Electronic Colloquium on Computational Complexity, Report No. 185 (2018)

On the Testability of Graph Partition Properties

Yonatan Nakar * Dana Ron †

Abstract

In this work we study the testability of a family of graph partition properties that generalizes a family previously studied by Goldreich, Goldwasser, and Ron (*Journal of the ACM, 1998*). While the family studied by Goldreich et al. includes a variety of natural properties, such as k-colorability and containing a large cut, it does not include other properties, such as split graphs, and more generally (p, q)-colorable graphs, that should clearly be regarded as graph partition properties. The generalization we consider better captures the idea of partition properties by allowing to impose constraints on the edge-densities within and between parts (relative to the sizes of the parts). We denote the family studied in this work by $\mathcal{GPP}$.

We first show that all properties in $\mathcal{GPP}$ have a testing algorithm whose query complexity is polynomial in $1/\epsilon$, where ϵ is the given proximity parameter (and there is no dependence on the size of the graph). As the testing algorithm has two-sided error, we next address the question of which properties in $\mathcal{GPP}$ can be tested with one-sided error and query complexity polynomial in $1/\epsilon$. We answer this question by establishing a characterization result. Namely, we define a subfamily $\mathcal{GPP}_{0,1}$ of $\mathcal{GPP}$ and show that every property $P \in \mathcal{GPP}_{0,1}$ is testable by a one-sided error algorithm that has query complexity $\mathrm{poly}(1/\epsilon)$ and that if $P \in \mathcal{GPP} \backslash \mathcal{GPP}_{0,1}$ then it cannot have a one-sided error testing algorithm whose query complexity is independent of the input graph's size.

*Tel Aviv University, *yonatannakar@mail.tau.ac.il.*
†Tel Aviv University, *danaron@tau.ac.il.*

ISSN 1433-8092

Contents

1 Introduction

In graph property testing, the goal is to decide whether a graph satisfies a prespecified property P or is far from satisfying P. To this end, the testing algorithm is given query access to the adjacency matrix of the input graph so that the algorithm can check whether there is an edge between any given pair of vertices.[1] A graph G over n vertices is said to be ϵ-far from satisfying P if it is necessary to add or delete more than ϵn^2 edges in order to turn G into a graph satisfying P. A *tester* for a graph property P is a randomized algorithm, which given query access to the graph, distinguishes with high constant probability between the case where G satisfies P and the case where G is ϵ-far from satisfying P. The tester should make the distinction between the two cases by observing a very small portion of the input graph. In other words, the tester must have sublinear query complexity.

We focus on properties that can be tested with no dependence on n. In particular, the query complexity of the testers we consider depends only on the proximity parameter ϵ, and the decisions of the testers do not depend on n as well. We call such graph properties *input-size oblivious testable*. Alon et al. [3] presented a complete characterization of input-size oblivious testable graph properties. Independently, Borgs et al. [10] obtained an analytic characterization of such properties through the theory of graph limits. However, while the query complexity of the tester emerging from the characterization of Alon et al. does not depend on the graph size, it could be super-polynomial in $\frac{1}{\epsilon}$. For example, the property of being triangle-free is input-size oblivious testable, but the query complexity of the best known tester for triangle-freeness is a tower function of $\frac{1}{\epsilon}$ [11]. Further, there exists a super-polynomial lower bound on the query complexity of testing triangle-freeness [1, 7]. Naturally, we strive to design testers with query complexity that is polynomial in $\frac{1}{\epsilon}$.

In this paper we consider a family of graph partition properties. This family of properties, which will be defined shortly, generalizes a family of graph partition properties that was introduced by Goldreich, Goldwasser, and Ron [13]. Examples of properties covered by their framework include bipartiteness, k-colorability, and the property of having a cut of at least βn^2 edges for some $\beta \in [0, 1]$. Their framework, while fairly general, lacks an ingredient that is necessary for specifying many natural graph partition properties such as split graphs (or more generally (p, q)-colorable graphs), probe complete graphs, and bisplit graphs.[2]

Given a graph $G = (V, E)$, a partition $(V_1, ..., V_k)$ of V, and a pair of parts V_i, V_j (possibly $i = j$) we denote by $e_G(V_i, V_j)$ the number of edges in G between the part V_i and part V_j (if $i = j$, then the notation refers to the number of edges within the part). Following the definitions in [13], the notation $e_G(V_i, V_j)$ counts each edge twice (both (u, v) and (v, u)) and when $i = j$ we also allow self-loops. That is, $e_G(V_i, V_j)$ counts the number of ones in the adjacency matrix representing the graph G. Also, we denote by $\overline{e}_G(V_i, V_j)$ the number of nonedges between the two parts.

