Phi$ is $\mathcal{F}_\delta$-closed. It follows that $\mathcal{F}_\delta\left(\mathcal{F}_\delta\left(\bigcap_{\Phi\in\mathcal{I}} \Phi\right)\right) = \bigcap_{\Phi\in\mathcal{I}} \Phi$, and relying on Eq. (A.2), we get that

$$\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) \subseteq \bigcap_{\Phi\in\mathcal{I}} \Phi\,. \quad \blacksquare$$

For an intuitive grasp of closure operators one may think of the *convex hull* of a body in Euclidean geometry or of the *topological closure* of a set in a topological space. We warn, however, that in some fields additional conditions are added to the basic three in Definition D.1, resulting in special classes of closure operators; the operator $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ does not necessarily belong to these special classes of operators. In particular, as shown next, $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ is not the convex hull operator in Euclidean spaces, is not a topological (i.e., Kuratowski) closure operator, and does not satisfy the conditions of closure operators used in matroid theory.

The *convex hull* operator in Euclidean spaces maps any set to the unique minimal convex set containing it. The following claim states that in Euclidean spaces the operator $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ *is not* the convex hull operator.

Claim D.4 *($\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ is not the convex hull operator). There exists a set $\Pi \subseteq \mathbb{R}^n$ such that $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ is not the convex hull of Π.*

Proof. Let $\Pi = \{x, y\}$ such that $\Delta(x, y) > 2\delta$. Note that the convex hull of Π contains the entire line segment between x and y. However, there exists a point z on this line segment such that $\Delta(z, x) \geq \delta$ and $\Delta(z, y) \geq \delta$. Thus, $z \in \mathcal{F}_\delta(\Pi)$, which implies that $z \notin \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$. It follows that the line segment between x and y is not contained in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$, and thus $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ is not the convex hull of Π. $\blacksquare$

Closure operators in topology are called *Kuratowski closure operators*, and satisfy the three conditions in Definition D.1 as well as the following additional condition: For $\Pi, \Pi' \subseteq \Omega$ it holds that $cl(\Pi) \cup cl(\Pi') = cl(\Pi \cup \Pi')$. However, $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ does not satisfy this condition in general.

Claim D.5 *($\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ is not a Kuratowski closure operator). There exists a space Ω and $\delta > 0$ such that the operator $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ does not satisfy the Kuratowski axioms.*

Proof. Let Ω be a graph that is a simple path $x_1 - x_2 - x_3$, and let $\delta = 2$. Consider $\Pi = \{x_1\}$ and $\Pi' = \{x_3\}$. Then $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) = \Pi$ and $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi')) = \Pi'$; but $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi \cup \Pi')) = \Omega \neq \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) \cup \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi'))$. $\blacksquare$

Closure operators in *matroid theory* (see, e.g., [GM12]) satisfy the three conditions in Definition D.1 as well as an additional fourth condition. We now define this fourth condition, and show that the operator $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ does not satisfy it in general.

Definition D.6 *(MacLane-Steinitz exchange property). A closure operator $cl : \mathcal{P}(\Omega) \to \mathcal{P}(\Omega)$ satisfies the MacLane-Steinitz exchange property if it meets the following condition: If there exist $\Pi \subseteq \Omega$ and $x, y \in \Omega$ such that $x \in cl(\Pi \cup \{y\}) \setminus cl(\Pi)$, then $y \in cl(\Pi \cup \{x\})$.*

Claim D.7 *($\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ does not satisfy the MacLane-Steinitz exchange property). There exists a space Ω and $\delta > 0$ such that the operator $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ does not satisfy the MacLane-Steinitz exchange property.*

Proof. Let Ω be a graph that is a simple path $x - y - z$, and let $\delta = 2$ and $\Pi = \emptyset$. Note that $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) = \Pi = \emptyset$, and $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi \cup \{y\})) = \Omega \ni x$, which implies that $x \in \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi \cup \{y\})) \setminus \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$. However, it holds that $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi \cup \{x\})) = \{x\} \not\ni y$. ■

