Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 72 (2015)

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

Roei Tell
Department of Computer Science
Weizmann Institute of Science
roei.tell@weizmann.ac.il

August 4, 2015

Abstract

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. 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 reject 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 study sets of the form $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$, and apply this study to investigate dual problems in property testing. In particular, we present conditions on a metric space, on δ, and on a set Π that are sufficient and/or necessary in order for the equality $\Pi = \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ to hold. Using these conditions, we derive 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, Closure Operator.

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., [Gol10b]). 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.

Key contributions. The contributions in this work consist of two parts. First, 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

[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.

number of queries, unless the problem is trivial to begin with; this stands in sharp contrast to testing standard problems with one-sided error (see discussion in the end of Section 1.2.1). 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.

The second topic in the paper is the generic study of *sets of the form* $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ in metric spaces. We present several necessary and/or sufficient conditions for a set to be $\mathcal{F}_\delta$-closed; some of these conditions are applicable in general metric spaces, whereas others apply only in specific classes of metric spaces (e.g., in graphs). Two interesting general observations in this context are that (1) the condition of being $\mathcal{F}_\delta$-closed can be presented as a collection of local conditions, where each local condition depends only on a δ-neighborhood in the space (see discussion after Theorem 1.15); and (2) the operator $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ has the structure of a *closure operator* on the power set of Ω (see Section 1.3.1 for details on the latter).

Organization. The rest of the introduction surveys our results, and is organized as follows. In Section 1.1 we set the stage for the rest of the paper, by asserting the existence, and in some sense the abundance, of sets that are not $\mathcal{F}_\delta$-closed. Section 1.2 presents our main results regarding dual problems in property testing. In Section 1.3 we study the generic question of identifying $\mathcal{F}_\delta$-closed sets in metric spaces. In Section 1.4 we describe, in high-level, the techniques used to obtain several results regarding dual problems in property testing.

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

As mentioned in the beginning of the introduction, 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, since for any $\Pi \subseteq \Omega$ it holds that $\mathcal{F}_\delta(\Pi) = \Omega \setminus \Pi$. In contrast, 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 $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) = \Omega$.

The following theorem, which we prove in Section 3.3, asserts that for 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). 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 $\mathcal{F}_\delta(\Pi) = \emptyset$.) In addition, consider a metric space in which there exist N pairwise-disjoint δ-neighborhoods, each containing at most $\log(N)$ points; in such a space, most sets are not $\mathcal{F}_\delta$-closed (for exact statements see Propositions 3.11 and 3.12).

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, in Section 3.4 we show that 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. There even exist spaces Ω, parameters $\delta > 0$, and sets $\Pi \subseteq \Omega$ such that *all* points in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) \setminus \Pi$ are almost δ-far 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: Specifically, the complexity of a dual problem is lower bounded by the complexity of the original problem (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 usually 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 current section corresponds to Section 5 of the text, and is organized as follows. In Section 1.2.1 we state general results regarding the query complexity of dual problems; in particular, we show a strong *separation between standard testing problems and dual problems* with respect to testing with one-sided error. We then study specific natural dual problems, corresponding to well-known properties: In Section 1.2.2 we focus on properties of functions, in Section 1.2.3 we focus on properties of distributions, and in Section 1.2.4 we focus on graph properties.

1.2.1 General results regarding the query complexity of dual problems

The query complexity of any dual problem is closely related to the query complexity of its original problem. First, 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. (This is since the promise problem that corresponds to the original problem is $(\Pi, \mathcal{F}_{\epsilon\cdot n}(\Pi))$, whereas the promise problem for the dual is $(\mathcal{F}_{\epsilon\cdot n}(\Pi), \mathcal{F}_{\epsilon\cdot n}(\mathcal{F}_{\epsilon\cdot n}(\Pi))) \supseteq (\mathcal{F}_{\epsilon\cdot n}(\Pi), \Pi)$.) Thus:

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.*

Needless to say, *if the dual problem is equivalent to its original problem, then their query complexities are identical*.

Building on Observation 1.4, in Section 5.1 we show a lower bound for testing dual problems with *one-sided error*, regardless of whether the dual problem is equivalent to its original. Recall that in property testing, testers with one-sided error are ones that 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 $\Omega(n)$.*

It follows that testing the dual problem of a (non-empty) property with one-sided error and query complexity $o(n)$ is possible 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, since for any $\epsilon > 0$ and sufficiently large n it holds that $\mathcal{F}_{\epsilon \cdot n}(\Pi_n) = \emptyset$.

The fact that testing dual problems with one-sided error is either trivial or requires a linear number of queries stands in *sharp contrast to standard property testing problems*. This is since in standard property testing problems, essentially for any sub-linear function $q : \mathbb{N} \to \mathbb{N}$, there exists a property of Boolean functions such that the query complexity of testing it 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. The following theorem, which is proved in Section 5.2, asserts that for 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.*

One fundamental problem in this field that involves testing error-correcting codes is the problem of *linearity testing* [BLR90], which consists of testing whether a function $\varphi : G \to H$, where G and H are groups, is a group homomorphism. The most well-known specific case of linearity testing consists of testing the set of linear functions $\varphi : \{0,1\}^n \to \{0,1\}$, which indeed induces an error-correcting code (i.e., the Hadamard code). For general groups, Guo [Guo15] showed sufficient conditions on G and on H such that the set of homomorphisms $G \to H$ induces an error-correcting code. Another fundamental problem that induces an error-correcting code is that of *low-degree testing* [RS96], which consists of testing the set of low-degree multivariate polynomials over a finite field.

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 5.3 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. In Section 5.4 we consider the dual problem of this problem, which consists of testing whether a distribution is *far* from the predetermined distribution.

When considering the worst-case, over all families of distributions, the distribution 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 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 5.13).

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 [GGR98], 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.

In Section 5.5 we consider several dual problems in this model, corresponding to well-known testing problems.

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 graphs that are very sparse. 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. In Section 5.6 we consider several dual problems in this model, corresponding to well-known testing problems.

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 nevertheless 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 $\mathcal{F}_\delta$-closed sets and the operator $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$

Our results in this section are intended to facilitate the analysis of sets of the form $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$, and in particular to simplify the identification of sets that are $\mathcal{F}_\delta$-closed.

1.3.1 General metric spaces

The following are several equivalent *characterizations of all $\mathcal{F}_\delta$-closed sets* in any metric space Ω and for any $\delta > 0$. A more extensive list of such characterizations appears in Theorem 3.2 in Section 3.1.

Theorem 1.15 *(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 $\Pi = \mathcal{F}_\delta(\Pi')$.*
4. *There exists $\Pi' \subseteq \Omega$ such that $\Pi = \bigcap_{x \in \Pi'} \mathcal{F}_\delta(\{x\})$.*

Condition (2) of Theorem 1.15 is the basic technical tool that we use to analyze $\mathcal{F}_\delta$-closed sets when lacking a more convenient tool for the specific case. Interestingly, this condition is actually a *collection of local conditions*, where by "local" we mean that each condition depends only on a ball of radius 2δ in Ω.[3] Thus, if Π violates one of these conditions, then it is not $\mathcal{F}_\delta$-closed, and otherwise it is $\mathcal{F}_\delta$-closed.

Condition (3) of Theorem 1.15 implies, in particular, that all sets of the form $\mathcal{F}_\delta(\Pi')$, for some $\Pi' \subseteq \Omega$, are $\mathcal{F}_\delta$-closed. Thus, it is always true that $\mathcal{F}_\delta(\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))) = \mathcal{F}_\delta(\Pi)$, which implies that repeated applications of the operator $\mathcal{F}_\delta$ on a set Π yield a sequence that consists only of the sets Π, $\mathcal{F}_\delta(\Pi)$, and $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$. Moreover, if Π is $\mathcal{F}_\delta$-closed, then the sequence consists only of Π and $\mathcal{F}_\delta(\Pi)$.

Condition (4) of Theorem 1.15 implies that the potentially small collection $\{\mathcal{F}_\delta(\{x\})\}_{x\in\Omega}$ "generates" the collection of all $\mathcal{F}_\delta$-closed sets (i.e., a set is $\mathcal{F}_\delta$-closed if and only if it is an intersection of sets from $\{\mathcal{F}_\delta(\{x\})\}_{x\in\Omega}$).

The operator $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ is a closure operator. For a space Ω and parameter $\delta > 0$, consider the operator $\mathcal{F}_\delta \circ \mathcal{F}_\delta$ (i.e., $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$) on the power set of Ω. In Section 3.2 we show that this operator satisfies the following:

Proposition 1.16 *(structural results regarding $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$). For any Ω, $\delta > 0$, and $\Pi, \Pi' \subseteq \Omega$ it holds that:*

1. *(extensiveness) $\Pi \subseteq \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$.*
2. *(upwards monotonicity) If $\Pi \subseteq \Pi'$ then $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) \subseteq \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi'))$.*
3. *(idempotency) $\mathcal{F}_\delta^{(4)}(\Pi) = \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ (where $\mathcal{F}_\delta^{(4)}$ means four applications of $\mathcal{F}_\delta$).*

The three assertions in Proposition 1.16 suffice to deduce that the operator $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ is a *closure operator* (or *hull operator)* on the power set of Ω, a well-studied notion in many mathematical fields including algebra, topology, and matroid theory (see, e.g., [KD06, Chp. 2] or [vdV93, Chp. 1]). A closure operator is characterized by a corresponding collection of closed sets, which are the sets in its image; in our case, this is exactly the collection of $\mathcal{F}_\delta$-closed sets. A general result about closure operators, which holds also in the specific case of $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$, is that the closure of a set Π (i.e., the image of the set under the operator) is the unique intersection of all closed sets that contain the set Π (see Proposition 3.6).

1.3.2 Graphical metric spaces

If the metric space Ω is an undirected connected graph equipped with the shortest path metric, then we call it a *graphical* metric space. In this section we show several conditions that are either necessary or sufficient to deduce that a set in a graphical space is $\mathcal{F}_\delta$-closed. We also study these conditions in the special case of the Boolean hypercube, since the latter is important for property testing and since it belongs to several interesting graph classes.

One *necessary* condition for a set (in a graphical space) to be $\mathcal{F}_\delta$-closed is that, loosely speaking, it does not "fully enclose" some vertex $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$. More precisely, if a set Π is

[3] Each condition depends on a ball of radius 2δ, since Condition (2) requires the existence of $z \in \mathcal{F}_\delta(\Pi)$ such that $\Delta(z, x) < \delta$, which holds if z is in the open radius-δ ball around x and the open radius-δ ball around z does not intersect with Π.

$\mathcal{F}_\delta$-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 adjacent to Π (see Proposition 4.1). However, this necessary condition is not a sufficient one: There exist graphs, values of $\delta > 0$ and sets that satisfy this condition but are *not* $\mathcal{F}_\delta$-closed. Moreover, the condition is not a sufficient one even in the special case of the Boolean hypercube (see Proposition 4.3).

The following *sufficient* condition for a set in a graphical space to be $\mathcal{F}_\delta$-closed, which we study in Section 4.1.2, is a strengthening of the aforementioned necessary condition.

Definition 1.17 *(strongly $\mathcal{F}_\delta$-closed). For a graphical Ω and $\delta > 0$, a set $\Pi \subseteq \Omega$ is* strongly $\mathcal{F}_\delta$-closed *if every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ lies on a* shortest path *(i.e., a path of length δ) from Π to $\mathcal{F}_\delta(\Pi)$.*

Indeed, as implied by its name, a set that is strongly $\mathcal{F}_\delta$-closed is $\mathcal{F}_\delta$-closed (see the discussion after Proposition 4.6). An equivalent definition of being strongly $\mathcal{F}_\delta$-closed is as follows: A set Π 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' such that $\Delta(x', \Pi) = \Delta(x, \Pi) + 1$.

The condition of being strongly $\mathcal{F}_\delta$-closed might be more convenient to evaluate in some cases, compared to the characterizations in Theorem 1.15, since it might be easier to argue about the immediate neighbors of $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ instead of about the δ-neighborhood of x (i.e., about a vertex $z \in \mathcal{F}_\delta(\Pi)$ such that $\Delta(x, z) < \delta$) as is required in Condition (2) of Theorem 1.15. However, being strongly $\mathcal{F}_\delta$-closed is not a necessary condition for being $\mathcal{F}_\delta$-closed: There exist graphical spaces Ω, parameters $\delta > 0$ and subsets $\Pi \subseteq \Omega$ such that Π is $\mathcal{F}_\delta$-closed but not strongly $\mathcal{F}_\delta$-closed. Furthermore, such sets exist even in the special case of the Boolean hypercube.

Proposition 1.18 *(strongly $\mathcal{F}_\delta$-closed is not a necessary condition for $\mathcal{F}_\delta$-closed in the Boolean hypercube). 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.*

Nevertheless, there exists graphs and values of $\delta > 0$ such that every $\mathcal{F}_\delta$-closed set in the graph is also strongly $\mathcal{F}_\delta$-closed. In Section 4.1.3 and Appendix C we briefly study the question of for which graphs (and for which values of $\delta > 0$) does it holds that a set is $\mathcal{F}_\delta$-closed if and only if it is strongly $\mathcal{F}_\delta$-closed. In particular, we observe that this holds for any graph when $\delta = 2$ (but not when $\delta \geq 3$). We also show that there exist graph families such that for every $\delta > 0$, every $\mathcal{F}_\delta$-closed set in the graph is also strongly $\mathcal{F}_\delta$-closed; these graph families include simple paths, cycles, and all $2 \times n$ grids.

A different direction of study, which we present in Section 4.1.4, is as follows: Instead of fixing δ and asking which sets are $\mathcal{F}_\delta$-closed, we ask, for a fixed set $\Pi \subseteq \Omega$, what are the values of δ for which Π is strongly $\mathcal{F}_\delta$-closed, $\mathcal{F}_\delta$-closed, or not $\mathcal{F}_\delta$-closed. Interestingly, for any set Π in a graphical space with bounded diameter, the values of δ for which Π is $\mathcal{F}_\delta$-closed constitute a single bounded interval. This interval starts at $\delta = 1$ (since every set is $\mathcal{F}_1$-closed), and for any set Π we denote the right-end of this interval by $\delta^{\mathsf{C}}(\Pi)$ (i.e., $\delta^{\mathsf{C}}(\Pi)$ is the maximal value for which Π is $\mathcal{F}_\delta$-closed). A similar claim holds for values of δ for which Π is strongly $\mathcal{F}_\delta$-closed. That is –

Proposition 1.19 *(values of δ for which a set is $\mathcal{F}_\delta$-closed and strongly $\mathcal{F}_\delta$-closed). For a graphical Ω with bounded diameter and a non-trivial $\Pi \subseteq \Omega$, there exist two integers $\delta^{\mathsf{C}}(\Pi)$ and $\delta^{\mathsf{SC}}(\Pi)$ such that $\delta^{\mathsf{SC}}(\Pi) \leq \delta^{\mathsf{C}}(\Pi)$ and for every integer $\delta > 0$ it holds that*

1. Π *is* $\mathcal{F}_\delta$*-closed if and only if* $\delta \in [1, \delta^{\mathsf{C}}(\Pi)]$.
2. Π *is strongly* $\mathcal{F}_\delta$*-closed if and only if* $\delta \in [1, \delta^{\mathsf{SC}}(\Pi)]$.

In contrast, if the space Ω is not graphical, then a statement analogous to Item (1) in Proposition 1.19 does not necessarily hold (see Proposition 4.13, and also recall that the notion of strongly $\mathcal{F}_\delta$-closed sets was not defined for non-graphical metric spaces).

1.3.3 The Boolean hypercube

In the Boolean hypercube, for any fixed set Π, we can obtain a lower bound for $\delta^{\mathsf{SC}}(\Pi)$ and an upper bound for $\delta^{\mathsf{C}}(\Pi)$, using coding-theoretic features of Π. In Section 4.2 we show such bounds, and demonstrate that, in general, the bounds we show are far from being tight.

In particular, $\delta^{\mathsf{C}}(\Pi)$ is smaller than the covering radius of Π, that is the minimal $\delta > 0$ such that every string x satisfies $\Delta(x, \Pi) \leq \delta$. On the other hand, $\delta^{\mathsf{SC}}(\Pi)$ is greater or equal to the unique decoding distance of Π. In fact, we prove a stronger statement, as follows. A set Π is called (δ, L)-list-decodable if every Hamming ball of radius δ contains at most L elements from Π. Then:

Proposition 1.20 *($\left(\delta, \frac{n}{\delta} - 1\right)$-list-decodable codes are strongly $\mathcal{F}_\delta$-closed). For a non-trivial set Π in the n-dimensional Boolean hypercube and $\delta > 0$, if Π is $\left(\delta, \frac{n}{\delta} - 1\right)$-list-decodable, then Π is strongly $\mathcal{F}_\delta$-closed.*

Again, this sufficient condition for being strongly $\mathcal{F}_\delta$-closed is not a necessary one: There exist sets that are strongly $\mathcal{F}_\delta$-closed for all values of $\delta \in [n-1]$, but are not even $(1, n)$-list-decodable. Nevertheless, in general, the requirement in Proposition 1.20 that every Hamming ball contains at most $\frac{n}{\delta} - 1$ elements cannot be significantly relaxed (see Proposition 4.18).

1.3.4 Digest

Figure 1 presents a summary of the *sufficient* conditions for a set to be $\mathcal{F}_\delta$-closed that were presented in Section 1.3.

Relevant to the Boolean Hypercube		Relevant to Graphical Spaces		Relevant to General Metric Spaces
$\left(\delta, \frac{n}{\delta} - 1\right)$-list-decodable	$\overset{\Longrightarrow}{\not\Longleftarrow}$	strongly $\mathcal{F}_\delta$-closed	$\overset{\Longrightarrow}{\not\Longleftarrow}$	$\mathcal{F}_\delta$-closed

Figure 1: Summary of the main conditions presented in Section 1.3

We point out the interesting fact that the three conditions in Figure 1 can be presented as collections of local conditions, where "local" conditions are ones that depend only on the behavior of Π in a local neighborhood of Ω. While the local conditions implied by the characterization of $\mathcal{F}_\delta$-closed sets in Condition (2) of Theorem 1.15 depend on balls of radius

2δ, the sufficient (but not necessary) conditions in Definition 1.17 and Proposition 1.20 imply local conditions that depend only on balls of radius δ.

1.4 Our techniques

This section focuses on our techniques for proving claims regarding dual problems in property testing (i.e., the claims in Section 1.2). In comparison, the proofs for the claims of Section 1.3 are easier, and some are straightforward. We note, however, that some constructions for counter-examples in Section 1.3 are quite evasive, and it seems a-priori non-obvious that a counter-example should even exist in these cases (see, e.g., Proposition 1.18).

The lower bound regarding testing dual problems with one-sided error (i.e., Theorem 1.5) stems from a similar lower bound with respect to testing standard problems with *perfect soundness*; that is, testing a property such that "no" inputs are *always* rejected. The query complexity of testing standard problems with perfect soundness is linear, unless the problem is trivial (i.e., unless $\mathcal{F}_\delta(\Pi_n) = \emptyset$ for a sufficiently large n; see Proposition 5.6). The lower bound regarding dual problems follows, since the query complexity of testing a dual problem with one-sided error is lower bounded by the query complexity of testing a standard problem with perfect soundness.

In testing specific dual problems, we rely on one of two general techniques. The first, which we apply in the cases of error-correcting codes (Theorem 1.6), monotone functions (Theorem 1.7), and distribution identity testing (Theorem 1.9), 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). The latter is done relying on the characterizations of $\mathcal{F}_\delta$-closed sets and on the sufficient conditions for a set to be $\mathcal{F}_\delta$-closed, described in Section 1.3.

The second technique is useful when the dual problem is different from the original one. Specifically, for the three dual problems that we solve in the context of graph property testing (k-colorability in the dense graph model, and connectivity and cycle-free graphs in the bounded-degree model), we reduce the dual problem to the problem of *tolerant testing*, introduced by Parnas, Ron, and Rubinfeld [PRR06]: 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. These are structural results regarding the property Π_n, which are of independent interest.

Then, we need to show that the corresponding tolerant testing problem can be efficiently solved. In the case of k-colorability in the dense graph model, the tolerant testing problem was solved by Fischer and Newman [FN07]; in the case of connected graphs in the bounded-degree model, we solve the tolerant testing problem ourselves (see Lemma 5.27 and Section 5.6.1.3); and in the case of cycle-free graphs in the bounded-degree model, the tolerant testing problem was solved by Marko and Ron [MR06].

We stress two points regarding the technique of reducing dual problems to tolerant testing problems. First, as mentioned in Section 1.1, it is *not* true in general that points in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$ are $(\alpha \cdot \delta)$-close to Π_n, for some $\alpha < 1$, and this is not even true for all sets in the Boolean hypercube. In fact, the proofs that this holds for the three specific properties mentioned above are not straightforward, and we were so far unable to prove that this holds in several other cases (e.g., for the property of graphs containing a large clique). Secondly, 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 [VV14] imply that the complexity of the corresponding tolerant testing problem is $\tilde{\Theta}(n)$.