Each property in the family of *Graph Partition Properties* considered in this work, is defined by an integer parameter k and $O(k^2)$ additional parameters in $[0, 1]$. Informally, a graph has the property if its vertices can be partitioned into k subsets such that the sizes of the subsets and the number of edges between pairs of subsets and within the subsets obey the constraints defined by the parameters of the property. Formally, a Graph Partition Property P is parameterized by an integer k denoting the number of parts and by the following parameters in the interval $[0, 1]$:

[1] Here we refer to what is known as the Dense Graph Model or the Adjacency Matrix Model [13].

[2] A graph is a *split graph* if it can be partitioned into an independent set and a clique. A graph is (p, q)-*colorable* if it can be partitioned into p cliques and q independent sets. A graph is *probe-complete* if it can be partitioned into an independent set and a clique such that every vertex in the independent set is adjacent to every vertex in the clique. A graph is *bisplit* if it can be partitioned into an independent set and a bi-clique.

1. Bounds on each part's size: for each $1 \le i \le k$ we have ρ_i^L, ρ_i^U s.t. part V_i must satisfy $\rho_i^L n \le |V_i| \le \rho_i^U n$.

2. Absolute bounds on the number of edges within each part: for each $1 \le i \le k$ we have ρ_{ii}^L, ρ_{ii}^U s.t. part V_i must satisfy $\rho_{ii}^L n^2 \le e_G(V_i, V_i) \le \rho_{ii}^U n^2$.

3. Absolute bounds on the number of edges between each pair of parts: for each pair $1 \le i, j \le k$ we have ρ_{ij}^L, ρ_{ij}^U s.t. the pair of parts V_i, V_j must satisfy $\rho_{ij}^L n^2 \le e_G(V_i, V_j) \le \rho_{ij}^U n^2$.

4. Relative bounds on the number of edges within each part: for each $1 \le i \le k$ we have $\alpha_{ii}^L, \alpha_{ii}^U$ s.t. part V_i must satisfy $\alpha_{ii}^L |V_i|^2 \le e_G(V_i, V_i) \le \alpha_{ii}^U |V_i|^2$.

5. Relative bounds on the number of edges between each pair of parts: for each pair $1 \le i, j \le k$ we have $\alpha_{ij}^L, \alpha_{ij}^U$ s.t. the pair of parts V_i, V_j must satisfy $2\alpha_{ij}^L |V_i| \cdot |V_j| \le e_G(V_i, V_j) \le 2\alpha_{ij}^U |V_i| \cdot |V_j|$.

The original graph partition framework that was introduced in [13] includes only Items 1–3. The absence of relative edge bounds makes the original framework weaker than the general framework we consider in this paper. In particular, using the original framework, one cannot express the notion of parts being cliques or the notion of a pair of parts being fully connected to each other. More generally, our framework enhances the expressive power of the original framework by adding the notion of edge densities,[3] a notion that does not exist in the original framework. We denote the class of graph partition properties (as defined above) by $\mathcal{GPP}$. We denote the class of graph partition properties that have no relative bounds on the number of edges (the one introduced in [13]) by $\mathcal{GPP}_{NR}$ (NR stands for Non-Relative).

We say that a graph property is poly$\left(\frac{1}{\epsilon}\right)$-testable if it is input-size oblivious testable and the tester's query complexity is polynomial in $\frac{1}{\epsilon}$. All the properties in the class $\mathcal{GPP}_{NR}$ are poly$\left(\frac{1}{\epsilon}\right)$-testable [13]. In this work, we first show how to use the algorithm presented in [13] as a subroutine to devise a tester for all the partition properties covered by our generalized framework, thus obtaining the following theorem:

Theorem 1. *Every property $P \in \mathcal{GPP}$ is poly$\left(\frac{1}{\epsilon}\right)$-testable.*

While the query complexity of the tester implied by 1 is a polynomial function of $\frac{1}{\epsilon}$ as desired, it has the disadvantage of having two-sided error (just like the algorithm described in [13]). A tester has one-sided error if, whenever a graph G satisfies P, the tester determines this with probability 1. Clearly, a one-sided error tester is preferable to a two-sided error tester because a one-sided error tester is capable of providing a witness demonstrating that the property is not satisfied by the input graph. Combining the two desired features of polynomial dependence on $\frac{1}{\epsilon}$ and having one-sided error leads to the definition of *easily testable graph properties* (as defined in e.g. [8, 4, 12]).