Finally, let us point to an interesting property that $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ *does* admit: Namely, $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ is the composition of another operator with itself; that is, $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ is the composed operator $\mathcal{F}_\delta \circ \mathcal{F}_\delta$. Moreover, the collection of closed sets under $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ is identical to the image of the composed operator (since by Theorem A.2, it holds that $\{\mathcal{F}_\delta(\Pi)\}_{\Pi \subseteq \Omega} = \{\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))\}_{\Pi \subseteq \Omega}$). This property seems distinct amongst the closure operators we are familiar with.

Appendix E $\mathcal{F}_\delta$-tight spaces

Recall that in general metric spaces and for a general $\delta > 0$, being strongly $\mathcal{F}_\delta$-closed is not a necessary condition for being $\mathcal{F}_\delta$-closed (as demonstrated in Figure 2 and Proposition A.7). However, there exist graphs and values of $\delta > 0$ for which this sufficient condition is also necessary. We call such spaces $\mathcal{F}_\delta$-*tight*; that is

Definition E.1 *($\mathcal{F}_\delta$-tight spaces). For a graphical space Ω and $\delta > 0$, we say that Ω is* $\mathcal{F}_\delta$-tight *if every $\mathcal{F}_\delta$-closed set in Ω is also strongly $\mathcal{F}_\delta$-closed.*

Thus, in $\mathcal{F}_\delta$-tight spaces, a set is $\mathcal{F}_\delta$-closed *if and only if* it is strongly $\mathcal{F}_\delta$-closed. In the current appendix we present an initial exploration of this notion. We first observe that *every* graph is $\mathcal{F}_1$-tight (since every set is both $\mathcal{F}_\delta$-closed and strongly $\mathcal{F}_\delta$-closed for $\delta = 1$, because the condition in Definition 2.4 holds vacuously). The following proposition asserts that every graph is also $\mathcal{F}_2$-tight.

Proposition E.2 *(all graphs are $\mathcal{F}_2$-tight). Every graphical space is $\mathcal{F}_2$-tight.*

Proof. Let $\Pi \subseteq \Omega$ be a set that is $\mathcal{F}_2$-closed. Relying on Definition 2.4, we show that every $x \notin \Pi \cup \mathcal{F}_2(\Pi)$ lies on a 2-path from Π to $\mathcal{F}_2(\Pi)$; that is, x has a neighbor in $\mathcal{F}_2(\Pi)$. Since Π is $\mathcal{F}_2$-closed, by Proposition A.3, every $x \notin \Pi \cup \mathcal{F}_2(\Pi)$ lies on a path to $\mathcal{F}_2(\Pi)$ such that every vertex v subsequent to x in the path satisfies $\Delta(v, \Pi) \geq 2$. Thus, the vertex subsequent to x on the path is a neighbor of x in $\mathcal{F}_2(\Pi)$. ■

However, not all graphical spaces are $\mathcal{F}_3$-tight, as demonstrated by the example in Figure 2. Nevertheless, every graphical space is $\mathcal{F}_\delta$-tight for values of δ that are larger than the diameter of the graph (since every $\Pi \subseteq \Omega$ satisfies $\mathcal{F}_\delta(\Pi) = \emptyset$ and $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) = \Omega$, and therefore the only $\mathcal{F}_\delta$-closed set is $\Pi = \Omega$, which is also strongly $\mathcal{F}_\delta$-closed). A natural consequent question is therefore:

For which graphs G and values of $\delta \in [3, \mathrm{diam}(G)]$ does it hold that G is $\mathcal{F}_\delta$-tight?

As an initial step towards tackling this question, we show several simple graph families that are $\mathcal{F}_\delta$-tight for *every* $\delta > 0$.