The general technical question underlying both techniques outlined above is the following: *Given a metric space Σ^n, a set $\Pi_n \subseteq \Sigma^n$, a parameter $\delta > 0$, and a point x that satisfies some requirements regarding its distance from Π_n, does there exist a point z such that $\Delta(x,z) < \delta$ and $\Delta(z,\Pi_n) \geq \delta$?* In most cases, given a point x that satisfies some distance requirement from Π_n, we show how to explicitly modify x to a corresponding suitable z. Our modification of x to z capitalizes on structural features of objects in the relevant metric space that satisfy the specific distance requirement. For example, when relying on Condition (2) of Theorem 1.15 to show that a set Π_n is $\mathcal{F}_\delta$-closed, we start from a point $x \notin \Pi_n \cup \mathcal{F}_\delta(\Pi_n)$, and modify it into $z \in \mathcal{F}_\delta(\Pi_n)$ such that $\Delta(x,z) < \delta$. Similarly, to reduce a dual problem to the corresponding tolerant testing problem (i.e., to prove that $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n)) \subseteq \{y : \Delta(y,\Pi_n) \leq \alpha \cdot \delta\}$), we start with x such that $\Delta(x,\Pi_n) \in (\alpha \cdot \delta, \delta)$, and modify it into $z \in \mathcal{F}_\delta(\Pi_n)$ such that $\Delta(x,z) < \delta$, which implies that $x \notin \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$.

2 Preliminaries

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 $B[x,\delta) \stackrel{\text{def}}{=} \{y : \Delta(x,y) < \delta\}$.

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.

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 $\Pi = \{\Pi_n\}_{n\in\mathbb{N}}$).

3 Sets of the form $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ and $\mathcal{F}_\delta$-closed sets

In this section we study the basic properties of sets of the form $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$. Motivated by applications in property testing, we focus on sets that satisfy $\Pi = \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$, which by Definition 1.1 are called *$\mathcal{F}_\delta$-closed sets*.

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 3.1 *(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 $\Delta(z,\Pi) \leq \Delta(z,x) < \delta$, which contradicts $z \in \mathcal{F}_\delta(\Pi)$. ■

However, as mentioned in the introduction, not every set Π satisfies $\Pi = \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$; that is, not every set is $\mathcal{F}_\delta$-closed.

In Section 3.1 we characterize the sets that are $\mathcal{F}_\delta$-closed in any metric space. Section 3.2 is a detour, in which we give additional insight into the relationship between any set Π

and $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$, by showing that the operator $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ satisfies the axioms of a *closure operator* (or *hull operator*).[4] In Section 3.3 we study sets that are not $\mathcal{F}_\delta$-closed, and in particular demonstrate their existence and lower bound the fraction of such sets in two classes of metric spaces. And in Section 3.4 we study the distance of points in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ from Π.

3.1 Characterizations of $\mathcal{F}_\delta$-closed sets

For a fixed Ω and $\delta > 0$, which are the $\mathcal{F}_\delta$-closed sets in Ω? The following theorem presents several equivalent characterizations of the $\mathcal{F}_\delta$-closed sets for any fixed Ω and δ.

Theorem 3.2 *(characterizations of $\mathcal{F}_\delta$-closed sets, extending Theorem 1.15). 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 3.2.1 *(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 3.2.2 *(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 3.2.1,

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

We now prove the equivalences of Conditions (1)–(6).

(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 $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$.

[4] This material will not be used in the rest of the paper.

(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 3.1 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 $\Pi'' = \mathcal{F}_\delta(\Pi')$.

(4) $\implies$ (1) Let $\Pi = \mathcal{F}_\delta(\Pi'')$ for some $\Pi'' \subseteq \Omega$. By Fact 3.1 it holds that $\Pi'' \subseteq \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi''))$, whereas by Fact 3.2.2, 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 3.1 again, we know that $\Pi \subseteq \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$, and thus $\Pi = \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$.

(4) $\iff$ (5) By Fact 3.2.1.

(5) $\iff$ (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. ■

In the introduction, following the statement of Theorem 1.15, we commented on the implications of some of these characterizations. Here, we add several additional comments. First, 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: $\mathcal{F}_\delta$-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 (4) and (3) implies that $\{\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))\}_{\Pi\subseteq\Omega} = \{\mathcal{F}_\delta(\Pi)\}_{\Pi\subseteq\Omega}$. Moreover, the operator $\mathcal{F}_\delta$ is a bijection between these two collections: 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 $\mathcal{F}_\delta$.

Condition (2) in Theorem 3.2 is the basic technical condition that we will use to evaluate whether sets are $\mathcal{F}_\delta$-closed. As mentioned in the discussion after the statement of Theorem 1.15, 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. The negation of Condition (2) yields a more explicit description of a collection of conditions such that each condition corresponds to a specific ball in Ω.

Corollary 3.3 *(being $\mathcal{F}_\delta$-closed as a collection of local conditions). If, for some $x \in \Omega$, it holds that $x \notin \Pi$ and $B[x,\delta) \cap \Pi \neq \emptyset$ and $B[x,\delta) \cap \mathcal{F}_\delta(\Pi) = \emptyset$, then Π is not $\mathcal{F}_\delta$-closed. Otherwise, Π is $\mathcal{F}_\delta$-closed.*

Proof. By negating Condition (2) in Theorem 3.2 we get that Π is not $\mathcal{F}_\delta$-closed if and only if there exists $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ such that for every $z \in \mathcal{F}_\delta(\Pi)$ it holds that $\Delta(z,x) \geq \delta$. Note that:

- For every $x \notin \Pi$ it holds that $x \notin \mathcal{F}_\delta(\Pi)$ if and only if $B[x,\delta) \cap \Pi \neq \emptyset$.
- The condition that for every $z \in \mathcal{F}_\delta(\Pi)$ it holds that $\Delta(z,x) \geq \delta$ is equivalent to the condition that $B[x,\delta) \cap \mathcal{F}_\delta(\Pi) = \emptyset$. ■

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

The current section is a detour, which is intended to provide additional insight to the relationship between Π and $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$, for any Ω and $\Pi \subseteq \Omega$. The results in this section will not be used in the rest of the paper, and thus are not essential in order to read other sections.

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 3.4 *(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 3.5 *($\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ 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 3.1. Axiom (2) follows by applying Fact 3.2.2 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 3.2. ■

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 3.2 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 3.6 *($\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ is the unique minimal $\mathcal{F}_\delta$-closed set containing Π). For any Ω, $\delta > 0$ and $\Pi \subseteq \Omega$ it holds that*

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

For convenience, we include a proof of Proposition 3.6 in Appendix A. The proof follows the standard proof of the analogous fact for general closure operators.

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 3.4, resulting in special classes of closure operators. In Appendix A we show that the operator $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ does not belong to some of these classes of operators. In particular, $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ 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.

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

The focus of this section is proving the existence, and in some sense the abundance, of sets that are not $\mathcal{F}_\delta$-closed. The main result presented in this section is that for any Ω such that not all points in it 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. We further show a lower bound on 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 Ω and the other is when Ω is the Boolean hypercube.

First, for every Ω let us delineate two "extreme" settings for δ that collapse $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ to a trivial operator. In one setting, δ is too large and $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) \equiv \Omega$ for any non-empty Π; in this case all non-trivial sets are not $\mathcal{F}_\delta$-closed. In the other setting, δ is too small and $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) = \Pi$ for any $\Pi \subseteq \Omega$; that is, all sets are $\mathcal{F}_\delta$-closed.

Fact 3.7 *(if δ is too large then $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) \equiv \Omega$). For any Ω such that $\sup_{x,y\in\Omega}\{\Delta(x,y)\}$ is finite, if $\delta > \sup_{x,y\in\Omega}\{\Delta(x,y)\}$, then for every non-empty $\Pi \subseteq \Omega$ it holds that $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) = \Omega$.*

Proof. In this case, for any $\Pi \neq \emptyset$ it holds that $\mathcal{F}_\delta(\Pi) = \emptyset$, and thus $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) = \Omega$. ∎

Fact 3.8 *(if δ is too small then $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) \equiv \Pi$). For any Ω such that $\inf_{x\neq y}\{\Delta(x,y)\} > 0$, if $\delta \leq \inf_{x\neq y}\{\Delta(x,y)\}$, then for every $\Pi \subseteq \Omega$ it holds that $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) = \Pi$.*

Proof. In this case, for every $\Pi \subseteq \Omega$ it holds that $\mathcal{F}_\delta(\Pi) = \Omega \setminus \Pi$, and thus $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) = \Omega \setminus \mathcal{F}_\delta(\Pi) = \Omega \setminus (\Omega \setminus \Pi) = \Pi$. ∎

Following Facts 3.7 and 3.8, and disregarding for a moment the "boundary case" when $\delta = \sup_{x\neq y}\{\Delta(x,y)\}$, we restrict our investigation to settings of Ω and δ such that

$$\delta \in \left(\inf_{x\neq y\in\Omega}\{\Delta(x,y)\}, \sup_{x,y\in\Omega}\{\Delta(x,y)\} \right) \tag{3.1}$$

The following theorem shows that for *every* δ that satisfies Eq. (3.1) 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 3.9 *(Theorem 1.2, restated). For any Ω, if $\delta > 0$ satisfies Eq. (3.1), 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 3.2 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. (3.1) 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.)

Proposition 3.10 *(values of δ for which the notion of $\mathcal{F}_\delta$-closed sets is non-trivial). Let Ω such that the supremum in Eq. (3.1) is attained (i.e., there exist $u, v \in \Omega$ such that $\Delta(u,v) = \sup_{x,y\in\Omega}\{\Delta(x,y)\}$). Then, for every $\delta > 0$, it holds that*

$$\delta \in \left(\inf_{x\neq y\in\Omega}\{\Delta(x,y)\}, \max_{x,y\in\Omega}\{\Delta(x,y)\}\right] \tag{3.2}$$

if and only if *there exist non-trivial sets that are $\mathcal{F}_\delta$-closed and non-trivial sets that are not $\mathcal{F}_\delta$-closed.*

Proof. Assume that δ does not satisfy Eq. (3.2). If $\delta \leq \inf_{x\neq y\in\Omega}\{\Delta(x,y)\}$, then by Fact 3.8 all sets are $\mathcal{F}_\delta$-closed; and if $\delta > \max_{x,y\in\Omega}\{\Delta(x,y)\}$, then by Fact 3.7 all non-trivial sets are not $\mathcal{F}_\delta$-closed.

For the other direction, assume that δ satisfies Eq. (3.2). Then, we can construct a non-trivial set that is not $\mathcal{F}_\delta$-closed identically to the proof of Theorem 3.9; and for an $\mathcal{F}_\delta$-closed set we take u and v such that $\Delta(u,v) = \max_{x,y}\{\Delta(x,y)\}$ and let $\Pi = \mathcal{F}_\delta(\{u\}) \neq \emptyset$. ■

Theorem 3.9 implies that for any Ω and $\delta > 0$ that satisfies Eq. (3.1) 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 *fraction* of sets that are not $\mathcal{F}_\delta$-closed.

Proposition 3.11 *(lower bound on the fraction 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 Corollary 3.3, 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 3.11 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 3.11 is far from tight for some spaces. In particular, in the special case of the Boolean hypercube, Proposition 3.12 presents a tighter lower bound, relying on a simple argument tailored to this specific case.

Proposition 3.12 *(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

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

3.4 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. However, in this section, we show several examples demonstrating that points in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ might be almost δ-far from Π.

Proposition 3.13 *(points in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ are not necessarily close to Π). There exists a space Ω such that for every $\delta > 0$ there exists a set $\Pi \subseteq \Omega$ such that for every $\delta' < \delta$ it holds that $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ contains points that are δ'-far from Π.*

Proof. Let $\Omega = (0, \infty)$ with the usual metric of $\mathbb{R}$. For any $\delta > 0$, let $\Pi = \{\delta\}$. Since every $x \in (0, 2\delta)$ satisfies $\Delta(x, \Pi) = |x - \delta| < \delta$, then $\mathcal{F}_\delta(\Pi) \subseteq \Omega \setminus (0, 2\delta) = [2\delta, \infty)$. Now, for every positive $\delta' < \delta$, let $z = \delta - \delta' > 0$. Note that z satisfies $\Delta(z, \delta) = \delta'$ (i.e., z is δ'-far from Π). However, since $\mathcal{F}_\delta(\Pi) \subseteq [2\delta, \infty)$, it follows that $\Delta(z, \mathcal{F}_\delta(\Pi)) = |2\delta - z| = 2\delta - (\delta - \delta') > \delta$, and thus $z \in \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$. ■

The following proposition shows that this phenomenon, where points in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ are almost δ-far from Π, happens also in the special case where Ω is the Boolean hypercube.

Proposition 3.14 *(an analogue of Proposition 3.13 for the Boolean hypercube). Let $\Omega = H_n$ be the n-dimensional Boolean hypercube. Then for every $\delta \geq 2$ there exists a set $\Pi \subseteq H_n$ such that $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ contains points that are $(\delta - 1)$-far from Π.*

Proof. We show a set $\Pi \neq H_n$ such that $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) = H_n$ and there exist points that are $(\delta - 1)$-far from Π. Recall that for $x \in H_n$, we denote by $\|x\|_1$ the Hamming weight of x. Let Π be the set of strings with Hamming weight $\delta - 1$ or more; that is, $\Pi = \{x \in H_n : \|x\|_1 \geq \delta - 1\}$. Note that every $x \notin \Pi$ (i.e., every x such that $\|x\|_1 \leq \delta - 2$) satisfies $\Delta(x, \Pi) = (\delta - 1) - \|x\|_1 \leq \delta - 1$, and hence $\mathcal{F}_\delta(\Pi) = \emptyset$ and $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) = H_n$. In particular, it holds that the vertex $o = (0, ..., 0)$ (i.e., $\|o\|_1 = 0$) satisfies $o \in \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ whereas $\Delta(o, \Pi) = \delta - 1$. ■

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 3.15 *(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.

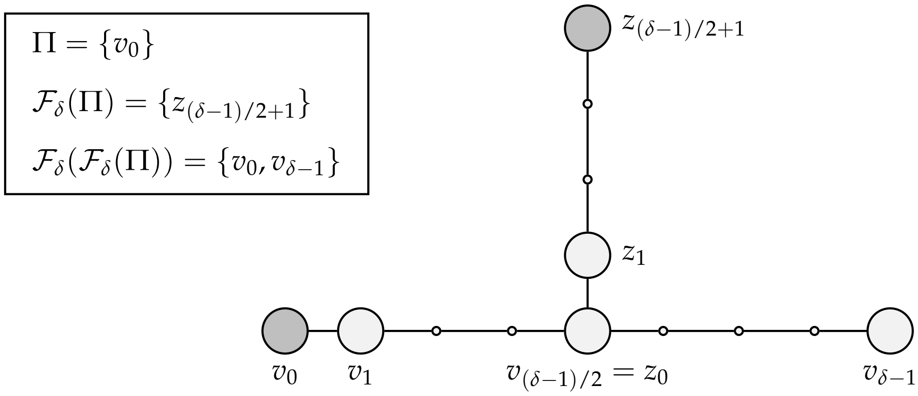

Figure 2: The space Ω.

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 $\Delta(v_{\delta-1}, \Pi) = \delta - 1$. ■

4 Evaluating whether a set is $\mathcal{F}_\delta$-closed in two special cases

Recall that Theorem 3.2 gives several sufficient and necessary conditions for a set to be $\mathcal{F}_\delta$-closed in a metric space. In this section we present several conditions that are either sufficient or necessary to deduce that a set is $\mathcal{F}_\delta$-closed, and that might be more convenient to evaluate for some sets than the characterizations in Theorem 3.2.

However, each of the conditions that we present applies only in a specific class of metric spaces: Some of them apply only in graphical spaces (see Section 4.1) and others apply only in the special case of the Boolean hypercube (see Section 4.2). Furthermore, all conditions we present are either sufficient or necessary, but not both.

4.1 Graphical spaces and strongly $\mathcal{F}_\delta$-closed sets

In this section we focus only on graphical spaces; recall that these 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 $\min_{x \neq y \in \Omega}\{\Delta(x, y)\} = 1$. Recall that Fact 3.8 states 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. Thus, in every graphical space, all sets are $\mathcal{F}_1$-closed. Accordingly, in this section we are mainly interested in integer values of $\delta \geq 2$.

In Section 4.1.1 we show a necessary condition for a set to be $\mathcal{F}_\delta$-closed in a graphical space. This necessary condition sets the stage for the subsequent section. In Section 4.1.2, which is the main part of our discussion of graphical spaces, we present a sufficient condition for a set to be $\mathcal{F}_\delta$-closed in a graphical space. We call sets that satisfy this sufficient condition *strongly $\mathcal{F}_\delta$-closed* sets. Section 4.1.3 is a detour, in which we explore spaces (and values of $\delta > 0$) for which the sufficient condition of being strongly $\mathcal{F}_\delta$-closed is also a necessary one. In Section 4.1.4 we show that for any fixed set in a graphical space, the values of δ for which the set is $\mathcal{F}_\delta$-closed (resp., strongly $\mathcal{F}_\delta$-closed) constitute a single interval.

4.1.1 Sets that "enclose" a vertex are not $\mathcal{F}_\delta$-closed

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 a set 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 4.1 *(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. 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 4.1.1. *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 4.1.1, 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. ■

By combining Proposition 4.1 and Fact 4.1.1, we get the following corollary, which sets the stage for Section 4.1.2. Loosely speaking, it states that for an $\mathcal{F}_\delta$-closed set Π, every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ lies on a path from Π to $\mathcal{F}_\delta(\Pi)$ that satisfies the following: The subpath from Π to x does not intersect $\mathcal{F}_\delta(\Pi)$ nor any neighbor of $\mathcal{F}_\delta(\Pi)$; and the subpath from x to $\mathcal{F}_\delta(\Pi)$ does not intersect Π nor any neighbor of Π.

Corollary 4.2 *(a corollary of Proposition 4.1). 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 $\Delta(v_i, \Pi) \geq 2$.*

Proposition 4.1 asserts that the condition specified in it (i.e., that 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 adjacent to Π) is a necessary condition for a set in a graphical space to be $\mathcal{F}_\delta$-closed. In some cases it is convenient to show that this condition is not met, and deduce that the set is not $\mathcal{F}_\delta$-closed; demonstrations for this technique appear in the proofs of Propositions 4.19, 5.20, 5.22, 5.23, 5.28, 5.31, 5.34, and C.2. Readers interested in further details regarding the condition in Proposition 4.1 are referred to Appendix B, where we show another condition that is equivalent to the condition in Proposition 4.1, which might be interesting by itself.

The condition in Proposition 4.1 is not sufficient to deduce that a set is $\mathcal{F}_\delta$-closed. To see this, consider the graph depicted in Figure 3 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 4.1. However, since $\mathcal{F}_3(\mathcal{F}_3(\{p\})) = \{p, v_1\}$, it follows that $\{p\}$ is not $\mathcal{F}_3$-closed.

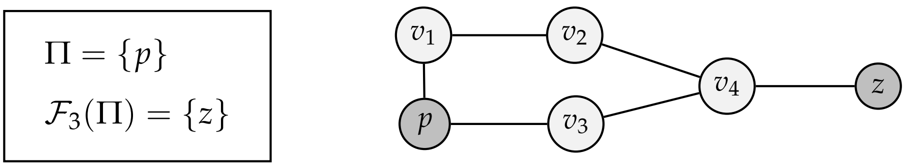

Figure 3: The singleton $\{p\}$ is not $\mathcal{F}_3$-closed, although the necessary condition stated in Proposition 4.1 is satisfied.

The following proposition demonstrates that, even in the special case of the Boolean hypercube, the necessary condition implied by Proposition 4.1 is not sufficient for a set to be $\mathcal{F}_\delta$-closed.

Proposition 4.3 *(the condition in Proposition 4.1 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 $4 \le \delta \le n-1$.

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

Combining these two facts, we deduce that if $|x| \ge 2$, 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 $\Delta(v,\Pi) \ge 2$: 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 $|x| \le 1$ 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 $|x| \le 1$ 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 3.2. 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 $\Delta(z,\emptyset) \ge \delta$. Assume towards a contradiction that there exists $z \in \mathcal{F}_\delta(\Pi)$ such that $\Delta(z,\emptyset) \le \delta-1$, which implies that $|z| \le \delta-1$.

- If $|z| \le \delta-2$, then we can remove all elements from z, and add the element 1, to obtain $\{1\} \in \Pi$. Therefore $\Delta(z,\Pi) \le \Delta(z,\{1\}) \le |z|+1 \le \delta-1$, which contradicts $z \in \mathcal{F}_\delta(\Pi)$.
- If $|z| = \delta-1 \ge 3$, 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 $z \cap p \ne \emptyset$, 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 $\Delta(z,\Pi) \le \Delta(z,p) \le \delta-2$, which contradicts $z \in \mathcal{F}_\delta(\Pi)$.

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

4.1.2 Strongly $\mathcal{F}_\delta$-closed sets

In Corollary 4.2 we showed the following necessary condition for a set to be $\mathcal{F}_\delta$-closed: If a set Π is $\mathcal{F}_\delta$-closed, then for every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$, there exists a path from Π to x that does not intersect $\mathcal{F}_\delta(\Pi)$ (nor any of its neighbors), and a path from x to $\mathcal{F}_\delta(\Pi)$ that does not intersect Π (nor any of its neighbors). While each of these two paths is actually a shortest path, their combination is not necessarily a shortest path from Π to $\mathcal{F}_\delta(\Pi)$. In this section, we prove that if every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ lies on a *shortest path* from Π to $\mathcal{F}_\delta(\Pi)$, then Π is $\mathcal{F}_\delta$-closed. We also show that this sufficient condition is, unfortunately, not a necessary one.

We start by presenting several equivalent formulations for the latter condition, which we call being *strongly $\mathcal{F}_\delta$-closed*.