Definition 1. *A graph property P is* easily-testable *if P is poly$\left(\frac{1}{\epsilon}\right)$-testable and the tester has one-sided error.*

An example of an easily testable graph partition property is the property of being k-colorable [13, 5]. In this paper we address the question of characterizing the easily testable graph partition properties. We show that every graph partition property belonging to a restricted subset of $\mathcal{GPP}$, which we denote by $\mathcal{GPP}_{0,1}$ (and formally define below), is easily testable, and every graph partition property $P \notin \mathcal{GPP}_{0,1}$ is not easily testable. That is, while Theorem 1 implies

[3] From now on, when using the term edge density, we refer to the fraction of edges between the parts (or within the part) relative to the number of vertex pairs between the parts (or within the part).

that every graph partition property P is poly$\left(\frac{1}{\epsilon}\right)$-testable, only those properties in $\mathcal{GPP}_{0,1}$ are poly$\left(\frac{1}{\epsilon}\right)$-testable with one-sided error. An analogous result was established for the class $\mathcal{GPP}_{NR}$ by Goldreich and Trevisan [14]. However, as the class $\mathcal{GPP}$ is more general, the class $\mathcal{GPP}_{0,1}$ contains properties that are not covered by the class $\mathcal{GPP}_{NR}$. We build on some techniques used in [14] to establish our characterization, but our characterization does not result from [14] and we rely on different ideas to arrive at it.

The class $\mathcal{GPP}_{0,1}$ is a subclass of $\mathcal{GPP}$ for which the following holds. For every property $P \in \mathcal{GPP}_{0,1}$, there are no absolute bounds on the number of edges between or within parts. If P has a constraint on the edge density between a pair of parts, or within a part, the constraint must be either that the edge density is exactly 0 or that the edge density is exactly 1. In addition, P does not constrain the sizes of the parts. Formally, P is parameterized by an integer k denoting the number of parts and by a function $d_P : [k] \times [k] \to \{0, 1, \perp\}$ denoting the relative edge density that P imposes on parts i and j:

$$d_P\left(i,j\right) = \begin{cases} 1 & \text{if every vertex in part } i \text{ should be connected to every vertex of part } j \\ 0 & \text{if there are no edges between part } i \text{ and part } j \\ \perp & \text{if any number of edges between part } i \text{ and } j \text{ is allowed} \end{cases}$$

Possibly, $i = j$ in which case it is the edge density within a single part. That is, if $d_P\left(i,i\right)$ is 0 or 1 then P forces part i to be an independent set or a clique respectively. It is clear from the definition that there are graph partition properties $P \in \mathcal{GPP}_{0,1}$ that are not part of the class $\mathcal{GPP}_{NR}$ (split graphs for instance).

The main result of our paper is a characterization of the easily testable graph partition properties.

Theorem 2. *A graph property $P \in \mathcal{GPP}$ is easily testable if and only if $P \in \mathcal{GPP}_{0,1}$.*

Recall that a property is easily testable if it is testable by a one-sided error input-size oblivious tester whose query complexity is polynomial in $\frac{1}{\epsilon}$. If we remove the requirement that the dependence on $\frac{1}{\epsilon}$ is polynomial, then the property is said to be *strongly testable*. Alon and Shapira [9] define the notion of a property being *semi-hereditary* (which is a certain relaxation of being hereditary), and show that a graph property P is strongly testable if and only if P is semi-hereditary. Since the properties in $\mathcal{GPP}_{0,1}$ are clearly hereditary, and therefore semi-hereditary, the condition of Alon and Shapira implies that they are strongly testable. However, this is not enough to prove the "if" part of Theorem 2, because being strongly testable does not mean that the tester's query complexity is poly$\left(\frac{1}{\epsilon}\right)$.