Proposition E.3 *(graphs that are $\mathcal{F}_\delta$-tight for every $\delta > 0$). The following graphs are $\mathcal{F}_\delta$-tight, for every $\delta > 0$:*

1. *A complete graph on $n \geq 2$ vertices.*
2. *A path on $n \geq 2$ vertices.*
3. *A cycle on $n \geq 2$ vertices.*
4. *A $2 \times n$ grid (i.e., a grid with two rows and n columns), for any $n \geq 2$.*
5. *A circular ladder graph on $2n \geq 4$ vertices; that is, the graph that is comprised of two cycles on n vertices such that for every $i \in [n]$, the i^{th} vertices in both cycles are connected by an edge.*

Item (1) of Proposition E.3 follows as a corollary of the fact that every graphical space is $\mathcal{F}_\delta$-tight for values of δ larger than the diameter of the graph. To prove the other items of Proposition E.3 we will first need the following corollary of Proposition A.3, of Fact A.3.1, and of Item (4) of Theorem A.2.

Corollary E.4 *(a corollary of Proposition A.3). For a graphical Ω, and $\delta \geq 2$, let $\Pi \subseteq \Omega$ be an* $\mathcal{F}_\delta$-closed set. *Then, for every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$, there exists a path $v_0, v_1, ..., v_m = x, ..., v_l$ such that:*

1. *$v_0 \in \Pi$, and for every $i \in [0, m-1]$ it holds that $\Delta(v_i, \mathcal{F}_\delta(\Pi)) \geq 2$.*
2. *$v_l \in \mathcal{F}_\delta(\Pi)$, and for every $i \in [m+1, l]$ it holds that $\Delta(v_i, \Pi) \geq 2$.*

We now prove Items (2) and (3) of Proposition E.3. An intuitive reason that a single proof suffices for both the path and the cycle is that being $\mathcal{F}_\delta$-closed (resp., strongly $\mathcal{F}_\delta$-closed) is a local phenomenon, and the local neighborhoods in both graphs are very similar.

Proposition E.5 *(Items (2) and (3) of Proposition E.3). Let G_n be either a simple path on $n \geq 2$ vertices or a cycle on $n \geq 2$ vertices. Then, for every $\delta > 0$ it holds that G_n is $\mathcal{F}_\delta$-tight.*

Proof. It suffices to prove that G_n is $\mathcal{F}_\delta$-tight for $\delta \geq 3$. Let $\delta \geq 3$, let $\Pi \subseteq G_n$ be an $\mathcal{F}_\delta$-closed set, and let $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$. According to Corollary E.4, there exists a path from x to Π that does not intersect $\mathcal{F}_\delta(\Pi)$, and a path from x to $\mathcal{F}_\delta(\Pi)$ that does not intersect Π. Without loss of generality, we can assume that both are simple paths. Now, note that a simple path from x to any set can only be one of two paths: The path obtained by walking from x constantly to one direction, and the path obtained by walking from x constantly to the other direction. Thus, in one of these paths, the first vertex from $\Pi \cup \mathcal{F}_\delta(\Pi)$ that we encounter is from Π, and in the other, the first vertex from $\Pi \cup \mathcal{F}_\delta(\Pi)$ that we encounter is from $\mathcal{F}_\delta(\Pi)$ (otherwise there would not exist two paths as in Corollary E.4).

Let x' be the neighbor of x to the side in which the first vertex from $\Pi \cup \mathcal{F}_\delta(\Pi)$ that we encounter is from $\mathcal{F}_\delta(\Pi)$. To see that $\Delta(x', \Pi) = \Delta(x, \Pi) + 1$, note that a shortest path from x' to Π can be one of two paths: The path obtained by walking constantly to the direction of x, and the path obtained by walking constantly to other direction. When walking constantly to the direction of x, the first vertex subsequent to x' on the path is x itself; such a path is necessarily longer than a shortest path from x to Π. Conversely, when going to the other direction, the first vertex from $\Pi \cup \mathcal{F}_\delta(\Pi)$ that we encounter is from $\mathcal{F}_\delta(\Pi)$; since the distance of such a vertex from Π is at least δ, such a path is of length at least $\delta \geq \Delta(x, \Pi) + 1$ (where the inequality is since $x \notin \mathcal{F}_\delta(\Pi)$). It follows that $\Delta(x', \Pi) = \Delta(x, \Pi) + 1$. ■