Definition 4.4 *(Definition 1.17, restated). For a graphical Ω and $\delta > 0$, a set $\Pi \subseteq \Omega$ is* strongly $\mathcal{F}_\delta$-closed *if every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ lies on a* shortest path *(i.e., a path of length δ) from Π to $\mathcal{F}_\delta(\Pi)$.*

Proposition 4.5 *(strongly $\mathcal{F}_\delta$-closed, equivalent formulation). 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 $z \in \mathcal{F}_\delta(\Pi)$ such that $\Delta(x, z) = \delta - \Delta(x, \Pi)$.*[5]

Proof. We first show that Definition 4.4 implies the condition in Proposition 4.5. Assume that every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ lies on a path of length δ from Π to $\mathcal{F}_\delta(\Pi)$. Let $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$. If $\Delta(x, \mathcal{F}_\delta(\Pi)) > \delta - \Delta(x, \Pi)$, then any path from Π to $\mathcal{F}_\delta(\Pi)$ that passes through x is of length at least $\Delta(\Pi, x) + \Delta(x, \mathcal{F}_\delta(\Pi)) > \delta$, which contradicts the hypothesis. Also, if $\Delta(x, \mathcal{F}_\delta(\Pi)) < \delta - \Delta(x, \Pi)$, then there exists a path from Π to $\mathcal{F}_\delta(\Pi)$ of length $\Delta(\Pi, x) + \Delta(x, \mathcal{F}_\delta(\Pi)) < \delta$, which is a contradiction. Hence $\Delta(x, \mathcal{F}_\delta(\Pi)) = \delta - \Delta(x, \Pi)$, which implies that there exists $z \in \mathcal{F}_\delta(\Pi)$ such that $\Delta(x, z) = \delta - \Delta(x, \Pi)$.

For the other direction, assume that for every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ there exists $z \in \mathcal{F}_\delta(\Pi)$ such that $\Delta(x, z) = \delta - \Delta(x, \Pi)$. Let $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$, and let $z \in \mathcal{F}_\delta(\Pi)$ be the vertex that exists by the hypothesis. Now, let $p \in \Pi$ such that $\Delta(p, x) = \Delta(\Pi, x)$. Then, a shortest path from p to x, combined with a shortest path from x to z, yields a path of length $\Delta(p, x) + \Delta(x, z) = \delta$ between Π and $\mathcal{F}_\delta(\Pi)$ that passes through x. ∎

Proposition 4.6 *(strongly $\mathcal{F}_\delta$-closed, equivalent formulation). 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 $\Delta(x', \Pi) = \Delta(x, \Pi) + 1$.*

Proof. Assume that for every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ there exists a neighbor x' of x such that $\Delta(x', \Pi) = \Delta(x, \Pi) + 1$. We show that for every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ there exists $z \in \mathcal{F}_\delta(\Pi)$ such that $\Delta(x, z) = \delta - \Delta(x, \Pi)$, and rely on Proposition 4.5 to deduce that Π is strongly $\mathcal{F}_\delta$-closed.

Let $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ and denote $x_0 = x$. By the hypothesis, there exists 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 $\Delta(x, \Pi) =$

[5] This condition can be generalized to non-graphical metric spaces. However, in general metric spaces, the easier-to-evaluate condition in Proposition 4.6 would not be applicable. We thus do not define the generalization in the current paper.

$\delta - 1$ 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 $\Delta(x_2, \Pi) = \Delta(x_1, \Pi) + 1$. This way we repeatedly apply this step such that for the i^{th} application it holds 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.

For the other direction, assume that Π is strongly $\mathcal{F}_\delta$-closed. Then, by Proposition 4.5, for every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ there exists $z \in \mathcal{F}_\delta(\Pi)$ such that $x = x_0, x_1, ..., x_{\delta - \Delta(\Pi,x)} = z$ is a path of length $\delta - \Delta(x, \Pi)$ between x and z. Hence it must be that $\Delta(x_1, \Pi) > \Delta(x, \Pi)$, otherwise there exists a path between z and Π of length at most

$$\Delta(z, x_1) + \Delta(x_1, \Pi) = \delta - \Delta(\Pi, x) - 1 + \Delta(x_1, \Pi) \leq \delta - 1$$

which contradicts $z \in \mathcal{F}_\delta(\Pi)$. Therefore, since $\Delta(x_1, \Pi) > \Delta(x, \Pi)$ and $\Delta(x_1, \Pi) \leq \Delta(x, \Pi) + 1$, it follows that $\Delta(x_1, \Pi) = \Delta(x, \Pi) + 1$. ■

Recall that Condition (2) of Theorem 3.2 asserts 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 $\Delta(x, z) < \delta$. Comparing this condition to Proposition 4.5, it follows that if a set is strongly $\mathcal{F}_\delta$-closed then it is $\mathcal{F}_\delta$-closed. However, the condition in Proposition 4.6 seems more convenient to evaluate in some cases: 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 a neighbor of x that is farther away from Π than x. Demonstrations for this technique appear in the proofs of Propositions 4.15, 4.17, 4.19, 5.10, and C.2.

While being strongly $\mathcal{F}_\delta$-closed is a sufficient condition for a set to be $\mathcal{F}_\delta$-closed, it is not a necessary condition. To see this, consider the graph depicted in Figure 4, 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.

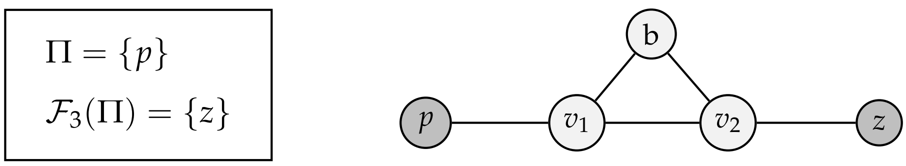

Figure 4: The singleton $\{p\}$ is $\mathcal{F}_3$-closed but not strongly $\mathcal{F}_3$-closed.

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

Proposition 4.7 *(Proposition 1.18, restated). 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 4.3, 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 4.7.1. Π *is not strongly* $\mathcal{F}_\delta$*-closed.*

Proof. Note that $\Delta(\emptyset, \Pi) = \delta - 1 \in (0, \delta)$, hence $\emptyset \notin \Pi \cup \mathcal{F}_\delta(\Pi)$. Relying on Proposition 4.6, 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. $\square$

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

Fact 4.7.2 *(all sets of size at least $2\cdot\delta-1$ are in $\mathcal{F}_\delta(\Pi)$). There exists $z \subseteq [n]$ satisfying $|z| \geq 2\cdot\delta - 1$. 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 $\delta - 1$. $\square$

Fact 4.7.3 *(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 $\Delta(z, \Pi) \geq \delta$. $\square$

It is thus left to show that Π is $\mathcal{F}_\delta$-closed. To do this we rely on Condition (2) from Theorem 3.2: 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 4.7.2 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 4.7.2, 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 4.7.3 (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 4.7.3, 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. ■

4.1.3 Detour: $\mathcal{F}_\delta$-tight spaces

In Figure 4 and Proposition 4.7, we presented two graphs (and values of δ) for which being strongly $\mathcal{F}_\delta$-closed is not a necessary condition for being $\mathcal{F}_\delta$-closed. 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 4.8 *($\mathcal{F}_\delta$-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 section we present an initial exploration of this notion.

First, observe that *every* graph is $\mathcal{F}_1$-tight: This is true since every set in a graphical space is strongly $\mathcal{F}_1$-closed (since for $\delta = 1$, the condition in Definition 4.4 holds vacuously). Thus, all sets in graphical spaces are both $\mathcal{F}_1$-closed and strongly $\mathcal{F}_1$-closed. The following proposition states that every graph is also $\mathcal{F}_2$-tight.

Proposition 4.9 *(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 4.4, 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 4.1, 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 4. Nevertheless, the following proposition asserts that every graphical space is $\mathcal{F}_\delta$-tight for values of δ that are larger than the diameter of the graph.

Proposition 4.10 *(graphs with diameter d are $\mathcal{F}_\delta$-tight for every $\delta > d$). Let Ω be a graphical space with diameter d. Then, for every $\delta > d$ it holds that Ω is $\mathcal{F}_\delta$-tight.*

Proof. Observe that for $\delta > d$, any $\Pi \subseteq \Omega$ satisfies $\mathcal{F}_\delta(\Pi) = \emptyset$ and $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) = \Omega$. Thus, the only $\mathcal{F}_\delta$-closed set is $\Pi = \Omega$, and this set is also strongly $\mathcal{F}_\delta$-closed. ■

Overall, we showed that every graph is $\mathcal{F}_1$-tight and $\mathcal{F}_2$-tight, but not necessarily $\mathcal{F}_3$-tight; and that every graph is $\mathcal{F}_\delta$-tight for values of δ that are larger than the diameter of the graph. A consequent question is therefore:

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

Indeed, this seems to be an interesting combinatorial question. We pose it as an open question in Section 6, and as an initial step towards tackling it, we show several simple graph families that are $\mathcal{F}_\delta$-tight for *every* $\delta > 0$.

Proposition 4.11 *(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 $n \geq 2$.*

The proof of Proposition 4.11 appears in Appendix C. Following Item (4), a natural question is whether the $n \times n$ grid is also $\mathcal{F}_\delta$-tight for every $\delta > 0$.

4.1.4 The values of δ for which a set is $\mathcal{F}_\delta$-closed

For a fixed set $\Pi \subseteq \Omega$, what are the values of δ for which Π is strongly $\mathcal{F}_\delta$-closed, or just $\mathcal{F}_\delta$-closed? The following proposition shows that for any set Π in a graphical space with bounded diameter, the values of δ for which Π is $\mathcal{F}_\delta$-closed constitute a single bounded interval; ditto for values of δ for which Π is strongly $\mathcal{F}_\delta$-closed.

Proposition 4.12 *(Proposition 1.19, restated). For a graphical Ω with bounded diameter and a non-trivial $\Pi \subseteq \Omega$, there exist two integers $\delta^{\mathsf{C}}(\Pi)$ and $\delta^{\mathsf{SC}}(\Pi)$ such that $\delta^{\mathsf{SC}}(\Pi) \leq \delta^{\mathsf{C}}(\Pi)$ and for every integer $\delta > 0$ it holds that*

1. *Π is $\mathcal{F}_\delta$-closed if and only if $\delta \in [1, \delta^{\mathsf{C}}(\Pi)]$.*
2. *Π is strongly $\mathcal{F}_\delta$-closed if and only if $\delta \in [1, \delta^{\mathsf{SC}}(\Pi)]$.*

Proof. Let $\Pi \subseteq \Omega$ such that $\Pi \notin \{\emptyset, \Omega\}$. The proposition will essentially follow from the following claim:

Claim 4.12.1. *For $\delta > 1$, if Π is $\mathcal{F}_\delta$-closed (resp., strongly $\mathcal{F}_\delta$-closed), then Π is $\mathcal{F}_{\delta-1}$-closed (resp., strongly $\mathcal{F}_{\delta-1}$-closed).*

Proof. We first prove the statement regarding $\mathcal{F}_\delta$-closed sets, and then prove the statement regarding strongly $\mathcal{F}_\delta$-closed sets in a similar way.

Assuming that Π is $\mathcal{F}_\delta$-closed, we rely on Condition (2) from Theorem 3.2, and show that for every $x \notin \Pi \cup \mathcal{F}_{\delta-1}(\Pi)$ there exists $z \in \mathcal{F}_{\delta-1}(\Pi)$ such that $\Delta(x, z) \leq \delta - 2$. If $\Omega = \Pi \cup \mathcal{F}_{\delta-1}(\Pi)$ then the claim vacuously holds. Otherwise, let $x \notin \Pi \cup \mathcal{F}_{\delta-1}(\Pi)$. Since $\mathcal{F}_\delta(\Pi) \subseteq \mathcal{F}_{\delta-1}(\Pi)$ it follows that $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$. Since Π is $\mathcal{F}_\delta$-closed, and relying on Condition (2) of Theorem 3.2 again, there exists $z' \in \mathcal{F}_\delta(\Pi)$ such that $\Delta(x, z') \leq \delta - 1$. Let $x = x_0, x_1, ..., x_{k-1}, x_k = z'$ be a path of length $k \leq \delta - 1$ from x to z'. Since $z' \in \mathcal{F}_\delta(\Pi)$ it follows that $\Delta(x_{k-1}, \Pi) \geq \delta - 1$, otherwise $\Delta(z', \Pi) \leq \Delta(z', x_{k-1}) + \Delta(x_{k-1}, \Pi) \leq \delta - 1$. Thus, $x_{k-1} \in \mathcal{F}_{\delta-1}(\Pi)$ and $\Delta(x, x_{k-1}) \leq k - 1 \leq \delta - 2$.

To prove the statement regarding strongly $\mathcal{F}_\delta$-closed sets, we rely on Proposition 4.5. Assuming that Π is strongly $\mathcal{F}_\delta$-closed, for $x \notin \Pi \cup \mathcal{F}_{\delta-1}(\Pi)$ we show $z \in \mathcal{F}_{\delta-1}(\Pi)$ such that $\Delta(x, z) = (\delta - 1) - \Delta(x, \Pi)$. Similar to the previous proof, it holds that $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$, and by Proposition 4.5 there exists a path $x = x_0, x_1, ...x_{k-1}, x_k = z'$ such that $z' \in \mathcal{F}_\delta(\Pi)$ and $k = \delta - \Delta(x, \Pi)$. Since $z' \in \mathcal{F}_\delta(\Pi)$ it follows that $\Delta(x_{k-1}, \Pi) \geq \delta - 1$. Thus, $x_{k-1} \in \mathcal{F}_{\delta-1}(\Pi)$ and $\Delta(x, x_{k-1}) = (\delta - 1) - \Delta(x, \Pi)$. □

It follows that the integer values of δ for which a non-trivial set Π is $\mathcal{F}_\delta$-closed (resp., strongly $\mathcal{F}_\delta$-closed) constitute a continuous interval. To see that the interval for which Π is $\mathcal{F}_\delta$-closed is upper-bounded, note that for any δ larger than the diameter of Ω, which is upper-bounded according to the hypothesis, it holds that $\mathcal{F}_\delta(\Pi) = \emptyset$, and thus $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) = \Omega \neq \Pi$ and Π is not $\mathcal{F}_\delta$-closed. Moreover, since for any $\delta > 0$, if Π is strongly $\mathcal{F}_\delta$-closed then Π is $\mathcal{F}_\delta$-closed, we get that the interval for which Π is strongly $\mathcal{F}_\delta$-closed is also upper-bounded, and that $\delta^{\mathsf{SC}}(\Pi) \leq \delta^{\mathsf{C}}(\Pi)$. To see that both intervals are lower bounded by 1, note that every set is strongly $\mathcal{F}_1$-closed, since the condition in Definition 4.4 holds vacuously. ■

The following proposition shows that a statement analogous to Item (1) in Proposition 4.12 does *not* hold in general metric spaces.

Proposition 4.13 *(a statement analogous to Proposition 4.12 does not hold in general metric spaces). There exists a non-graphical metric space Ω and a set $\Pi \subseteq \Omega$ such that the values of δ for which Π is $\mathcal{F}_\delta$-closed in Ω* do not *lie in a single interval.*

Proof. Let $\Omega = \{0, 1, 3\}$ with the standard metric of $\mathbb{R}$, and let Π be the singleton $\{0\}$. Then:

- For $\delta = 1$ it holds that $\mathcal{F}_1(\{0\}) = \{1, 3\}$ and $\mathcal{F}_1(\mathcal{F}_1(\{0\})) = \{0\}$, and thus $\{0\}$ is $\mathcal{F}_\delta$-closed.
- For $\delta = 2$ it holds that $\mathcal{F}_2(\{0\}) = \{3\}$ and $\mathcal{F}_2(\mathcal{F}_2(\{0\})) = \{0, 1\}$, and thus $\{0\}$ is not $\mathcal{F}_2$-closed.
- For $\delta = 3$ it holds that $\mathcal{F}_3(\{0\}) = \{3\}$ and $\mathcal{F}_3(\mathcal{F}_3(\{0\})) = \{0\}$, and thus $\{0\}$ is $\mathcal{F}_3$-closed. ■

The counter-example in the proof of Proposition 4.13 is indeed quite artificial. Note that the proof of Proposition 4.13 demonstrates that, for a fixed $\Pi \subseteq \Omega$, the operator $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ is not necessarily monotone with respect to δ.

4.2 The Boolean hypercube and list-decodable codes

In the current section we focus solely on the n-dimensional Boolean hypercube $\Omega = H_n$, and continue studying the question from Section 4.1.4: For every fixed set $\Pi \subseteq H_n$, we want to find the values of δ for which Π is strongly $\mathcal{F}_\delta$-closed, or just $\mathcal{F}_\delta$-closed. In particular, for every fixed $\Pi \subseteq H_n$, we will rely on coding-theoretic features of Π (i.e., view Π as an error-correcting code), to obtain a lower bound for $\delta^{\mathtt{SC}}(\Pi)$ and an upper bound for $\delta^{\mathtt{C}}(\Pi)$. We will also show that these bounds are, in general, far from being tight.

4.2.1 Motivation: Two simple observations

We state two simple observations that motivate the use of the coding-theoretic features of a set Π to bound $\delta^{\mathtt{SC}}(\Pi)$ and $\delta^{\mathtt{C}}(\Pi)$. By standard coding theory terminology, the covering radius of a set Π is the minimum $\delta > 0$ such that every $x \in H_n$ satisfies $\Delta(x, \Pi) \leq \delta$. The first observation is that for any non-trivial set Π and δ larger than the covering radius of Π, it holds that $\mathcal{F}_\delta(\Pi) = \emptyset$, which implies that Π is not $\mathcal{F}_\delta$-closed. Therefore, $\delta^{\mathtt{C}}(\Pi)$ is upper-bounded by the covering radius of Π.

Observation 4.14 *($\delta^{\mathtt{C}}(\Pi)$ is upper-bounded by the covering radius of Π). For any non-trivial $\Pi \subseteq H_n$, let $\delta^{\mathtt{CR}}(\Pi)$ be the* covering radius *of Π; that is, the minimal $\delta \geq 0$ such that every $x \in H_n$ satisfies $\Delta(x, \Pi) \leq \delta$. Then, $\delta^{\mathtt{C}}(\Pi) < \delta^{\mathtt{CR}}(\Pi)$.*

Another standard term from coding theory is the unique decoding distance of a set Π, that is $d = \frac{1}{2} \cdot \min_{x \neq y \in \Pi}\{\Delta(x, y)\}$. Then, the second simple observation is the following:

Proposition 4.15 *($\delta^{\mathtt{SC}}(\Pi)$ is lower-bounded by the unique decoding distance of Π). For any non-trivial $\Pi \subseteq H_n$ such that $|\Pi| \geq 2$, let $d = \frac{1}{2} \cdot \min_{x \neq y \in \Pi}\{\Delta(x, y)\}$ be the* unique decoding distance *of Π. Then, $\delta^{\mathtt{SC}}(\Pi) \geq d$.*

Proof. We prove that Π is strongly $\mathcal{F}_d$-closed, relying on Proposition 4.6: For every $x \notin \Pi \cup \mathcal{F}_d(\Pi)$, we show a neighbor x' of x such that $\Delta(x', \Pi) = \Delta(x, \Pi) + 1$. Let $x \notin \Pi \cup \mathcal{F}_d(\Pi)$, and note that it is in the $(d-1)$-neighborhood of exactly one $p \in \Pi$. By flipping a bit $i \in [n]$ such that $x_i = p_i$, we obtain a neighbor x' of x such that either $x' \in \mathcal{F}_d(\Pi)$ (and $\Delta(x, \Pi) = d - 1$), or x' is still in the $(d-1)$-neighborhood of p, in which case $\Delta(x', \Pi) = \Delta(x', p) = \Delta(x, p) + 1$. Either way, x' is farther from Π compared to x. ■

4.2.2 List-decodable codes

In this section we show a lower bound on $\delta^{\mathsf{SC}}(\Pi)$ that is potentially larger than the one shown in Proposition 4.15. Loosely speaking, it is intuitive to expect that if the set Π is very sparse in a neighborhood of x, then we can find a neighbor x' of x that is farther from Π. Accordingly, we expect that if Π is sparse in *every* neighborhood of Ω, then it will be strongly $\mathcal{F}_\delta$-closed. Such "locally sparse" sets are known in coding theory as *list-decodable codes*.[6]

Definition 4.16 *(list-decodable codes). For a non-empty $\Pi \subseteq H_n$ and $\delta, L \in \mathbb{N}$, we say that Π is (δ, L)-*list-decodable *if for every $x \in H_n$ it holds that $|\Pi \cap B[x, \delta]| \leq L$, where $B[x, \delta]$ is the closed Hamming ball of radius δ around x. The number δ is referred to as the* decoding radius*, whereas L is referred to as the* list size.

We now show that for any set Π and $\delta > 0$, if Π is $\left(\delta, \frac{n}{\delta} - 1\right)$-list-decodable, then it is strongly $\mathcal{F}_\delta$-closed. It follows that the maximal $\delta > 0$ such that Π is $(\delta, \frac{n}{\delta} - 1)$-list-decodable lower bounds $\delta^{\mathsf{SC}}(\Pi)$.