Therefore, to prove the "if" direction we give a poly$\left(\frac{1}{\epsilon}\right)$ one-sided error testing algorithm for the property. An alternative proof to the "if" direction follows from [6]. We've chosen to include our direct and self-contained proof because we believe our proof may be used to derive generalizations to the claim, and thus potentially capture other graph properties that don't fall under our specific definition. As for the "only if'' direction, we could use [9] to get that if a property P in $\mathcal{GPP}$ is easily testable (and hence strongly testable), then it is semi-hereditary. We would then need to prove that if $P \in \mathcal{GPP}$ is semi-hereditary, then $P \in \mathcal{GPP}_{0,1}$. Establishing this claim would be essentially the same as our direct proof that if a property P in $\mathcal{GPP}$ is easily testable, then $P \in \mathcal{GPP}_{0,1}$, and would be based on the same proof ingredients.

We next give a brief summary of each of our results.

1.1 A Two-Sided Error Tester for properties in $\mathcal{GPP}$

In order to prove the existence of a (two-sided error) poly$\left(\frac{1}{\epsilon}\right)$-testing algorithm for $\mathcal{GPP}$ we show how to reduce the problem of testing properties in $\mathcal{GPP}$ to testing properties in $\mathcal{GPP}_{NR}$. Recall

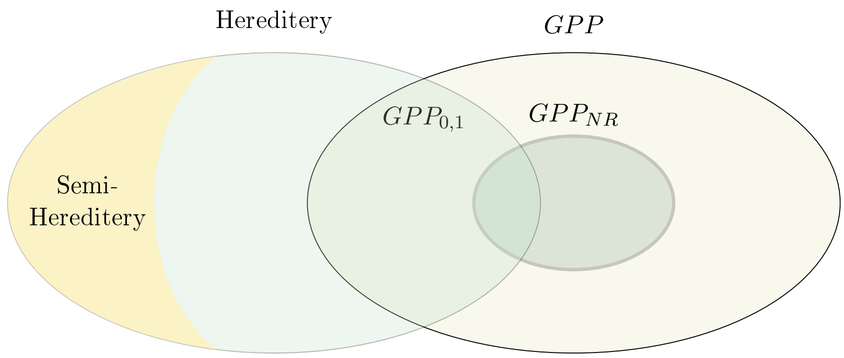

Figure 1.1: Inclusion relations among the graph partition properties

that the difference between properties in $\mathcal{GPP}$ and properties in $\mathcal{GPP}_{NR}$ is that the former include edge density constraints (that are relative to the sizes of the parts), while the latter include only absolute constraints on the sizes of the parts and the number of edges between/within them. We next give the high-level idea of the reduction.

Given a property $P \in \mathcal{GPP}$, we define a collection of properties in $\mathcal{GPP}_{NR}$, by *discretizing* P and replacing the edge-density constraints with absolute constraints on the number of edges. We then run the testing algorithm of [13], denoted $\mathcal{A}$, on G and each property in the constructed collection, with distance parameter $\frac{\epsilon}{2}$. If $\mathcal{A}$ accepts for at least one of these properties, then we accept, and otherwise we reject. The definition of the collection is such that if G satisfies P, then G satisfies at least one of the properties in the collection, so that our algorithm accepts with high constant probability. In order to show that if G is ϵ-far from P, then G is $\frac{\epsilon}{2}$-far from *every* property in the collection, we prove the contrapositive statement. That is, if for at least one of the properties P' in the collection, G is $\frac{\epsilon}{2}$-close to P', then G is $\frac{\epsilon}{2}$-close to P. While the first part of the analysis (regarding G that satisfies P) is fairly immediate, the second part (regarding G that is $\frac{\epsilon}{2}$-far from P) requires a more subtle analysis. Essentially, we need to show how to "fix" G (remove/add edges), so as to obtain a graph that satisfies P. This requires showing the existence of a partition $(V_1, \ldots, V_k)$ that obeys all the constraints defined by P, while closeness to P' only ensures the existence of a partition $(V'_1, \ldots, V'_k)$ that "almost" satisfies P'.