One can view a simple path on n vertices as a grid with one row and n columns; that is, view the n-path as the $1 \times n$ grid. A consequent natural question is the following:

Is the $n \times n$ grid $\mathcal{F}_\delta$-tight for every $\delta > 0$?

We present an initial step towards answering this question. In particular, the following proposition asserts that the graph with two rows and n columns (i.e., the $2 \times n$ grid) is also $\mathcal{F}_\delta$-tight for every $\delta > 0$. Similar to the proof of Proposition E.5, a nearly identical proof applies both to the $2 \times n$ grid and to the circular ladder graph on $2n$ vertices.

Proposition E.6 *(Items (4) and (5) of Proposition E.3). Let $G_{2,n}$ be either the $2 \times n$ grid or the circular ladder graph on $2n$ vertices. Then, for every $\delta > 0$ it holds that $G_{2,n}$ is $\mathcal{F}_\delta$-tight.*

Proof. We prove the claim for the case in which $G_{2,n}$ is the $2 \times n$ grid. The proof for the circular ladder graph is nearly identical, but slightly more cumbersome in terms of notation; we will explicitly note the single place in which there is a minor difference. For $i \in \{1, 2\}$, we denote the vertices in the i^{th} row of $G_{2,n}$ by $v_{i,1}, ..., v_{i,n}$. Also, we define the left and right directions in the graph in the natural way (i.e., within a fixed row $i \in \{1, 2\}$, the left direction is towards $v_{i,1}$, and the right direction is towards $v_{i,n}$).

Note that it suffices to prove that $G_{2,n}$ is $\mathcal{F}_\delta$-tight for $\delta \geq 3$. Let $\delta \geq 3$, let $\Pi \subseteq G_{2,n}$ be an $\mathcal{F}_\delta$-closed set, and let $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$. We will show a neighbor x' of x such that $\Delta(x', \Pi) = \Delta(x, \Pi) + 1$. Without loss of generality, assume that $x = v_{1,j}$, for $j \in [n]$.

High-level overview. The proof is based on a case analysis. In particular, it consists of three cases, depending on the neighborhood of x. The first case is when the vertex beneath x (i.e., the vertex $v_{2,j}$) is in $\mathcal{F}_\delta(\Pi)$. In this case, the vertex beneath x is a neighbor of x that is farther from Π (since $x \notin \mathcal{F}_\delta(\Pi)$). The second case is when the vertex beneath x is in Π. In this case, since Π is $\mathcal{F}_\delta$-closed, Proposition A.3 implies that there exists a path from x to $\mathcal{F}_\delta(\Pi)$ such that any vertex subsequent to x on the path is neither in Π nor adjacent to Π. The vertex immediately subsequent to x on the path is a neighbor of x that is farther from Π (since, in this case, x is adjacent to $v_{2,j} \in \Pi$).

The third and last case, in which the vertex beneath x is not in $\Pi \cup \mathcal{F}_\delta(\Pi)$, will be the main focus of our proof. In this case, we will rely on Corollary E.4 to show that when walking constantly from x to one horizontal direction (say, to the left), we reach a column in which there is a vertex from Π before reaching any column in which there is a vertex from $\mathcal{F}_\delta(\Pi)$; and when walking constantly from x to the other horizontal direction (say, to the right), we reach a column in which there is a vertex from $\mathcal{F}_\delta(\Pi)$ before reaching any column in which there is a vertex from Π. We prove that the neighbor of x to the right (i.e., to the direction in which we reach a column with a vertex from $\mathcal{F}_\delta(\Pi)$) is farther from Π, compared to x. The proof of the latter fact will rely on a more fine-grained case analysis as well as on Condition (2) of Theorem A.2.