Proposition 4.17 *(Proposition 1.20, extended). For any non-empty $\Pi \subseteq H_n$, let $\delta^{\mathsf{LD}}(\Pi)$ be the maximal $\delta \in [n]$ such that Π is $\left(\delta, \frac{n}{\delta} - 1\right)$-list-decodable. If no such $\delta \in [n]$ exists, let $\delta^{\mathsf{LD}}(\Pi) = 0$. Then, $\delta^{\mathsf{SC}}(\Pi) \geq \delta^{\mathsf{LD}}(\Pi)$.*

Two preliminary comments are in order. First, note that if the unique decoding distance of Π is $d \leq \frac{n}{2}$, then Π is $(d, \frac{n}{d} - 1)$-list-decodable. In this case, $\delta^{\mathsf{LD}}(\Pi)$ is a potentially larger lower bound on $\delta^{\mathsf{SC}}(\Pi)$ than d. Second, note that δ^{LD} is not a standard quantity: In a typical setting, one usually fixes a target list size, and is interested in the maximal decoding radius, for that list size.[7] In contrast, in the definition of $\delta^{\mathsf{LD}}(\Pi)$, the allowed list size decreases as the decoding radius increases.

Proof of Proposition 4.17. For a set $\Pi \subseteq H_n$ and $\delta > 0$ such that Π is $\left(\delta, \frac{n}{\delta} - 1\right)$-list-decodable, we show that Π is strongly $\mathcal{F}_\delta$-closed. Relying on Proposition 4.6, for $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$, we need to show a neighbor x' of x such that $\Delta(x', \Pi) = \Delta(x, \Pi) + 1$.

High-level overview. We will prove that there exists a coordinate $i \in [n]$ such that all vertices $p \in \Pi$ satisfying $\Delta(p, x) \leq \Delta(x, \Pi) + 1$ also satisfy $p_i = x_i$. Thus, by flipping the i^{th} bit of x, we obtain a neighbor x' of x such that for every $p \in \Pi$ it holds that $\Delta(x', p) \geq \Delta(x, \Pi) + 1$. This is true since, if $\Delta(x, p) \leq \Delta(x, \Pi) + 1$, then x' is farther from p than x (because $x'_i \neq p_i$, whereas $x_i = p_i$), and thus $\Delta(x', p) \geq \Delta(x, p) + 1 \geq \Delta(x, \Pi) + 1$. On the

[6] While many texts define list-decodability using relative distance (see, e.g., [Vad12]), for coherency with the rest of the current text we use the notion of absolute distance.

[7] A typical setting of parameters in the study of list-decodable codes (at least within the TCS community) would allow for a list size of $\mathrm{poly}(n)$.

other hand, if $\Delta(x,p) \geq \Delta(x,\Pi)+2$, then, since x' cannot be closer to p by more than one unit, compared to x, we get that $\Delta(x',p) \geq \Delta(x,p)-1 \geq \Delta(x,\Pi)+1$.

The actual proof. Denote by Π_x the set of vertices in Π whose distance from x is either $\Delta(x,\Pi)$ or $\Delta(x,\Pi)+1$; that is, $\Pi_x = \{p \in \Pi : \Delta(x,p) = \Delta(x,\Pi) \vee \Delta(x,p) = \Delta(x,\Pi)+1\}$. Similar to previous proofs, we identify every $v \in \{0,1\}^n$ with the corresponding subset of $[n]$ (i.e., $i \in [n]$ such that $v_i = 1$). In addition, for any set $S \subseteq H_n$, let $\mathtt{sd}(x,S) = \bigcup_{s\in S} \mathtt{sd}(x,s)$.

We first prove that $|\mathtt{sd}(x,\Pi_x)| \leq n-1$, which implies that there exists $i \in [n]$ such that for every $p \in \Pi_x$ it holds that $p_i = x_i$. Since the distance of any $p \in \Pi$ from x is at least $\Delta(x,\Pi)$, it holds that $\Pi_x = B[x,\Delta(x,\Pi)+1] \cap \Pi$. Since $\Delta(x,\Pi) \leq \delta-1$ (because $x \notin \mathcal{F}_\delta(\Pi)$), it holds that $B[x,\Delta(x,\Pi)+1] \subseteq B[x,\delta]$, and thus

$$\Pi_x \subseteq B[x,\delta] \cap \Pi\,. \tag{4.1}$$

By our hypothesis, it holds that $|B[x,\delta]\cap\Pi| \leq \left(\frac{n}{\delta}-1\right)$. Also, for every $z \in B[x,\delta]$ it holds that $|\mathtt{sd}(x,z)| = \Delta(x,z) \leq \delta$. Combining these facts, and relying on Eq. (4.1), we get that

$$\begin{aligned}
|\mathtt{sd}(x,\Pi_x)| &\leq |\mathtt{sd}(x,B[x,\delta]\cap\Pi)| \\
&\leq \left(\frac{n}{\delta}-1\right) \cdot \max_{z\in B[x,\delta]\cap\Pi}\{|\mathtt{sd}(z,x)|\} \\
&\leq \left(\frac{n}{\delta}-1\right) \cdot \delta \\
&\leq n-1\,.
\end{aligned}$$

Thus, there exists $i \in [n]$ such that for every $p \in \Pi_x$ it holds that $x_i = p_i$. By flipping this coordinate in x we obtain x' such that the following hold:

- For every $p \in \Pi_x$ it holds that $x_i = p_i$, whereas $x'_i \neq p_i$. Therefore, $\Delta(x',p) = \Delta(x,p)+1$. Since $\Delta(x,\Pi) \leq \Delta(x,p)$, we get that $\Delta(x',p) \geq \Delta(x,\Pi)+1$.
- For every $p \in \Pi \setminus \Pi_x$ it holds that $\Delta(x,p) \geq \Delta(x,\Pi)+2$. Relying on the triangle inequality, we get that $\Delta(x,p) \leq \Delta(x',p)+1$, which implies that $\Delta(x',p) \geq \Delta(x,p) - 1 \geq \Delta(x,\Pi)+1$.

Therefore, the distance of x' from every $p \in \Pi$ is at least $\Delta(x,\Pi)+1$. ■

It is natural to ask whether the requirement on the list size (of $\frac{n}{\delta}-1$) in Proposition 4.17 can be relaxed. The following proposition states that the list size condition is tight up to a constant multiplicative factor with respect to the conclusion that the set is strongly $\mathcal{F}_\delta$-closed, and tight up to a linear additive term (in n) with respect to the conclusion that the set is $\mathcal{F}_\delta$-closed. Actually, we show that there exist relatively small sets that are not strongly $\mathcal{F}_\delta$-closed (resp., $\mathcal{F}_\delta$-closed), while noting that every set of size k is (δ,k)-list-decodable for every $\delta > 0$.

Proposition 4.18 *(on the tightness of the list size in the condition of Proposition 4.17).*

1. *(tightness with respect to being strongly $\mathcal{F}_\delta$-closed). For every $n \geq 9$ and $1 \leq \delta \leq n/2$ such that $\delta-1$ divides n, there exists a set of cardinality $\frac{n}{\delta-1}$ that is not strongly $\mathcal{F}_\delta$-closed.*

2. *(tightness with respect to being $\mathcal{F}_\delta$-closed). For every $n \geq 3$ and $2 \leq \delta \leq n$, there exists a set of cardinality $n - \delta + 2$ that is not $\mathcal{F}_\delta$-closed.*

Proof. In this proof we again identify every $v \in \{0,1\}^n$ with the corresponding subset of $[n]$ (i.e., $i \in [n]$ such that $v_i = 1$). For the first statement, we can use the construction from the proof of Proposition 4.7. In particular, the set Π is a collection of $\frac{n}{\delta-1}$ sets that form an equipartition of $[n]$. In the proof of Proposition 4.7 we showed that such a set is not strongly $\mathcal{F}_\delta$-closed.

For the second statement, we use a variation of the construction in the proof of Proposition 4.3. Let δ be as in the statement, and let

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

To show that Π is not $\mathcal{F}_\delta$-closed, we rely on Condition (2) from Theorem 3.2: In particular, since $\delta \geq 2$, it holds that $\emptyset \notin \Pi \cup \mathcal{F}_\delta(\Pi)$, and we show that there does not exist $z \in \mathcal{F}_\delta(\Pi)$ such that $\Delta(\emptyset, z) \leq \delta - 1$. Let z be such that $\Delta(z, \emptyset) \leq \delta - 1$, implying that $|z| \leq \delta - 1$.

- If $|z| \leq \delta - 2$, then we can remove all its elements, and add the element 1, to obtain the set $\{1\} \in \Pi$. Thus, $\Delta(z, \Pi) \leq \Delta(z, \{1\}) \leq |z| + 1 \leq \delta - 1$, which implies that $z \notin \mathcal{F}_\delta(\Pi)$.
- If $|z| = \delta - 1$, since $|\bigcup_{p\in\Pi} p| = n - \delta + 2$ and z contains $\delta - 1$ elements from $[n]$, it follows that z intersects the set $\bigcup_{p\in\Pi} p$. Thus, $z \cap p = p$ for some $p \in \Pi$, implying that we can remove all the other elements from z to obtain $p \in \Pi$. Therefore $\Delta(z, \Pi) \leq \Delta(x, p) \leq \delta - 2$, which implies that $z \notin \mathcal{F}_\delta(\Pi)$. ■

4.2.3 The non-tightness of the bounds for δ^{SC} and δ^{C}

For any non-trivial $\Pi \subseteq H_n$, recall that Observation 4.14 implies that $\delta^{\mathrm{C}}(\Pi) < \delta^{\mathrm{CR}}(\Pi)$, whereas Proposition 4.17 implies that $\delta^{\mathrm{SC}}(\Pi) \geq \delta^{\mathrm{LD}}(\Pi)$. By combining these bounds with the fact that $\delta^{\mathrm{SC}}(\Pi) \leq \delta^{\mathrm{C}}(\Pi)$, and with the fact that both $\delta^{\mathrm{LD}}(\Pi)$ and $\delta^{\mathrm{CR}}(\Pi)$ are values in the interval $[0, n]$, we get the following bounds on δ^{SC} and on δ^{C}:

$$0 \leq \delta^{\mathrm{LD}}(\Pi) \leq \delta^{\mathrm{SC}}(\Pi) \leq \delta^{\mathrm{C}}(\Pi) < \delta^{\mathrm{CR}}(\Pi) \leq n . \tag{4.2}$$

In particular, Eq. (4.2) implies the non-obvious fact that $\delta^{\mathrm{CR}}(\Pi) > \delta^{\mathrm{LD}}(\Pi)$.

The following proposition demonstrates that the bounds that δ^{LD} and δ^{CR} yield for δ^{SC} and δ^{C}, respectively, are, in general, far from being tight. In particular, the proposition asserts the existence of two sets, Π and Π', such that $\delta^{\mathrm{LD}}(\Pi) = \delta^{\mathrm{LD}}(\Pi') = 0$ (i.e., δ^{LD} is the lowest possible bound for both sets) and $\delta^{\mathrm{CR}}(\Pi) = \delta^{\mathrm{CR}}(\Pi') = n - 1$ (i.e., δ^{CR} is almost the highest possible bound for both sets), but Π and Π' vastly differ with respect to the values of $\delta > 0$ for which they are $\mathcal{F}_\delta$-closed.

Proposition 4.19 *(non-tightness of the bounds that δ^{LD} and of δ^{CR} yield for δ^{SC} and δ^{C}, respectively). For every $n \geq 2$, there exist two sets $\Pi, \Pi' \subseteq H_n$, such that $\delta^{\mathrm{LD}}(\Pi) = \delta^{\mathrm{LD}}(\Pi') = 0$ (i.e., both are not $(1, n-1)$-list-decodable), and $\delta^{\mathrm{CR}}(\Pi) = \delta^{\mathrm{CR}}(\Pi') = n - 1$, but:*

1. Π *is strongly* $\mathcal{F}_\delta$*-closed for every* $\delta \in [n-1]$.

2. Π' *is not* $\mathcal{F}_\delta$*-closed for every* $\delta \geq 2$.

Proof. Recall that for $x \in H_n$, we denote by $\|x\|_1$ the Hamming weight of x. Let $\Pi = \{p : \|p\|_1 \leq 1\}$; that is, Π is the set of strings with Hamming weight 0 or 1. For $o = (0, ..., 0)$ (i.e., $\|o\|_1 = 0$), let $\Pi' = \Pi \setminus \{o\}$; that is, Π' is the set of strings with Hamming weight 1.

To see that $\delta^{\mathrm{LD}}(\Pi) = \delta^{\mathrm{LD}}(\Pi') = 0$, note that in both cases, the radius-1 ball around the origin o contains at least n points from the set. Thus, both sets are not $(1, n-1)$-list-decodable. To see that $\delta^{\mathrm{CR}}(\Pi) = \delta^{\mathrm{CR}}(\Pi') = n-1$, note that every x such that $\|x\|_1 \geq 1$ satisfies $\Delta(x, \Pi) = \Delta(x, \Pi') = \|x\|_1 - 1 \leq n-1$, whereas for $z = (1, ..., 1)$ it holds that $\Delta(z, \Pi) = \Delta(z, \Pi') = n-1$.

To prove Item (1), we rely on Proposition 4.6: For $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$, we show a neighbor x' of x such that $\Delta(x', \Pi) = \Delta(x, \Pi) + 1$. In particular, let $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$, and note that any such x satisfies $\|x\|_1 \in [2, \delta] \subseteq [2, n-1]$. Let $i \in [n]$ such that $x_i = 0$. By flipping the i^{th} bit in x, we obtain x' such that $\Delta(x', \Pi) = \|x'\|_1 - 1 = \|x\|_1 = \Delta(x, \Pi) + 1$. To prove Item (2), note that every path from $o \notin \Pi' \cup \mathcal{F}_\delta(\Pi')$ to any other vertex, and in particular to $\mathcal{F}_\delta(\Pi')$, passes through some $p \in \Pi'$. Relying on Proposition 4.1, it follows that Π' is not $\mathcal{F}_\delta$-closed for any $\delta \geq 2$. ■

5 Applications for dual problems in property testing

In this section we apply the techniques for identifying $\mathcal{F}_\delta$-closed sets to study *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 $\mathcal{F}_{\epsilon \cdot n}(\Pi)$. Recall that we are interested either in an upper bound on the asymptotic query complexity (as a function of n) for *every* constant $\epsilon > 0$, or in a lower bound for *some* constant $\epsilon > 0$. Thus, for a property $\Pi = \{\Pi_n\}_{n \in \mathbb{N}}$, we usually refer to the *dual problem of the problem of testing* Π, or in short to *the dual problem of* Π, without specifying a parameter $\epsilon > 0$.

Definition 5.1 *(Definition 1.3, restated). 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.

In Section 5.1 we state and prove general results regarding the query complexity of dual problems. In Sections 5.2 – 5.6 we study several classes of natural dual problems: We identify dual problems that are equivalent to the original problems as well as dual problems that are different from their original problems, and prove bounds on their query complexity.

5.1 General results regarding the query complexity of dual problems

The following proposition holds for any dual problem, regardless of whether it is equivalent to its original problem or not. Towards its statement we extend Definition 2.1, by defining two special types of testers:

Definition 5.2 *(extending Definition 2.1 for testers with one-sided error and for testers with perfect soundness). 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.

While the first notion (i.e., one-sided error) is a standard notion in property testing, the second notion (i.e., perfect soundness) is not standard, and we introduce it mainly as an auxiliary notion. The *query complexity* of ϵ-testing Π with one-sided error (or with perfect soundness) is defined in the straightforward way.

Proposition 5.3 *(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).*

Proof. For $\Pi \subseteq \Sigma^n$ and $\epsilon > 0$, let T be an ϵ-tester for $\mathcal{F}_{\epsilon \cdot n}(\Pi)$. Then, T accepts every $x \in \mathcal{F}_{\epsilon \cdot n}(\Pi)$, with high probability, and rejects every $x \in \mathcal{F}_{\epsilon \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi))$, with high probability. By Fact 3.1, it holds that $\Pi \subseteq \mathcal{F}_{\epsilon \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi))$. Hence, the tester T', obtained by complementing the output of T, accepts every $x \in \mathcal{F}_{\epsilon \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi)) \supseteq \Pi$, with high probability, and rejects every $x \in \mathcal{F}_{\epsilon \cdot n}(\Pi)$, with high probability. Thus, T' is an ϵ-tester for Π. It follows that for every Π and $\epsilon > 0$, the query complexity of ϵ-testing Π is upper-bounded by the query complexity of ϵ-testing $\mathcal{F}_{\epsilon \cdot n}(\Pi)$.

For the "moreover" statement, note that for every $x \in \Sigma^n$, the probability that T accepts (resp., rejects) x equals the probability that T' rejects (resp., accepts) x. Therefore a tester T with one-sided error (resp., with perfect soundness) yields a tester T' with perfect soundness (resp., with one-sided error). ■

The proof of Proposition 5.3 relied on the fact that an ϵ-tester for $\mathcal{F}_{\epsilon \cdot n}(\Pi)$ always yields an ϵ-tester for Π. The converse statement, however, is not true.

Observation 5.4 *(ϵ-testers for Π do not necessarily yield testers for $\mathcal{F}_{\epsilon \cdot n}(\Pi)$). Let Σ be a set and $\epsilon > 0$. Then, for every $\Pi \subseteq \Sigma^n$ that is not $\mathcal{F}_{\epsilon \cdot n}$-closed, there exists an ϵ-tester T for Π such that complementing the output of T does not yield an ϵ-tester for $\mathcal{F}_{\epsilon \cdot n}(\Pi)$.*

Proof. Let T be a trivial tester that on input $x \in \Sigma^n$ makes all possible n queries and accepts if and only if $x \in \Pi$, and let T' be the tester that is obtained by complementing the output of T. Since Π is not $\mathcal{F}_{\epsilon \cdot n}$-closed, there exists $y \in \mathcal{F}_{\epsilon \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi)) \setminus \Pi$, whereas T rejects $y \notin \Pi$. Thus, T' accepts y although $y \in \mathcal{F}_{\epsilon \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi))$, implying that T' is not an ϵ-tester for $\mathcal{F}_{\epsilon \cdot n}(\Pi)$. ■

We stress that Observation 5.4 only says that an ϵ-tester for $\mathcal{F}_{\epsilon \cdot n}(\Pi)$ is not necessarily obtained by a specific modification (complementation of the output) to an arbitrary ϵ-tester for Π. In particular, Observation 5.4 does not imply anything about the query complexity of ϵ-testing $\mathcal{F}_{\epsilon \cdot n}(\Pi)$. However, if Π is $\mathcal{F}_{\epsilon \cdot n}$-closed, then the problem of ϵ-testing Π and the problem of ϵ-testing $\mathcal{F}_{\epsilon \cdot n}(\Pi)$ are essentially equivalent.

Observation 5.5 *(problems that are equivalent to their dual problems). If the problem of testing a property is equivalent to its dual problem (according to Definition 5.1), then their query complexities are identical.*

We now show a general lower bound on testing dual problems with *one-sided error*. First, we need the following proposition from our prior work [Tel14, Apdx. A].[8]

Proposition 5.6 *(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 key observation is as follows. If there exists an ϵ-tester with query complexity q for Π, then the tester accepts some input $p \in \Pi$, with positive probability, after making q queries. Fix random coins r such that the tester accepts p when using these coins. Then, whenever the tester uses the random coins r, it will also accept any other input that agrees with p on the relevant q coordinates. Since the tester has perfect soundness, every such input cannot be $(\epsilon \cdot n)$-far from Π. More formally,

Claim 5.6.1. *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 5.6.1,

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

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

[8]The said appendix is unrelated to the rest of [Tel14], and will be omitted from [Tel14] in future versions of it.

By combining Proposition 5.6 and Proposition 5.3 we get the following corollary.

Corollary 5.7 *(Theorem 1.5, restated). 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 $\Omega(n)$.*

It follows that *dual problems* can be tested with one-sided error and query complexity $o(n)$ 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, since for any $\epsilon > 0$ and sufficiently large n it holds that $\mathcal{F}_{\epsilon\cdot n}(\Pi_n) = \emptyset$, and thus the property can be tested without querying the input at all.

5.2 Testing duals of error-correcting codes

In the n-dimensional Boolean 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 $\min_{x,y\in\Pi_n}\{\Delta(x,y)\} \geq \zeta \cdot n$. Proposition 4.15 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. Therefore:

Theorem 5.8 *(Theorem 1.6, restated). For any* error-correcting code with constant relative distance, *the problem of testing the code is equivalent to its dual problem.*

Several fundamental problems in property testing involve testing such codes, and so Theorem 5.8 is particularly appealing for the duals of these problems. For example, the following well-known problems involve testing error-correcting codes:

1. The problem of *linearity testing* [BLR90], which consists of testing whether a function $\varphi : G \to H$, where G and H are groups, is a group homomorphism. The most well-known specific case of linearity testing consists of testing the set of linear functions $\varphi : \{0,1\}^n \to \{0,1\}$, which indeed induces an error-correcting code (i.e., the Hadamard code). For general groups, Guo [Guo15] showed sufficient conditions on G and on H such that the set of homomorphisms $G \to H$ induces an error-correcting code. Theorem 5.8 implies that in these cases, the BLR tester can be used to test whether a function is *far from being a group homomorphism* with $O(1)$ queries (by complementing the tester's output). For results regarding its complexity, see, e.g., [BLR90, BGLR93, BS94, BCH+96, KLX10].

2. The problem of *low-degree testing* [RS96], which consists of testing whether a multivariate function over a finite field is a low-degree polynomial. Theorem 5.8 implies that this problem is equivalent to the problem of testing whether a function is *far from being a low-degree polynomial*. For results regarding its complexity, see, e.g., [AKK+03, KR06, JPRZ09, HSS13, GHS15].