The actual proof. The overview showed how to handle the cases in which $v_{2,j} \in \Pi$ or $v_{2,j} \in \mathcal{F}_\delta(\Pi)$. Thus, we focus on proving the case in which

$$v_{2,j} \notin \Pi \cup \mathcal{F}_\delta(\Pi) \ . \tag{E.1}$$

We start by limiting our analysis to a local neighborhood in the graph $G_{2,n}$, and introducing some additional notation. These will rely on the following observation:

Claim E.6.1. *There exists a column to the left of column j with a vertex from $\Pi \cup \mathcal{F}_\delta(\Pi)$, and a column to the right of column j with a vertex from $\Pi \cup \mathcal{F}_\delta(\Pi)$. Moreover, the first such column that we encounter when walking from x to one direction (i.e., to the left or to the right) contains a vertex from Π, and the first such column that we encounter when walking from x to the other direction contains a vertex from $\mathcal{F}_\delta(\Pi)$.*

Proof. Since Π is $\mathcal{F}_\delta$-closed, and relying on Corollary E.4, there exists a path from x to Π (resp., to $\mathcal{F}_\delta(\Pi)$) such that any vertex subsequent to x on the path is neither in $\mathcal{F}_\delta(\Pi)$ (resp., in Π) nor adjacent to $\mathcal{F}_\delta(\Pi)$ (resp., to Π). Also note that column j does not contain a vertex from $\Pi \cup \mathcal{F}_\delta(\Pi)$ (since $x = v_{1,j} \notin \Pi \cup \mathcal{F}_\delta(\Pi)$, and relying on Eq. (E.1)). Thus, both paths that exist according to Corollary E.4 end in columns either to the right or to the left of column j.

Now, observe that a column in the graph cannot contain one vertex from Π and another vertex from $\mathcal{F}_\delta(\Pi)$ (since $\delta \geq 3$, and the vertices in the column are adjacent). Also note that if a column contains a vertex from a set Π', then any path going through the column intersects Π' or a vertex adjacent to Π'. Therefore, the path from x to Π cannot intersect a column in which there is a vertex from $\mathcal{F}_\delta(\Pi)$, and the path from x to $\mathcal{F}_\delta(\Pi)$ cannot intersect a column in which there is a vertex from Π. The claim follows. $\square$

Denote by $j_R \in [n]$ the first column to the right of column j such that one of the vertices in the column is in $\Pi \cup \mathcal{F}_\delta(\Pi)$; that is, $j_R = \min\{j' > j : \exists i \in \{1,2\}, v_{i,j'} \in \Pi \cup \mathcal{F}_\delta(\Pi)\}$. Similarly, denote $j_L = \max\{j' < j : \exists i \in \{1,2\}, v_{i,j'} \in \Pi \cup \mathcal{F}_\delta(\Pi)\}$. Also, denote by i_R the row of the vertex in column j_R that is in $\Pi \cup \mathcal{F}_\delta(\Pi)$ (or $i_R = 1$, if both vertices in column j_R are in $\Pi \cup \mathcal{F}_\delta(\Pi)$); that is, $i_R = \min\{i \in \{1,2\} : v_{i,j_R} \in \Pi \cup \mathcal{F}_\delta(\Pi)\}$. Denote i_L in an analogous way. Without loss of generality, assume that $v_{i_L,j_L} \in \Pi$ and that $v_{i_R,j_R} \in \mathcal{F}_\delta(\Pi)$. The rest of the proof will focus only on columns $j_L, ..., j_R$ in the graph.[13]

[13] In the case of the circular ladder graph, the argument is slightly different in terms of notation. Assume that the vertices of the graph are organized in two rows of n vertices, similar to the grid, such that the left-most

Now, let $x' = v_{1,j+1}$ be the vertex to the right of x (indeed, it is possible that $x' = v_{1,j_R}$, in case $j_R = j+1$). We will prove that $\Delta(x', \Pi) = \Delta(x, \Pi) + 1$. Figure 5 depicts the relevant part of the graph, reflecting some of our assumptions and notations at this point.

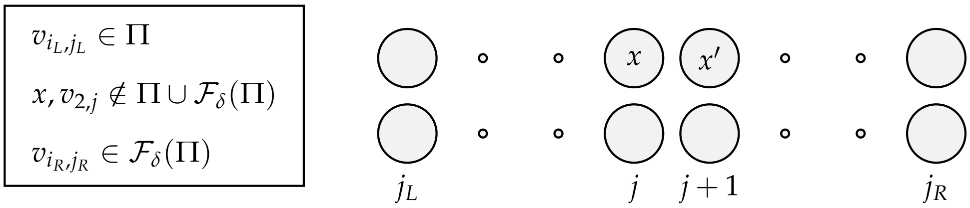

Figure 5: The relevant part of the graph $G_{2,n}$, reflecting our assumptions and notations at this point (as well as an additional, unjustified assumption that $j_R \neq j+1$). Note that columns $j_L+1, ..., j_R-1$ do not contain vertices from $\Pi \cup \mathcal{F}_\delta(\Pi)$.

Before proceeding, let us define one more term. For any two vertices $v_{i',j'}$ and $v_{i'',j''}$ in the graph, a path from $v_{i',j'}$ to $v_{i'',j''}$ is called a *straight simple path* if it is comprised of a shortest path from $v_{i',j'}$ to $v_{i',j''}$, and then (if $i' \neq i''$) a step from $v_{i',j''}$ to $v_{i'',j''}$. That is, we first walk "within the row", and then, if needed, conclude with a step to the other row. We will frequently rely on the following simple observation: If there exists a path of length k between two vertices in the graph, then there exists a straight simple path of length k between the vertices. Thus, for any vertex $v_{i',j'}$ and set $\Pi' \subseteq G_{2,n}$, *to prove that $\Delta(v_{i',j'}, \Pi') \geq k$, it suffices to prove that any straight simple path from $v_{i',j'}$ to Π' is of length at least k.*

To prove that $\Delta(x', \Pi) = \Delta(x, \Pi) + 1$, we show that any straight simple path from x' to Π is of length at least $\Delta(x, \Pi) + 1$. Note that, since $v_{2,j+1} \notin \Pi$, such a path starts by walking from x' either to the left or to the right (where $v_{2,j+1} \notin \Pi$ is since the first column to the right of column j with a vertex from $\Pi \cup \mathcal{F}_\delta(\Pi)$ contains a vertex from $\mathcal{F}_\delta(\Pi)$, so it cannot contain a vertex from Π).

Any straight simple path from x' to Π that starts by walking to the left passes through x, and is therefore longer than a shortest path from x to Π. Hence, to prove that $\Delta(x', \Pi) = \Delta(x, \Pi) + 1$, it suffices to show that any straight simple path from x' to Π that starts by walking to the right is of length at least $\Delta(x, \Pi) + 1$. Note that such a path passes through v_{1,j_R}, since there are no vertices from $\Pi \cup \mathcal{F}_\delta(\Pi)$ in columns $j, ..., j_R - 1$. Thus, the length of such a path is at least

$$\Delta(x', v_{1,j_R}) + \Delta(v_{1,j_R}, \Pi) \,. \tag{E.2}$$

Since $x \notin \mathcal{F}_\delta(\Pi)$, it holds that $\Delta(x, \Pi) + 1 \leq \delta$. Thus, the value of the expression in Eq. (E.2) can be smaller than $\Delta(x, \Pi) + 1$ only if it is at most $\delta - 1$. However, note that $\Delta(v_{1,j_R}, \Pi) \geq \delta - 1$, since there is a vertex from $\mathcal{F}_\delta(\Pi)$ in column j_R. Thus, the value of

and right-most vertices in each row are adjacent. In this case, it is possible that $j \in \{1, n\}$, and thus it does not necessarily hold that $j_R > j$ and $j_L < j$. However, since the rest of the proof will depend only on columns $j_L, ..., j_R$ in the graph, we may assume without loss of generality that $j_L < j < j_R$. This is the only place in which the proofs for the grid and for the circular ladder graphs differ.