Similarly, the problem of testing whether a Boolean function over $\{0,1\}^\ell$ is *far from being an s-sparse low-degree polynomial* is equivalent to its dual problem, and its query complexity is between $\Omega(s)$ and $O(\text{poly}(s))$ queries (see, e.g., [DLM+07, BO10, Gol10a, DLM+11, BBM12, BK12, Tel14]). For $d \in \mathbb{N}$, the original problem consists of testing whether a function is a degree-d polynomial with s non-zero coefficients. Note that the property of degree-d polynomials with s non-zero coefficients generalizes the property of "k-linearity" (i.e., of linear functions with k non-zero coefficients).

Note that in cases where these problems involve testing Boolean functions over $\{0,1\}^\ell$, the generated error-correcting code is in $\{0,1\}^{2^\ell}$. According to Corollary 5.7, the corresponding dual problems cannot be tested with one-sided error and $o(2^\ell)$ queries.

5.3 Testing functions that are far from monotone

Let $[n]$ be a partially ordered set,[9] 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 $f(x) \leq f(y)$. The problem of *testing monotone functions* was introduced by Goldreich *et al.* [GGL$^+$00], and various versions of it have been studied over the years (see, e.g., [DGL$^+$99, LR01, FLN$^+$02, ACCL07, RRS$^+$12, BCGSM12, CS13a, CS13b, CS14, CST14, CDST15, KMS15]).

Throughout this section, we identify every function $f : [n] \to \Sigma$ with a corresponding string $f \in \Sigma^n$. Recall the following standard definitions from poset theory: 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. The main result that we prove in this section is the following:

Proposition 5.9 *(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.*

In the special case of functions over the domain of the Boolean hypercube $\{0,1\}^\ell$, where $2^\ell = n$, Proposition 5.9 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 $\binom{\ell}{\lfloor \ell/2 \rfloor}$. Thus, 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 Proposition 5.9 as a special case.

Proof of Proposition 5.9. 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 Condition (2) of Theorem 3.2: 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 $\Delta(f, h) < \delta$.

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.

[9] 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 $\leq$.

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 5.9.3).

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 5.9.1). 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 5.9.2) 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 5.9.1. *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 $\mathcal{T}$. □

Corollary 5.9.2. *Let $[n]$, Σ and δ be as in Proposition 5.9. 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 5.9.1, 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. □

Let $\mathcal{C}$ be a collection of disjoint pairs for f, as in Corollary 5.9.2. 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 5.9.3. *For $h : [n] \to \Sigma$, if there exists a collection $\mathcal{C}$ of disjoint violating pairs for h having size ρ, then $\Delta(h, \Pi_n) \geq \rho$.*[10]

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 $\Delta(h, \Pi_n) = \Delta(h, g) \geq |\mathcal{C}|$. □

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

Detour: Boolean functions. 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. Although we are not aware of any implications of this fact with respect to property testing, we find it 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 5.9, but we will use an additional lemma, which is specific for Boolean functions, and was proved in [FLN+02].

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

Proposition 5.10 *(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 prove that Π_n is strongly $\mathcal{F}_\delta$-closed, by relying on Proposition 4.6: For $f \notin \Pi_n \cup \mathcal{F}_\delta(\Pi_n)$ we show a function f' such that $\Delta(f,f') = 1$ and $\Delta(f',\Pi_n) = \Delta(f,\Pi_n)+1$. We will rely on the following lemma.

Lemma 5.10.1 *(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 5.9.3 and Lemma 5.10.1, we get the following corollary:

Corollary 5.10.2. *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 5.10.2, there exists a collection $\mathcal{V}$ of disjoint violating pairs for f, such that $|\mathcal{V}| = \Delta(f,\Pi_n) < \delta$. According to Lemma 5.9.1, 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 5.10.2 again, we get that $\Delta(f',\Pi_n) = \Delta(f,\Pi_n)+1$. ■

Implications on testing. Proposition 5.9 implies the following:

Theorem 5.11 *(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 addition, the proof of Proposition 5.9 shows that for a poset P_n and a range Σ_n as in Theorem 5.11, there always exist functions that are $\Omega(n)$-far from being monotone. Thus, according to Corollary 5.7, *testing the dual problem with one-sided error requires $\Omega(n)$ queries.* Note that in the case of functions over the Boolean hypercube $\{0,1\}^\ell$, where $n = 2^\ell$, this lower bound is $\Omega(2^\ell)$.

We explicitly state lower- and upper-bounds on the query complexity of testing functions that are far from monotone over the *Boolean hypercube* $\{0,1\}^\ell$, relying on known results regarding the standard problem. For Boolean functions, a recent upper bound of $\tilde{O}(\sqrt{\ell})$ was given by by Khot, Minzer, and Safra [KMS15], and a lower bound of $\Omega(\ell^{1/2-o(1)})$ for non-adaptive testers was proved by Chen *et al.* [CDST15]. For functions to a general range Σ, a lower bound of $\Omega\left(\min\left\{|\Sigma|^2, \ell\right\}\right)$ was proved by Blais, Brody, and Matulef [BBM12], and an upper bound of $O(\ell/\epsilon)$ was proved by Chakrabarty and Seshadhri [CS13b]. Results regarding testing functions that are far from monotone over general posets can be derived relying on, e.g., [DGL+99, FLN+02, CS13b, CS14].

5.4 Testing distributions that are far from a known distribution

An important sub-field of property testing is the one of testing *properties of distributions*, initiated by Batu *et al.* [BFR+13] (for recent surveys, see [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 $\|\mathbf{I} - \mathbf{D}\|_1 = \sum_{i \in [n]} |\mathbf{I}(i) - \mathbf{D}(i)| \geq \epsilon$.

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 between them).

We consider the dual problem, in which, for a fixed $\mathbf{D}$, an ϵ-tester needs to distinguish between the case $\mathbf{I} \in \mathcal{F}_\epsilon(\{\mathbf{D}\})$ and the case $\mathbf{I} \in \mathcal{F}_\epsilon(\mathcal{F}_\epsilon(\{\mathbf{D}\}))$. The main question in this section is for which families of distributions $\{\mathbf{D}_n\}_{n \in \mathbb{N}}$, where $\mathbf{D}_n$ is a distribution over $[n]$, the problem of testing the property $\{\{\mathbf{D}_n\}\}_{n \in \mathbb{N}}$ is equivalent to its dual problem. More explicitly, we ask for which families of distributions does it hold 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 5.1).

While in $\mathbb{R}^n$ with the *Euclidean metric*, every singleton is $\mathcal{F}_\delta$-closed for every $\delta > 0$, the following proposition shows that the analogous fact is not true in the simplex with the ℓ_1 norm.

Proposition 5.12 *(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 Condition (2) of Theorem 3.2, 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

$$\begin{aligned} \Delta(\mathbf{Z}, \mathbf{D}_n) &= \mathbf{Z}(1) - \mathbf{D}(1) + \sum_{i=2}^{n} |\mathbf{Z}(i) - \mathbf{D}_n(i)| \\ &\leq \frac{1}{n} + \sum_{i=2}^{n} \mathbf{Z}(i) + \sum_{i=2}^{n} \mathbf{D}_n(i) \\ &< \frac{3}{n} \end{aligned}$$

and thus $\Delta(\mathbf{Z}, \mathbf{D}_n) < \delta$, implying that $\mathbf{Z} \notin \mathcal{F}_\delta(\{\mathbf{D}_n\})$. This completes the proof in the case of $\Delta(\mathbf{Z}, \mathbf{D}_n) > 1 - \frac{1}{n}$. Otherwise, $\mathbf{Z}(1) \leq 1 - \frac{1}{n}$. For this case we use the following fact:

Fact 5.12.1. *For $a, b \in \mathbb{R}^+$ it holds that $b - |b - a| \geq -a$.*

Proof. Relying on the triangle inequality and on the fact that $a, b \geq 0$, we get that

$$|b - a| \leq |b| + |a| = b + a$$

and by rearranging we get that $b - |b - a| \geq -a$. □

Now, 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

$$\begin{aligned} \Delta(\mathbf{Z}, \mathbf{X}) - \Delta(\mathbf{Z}, \mathbf{D}_n) &= \sum_{i=1}^{n} \Big(|\mathbf{Z}(i) - \mathbf{X}(i)| - |\mathbf{Z}(i) - \mathbf{D}_n(i)| \Big) \\ &= \frac{1}{n} + \sum_{i=2}^{n} \left(\mathbf{Z}(i) - \left| \mathbf{Z}(i) - \frac{1}{n(n-1)} \right| \right) \\ &\geq \frac{1}{n} - (n-1) \cdot \frac{1}{n(n-1)} && \text{(by Fact 5.12.1)} \\ &= 0 \,. \end{aligned}$$

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 $\Delta(\mathbf{Z}, \mathbf{X}) - \Delta(\mathbf{Z}, \mathbf{D}_n) < 0$). ■

Nevertheless, the following two propositions show that for many natural distributions, the singleton induced by the fixed distribution is $\mathcal{F}_\delta$-closed for every sufficiently small $\delta > 0$. In these cases, 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 refers to distribution families in which each support element has $\Omega(1)$ probability mass. We start by proving the latter proposition, since the proof is much simpler and both proofs rely on similar ideas.

Proposition 5.13 *(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 $\mathbf{D}_n(i) \geq \rho$ 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 Condition (2) of Theorem 3.2: 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 $\mathbf{X} \neq \mathbf{D}_n$ 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

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

and thus $\mathbf{X}(j) > \mathbf{D}_n(j) - \rho/2 \geq \rho/2$, where the last inequality is by the hypothesis that $\mathbf{D}_n(j) \geq \rho$. Similarly, $\mathbf{X}(i) - \mathbf{D}_n(i) < \rho/2$. Now, note that $\mathbf{D}_n(i) \leq 1 - \rho$: 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 $\mathbf{D}_n(k) \leq 1 - \rho$. 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 following proposition shows an arguably broader class of distributions that induce $\mathcal{F}_\delta$-closed properties. Although the proof is technically more involved, the basic idea is similar to the one in the proof of Proposition 5.13: For $\mathbf{X} \notin \{\mathbf{D}_n\} \cup \mathcal{F}_\delta(\{\mathbf{D}_n\})$, we explicitly construct $\mathbf{Z} \in \mathcal{F}_\delta(\{\mathbf{D}_n\})$ such that $\Delta(\mathbf{X}, \mathbf{Z}) < \delta$, by modifying $\mathbf{X}$ on carefully chosen coordinates.

Proposition 5.14 *(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 $\mathbf{D}_n(i) \leq \frac{\delta}{30}$. We prove that $\Pi = \{\mathbf{D}_n\}$ is $\mathcal{F}_\delta$-closed, relying on Condition (2) of Theorem 3.2: 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 $\texttt{Farther}(i)$ be as large as possible, compared to $\texttt{Change}(i)$.

For the construction itself we will rely on the following lemma, which we prove:

Lemma 5.14.1. *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,* $\Pr_{j \sim \mathbf{X}}[j \in \texttt{LIGHT}] > \frac{1}{2}$.

(The term $\texttt{LIGHT}$ is used since the elements in this set will have upper bounded probabilistic mass; see the exact definition in the actual proof below).

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 $i^{\text{UP}} \in [n]$ such that $\mathbf{Z}_{i^{\text{UP}}} > \mathbf{D}_{i^{\text{UP}}}$. We increase the probabilistic mass of $\mathbf{Z}_{i^{\text{UP}}}$ by Δ, and since after the modification it holds that $\mathbf{Z}_{i^{\text{UP}}} > \mathbf{X}_{i^{\text{UP}}} > \mathbf{D}_{i^{\text{UP}}}$, we get that $\texttt{Farther}(i^{\text{UP}}) = \texttt{Change}(i^{\text{UP}})$. 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,

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

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 5.14. We start by proving the two items of Lemma 5.14.1 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 5.14.2 *(Item 1 in Lemma 5.14.1). 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 5.14.3 *(Item 2 in Lemma 5.14.1). 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

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

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

Proof. According to Claim 5.14.3, and since every $i \in [n]$ satisfies $\mathbf{D}_i \leq \frac{\delta}{30}$, 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

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

We construct S by initiating $S = \emptyset$, and adding elements from $\texttt{LIGHT} \setminus \{i\}$ to S until $\sum_{j \in S} \mathbf{X}_j \geq \frac{1}{3} \cdot \delta$. 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 $\sum_{j \in S} \mathbf{X}_j \geq \frac{1}{3} \cdot \delta$. 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 5.14) 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 $\Delta(\mathbf{X}, \mathbf{D}) \geq \frac{2}{3} \cdot \delta$. In this case $\alpha \geq \frac{2}{3}$, 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 5.14.5. *(construction of the distribution $\mathbf{Z}$ when $\Delta(\mathbf{X}, \mathbf{D}) \geq \frac{2}{3} \cdot \delta$).*

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 $\frac{1}{3} \cdot \delta \leq \sum_{j \in S} \mathbf{X}_j < \frac{1}{2} \cdot \delta$.
   (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 5.14.4, 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 $i^{\text{UP}} = \text{argmax}_{i \in [n]}\{\mathbf{X}_i - \mathbf{D}_i\}$, then $\mathbf{Z}_{i^{\text{UP}}} > \mathbf{X}_{i^{\text{UP}}} > \mathbf{D}_{i^{\text{UP}}}$, implying that $\texttt{Farther}(i^{\text{UP}}) = \texttt{Change}(i^{\text{UP}}) = \Delta$. Furthermore, since for every $j \in S$ it holds that j and $\mathbf{Z}$ satisfy the conditions in Claim 5.14.2, then for every $j \in S$ it holds that $\texttt{Farther}(j) \geq 0$. Thus,

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

and $\texttt{Change}(i^{\text{UP}}) = \Delta \geq \frac{1}{3} \cdot \delta \geq \delta - \Delta(\mathbf{X}, \mathbf{D})$. 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 5.14.6. *(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` $i^{\text{UP}} = \text{argmax}_{i \in [n]}\{\mathbf{X}_i - \mathbf{D}_i\}$ `set` $\mathbf{Z}_{i^{\text{UP}}} = \mathbf{X}_{i^{\text{UP}}} + \Delta$.

3. `(decrease` Δ `mass)`

   (a) `Let` $S = \emptyset$.

   (b) `While` $\sum_{j \in S} \mathbf{X}_j < \Delta$ `do` $S \leftarrow \text{argmax}_{i \in \texttt{LIGHT} \setminus (S \cup \{i^{\text{UP}}\})}\{\mathbf{X}_i\}$.

   (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 5.14.6 satisfies, which we will later rely on.

Claim 5.14.7. *Construction 5.14.6 is well-defined, and it produces a distribution $\mathbf{Z}$ such that:*

1. *For $i^{\text{UP}} \in [n]$ it holds that $\mathbf{Z}_{i^{\text{UP}}} = \mathbf{X}_{i^{\text{UP}}} + \Delta$ and $\mathbf{X}_{i^{\text{UP}}} > \mathbf{D}_{i^{\text{UP}}}$.*

2. *For $S \subseteq \texttt{LIGHT}$ 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 5.14.7, 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 $\mathbf{Z}_{i^{\mathrm{UP}}} > \mathbf{X}_{i^{\mathrm{UP}}} > \mathbf{D}_{i^{\mathrm{UP}}}$, it follows that $\mathtt{Farther}(i^{\mathrm{UP}}) = \mathtt{Change}(i^{\mathrm{UP}})$. Also, since for every $j \in S$ it holds that j and $\mathbf{Z}$ satisfy the conditions in Claim 5.14.2, it follows that $\mathtt{Farther}(j) \geq \frac{1-\alpha}{1+\alpha}\cdot\mathtt{Change}(j)$. Therefore,

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

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 5.14.7.

Proof of Claim 5.14.7. To see that Construction 5.14.6 is well-defined, note that according to Fact 5.14.4 there is sufficient probability mass in $\mathtt{LIGHT}\setminus\{i^{\mathrm{UP}}\}$ in order for the loop in Step (3b) of Construction 5.14.6 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 5.15 *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 $\Delta > \frac{1}{2}\cdot\frac{1}{3} = \frac{1}{6}$. Also, since $S \subseteq \mathtt{LIGHT}$, 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

$$|S| \geq \frac{\Delta}{\max_{i\in S}\{\mathbf{X}_i\}} > \frac{\frac{1}{6}\cdot\delta}{\delta/20} > 3\,.$$

Therefore,

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

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 $\mathbf{Z}_i \geq 0$. □

This completes the proof of Proposition 5.14. ■

Implications on testing. Proposition 5.14 implies the following:

Theorem 5.16 *(Theorem 1.9, restated). Let $\{\mathbf{D}_n\}_{n \in \mathbb{N}}$ be a family of distributions such that $lim_{n \to \infty} \|\mathbf{D}_n\|_\infty = 0$. Then, the problem of testing whether an input distribution $\mathbf{I}_n$ is identical to $\mathbf{D}_n$ is equivalent to its dual problem.*

In particular, the problem of testing whether an input distribution is uniform is equivalent to its dual problem. Also, according to Proposition 5.13, for any distribution $\mathbf{D}$ such that the probabilistic mass of each support element is $\Omega(1)$, the problem of testing whether an input distribution $\mathbf{I}$ is identical to $\mathbf{D}$ is equivalent to its dual problem.

The query complexity of the distribution testing problem is $\tilde{\Theta}(\sqrt{n})$. For the uniform distribution, a lower bound of $\Omega(\sqrt{n})$ is not hard to obtain, by analyzing the collision probability and relying on the birthday paradox (see, e.g., [Can15, Sec 3.2.1]); and an upper bound of $O(\sqrt{n})$ was implicitly proved by Goldreich and Ron in [GR00, GR02]. An upper bound of $\tilde{O}(\sqrt{n})$ for arbitrary distributions was proved by Batu *et al.* [BFF+01], and a fine-grained analysis was recently given by Valiant and Valiant [VV14], who showed tight bounds on the complexity of this problem on a distribution-by-distribution basis.

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 Theorem 5.16 and in Proposition 5.13, the query complexity of the dual problem is $\tilde{O}(\sqrt{n})$, and is also upper bounded by the finer upper bound given by [VV14].

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

Property testing in the *dense graph model* was initiated by Goldreich, Goldwasser, and Ron [GGR98]. 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.

In this model, 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.

Loosely speaking, we show that the following dual problems in the dense graph model are different from their original problems:

- **k-colorability** (cf., [GGR98]): Testing whether a graph is *far from being k-colorable*.
- **ρ-clique** (cf., [GGR98]): For $\rho \in (0,1)$, testing whether a graph on v vertices is *far from having clique of size $\rho \cdot v$*.
- **Isomorphism testing** (cf., [Fis05, FM08]): For a graph G that is explicitly known in advance, testing whether an input graph H is *far from being isomorphic to G*.

Nevertheless, we show that the query complexity of testing whether a graph is *far from being k-colorable* is $O(1)$, where the O-notation hides a huge dependence on the proximity parameter ϵ.

5.5.1 A general result regarding dual problems in the dense graph model

In this section we present a result that can be used to prove that the complexity of some dual problems in the dense graph model is $O(1)$. The following definition is adapted from [FN07], which follows [PRR06].

Definition 5.17 *((α,ϵ)-estimation tester; cf. Definition 2.1, and [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 in the straightforward way, analogously to Definition 2.1. Fischer and Newman [FN07] proved the following result.

Theorem 5.18 *(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 $O(1)$.*

The following is a corollary of Theorem 5.18 that is interesting in the context of dual problems in the dense graph model.

Corollary 5.19 *(a sufficient condition for a dual problem to be testable with $O(1)$ queries). Let $\Pi = \{\Pi_n\}_{n\in\mathcal{N}}$ be a property of graphs in the dense graph model with query complexity $O(1)$. If for every sufficiently small $\epsilon > 0$ there exists $\alpha \in (0,1)$ such that for every sufficiently large n it holds that $\mathcal{F}_{\epsilon\cdot n}(\mathcal{F}_{\epsilon\cdot n}(\Pi_n)) \subseteq \{x : \Delta(x,\Pi_n) \leq (\alpha\cdot\epsilon)\cdot n\}$, then the query complexity of the dual problem of Π is $O(1)$.*

Proof. For any $\epsilon > 0$, let $\alpha \in (0,1)$ such that for a sufficiently large n it holds that $\mathcal{F}_{\epsilon \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n)) \subseteq \{x : \Delta(x, \Pi_n) \leq (\alpha \cdot \epsilon) \cdot n\}$. Since the query complexity of Π is $O(1)$, Theorem 5.18 implies that there exists an (α, ϵ)-estimation tester T for Π with query complexity $O(1)$. The point is that for a sufficiently large n it holds that T accepts, with high probability, every $x \in \Sigma^n$ such that $\Delta(x, \Pi_n) \leq (\alpha \cdot \epsilon) \cdot n$, and rejects, with high probability, every $x \in \mathcal{F}_{\epsilon \cdot n}(\Pi_n)$. Since $\mathcal{F}_{\epsilon \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n)) \subseteq \{x : \Delta(x, \Pi_n) \leq (\alpha \cdot \epsilon) \cdot n\}$, complementing the output of T yields an ϵ-tester for $\mathcal{F}_{\epsilon \cdot n}(\Pi_n)$ with query complexity $O(1)$. ■

Note that the tester for dual problems obtained by using Corollary 5.19 has *two-sided error*, since the estimation tester given by [FN07] has a two-sided error. This two-sided error cannot be eliminated; that is, Corollary 5.19 cannot yield a tester with one-sided error in general. This is the case since there exist dual problems that are not trivial (i.e., such that $\mathcal{F}_\delta(\Pi_n) \neq \emptyset$) to which Corollary 5.19 applies (see, e.g., Proposition 5.21); but, according to Corollary 5.7, testing such problems with one-sided error requires a linear number of queries.

5.5.2 Testing the property of being far from k-colorable in the dense graph model

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 $O(1)$, relying on Corollary 5.19.

Proposition 5.20 *(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 rely on Proposition 4.1, which asserts that if Π_n is $\mathcal{F}_\delta$-closed, then for every $G \notin \Pi_n \cup \mathcal{F}_\delta(\Pi_n)$ there exists 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. In particular, 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. Thus, for any $\delta \geq 2$, there does not exist a path as above from $G \notin \Pi_n \cup \mathcal{F}_\delta(\Pi_n)$ to $\mathcal{F}_\delta(\Pi_n)$, which 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 (i.e., removing an edge from the $(k+1)$-clique) 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 need 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. (This is true since after removing the edge, the vertices of the (former) clique can be colored using k colors; and to extend this coloring to the rest of the graph, note that the additional edge either connects a vertex from the clique and a vertex from outside the clique, or connects two vertices from outside the clique. In both cases, we can extend the k-coloring of the clique to a k-coloring of the rest of the graph.) ■

The query complexity of the dual problem. A tester for the original problem with query complexity $O(1)$ was given by Goldreich, Goldwasser, and Ron [GGR98]. Note that the query complexity of testing whether a graph on v vertices is far from being k-colorable with *one-sided error* is $\Omega(n) = \Omega\left(\binom{v}{2}\right)$. This is true since for every $v \in \mathbb{N}$ and $n = \binom{v}{2}$ there exist graphs over v vertices that are $\Omega(n)$-far from being k-colorable (e.g., the complete graph), and relying on Corollary 5.7.

We now show that *the query complexity of the dual problem is also* $O(1)$. To do this, we rely on Corollary 5.19: This requires proving that for every sufficiently small $\epsilon > 0$ there exists $\alpha \in (0,1)$ such that for every sufficiently large $n \in \mathcal{N}$ it holds that $\mathcal{F}_{\epsilon \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n)) \subseteq \{G : \Delta(G, \Pi_n) \leq (\alpha \cdot \epsilon) \cdot n\}$.

Proposition 5.21 *(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 $\mathcal{F}_{\epsilon \cdot n}(\mathcal{F}_{\epsilon \cdot n}(\Pi_n)) \subseteq \{G : \Delta(G, \Pi_n) \leq (\alpha \cdot \epsilon) \cdot n\}$.*

Note that Corollary 5.19 only requires that for every (sufficiently small) $\epsilon > 0$ there exists $\alpha \in (0,1)$ such that the statement holds, whereas we show that there exists a single $\alpha \in (0,1)$ that suffices for every (sufficiently small) $\epsilon > 0$.

Proof. We start with a high-level overview, and then proceed to the actual proof.

High-level overview. Let $\alpha = 1 - \frac{1}{\binom{k+1}{2}}$. To prove the proposition, we show that for every sufficiently small $\epsilon > 0$, and sufficiently large n, and $\delta = \epsilon \cdot n$, every graph $G \in \{0,1\}^n$ such that $\Delta(G, \Pi_n) > \alpha \cdot \delta$ satisfies $G \notin \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$. Observe that if $\Delta(G, \Pi_n) \geq \delta$, then $G \in \mathcal{F}_\delta(\Pi_n)$, which implies that $G \notin \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$. Thus, it suffices to show that any graph G such that $\Delta(G, \Pi_n) \in (\alpha \cdot \delta, \delta)$ satisfies $G \notin \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$.

To do this, for any graph G such that $\Delta(G, \Pi_n) \in (\alpha \cdot \delta, \delta)$, we construct a graph $H \in \mathcal{F}_\delta(\Pi_n)$ such that $\Delta(G, H) < \delta$, which implies that $G \notin \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$. We first show that for any such graph G, there exists a collection $\mathcal{I}$ of $\delta - \Delta(G, \Pi_n)$ independent sets of size $(k+1)$ in G such that every two independent sets in $\mathcal{I}$ share at most one common vertex (see Lemma 5.21.2). We also show that for every independent set in $\mathcal{I}$, if we add $\binom{k+1}{2}$ edges to it, turning it to a $(k+1)$-clique, we obtain a graph that is farther away from Π_n (see Claim 5.21.3). Accordingly, we change every independent set in $\mathcal{I}$ to a $(k+1)$-clique, obtaining a graph H.

Note that $\Delta(H, \Pi_n) = \Delta(G, \Pi_n) + |\mathcal{I}| = \delta$, and thus $H \in \mathcal{F}_\delta(\Pi_n)$. To see that $\Delta(G, H) < \delta$, note that for every set in $\mathcal{I}$ we added $\binom{k+1}{2}$ edges to G to obtain H. Thus, we get that $\Delta(G, H) = \binom{k+1}{2} \cdot |\mathcal{I}| = \binom{k+1}{2} \cdot (\delta - \Delta(G, \Pi_n))$. Now, by our choice of α, we have $\binom{k+1}{2} = \frac{1}{1-\alpha}$, whereas by the hypothesis regarding G we have $\delta - \Delta(G, \Pi_n) < (1-\alpha)\cdot\delta$. Therefore, it holds that $\Delta(G,H) = \frac{1}{1-\alpha}\cdot(\delta - \Delta(G,\Pi_n)) < \delta$.

The core of the proof is showing that the collection $\mathcal{I}$ exists (i.e., Lemma 5.21.2, which relies on Claim 5.21.1). This is shown as follows. Let G be a graph on v vertices such that $\Delta(G,\Pi_n) \in (\alpha\cdot\delta, \delta)$. Since $\Delta(G,\Pi_n) < \delta$, it follows that for some k-partition of G, there exists a cell U in the partition such that the number of vertices in U is at least v/k and the number of edges between them is at most δ. Since δ is very small, the subgraph induced by the vertices of U is very sparse. Relying on a well-known result of Bollobás [Bol76], we show that such a sparse graph contains $\Omega(n) = \Omega\left(\binom{v}{2}\right)$ independent-sets of size $k+1$, such that each pair of sets share at most one common vertex.

Indeed, for this argument to work we must set $\epsilon > 0$ to be sufficiently small such that $\delta = \epsilon\cdot n$ will satisfy two conditions: First, δ should be sufficiently small in order for U to be sparse enough; and second, the exact number of edge-disjoint cliques, which was hidden in the Ω-notation, should be at least $(1-\alpha)\cdot\delta$.

The actual proof. Throughout the proof, it will be convenient to think of the number of vertices, denoted by v, as the primary asymptotic parameter (recall that $n = \binom{v}{2}$). We need to prove the statement of the proposition for every "sufficiently small" $\epsilon > 0$; to define what "sufficiently small" means, we will need the following claim.

Claim 5.21.1 *(very dense graphs contain $\Omega(n)$ edge-disjoint $(k+1)$-cliques). There exists $\rho \in (0,1)$ such that any graph on v vertices with $\rho\cdot\binom{v}{2}$ edges contains $\Omega\left(\binom{v}{2}\right)$ edge-disjoint $(k+1)$-cliques.*

Proof. The claim follows as a corollary of a well-known theorem by Bollobás, which we now describe. A decomposition of a graph G is a collection of edge-disjoint subgraphs of G such that every edge of G belongs to exactly one subgraph in the collection. Bollobás [Bol76] showed that for every $k \geq 2$, there exists $e(k) \in (0,1)$ such that any graph on v vertices can be decomposed to a collection $\mathcal{C}$ of subgraphs, satisfying:

1. Every subgraph in $\mathcal{C}$ is either a single edge or a clique on $k+1$ vertices.
2. $|\mathcal{C}| \leq e(k)\cdot\binom{v}{2}$. [11]

Let G be a graph on v vertices with m edges. Let $\mathcal{C}$ be the decomposition of G that exists according to the above. Since the edges of $|\mathcal{C}|$ subgraphs cover the m edges of G, and each subgraph is either a single edge or a $(k+1)$-clique, it follows that at least $\frac{m-|\mathcal{C}|}{\binom{k+1}{2}}$ of the subgraphs in $\mathcal{C}$ are $(k+1)$-cliques. Thus, for any constant $\rho > e(k)$, if G contains $m = \rho\cdot\binom{v}{2}$ edges, then it contains $\frac{m-|\mathcal{C}|}{\binom{k+1}{2}} \geq \frac{\rho-e(k)}{\binom{k+1}{2}}\cdot\binom{v}{2} = \Omega\left(\binom{v}{2}\right)$ edge-disjoint $(k+1)$-cliques. □

Now, let $\alpha = 1 - \frac{1}{\binom{k+1}{2}}$. According to Claim 5.21.1, there exist $\rho > 0$ and $\xi > 0$ such that every graph on v/k vertices with $\rho\cdot\binom{v/k}{2}$ edges contains at least $\xi\cdot\binom{v}{2}$ edge-disjoint $(k+1)$-cliques. Let $\epsilon > 0$ be sufficiently small such that for a sufficiently large $v \in \mathbb{N}$ and $n = \binom{v}{2}$ it

[11] Actually, the original result of Bollobás asserts that $|\mathcal{C}| \leq t_k(v)$, where $t_k(v) \leq \frac{1}{2}\left(1-\frac{1}{k}\right)\cdot v^2$ is the $(v,k+1)$-Turán number. For our purposes it will be more convenient to use a fraction of $\binom{v}{2}$.

holds that $\delta = \epsilon \cdot n$ satisfies

$$\delta < \min \left\{ (1-\rho) \cdot \binom{v/k}{2},\ \frac{\xi}{1-\alpha} \cdot \binom{v}{2} \right\} . \tag{5.4}$$

Let $v \in \mathbb{N}$ be sufficiently large, and let $n = \binom{v}{2}$ and $\delta = \epsilon \cdot n$. According to the overview, it suffices to construct, for any graph G with v vertices satisfying $\Delta(G, \Pi_n) \in (\alpha \cdot \delta, \delta)$, a corresponding graph $H \in \mathcal{F}_\delta(\Pi_n)$ such that $\Delta(G, H) < \delta$ (because this implies that $G \notin \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n)))$.

In order to construct H, we first need to define some terminology. For any graph $G = ([v], E)$ and a k-partition P of $[v]$, we call $(u, w) \in [v] \times [v]$ a *violating pair* for P if u and w are adjacent and are in the same cell of the partition. Note that the distance of G from being k-colorable is the minimum, over all k-partitions P of $[v]$, of the number of violating pairs for P. The following lemma establishes the existence of a collection $\mathcal{I}$ of independent sets in G, each of size $k+1$, as in the high-level overview.

Lemma 5.21.2. *Let G be a graph on v vertices satisfying $\Delta(G, \Pi_n) \in (\alpha \cdot \delta, \delta)$. Then, there exists a collection $\mathcal{I}$ of independent sets in G, such that $|\mathcal{I}| = \delta - \Delta(G, \Pi_n)$, each set consists of $k+1$ vertices, and every two independent sets $c_1, c_2 \in \mathcal{C}$ share at most one common vertex.*

Proof. Since $\Delta(G, \Pi_n) < \delta$, there exists a k-partition of the vertices of G with less than δ violating edges. Let U be the cell in the partition with the maximal number of vertices, and note that $|U| \geq v/k$, and that the number of edges with both end-points in U is less than δ. Without loss of generality, assume that $|U| = v/k$ (since we can remove vertices from U, and the number of edges between its vertices will still be less than δ). Relying on Eq. (5.4), the number of edges between the vertices of U is less than $(1-\rho) \cdot \binom{v/k}{2}$.

Let $\overline{G}$ be the complement graph of G, and $\overline{U}$ be the subgraph of $\overline{G}$ induced by the vertices of U. Note that the number of edges between vertices of $\overline{U}$ is more than $\rho \cdot \binom{v/k}{2}$, and thus (by our definition of ρ and Claim 5.21.1) there exist at least $\xi \cdot \binom{v}{2}$ edge-disjoint $(k+1)$-cliques in $\overline{U}$. According to Eq. (5.4), it holds that $\delta < \frac{\xi}{1-\alpha} \cdot \binom{v}{2}$, and hence $\xi \cdot \binom{v}{2} > (1-\alpha) \cdot \delta$. Since $\Delta(G, \Pi_n) > \alpha \cdot \delta$, it holds that $(1-\alpha) \cdot \delta > \delta - \Delta(G, \Pi_n)$.

It follows that there exists a collection of more than $\delta - \Delta(G, \Pi_n)$ independent sets, each of size $k+1$, in U, corresponding to the $(k+1)$-cliques in $\overline{U}$. Since the cliques were edge-disjoint, every two independent sets in the collection share at most one common vertex. $\square$

Let $\mathcal{I}$ be a collection of $\delta - \Delta(G, \Pi_n)$ independent sets in G as in Lemma 5.21.2. We modify G into H by adding, for each independent set in $\mathcal{I}$, edges between all pairs of vertices in the set. For each set in $\mathcal{I}$, we added $\binom{k+1}{2} = \frac{1}{1-\alpha}$ edges to G. Overall, the number of edges we added to G to obtain H is $|\mathcal{I}| \cdot \binom{k+1}{2} = (\delta - \Delta(G, \Pi_n)) \cdot \frac{1}{1-\alpha} < \delta$, where the last inequality relied on the fact that $\delta - \Delta(G, \Pi_n) < (1-\alpha) \cdot \delta$ (because $\Delta(G, \Pi_n) > \alpha \cdot \delta$). Therefore, $\Delta(G, H) < \delta$. To conclude the proof it is left to show that $H \in \mathcal{F}_\delta(\Pi_n)$.

Claim 5.21.3. *For a graph G on v vertices, let I be an independent set of size $k+1$ in G. Let G' be the graph obtained by adding to G all edges connecting pairs of vertices in I (i.e., turning I from an independent set to a clique). Then $\Delta(G', \Pi_n) \geq \Delta(G, \Pi_n) + 1$.*

Proof. For any k-partition P of the vertices of G, the number of violating pairs for P in G' is larger than the number of violating pairs 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 pair for P in G', whereas no edges were removed when modifying G to G' (and thus all violating pairs for P in G are also violating pairs for P in G'). The claim follows. □

To see that $\Delta(H, \Pi_n) \geq \delta$, assume that we sequentially turn each independent set in $\mathcal{I}$ to a $(k+1)$-clique. Since every two independent sets in $\mathcal{I}$ share at most one common vertex, after turning each independent set to a clique, all the remaining sets in $\mathcal{I}$ are still independent sets. Thus, repeatedly invoking Claim 5.21.3 (after turning each independent set to a clique), it holds that $\Delta(H, \Pi_n) \geq \Delta(G, \Pi_n) + |\mathcal{I}| = \delta$. ■

5.5.3 Testing the property of being far from having a large clique in the dense graph model

In this section we study the dual problem of ρ-clique: 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. We show that this problem is different from its original problem.

Proposition 5.22 *(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 $\rho \cdot v$. Similar to the proof of Proposition 5.20, we show that Π is not $\mathcal{F}_\delta$-closed, relying on the necessary condition in Proposition 4.1. In particular, we show a graph G such that $\Delta(G, \Pi) = 1$, and all neighbors of G are either in Π or adjacent to Π. It follows that 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 $\mathcal{F}_\delta(\Pi)$ such that every graph subsequent to G on the path is neither in Π nor adjacent to Π. Relying on Proposition 4.1, this implies 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 $\Delta(G', \Pi) = 1$. ■

Implications on testing. Similar to the problem of testing k-colorability, a tester for the original problem of ρ-clique with query complexity $O(1)$ was given by Goldreich, Goldwasser, and Ron [GGR98]. However, in the case of ρ-clique it is not clear whether this upper bound also holds for the dual problem. Nevertheless, according to Corollary 5.7, 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 with one-sided error requires $\Omega(n)$ queries.

5.5.4 Testing the property of being far from isomorphic to a graph in the dense graph model

The problem of *testing graph isomorphism* was introduced by Fischer [Fis05]. We study the dual problem of a well-known version of this problem: In the dual problem, for a graph G on v vertices that is predetermined and explicitly known in advance, the problem consists of ϵ-testing the set of graphs that are $\left(\epsilon \cdot \binom{v}{2}\right)$*-far from being isomorphic to* G. We show that the dual problem is different from the original problem.

Proposition 5.23 *(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 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. Note that Π_n is exactly the set of vectors with Hamming weight 1, since each of these vectors 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 (since isomorphic copies of a graph have the same number of edges). However, $\Pi_n = B[\emptyset, 1] \setminus \{\emptyset\}$ is a property that we already considered in the proof of Proposition 4.19, where we proved that it is not $\mathcal{F}_2$-closed, relying on Proposition 4.1: We showed that there does not exist a path from $\emptyset \notin \Pi_n \cup \mathcal{F}_\delta(\Pi_n)$ to $\mathcal{F}_\delta(\Pi_n)$. ■

Fischer and Matsliah proved [FM08] that the query complexity of this version of the graph isomorphism is $\tilde{\Theta}(\sqrt{v})$. We deduce that the query complexity of the dual problem is lower bounded by $\Omega(\sqrt{v})$. Also, according to Corollary 5.7, and since the testing problem is not trivial, testing the dual problem with one-sided error requires $\Omega(n)$ queries.

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

Property testing in the *bounded-degree model* was initiated by Goldreich and Ron [GR02]. 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.[12]

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.

Loosely speaking, we show that the following dual problems in the bounded-degree model are different from their original problems:

- **Connectivity**: For any $d \geq 3$, testing whether a graph is *far from being connected*.
- **Cycle-free graphs:** For any $d \geq 3$, testing whether a graph is *far from being cycle-free*.
- **Bipartiteness**: For any $k \geq 2$ and $d \geq k+1$, testing whether a graph is *far from being bipartite*.

Nevertheless, we show that the query complexity of testing whether a graph is *far from being connected* and of testing whether a graph is *far from being cycle-free* is $O(1)$, as is the case for the corresponding original problems.

5.6.1 Testing the property of being far from connected in the bounded-degree model

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 $\mathrm{poly}(1/\epsilon)$, 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 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 $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$.

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

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 5.6.1.1, might be of independent interest. Then, in Section 5.6.1.2, 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 5.6.1.2 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 5.6.1.3 we show that the latter task can be done, using only $O(1)$ queries.[13] This fact relies again on the combinatorial formula from Section 5.6.1.1; 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; that is, $\mathtt{fd}(c) = \sum_{i \in c} \mathtt{fd}(i)$. 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 $C^k = C^k(G)$.

5.6.1.1 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 $d \geq 2$.

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

$$\mathtt{wc}(G) \stackrel{\text{def}}{=} 2 \cdot \left|C^0(G)\right| + \left|C^{2+}(G)\right| - 1 \tag{5.5}$$

[13] 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 5.27 and [MR06, Sec. 2.1]).

be the *weighted count* of connected components in G. We will see (in Lemma 5.26) 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. (5.5) 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 5.24. *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. ∎

Claim 5.25. *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. ∎

We now prove that in the special case where d is even, the combinatorial formula in Eq. (5.5) indeed expresses the distance of a graph from being connected.

Lemma 5.26. *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 $\Delta(G, \Pi_n) \leq \texttt{wc}(G)$: We modify G to a connected graph, by adding and removing at most $\texttt{wc}(G)$ edges. For the modification, we first remove an edge from each connected component $c \in C^0$; according to Claim 5.25, 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 $2 \cdot |C^0| + |C^{2+}| - 1 = \texttt{wc}(G)$ edges in G to obtain a connected graph.

To show that $\Delta(G, \Pi_n) \geq \texttt{wc}(G)$, we fix an arbitrary connected graph $G' \in \Pi_n$, and show that $\Delta(G, G') \geq \texttt{wc}(G)$. Relying on Claim 5.24, 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 $2 \cdot |C^0(G)| + |C^{2+}(G)| - 1 = \texttt{wc}(G)$. Thus, for every $G' \in \Pi_n$ it holds that $\Delta(G, G') \geq \texttt{wc}(G)$, which implies that $\Delta(G, \Pi_n) \geq \texttt{wc}(G)$. ∎

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:

$$\texttt{wc}(G) \stackrel{\text{def}}{=} 2 \cdot |C^0| + \left|C^{1+}\right| - 1 + \max\left\{0, \left|C^{1+}\right| - 1 - \left\lceil \frac{\texttt{fd}(G)}{2} \right\rceil \right\} . \tag{5.6}$$

First observe that when $|C^1| = 0$ (and in particular, when d is even), the formula in Eq. (5.6) agrees with the formula in Eq. (5.5). This is true because in this case $|C^{1+}| = |C^{2+}|$ and the value of the right-most expression in Eq. (5.6) is zero (because $\texttt{fd}(G) \geq 2 \cdot |C^{2+}|$, which implies that $|C^{2+}| - \left\lceil \frac{\texttt{fd}(G)}{2} \right\rceil \leq 0$). The following lemma, which is the main result in this section, asserts that $\texttt{wc}(G)$ equals the distance of G from being connected also in the general case.

Lemma 5.27. *For any $d \geq 2$ and $n \in \mathbb{N}$, every graph G with maximal degree d over $[n]$ that is not connected satisfies $\Delta(G, \Pi_n) = \texttt{wc}(G)$.*

The proof of Lemma 5.27 relies mostly on ideas similar to the ideas in the proof of Lemma 5.26, but it is significantly more involved and tedious (reflecting the more complex expression for $\texttt{wc}(G)$). Readers that are not interested in the technical details can safely skip the proof, and continue reading from Section 5.6.1.2.