the expression in Eq. (E.2) is smaller than $\Delta(x,\Pi)+1$ only if the following conditions hold: $\Delta(x,\Pi)=\delta-1$, and $x'=v_{1,j_R}$ (i.e., $j_R=j+1$), and $\Delta(x',\Pi)=\delta-1$. We prove that this case, in fact, does not happen. More specifically, we prove that if $\Delta(x,\Pi)=\delta-1$, and $j_R=j+1$, and $\Delta(x',\Pi)=\delta-1$, then Π is not $\mathcal{F}_\delta$-closed, which is a contradiction.

Claim E.6.2. *Assuming that $v_{2,j}\notin\Pi\cup\mathcal{F}_\delta(\Pi)$, and $v_{i_L,j_L}\in\Pi$, and $\Delta(x,\Pi)=\delta-1$, and $j_R=j+1$, and $\Delta(x',\Pi)=\delta-1$, it follows that Π is not $\mathcal{F}_\delta$-closed.*

Assume, for a moment, that Claim E.6.2 holds. Then, the expression in Eq. (E.2) is lower bounded by $\Delta(x,\Pi)+1$, which implies that any straight simple path from x' to Π that starts by walking to the right is of length at least $\Delta(x,\Pi)+1$. It follows that $\Delta(x',\Pi)=\Delta(x,\Pi)+1$, which finishes the current and last case (in which $v_{2,j}\notin\Pi\cup\mathcal{F}_\delta(\Pi)$), and concludes the proof. Thus, to conclude the proof it is just left to prove Claim E.6.2.

Proof of Claim E.6.2. First note that since column $j_R=j+1$ contains a vertex from $\mathcal{F}_\delta(\Pi)$, and $\Delta(x',\Pi)=\delta-1$, it follows that $v_{2,j+1}\in\mathcal{F}_\delta(\Pi)$. Figure 6 depicts columns $j_L,\ldots,j+1=j_R$ of the graph, reflecting our assumptions at this point.

$v_{i_L,j_L}\in\Pi$

$\Delta(x,\Pi)=\Delta(x',\Pi)=\delta-1$

$v_{2,j}\notin\Pi\cup\mathcal{F}_\delta(\Pi)$

$v_{2,j+1}\in\mathcal{F}_\delta(\Pi)$

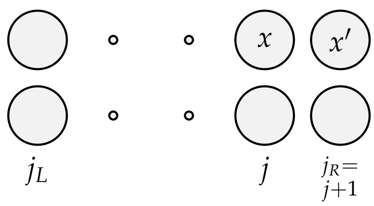

Figure 6: Columns $j_L,\ldots,j+1=j_R$ of the graph $G_{2,n}$, reflecting our assumptions at this point.

Fact E.6.2.1. *From the hypothesis of Claim E.6.2 it follows that $j-j_L=\delta-1$.*

Proof. To see that $j-j_L\geq\delta-1$, note that:

- If $v_{1,j_L}\in\Pi$, then, since $\Delta(x,\Pi)=\delta-1$, we get that $\delta-1=\Delta(x,\Pi)\leq\Delta(x,v_{1,j_L})=j-j_L$.
- If $v_{1,j_L}\notin\Pi$, then $v_{2,j_L}\in\Pi$ (since one of the vertices in column j_L is in Π). In this case, the distance of $v_{2,j_L}\in\Pi$ from $v_{2,j+1}\in\mathcal{F}_\delta(\Pi)$ is at least δ. Thus, $\delta\leq\Delta(v_{2,j_L},v_{2,j+1})=j+1-j_L$, which implies that $j-j_L\geq\delta-1$.

To see that $j-j_L\leq\delta-1$, assume otherwise, and note that it implies that $\Delta(x,\Pi)\geq\delta$, which contradicts $x\notin\mathcal{F}_\delta(\Pi)$. This is true since any straight simple path from x to Π that starts by walking to the right passes through x'; since $\Delta(x',\Pi)=\delta-1$, such a path is of length at least $\Delta(x,x')+\Delta(x',\Pi)=\delta$. Conversely, any straight simple path from x to Π that starts by walking to the left passes through v_{1,j_L}; if indeed $j-j_L\geq\delta$, then such a path is of length at least $\Delta(x,v_{1,j_L})+\Delta(v_{1,j_L},\Pi)\geq\delta$. ∎

To show that Π is not $\mathcal{F}_\delta$-closed, we rely on Condition (2) of Theorem A.2: We show a vertex $v' \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ such that there does not exist $z \in \mathcal{F}_\delta(\Pi)$ satisfying $\Delta(v', z) < \delta$. In particular, let $v' = v_{1,j_L+1}$ be the vertex to the right of v_{1,j_L}. Since there are no vertices from $\Pi \cup \mathcal{F}_\delta(\Pi)$ in columns $j_L + 1, ..., j$, it holds that $v' \notin \Pi \cup \mathcal{F}_\delta(\Pi)$. We show that $\Delta(v', \mathcal{F}_\delta(\Pi)) \geq \delta$, which implies that there does not exist $z \in \mathcal{F}_\delta(\Pi)$ satisfying $\Delta(v', z) < \delta$.

Fact E.6.2.2. *From the hypothesis of Claim E.6.2 it follows that $\Delta(v', \mathcal{F}_\delta(\Pi)) \geq \delta$.*

Proof. Note that $v_{2,j_L+1} \notin \mathcal{F}_\delta(\Pi)$, since columns $j_L + 1, ..., j$ do not contain vertices from $\Pi \cup \mathcal{F}_\delta(\Pi)$. Thus, any straight simple path from v' to $\mathcal{F}_\delta(\Pi)$ starts by walking either to the left or to the right. Any path that starts by walking from v' to the left goes through v_{1,j_L}. Since a vertex in column j_L is in Π, it holds that $\Delta(v_{1,j_L}, \Pi) \leq 1$, and thus $\Delta(v_{1,j_L}, \mathcal{F}_\delta(\Pi)) \geq \delta - 1$. Hence, any straight simple path from v' to $\mathcal{F}_\delta(\Pi)$ that starts by walking to the left is of length at least $\Delta(v', v_{1,j_L}) + \Delta(v_{1,j_L}, \mathcal{F}_\delta(\Pi)) \geq \delta$.

Conversely, any straight simple path from v' to $\mathcal{F}_\delta(\Pi)$ that starts by walking to the right passes through x' (since there are no vertices from $\Pi \cup \mathcal{F}_\delta(\Pi)$ in columns $j_L + 1, ..., j$). Relying on Fact E.6.2.1, and on the fact that $x' \notin \mathcal{F}_\delta(\Pi)$ (since $\Delta(x', \Pi) = \delta - 1$), any such path is of length at least $\Delta(v', x') + \Delta(x', \mathcal{F}_\delta(\Pi)) = (j + 1) - (j_L + 1) + 1 = \delta$. □

By Condition (2) of Theorem A.2, it follows that Π is not $\mathcal{F}_\delta$-closed, which concludes the proof of Claim E.6.2. □

As mentioned in the discussion after the statement of Claim E.6.2, the proof of the latter concludes the proof of Proposition E.6. ■

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il