Proof of Lemma 5.27. Let us begin with a short overview of the proof. Given a graph $G \notin \Pi_n$, we wish to show that $\Delta(G, \Pi_n) = \texttt{wc}(G)$. To show that $\Delta(G, \Pi_n) \leq \texttt{wc}(G)$, 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 5.26. 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 5.26, 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 5.27.2.

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. (5.6) is equivalent to the following definition:

$$\mathtt{wc}(G) \stackrel{\text{def}}{=} \begin{cases} 2 \cdot |C^0| + |C^{1+}| - 1 & \mathtt{aux}(G) \leq 0 \\ 2 \cdot (|C| - 1) - \left\lceil \frac{\mathtt{fd}(G)}{2} \right\rceil & \mathtt{aux}(G) > 0 \end{cases} . \tag{5.7}$$

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 5.25.)
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., $\texttt{fd}(c_0) = 0$), 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 5.25, and since $\texttt{fd}(c_0) = 0$ 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 $\texttt{wc}(G)$.

Claim 5.27.1. *On any input graph $G \notin \Pi_n$, Algorithm 1 makes $\texttt{wc}(G)$ edge modifications to G.*

Proof. First note that in Step (1) we remove $|C^0|$ edges, whereas in Step (2) we add $|C^0| + |C^{2+}| - 1$ 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 $\texttt{fd}^{(St3)}(c_0)$. In the beginning of Step (2), the vertices in c_0 had $\sum_{c \in C^{2+}} \texttt{fd}(c) + 2 \cdot |C^0|$ free degrees; and during Step (2) we added $|C^0| + |C^{2+}| - 1$ 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

$$\begin{aligned}\texttt{fd}^{(St3)}(c_0) &= \sum_{c \in C^{2+}} \texttt{fd}(c) + 2 \cdot |C^0| - 2 \cdot \left(|C^0| + |C^{2+}| - 1\right) \\ &= \texttt{fd}(G) - \left|C^1\right| - 2 \cdot |C^{2+}| + 2\,. \qquad (5.8)\end{aligned}$$

If $\texttt{fd}^{(St3)}(c_0) \geq |C^1|$, then the loop in Step (3) will end when the graph is connected; and otherwise, the loop will end after $\texttt{fd}^{(St3)}(c_0)$ 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. (5.8), it follows that:

$$\left\lceil \frac{|S|}{2} \right\rceil = \left\lceil \frac{\left|C^1\right| - \mathtt{fd}^{(St3)}(c_0)}{2} \right\rceil$$

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

$$= \left|C^{1+}\right| - 1 - \left\lceil \frac{\mathtt{fd}(G)}{2} \right\rceil \tag{5.9}$$

$$= \mathtt{aux}(G)\ . \tag{5.10}$$

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. (5.10)) 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. (5.7), 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

$$2\cdot\left|C^0\right| + \left|C^{2+}\right| - 1 + \mathtt{fd}^{(St3)}(c_0) + |S| + \left\lceil\frac{|S|}{2}\right\rceil$$

$$= 2\cdot\left|C^0\right| + \left|C^1\right| + \left|C^{2+}\right| - 1 + \left\lceil\frac{|S|}{2}\right\rceil \qquad \left(|S| = \left|C^1\right| - \mathtt{fd}^{(St3)}(c_0)\right)$$

$$= 2\cdot\left|C^0\right| + \left|C^{1+}\right| - 1 + \left(\left|C^{1+}\right| - 1 - \left\lceil\frac{\mathtt{fd}(G)}{2}\right\rceil\right) \qquad \text{(by Eq. (5.9))}$$

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

  Now, since $|S| > 0$, according to Eq. (5.10) it follows that $\mathtt{aux}(G) > 0$, which (according to Eq. (5.7)) 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 5.27.1. □

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 5.27.2. *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. (5.7), we have $\mathtt{wc}(G) = 2 \cdot |C^0| + |C^{1+}| - 1$. Relying on Claim 5.24, there exist $|C(G)| - 1 = |C^0(G)| + |C^{1+}(G)| - 1$ 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 $|C^0(G)|$. Therefore, the symmetric difference between the edge-sets of G and of G' is of size at least $2 \cdot |C^0| + |C^{1+}| - 1 = \mathtt{wc}(G)$, which finishes the first case.

Case 2: $\mathtt{aux}(G) > 0$. According to Eq. (5.7), we have $\mathtt{wc}(G) = 2 \cdot (|C(G)| - 1) - \left\lceil \frac{\mathtt{fd}(G)}{2} \right\rceil$. Relying on Claim 5.24, 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:

$$\mathtt{fd}(G'') = \mathtt{fd}(G) - 2 \cdot (|C(G)| - 1) \ . \tag{5.11}$$

Combining Eq. (5.11) 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. (5.11), 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 5.27.2 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 5.27. ■

5.6.1.2 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 5.28 *(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, we will rely on Proposition 4.1. Specifically, we will 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. Thus, for any $\delta \geq 2$, there does not exist a path from $G \notin \Pi_n \cup \mathcal{F}_\delta(\Pi_n)$ to $\mathcal{F}_\delta(\Pi_n)$ such that any graph subsequent to G on the path is neither in Π_n nor adjacent to Π_n. According to Proposition 4.1, this implies that Π_n is not $\mathcal{F}_\delta$-closed.

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 4.1, 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 5.24, 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 5.27 and on Eq. (5.6), 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, we now build on Lemma 5.27 to 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 5.29 *(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. We prove the counter-positive: Given a graph G such that $\Delta(G, \Pi_n) \geq (1 - \frac{1}{4d}) \cdot \delta$, we show that $\Delta(G, \mathcal{F}_\delta(\Pi_n)) \leq \delta - 1$, which implies that $G \notin \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$. This is done by modifying every such graph G to a graph $H \in \mathcal{F}_\delta(\Pi_n)$, with less than $\delta - 1$ edge additions/removals.

Recall, according to Lemma 5.27 and Eq. (5.6), that 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 to create new connected components, and then add edges within connected components, to decrease the number of free degrees to its original value. Specifically, we will repeatedly perform a basic modification step, which consists of isolating a small connected component, and then adding edges within the new component and between vertices in the original connected component (from which the new component was detached). This basic modification step will be depicted in the proof of the following claim.

Claim 5.29.1. *For every $G \notin \mathcal{F}_\delta(\Pi_n)$ there exists G' such that $\Delta(G, G') \leq 4 \cdot d$ and $\Delta(G', \Pi_n) \geq \Delta(G, \Pi_n) + 1$.*

Let us assume for a moment that Claim 5.29.1 is correct, and see how it implies Proposition 5.29. Given G such that $\Delta(G,\Pi_n) \in \left(\left(1-\frac{1}{4d}\right)\cdot\delta,\delta\right)$, we denote $G = G_0$, and start an iterative process of modifying the graph. In each iteration, given a graph $G_i \notin \mathcal{F}_\delta(\Pi_n)$, we rely on Claim 5.29.1 to obtain G_{i+1} such that $\Delta(G_i,G_{i+1}) \le 4\cdot d$ and $\Delta(G_{i+1},\Pi_n) \ge \Delta(G_i,\Pi_n)+1$. After at most $t \stackrel{\mathrm{def}}{=} \delta - \Delta(G,\Pi_n) < \frac{\delta}{4d}$ iterations, we obtain a graph that is δ-far from Π_n, and that disagrees with G on at most $4\cdot d\cdot t < \delta$ edges. Thus, it is only left to prove Claim 5.29.1.

Proof of Claim 5.29.1. 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 5.24, it follows that G is δ-far from being connected, which contradicts the hypothesis.

As a warm-up, 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 5.26, in this case the distance of a graph H from being connected is $\Delta(H,\Pi_n) = 2\cdot\left|C^0(H)\right| + \left|C^{2+}(H)\right| - 1$. 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 add d multiple edges between these two vertices. Overall, we removed at most $2\cdot d$ edges, and added d edges, and so $\Delta(G,G') \le 3\cdot d$. Note that, 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 no free degrees (i.e., $c \in C^0(G)$) whereas $c\setminus\{i,j\}$ has free degrees (i.e., $c\setminus\{i,j\} \in C^{2+}(G')$); or that c originally had free degrees in G and $c\setminus\{i,j\}$ has free degrees in G'. The reader can verify that in both cases it holds that $\Delta(G',\Pi_n) \ge \Delta(G,\Pi_n)+1$.

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 $m \le 2\cdot d$ (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, $\Delta(G,G') \le 4\cdot d$. Therefore we only need to prove that $\Delta(G',\Pi_n) \ge \Delta(G,\Pi_n)+1$. 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 5.29.1.1. *After the modification of G to G', the number of free degrees in the graph has not increased; that is,* $\mathtt{fd}(G') \le \mathtt{fd}(G)$.

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

$$\mathtt{fd}(G') - \mathtt{fd}(G) = m + \deg_G(i) + \deg_G(j) - 2\cdot\lfloor m/2\rfloor - 2\cdot d\,. \tag{5.12}$$

If $\deg_G(i) + \deg_G(j) < 2d$, then the expression in Eq. (5.12) 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. (5.12) equals zero. $\square$

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 5.29.1.1, this will now allow us to prove that $\Delta(G', \Pi_n) \geq \Delta(G, \Pi_n) + 1$.

For any graph H, denote $\varphi_1(H) = 2 \cdot \left|C^0(H)\right| + \left|C^{1+}(H)\right| - 1$ and $\varphi_2(H) \stackrel{\text{def}}{=} \left|C^{1+}(H)\right| - 1 - \left\lceil \frac{\mathtt{fd}(G)}{2} \right\rceil$. Then, according to Lemma 5.27, it holds that:

$$\Delta(H, \Pi_n) = \varphi_1(H) + \max\{0, \varphi_2(H)\} \ . \tag{5.13}$$

We prove that $\Delta(G', \Pi_n) \geq \Delta(G, \Pi_n) + 1$ by relying on Eq. (5.13), and considering three separate cases.

Case 1: $\left|C^0(G')\right| \geq \left|C^0(G)\right| + 2$. Note that $\left|C^1(G')\right| \geq \left|C^1(G)\right| - 1$, since the only way for G' to have less components with free degrees, compared to G, is if the component c had free degrees in G, but all the components that consist of vertices in c_0 in G' have no free degrees. Relying on this fact, and on Fact 5.29.1.1, it follows that $\varphi_2(G') \geq \varphi_2(G) - 1$. However, since $\left|C^0(G')\right| \geq \left|C^0(G)\right| + 2$, and relying again on the fact that $\left|C^1(G')\right| \geq \left|C^0(G)\right| - 1$, it follows that $\varphi_1(G') \geq \varphi_1(G) + 3$. Thus, $\Delta(G', \Pi_n) - \Delta(G, \Pi_n) \geq 2$.

Case 2: $\left|C^0(G')\right| = \left|C^0(G)\right|$. 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 $\left|C^{1+}(G')\right| \geq \left|C^{1+}(G)\right| + 1$. It follows that $\varphi_1(G') \geq \varphi_1(G) + 1$, and, relying on Fact 5.29.1.1, 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: $\left|C^0(G')\right| = \left|C^0(G)\right| + 1$. In this case it necessarily holds that $\left|C^{1+}(G')\right| \geq \left|C^{1+}(G)\right|$. 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 5.29.1.1) that $\varphi_2(G') \geq \varphi_2(G)$. Overall, we get that $\Delta(G', \Pi_n) \geq \Delta(G, \Pi_n) + 2$. $\square$

This completes the proof of Proposition 5.29. ■

A comment about non-simple graphs. Recall that in the preliminary definitions of the current section (i.e., Section 5.6.1), we assumed that the space of graphs we are dealing with also contains graphs with multiple edges and self-loops. Throughout Section 5.6.1.2, we relied on the assumption that such non-simple graphs exist in our metric space. Most notably, we relied on this assumption in Claim 5.29.1, which was the main step in proving Proposition 5.29. We believe that it is possible to prove a claim similar to Claim 5.29.1, and thus also obtain a result similar to Proposition 5.29, without relying on the existence of non-simple graphs, but it was not our focus in this work.

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

Proposition 5.28 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 5.29 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 5.27, for a given graph G, this is equivalent to estimating the following quantity:

$$2 \cdot \left|C^0\right| + \left|C^{1+}\right| - 1 + \max\left\{0, \left|C^{1+}\right| - 1 - \left\lceil \frac{\mathtt{fd}(G)}{2} \right\rceil \right\} . \tag{5.14}$$

We will see that *each* of the terms in Eq. (5.14) can be estimated up to an additive error of $\gamma \cdot n$, for any $\gamma > 0$, using only $\mathrm{poly}(1/\gamma)$ 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

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

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,

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

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 5.30 *(a tester for the dual problem of connectivity). Let $d \ge 2$, 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 5.29. 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. ■

5.6.2 Testing the property of being far from cycle-free in the bounded-degree model

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 $\mathrm{poly}(1/\epsilon)$, 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, that 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 connected, 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 by using the algorithm of [MR06] to estimate the distance of an input graph from the property.

Proposition 5.31 *(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. We will rely on Proposition 4.1. Specifically, we will 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. Thus, for any $\delta \geq 2$, there does not exist a path from $G \notin \Pi_n \cup \mathcal{F}_\delta(\Pi_n)$ to $\mathcal{F}_\delta(\Pi_n)$ such that any graph subsequent to G on the path is neither in Π_n nor adjacent to it. According to Proposition 4.1, this implies that Π_n is not $\mathcal{F}_\delta$-closed.

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. Relying on Proposition 4.1, we deduce that Π_n is not $\mathcal{F}_\delta$-closed. ■

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 5.32 *(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. To prove the lemma we show the counter-positive: Given a graph G such that $\Delta(G, \Pi_n) \in (\frac{2}{3} \cdot \delta, \delta)$, we show that G can be modified to a graph that is δ-far from Π_n by at most $\delta - 1$ modifications, which implies that $G \notin \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$. The main claim that we will need is the following:

Claim 5.32.1. *For every $G \notin \mathcal{F}_\delta(\Pi_n)$ there exists G' such that $\Delta(G, G') \leq 3$ and $\Delta(G', \Pi_n) = \Delta(G, \Pi_n) + 1$.*

Before proving Claim 5.32.1, let us assume for a moment that it is correct, and see how it implies Proposition 5.32. Given a graph $G = G_0$ such that $\Delta(G, \Pi_n) \in (\frac{2}{3} \cdot \delta, \delta)$, we start an iterative process of modifying G into a graph $H \in \mathcal{F}_\delta(\Pi_n)$. For $t \stackrel{\text{def}}{=} \delta - \Delta(G, \Pi_n) < \frac{1}{3} \cdot \delta$ iterations, we rely on Claim 5.32.1, to obtain a sequence of graphs $G_0, G_1, ..., G_t$ such that $\Delta(G_i, G_{i+1}) \leq 3$ and $\Delta(G_{i+1}, \Pi_n) = \Delta(G_i, \Pi_n) + 1$. After t iterations, we get that $\Delta(G_t, G_0) \leq 3 \cdot t < \delta$, and that $\Delta(G_t, \Pi_n) = \Delta(G, \Pi_n) + t = \delta$, as required. It is thus left to prove Claim 5.32.1.

Proof of Claim 5.32.1. 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., $|C(G)| \leq \frac{n}{6 \cdot d}$), 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 $\Delta(G, \Pi_n) \leq \delta = \epsilon \cdot d \cdot n$, we get that

$$|E(G)| = \Delta(G, \Pi_n) + n - |C(G)| < (1 + \epsilon \cdot d) \cdot n - |C(G)| \, . \tag{5.15}$$

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

Case 1: $|C(G)| \leq \frac{n}{6 \cdot d}$. Denote the number of vertices with degree d in G by m. Then, relying on Eq. (5.15), we get that

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

It follows that $m < \left(\frac{2}{d} + 2 \cdot \epsilon\right) \cdot n$, and since $d \geq 3$ 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. (5.15), 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 $d \geq 3$) and that satisfies $|C(G')| \geq |C(G)| - 2$ and $|E(G')| = |E(G)| + 3$. It follows that $\Delta(G', \Pi_n) \geq \Delta(G, \Pi_n) + 1$. □

This completes the proof of Proposition 5.32. ■

Proposition 5.32 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 5.33 *(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 $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi_n))$.*

5.6.3 Testing the property of being far from bipartite in the bounded-degree model

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 $\left(\epsilon \cdot \binom{n}{2}\right)$*-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 5.6.2), in the current section we also consider *only simple graphs*.

Proposition 5.34 *(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. Similar to the proof of Proposition 5.31, 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. Relying on Proposition 4.1, this implies that Π_n is not $\mathcal{F}_\delta$-closed.

The construction of G is identical to the one in the proof of Proposition 5.20: The graph G contains a single $(k+1)$ clique alongside $n-(k+1)$ isolated vertices. In the proof of Proposition 5.20 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. ■

Proposition 5.34 implies that the dual problem of testing k-colorability in the bounded-degree model is different from the original problem. For $k = 2$ (i.e., testing bipartiteness), the query complexity of the original 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.

6 Open questions

$\mathcal{F}_\delta$-tight spaces. A graph-theoretical problem we encountered during this work is the characterization of $\mathcal{F}_\delta$-tight spaces. Recall that, by Definition 4.8, a graphical space is $\mathcal{F}_\delta$-tight if every $\mathcal{F}_\delta$-closed set in it is also strongly $\mathcal{F}_\delta$-closed. That is, if for every $\mathcal{F}_\delta$-closed set Π and every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ it holds that x lies on a *shortest path* from Π to $\mathcal{F}_\delta(\Pi)$. In Section 4.1.3 we showed that all graphical spaces are $\mathcal{F}_\delta$-tight for $\delta = 1, 2$ and for values of δ larger than the diameter of the graph. We also showed that there exist spaces that are not $\mathcal{F}_\delta$-tight for $\delta = 3$. This leaves open the following general question.

Question 1 *($\mathcal{F}_\delta$-tight spaces). For which graphs G and values of $\delta \in [3, \text{diam}(G)]$ does it hold that G is $\mathcal{F}_\delta$-tight?*

In Proposition 4.11 in Section 4.1.3 we presented an initial exploration of this question, by showing several examples for graphs that are $\mathcal{F}_\delta$-tight for every $\delta > 0$.

Separation between dual problems and standard problems. Recall that, according to Proposition 5.3, the complexity of any dual problem is lower bounded by the complexity of the original problem. This leads to the following question:

Question 2 *(separation between dual problems and standard problems). Is there a property testing problem with query complexity that is significantly lower than the query complexity of its dual problem?*

A different interesting direction is to bound the query complexity of standard property testing problems by determining the query complexity of their dual problems. In particular, by Proposition 5.3, any upper bound on a dual problem implies an identical upper bound on the original problem.

Dual problems in the dense graph model. Recall that in the dense graph model, Corollary 5.19 states the following (relying on [FN07]): For a graph property $\Pi = \{\Pi_n\}_{n\in\mathcal{N}}$ that is testable with $O(1)$ queries, if for every $\epsilon > 0$ there exists $\alpha \in (0,1)$ such that for all sufficiently large n it holds that $\mathcal{F}_{\epsilon\cdot n}(\mathcal{F}_{\epsilon\cdot n}(\Pi_n)) \subseteq \{x : \Delta(x,\Pi_n) \leq (\alpha\cdot\epsilon)\cdot n\}$, then the dual problem is also testable with $O(1)$ queries.

Question 3 *(testable graph properties such that points in $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ are sufficiently close to Π). Let $\Pi = \{\Pi_n\}_{n\in\mathcal{N}}$ be a graph property in the dense graph model that is testable with $O(1)$ queries. Does it hold that for every $\epsilon > 0$ there exists $\alpha \in (0,1)$ such that for all sufficiently large n it holds that $\mathcal{F}_{\epsilon\cdot n}(\mathcal{F}_{\epsilon\cdot n}(\Pi_n)) \subseteq \{x : \Delta(x,\Pi_n) \leq (\alpha\cdot\epsilon)\cdot n\}$?*

An affirmative answer to this question would imply that, in the dense graph model, a dual problem is testable with $O(1)$ queries *if and only if* the original problem is testable with $O(1)$ queries.

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. This research was partially supported by the Israel Science Foundation (grant No. 671/13).

References

[ACCL07] Nir Ailon, Bernard Chazelle, Seshadhri Comandur, and Ding Liu. Estimating the distance to a monotone function. *Random Structures & Algorithms*, 31(3):371–383, 2007.

[AKK+03] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. Testing low-degree polynomials over GF(2). In *Proc. 7th International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*, pages 188–199, 2003.

[BBM12] Eric Blais, Joshua Brody, and Kevin Matulef. Property testing lower bounds via communication complexity. *Computational Complexity*, 21(2):311–358, 2012.

[BCGSM12] Jop Briët, Sourav Chakraborty, David García-Soriano, and Arie Matsliah. Monotonicity testing and shortest-path routing on the cube. *Combinatorica*, 32(1):35–53, 2012.

[BCH+96] Mihir Bellare, Don Coppersmith, Johan Håstad, Marcos Kiwi, and Madhu Sudan. Linearity testing in characteristic two. *IEEE Transactions on Information Theory*, 42(6, part 1):1781–1795, 1996.

[BFF+01] Tugkan Batu, Eldar Fischer, Lance Fortnow, Ravi Kumar, Ronitt Rubinfeld, and Patrick White. Testing random variables for independence and identity. In *Proc. 42nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 442–451, 2001.

[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.

[BGLR93] Mihir Bellare, Shafi Goldwasser, Carsten Lund, and Alexander Russell. Efficient probabilistically checkable proofs and applications to approximations. In *Proc. 25th Annual ACM Symposium on Theory of Computing (STOC)*, pages 294–304, 1993.

[BK12] Eric Blais and Daniel M. Kane. Tight bounds for testing k-linearity. In *Proc. 16th International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*, pages 435–446, 2012.

[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.

[BO10] Eric Blais and Ryan O'Donnell. Lower bounds for testing function isomorphism. In *Proc. 25th Annual IEEE Conference on Computational Complexity (CCC)*, pages 235–246, 2010.

[Bol76] Béla Bollobás. On complete subgraphs of different orders. *Mathematical Proceedings of the Cambridge Philosophical Society*, 79(1):19–24, 1976.

[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.

[BS94] Mihir Bellare and Madhu Sudan. Improved non-approximability results. In *Proc. 26th Annual ACM Symposium on Theory of Computing (STOC)*, pages 184–193, 1994.

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

[CDST15] Xi Chen, Anindya De, Rocco A. Servedio, and Li-Yang Tan. Boolean function monotonicity testing requires (almost) $n^{1/2}$ non-adaptive queries. In *Proc. 47th Annual ACM Symposium on Theory of Computing (STOC)*, 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.

[CS13a] Deeparnab Chakrabarty and C. Seshadhri. A $o(n)$ monotonicity tester for Boolean functions over the hypercube. In *Proc. 45th Annual ACM Symposium on Theory of Computing (STOC)*, pages 411–417, 2013.

[CS13b] Deeparnab Chakrabarty and C. Seshadhri. Optimal bounds for monotonicity and Lipschitz testing over hypercubes and hypergrids. In *Proc. 45th Annual ACM Symposium on Theory of Computing (STOC)*, pages 419–428, 2013.

[CS14] Deeparnab Chakrabarty and C. Seshadhri. An optimal lower bound for monotonicity testing over hypergrids. *Theory of Computing*, 10:453–464, 2014.

[CST14] Xi Chen, Rocco A. Servedio, and Li-Yang Tan. New algorithms and lower bounds for monotonicity testing. In *Proc. 55th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 286–295, 2014.

[DGL$^+$99] Yevgeniy Dodis, Oded Goldreich, Eric Lehman, Sofya Raskhodnikova, Dana Ron, and Alex Samorodnitsky. Improved testing algorithms for monotonicity. In *Proc. 3rd International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*, pages 97–108, 1999.

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

[DLM$^+$07] Ilias Diakonikolas, Homin K. Lee, Kevin Matulef, Krzysztof Onak, Ronitt Rubinfeld, Rocco A. Servedio, and Andrew Wan. Testing for concise representations. In *Proc. 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 549–558, 2007.

[DLM$^+$11] Ilias Diakonikolas, Homin K. Lee, Kevin Matulef, Rocco A. Servedio, and Andrew Wan. Efficiently testing sparse GF(2) polynomials. *Algorithmica*, 61(3):580–605, 2011.

[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.

[GHS15] Alan Guo, Elad Haramaty, and Madhu Sudan. Robust testing of lifted codes with applications to low-degree testing. *Electronic Colloquium on Computational Complexity: ECCC*, 22:43, 2015.

[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.

[Gol10a] Oded Goldreich. On testing computability by small width OBDDs. In *Proc. 14th International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*, pages 574–587, 2010.

[Gol10b] 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.

[GR00] Oded Goldreich and Dana Ron. On testing expansion in bounded-degree graphs. *Electronic Colloquium on Computational Complexity: ECCC*, 7:20, 2000.

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

[Guo15] Alan Guo. Group homomorphisms as error correcting codes. *Electronic Journal of Combinatorics*, 22(1):14, 2015.

[HSS13] Elad Haramaty, Amir Shpilka, and Madhu Sudan. Optimal testing of multivariate polynomials over small prime fields. *SIAM Journal of Computing*, 42(2):536–562, 2013.

[JPRZ09] Charanjit S. Jutla, Anindya C. Patthak, Atri Rudra, and David Zuckerman. Testing low-degree polynomials over prime fields. *Random Structures and Algorithms*, 35(2):163–193, 2009.

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

[KLX10] Tali Kaufman, Simon Litsyn, and Ning Xie. Breaking the ϵ-soundness bound of the linearity test over $\mathrm{GF}(2)$. *SIAM Journal of Computing*, 39(5):1988–2003, 2010.

[KMS15] Subhash Khot, Dor Minzer, and Muli Safra. On monotonicity testing and boolean isoperimetric type theorems. *Electronic Colloquium on Computational Complexity: ECCC*, 22:11, 2015.

[KR06] Tali Kaufman and Dana Ron. Testing polynomials over general fields. *SIAM Journal of Computing*, 36(3):779–802, 2006.

[LR01] Eric Lehman and Dana Ron. On disjoint chains of subsets. *Journal of Combinatorial Theory. Series A*, 94(2):399–404, 2001.

[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.

[RRS+12] Dana Ron, Ronitt Rubinfeld, Muli Safra, Alex Samorodnitsky, and Omri Weinstein. Approximating the influence of monotone boolean functions in $o(\sqrt{n})$ query complexity. *ACM Transactions on Computation Theory*, 4(4):11, 2012.

[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.

[Vad12] Salil P. Vadhan. *Pseudorandomness*. Foundations and Trends in Theoretical Computer Science. Now Publishers, 2012.

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

[VV14] Gregory Valiant and Paul Valiant. An automatic inequality prover and instance optimal identity testing. In *Proc. 55th Annual ACM Symposium on Theory of Computing (STOC)*, pages 51–60, 2014.

Appendix A: Additional results regarding the operator $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$

Following Proposition 3.5, in this appendix we explore additional properties of the operator $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$. More precisely, we prove that $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ *does not* admit some properties in general, and thus does not belong to some specific classes of closure operators. In particular, we show that $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ is not the convex hull operator in Euclidean spaces, is not a Kuratowski (topological) closure operator, and does not satisfy the axioms of closure operators from matroid theory. In the end of the appendix we repay a debt from Section 3.2, by including a proof for Proposition 3.6.

Before proving these results, 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 3.2, 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.

A.1 Properties that $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ does not admit

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 A.1 *($\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ 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. ■

Closure operators in topology are called *Kuratowski closure operators*, and satisfy the three conditions in Definition 3.4 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 A.2 *($\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ 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 $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi \cup \Pi')) = \Omega \neq \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi)) \cup \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi'))$. ■

Closure operators in *matroid theory* (see, e.g., [GM12]) satisfy the three conditions in Definition 3.4 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 A.3 *(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 A.4 *($\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$ 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 $\mathcal{F}_\delta(\mathcal{F}_\delta(\Pi \cup \{x\})) = \{x\} \not\ni y$. ∎

A.2 Proof of Proposition 3.6 from Section 3.2

In general, a closure operator maps any set Π to the unique smallest closed set containing Π. Proposition 3.6 from Section 3.2 asserts that this is indeed the case in the special case of the operator $\Pi \mapsto \mathcal{F}_\delta(\mathcal{F}_\delta(\Pi))$. For convenience, we now include a proof for the proposition.

Proposition A.5 *(Proposition 3.6, restated). 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 3.2, and was mentioned in the discussion after the proof of Theorem 3.2.

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 3.4 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 3.4 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

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

Appendix B: Sets with "holes" are not $\mathcal{F}_\delta$-closed

Recall that Proposition 4.1 presents a condition that is necessary for a set in a graphical space to be $\mathcal{F}_\delta$-closed: That for every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ there exists a path from x to $\mathcal{F}_\delta(\Pi)$ that does not intersect Π nor any vertices adjacent to Π. In this appendix we show a condition that is equivalent to the one in Proposition 4.1. Intuitively, we show that a set that contains a "small hole" is not $\mathcal{F}_\delta$-closed. Since this statement is still quite vague, we now describe it in further detail.

For any $\Psi \subseteq \Omega$, let the *vertex boundary* of Ψ be $\partial\Psi = \{x \in \Psi : \exists y \in \Omega \setminus \Psi, \Delta(x,y) = 1\}$; that is, $\partial\Psi$ consists of all vertices in Ψ with neighbors outside of Ψ. Also, the *interior* of Ψ is $\Psi \setminus \partial\Psi$, and consists of all vertices in Ψ such that all their neighbors are in Ψ. We now use these notations to describe a set Π with a "hole" in it. Consider some neighborhood $\Psi \subseteq \Omega$ such that two conditions hold: First, the interior of Ψ contains vertices that are not in Π; and second, the vertex boundary of Ψ satisfies $\partial\Psi \subseteq \Pi$. Thus, the interior of Ψ is "enclosed" by Π. In such a case we think of the interior of Ψ as a "hole" in Π, and of Ψ as a neighborhood of Ω in which Π contains a "hole". Figure 5 presents an example for such a case.

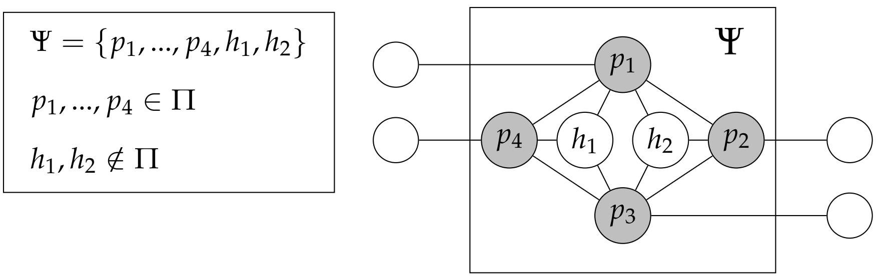

Figure 5: An example for a neighborhood Ψ in a graph and a set Π such that Π contains a "hole" in Ψ. The vertices $p_1, ..., p_4$ constitute $\partial\Psi$, and note that $\partial\Psi \subseteq \Pi$. The vertices h_1, h_2 constitute the interior of Ψ, and are not in Π. We think of the interior of Ψ (i.e., of $\{h_1, h_2\}$) as a "hole" in Π.

We now formally define what it means for a set Π to have a "hole of diameter $\delta - 1$". Note that in the examples described so far we required that $\partial\Psi \subseteq \Pi$; that is, that Π fully "encloses" the interior of Ψ. In the definition itself we relax this requirement, and only require that every $z \in \partial\Psi$ is adjacent to Π.

Definition B.1 *(sets with "holes of diameter $\delta - 1$"). For a graphical Ω, $\delta \geq 2$ and $\Pi \subseteq \Omega$, assume that there exists $\Psi \subseteq \Omega$ such that the following hold:*

1. *(the interior of Ψ is "enclosed" by Π) Every $z \in \partial\Psi$ satisfies $\Delta(z, \Pi) \leq 1$.*

2. *(the interior of Ψ contains a vertex not in Π) There exists $x \in \Psi \setminus \partial\Psi$ such that $x \notin \Pi$.*

3. *(the interior of Ψ is "$(\delta - 1)$-covered" by Π) Every $x \in \Psi$ satisfies $\Delta(x, \Pi) \leq \delta - 1$.*

Then we say that Π has a hole of diameter $\delta - 1$ *in* Ψ.

We now show that a set has a "hole of diameter $\delta - 1$" if and only if it does not satisfy the necessary condition for a set to be $\mathcal{F}_\delta$-closed that was presented in Proposition 4.1. Thus, sets that have a "hole of diameter $\delta - 1$" are not $\mathcal{F}_\delta$-closed. The existence of such a "hole" might be convenient to prove in some cases, since it only requires arguing about Π in a neighborhood Ψ of Ω, and not about $\mathcal{F}_\delta(\Pi)$.

Proposition B.2 *(the condition of not having "holes of diameter $\delta - 1$" is equivalent to the condition in Proposition 4.1). For a graphical Ω and $\delta \geq 2$ it holds that $\Pi \subseteq \Omega$ has a "hole of diameter $\delta - 1$", as in Definition B.1, if and only if there exists $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ such that for every path $x = v_0, v_1, ..., v_l = z$, where $z \in \mathcal{F}_\delta(\Pi)$, there exists $i \in [l]$ such that $\Delta(v_i, \Pi) \leq 1$.*

Proof. In one direction, assume that for $\Pi \subseteq \Omega$ and $\delta > 0$ there exists $\Psi \subseteq \Omega$ such that Ψ and δ satisfy conditions of Definition B.1. By Condition (2) of Definition B.1, there exists $x \in \Psi \setminus (\Pi \cup \partial\Psi)$. By Condition (3) of Definition B.1, it holds that $\Psi \cap \mathcal{F}_\delta(\Pi) = \emptyset$, and thus $x \notin \mathcal{F}_\delta(\Pi)$. We show that $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ satisfies the conditions of Proposition 4.1 (i.e., every path from x to $\mathcal{F}_\delta(\Pi)$ intersects Π or a vertex adjacent to Π).

Let $x = v_0, v_1, ..., v_l = z \in \mathcal{F}_\delta(\Pi)$ be a path from x to $\mathcal{F}_\delta(\Pi)$. Since $\Psi \cap \mathcal{F}_\delta(\Pi) = \emptyset$, it follows that $\mathcal{F}_\delta(\Pi) \subseteq \overline{\Psi}$. In particular, $z \notin \Psi$, and thus the path from x to z passes through $\partial\Psi$. Let $i \in \{0, ..., l\}$ such that $v_i \in \partial\Psi$. Since $x \notin \partial\Psi$ it follows that $v_i \neq x$, hence $i \in [l]$. By Condition (1) of Definition B.1, it holds that $\Delta(v_i, \Pi) \leq 1$.

For the other direction, let $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$ such that for every path $x = v_0, v_1, ..., v_l = z$, where $z \in \mathcal{F}_\delta(\Pi)$, there exists $i \in [l]$ such that $\Delta(v_i, \Pi) \leq 1$. We construct Ψ that satisfies the conditions of Definition B.1, as follows. Let $\mathcal{P}$ be the collection of all finite paths that start from x and end in some vertex in $\mathcal{F}_\delta(\Pi)$; note that these paths are not necessarily simple, and thus $\mathcal{P}$ is an infinite collection. For every path $x = v_0, v_1, ..., v_l = z \in \mathcal{F}_\delta(\Pi)$ in $\mathcal{P}$, let v_i be the first vertex in the path that satisfies $v_i \neq x$ and $\Delta(v_i, \Pi) \leq 1$. We define the path's *truncation* to be all vertices v_j in the path such that $j \leq i$. We define Ψ be the set of all vertices that are in truncations of paths in $\mathcal{P}$.

To see that Condition (1) of Definition B.1 holds, assume towards a contradiction that there exists $v \in \partial\Psi$ such that $\Delta(v, \Pi) > 1$. Since $v \in \Psi$, there exists a path $x = v_0, v_1, ..., v_r = v, v_{r+1}, ..., z \in \mathcal{F}_\delta(\Pi)$ such that for every $i \in [r]$ it holds that $\Delta(v_i, \Pi) > 1$. However, this implies that for any neighbor v' of v there exists a path $x = v_0, v_1, ..., v_r = v, v', v, v_{r+1}, ..., z$ such that for every $i \in [r]$ it holds that $\Delta(v_i, \Pi) > 1$, which implies that v' is in the truncation of that path. Thus $v' \in \Psi$. Since all of v's neighbors are in Ψ, it cannot be that $v \in \partial\Psi$.

To see that the vertex x that exists according to our hypothesis satisfies Condition (2) of Definition B.1, note that by the hypothesis $x \notin \Pi$, and that by the definition of Ψ it holds that $x \in \Psi$. Furthermore, since each of x's neighbors is in the truncation of *some* path from x

to $\mathcal{F}_\delta(\Pi)$ (e.g., a path from x to the neighbor, back to x, and then to $\mathcal{F}_\delta(\Pi)$), it follows that all of x's neighbors are in Ψ, hence $x \notin \partial\Psi$. Therefore $x \in \Psi \setminus (\Pi \cup \partial\Psi)$.

To see that Condition (3) of Definition B.1 holds, first note that by the hypothesis $x \notin \mathcal{F}_\delta(\Pi)$. Now, let $z \in \Psi$ such that $z \neq x$, and we show that $z \notin \mathcal{F}_\delta(\Pi)$. By the definition of Ψ it holds that z is in the truncation of some path from x to $\mathcal{F}_\delta(\Pi)$. Denote the prefix of such a path, leading from x to z, by $x = v_0, v_1, ..., v_l = z$, and note that for every $i \in [l-1]$ it holds that $\Delta(v_i, \Pi) > 1$ (since this is a prefix of a *truncation* of a path). However, if $z \in \mathcal{F}_\delta(\Pi)$, then this prefix is a path from x to $\mathcal{F}_\delta(\Pi)$ without a vertex in the path that is in Π or adjacent to Π, which contradicts the hypothesis. Therefore $z \notin \mathcal{F}_\delta(\Pi)$. ■

Appendix C: Examples of $\mathcal{F}_\delta$-tight spaces

Recall that in Section 4.1.3 we defined $\mathcal{F}_\delta$-*tight* spaces as follows:

Definition C.1 *(Definition 4.8, restated). 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.*

In this appendix we prove that several specific graphs (or, more accurately, graph families) are $\mathcal{F}_\delta$-tight for *every* $\delta > 0$. In particular, we prove the following proposition:

Proposition C.2 *(Proposition 4.11, extended). The following graphs are $\mathcal{F}_\delta$-tight, for every $\delta > 0$:*

1. *Any graph on $n \geq 2$ vertices with diameter at most 2 (and in particular, 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.*

Recall that in Section 4.1.3 we showed that every graphical space is $\mathcal{F}_1$-tight and $\mathcal{F}_2$-tight, and is $\mathcal{F}_\delta$-tight for values of δ larger than the diameter of the graph. Item (1) of Proposition C.2 follows as a corollary. We now prove Items (2) and (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 C.3 *(Items (2) and (3) of Proposition C.2). 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 $\delta \geq 3$, and let $\Pi \subseteq G_n$ be an $\mathcal{F}_\delta$-closed set. We prove that Π is strongly $\mathcal{F}_\delta$-closed, relying on Proposition 4.6: For every $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$, we show a neighbor x' of x such that $\Delta(x', \Pi) = \Delta(x, \Pi) + 1$.

Let $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$. According to Corollary 4.2, 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 4.2).

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 C.3, a nearly identical proof applies both to the $2 \times n$ grid and to the circular ladder graph on $2n$ vertices.

Proposition C.4 *(Items (4) and (5) of Proposition C.2). 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.*

The following proof of Proposition C.4 is quite tedious. In particular, the proof relies on elementary arguments and case analyses that are, in our opinion, not insightful. We hope to find a more insightful proof in the future.

Proof of Proposition C.4. 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 $\delta \geq 3$, and let $\Pi \subseteq G_{2,n}$ be an $\mathcal{F}_\delta$-closed set. We show that Π is strongly $\mathcal{F}_\delta$-closed, relying on Proposition 4.6: For $x \notin \Pi \cup \mathcal{F}_\delta(\Pi)$, we 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 4.1 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 4.2 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 3.2.

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

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

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 C.4.1. *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 4.2, 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. (C.1)). Thus, both paths that exist according to Corollary 4.2 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.[14]

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 6 depicts the relevant part of the graph, reflecting some of our assumptions and notations at this point.

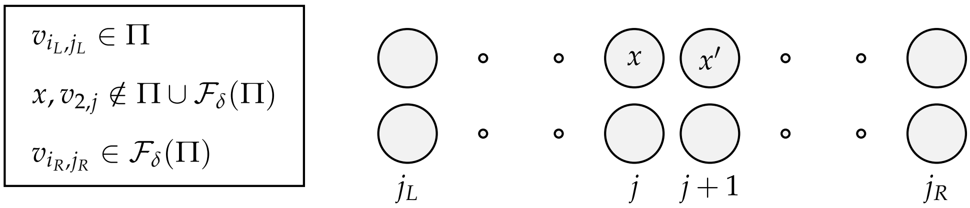

Figure 6: The relevant part of the graph $G_{2,n}$, reflecting our assumptions and notations at this point (as well as an additional, unjustified assumption that $j_R \neq j+1$). Note that columns $j_L + 1, ..., j_R - 1$ do not contain vertices from $\Pi \cup \mathcal{F}_\delta(\Pi)$.

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

[14]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 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.

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

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

Since $x \notin \mathcal{F}_\delta(\Pi)$, it holds that $\Delta(x, \Pi) + 1 \leq \delta$. Thus, the value of the expression in Eq. (C.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 the expression in Eq. (C.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 C.4.2. *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 C.4.2 holds. Then, the expression in Eq. (C.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 C.4.2.

Proof of Claim C.4.2. 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 7 depicts columns $j_L, ..., j + 1 = j_R$ of the graph, reflecting our assumptions at this point.

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

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

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

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

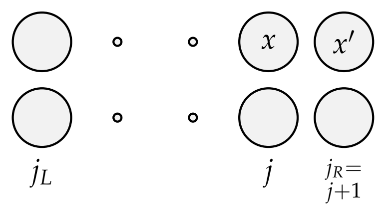

Figure 7: Columns $j_L, ..., j + 1 = j_R$ of the graph $G_{2,n}$, reflecting our assumptions at this point.

Fact C.4.2.1. *From the hypothesis of Claim C.4.2 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 3.2: 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 C.4.2.2. *From the hypothesis of Claim C.4.2 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 C.4.2.1, 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 3.2, it follows that Π is not $\mathcal{F}_\delta$-closed, which concludes the proof of Claim C.4.2. □

As mentioned in the discussion after the statement of Claim C.4.2, the proof of the latter concludes the proof of Proposition C.4. ■

ECCC ISSN 1433-8092
http://eccc.hpi-web.de