1.2 A One-Sided Error Tester for Properties in $\mathcal{GPP}_{0,1}$

The tester samples $\Theta\left(\frac{k\log(k)}{\epsilon^2}\right)$ vertices uniformly and independently at random, checks whether or not the induced subgraph satisfies P and answers accordingly.[4] Since all the graph partition properties in $\mathcal{GPP}_{0,1}$ are hereditary, it clearly holds that if a graph G satisfies P, then every induced subgraph of G also does. Hence, if $G \in P$, the suggested tester accepts with a probability of 1.

The heart of the proof is in showing that if G is ϵ-far from satisfying P, where $P \in \mathcal{GPP}_{0,1}$, then with high constant probability, the subgraph induced by the sample does not satisfy P. In other words, we would like to show that with high constant probability over the choice of the sample S, every partition $(S_1, \ldots, S_k)$ of the sample violates at least one of the constraints defined by the property P. That is, there is a pair (u, v), where $u \in S_i$ and $v \in S_j$ such that either $(u, v) \in E$ while $d_P(i, j) = 0$, or $(u, v) \notin E$ while $d_P(i, j) = 1$. Such a partition is said to be *invalid*. In order to prove this claim we extend the analysis of Alon and Krivelevich [5] for testing k-colorability. We next give a high-level description of the analysis.

[4] If the tested graph partition property is NP-hard to decide, then the running time is super-polynomial in the sample size, which is unavoidable assuming $P \neq NP$.

Given the sample S, we construct a k-ary tree. Each node in the tree corresponds to a partial partition of the sample. That is, a partition of a subset of the sample. In particular, each internal node corresponds to a *valid* partition (where the root corresponds to a trivial partition of the empty set). If an internal node corresponds to a partition $(S'_1, \dots, S'_k)$ of a subset S' of the sample, then its children correspond to all partitions of $S' \cup \{u\}$ that extend the partition $(S'_1, \dots, S'_k)$ for some sample point $u \in S \setminus S'$. That is, partitions of the form $(S'_1, \dots, S'_{i-1}, S'_i \cup \{u\}, S'_{i+1}, \dots, S'_k)$. Observe that if we obtain a tree for which all leaves correspond to invalid partitions (i.e., that violates some constraint of P), then there is no valid partition of S.

Consider a node x in the tree, corresponding to a partition $(S'_1, \dots, S'_k)$ of $S' \subset S$. For each vertex $v \notin S'$, let $0 \le a_x(v) \le k$ be the number of parts in the partition to which u can be added so that the resulting partition is valid, and let a_x be the sum of $a_x(v)$ taken over all $v \notin S'$. Observe that for the root of the tree, r (which corresponds to $S' = \emptyset$), $a_r = k \cdot n$, and if y is a child of x, then $a_y \le a_x$. If the partition corresponding to y is invalid, then $a_y = 0$. We show that with high constant probability over the choice of the sample, we can construct a tree for which the following holds. For every path in the tree, the value of a_x decreases in a relatively significant manner when comparing each node to its children. This allows us to show that we can obtain a tree in which all partitions corresponding to the leaves are invalid.

1.3 Easily Testable Graph Partition Properties Must be in $\mathcal{GPP}_{0,1}$

Our proof that if a property $P \in \mathcal{GPP}$ is easily testable, then P must be in $\mathcal{GPP}_{0,1}$ is the most technically involved part of this work. The proof consists of several steps, and we next give a high-level outline of these steps. We note that the proof uses the fact that easy testability implies strong testability. That is, we rely on the existence of a one-sided error tester for the property that is oblivious of the size of the graph, but we do not rely on the tester having complexity $\mathrm{poly}\left(\frac{1}{\epsilon}\right)$.

Recall that properties in $\mathcal{GPP}_{0,1}$ are defined by the following types of constraints over graph partitions. First, for each part, either the edge density within the part is unconstrained, or it is constrained in an extreme manner. The latter means that no edges are allowed within the part, or that there must be all possible edges. We say in such a case that the part is *homogeneous*. Similarly, for each pair of parts, either there is no constraint on the edge density between the parts, or it is extreme (no edges, or all edges). Here too we say in the latter case that the pair is homogeneous. Finally, as opposed to $\mathcal{GPP}$, there are no constraints on the sizes of the parts. Observe that the trivial property, that is, the property that contains all graphs, belongs to $\mathcal{GPP}_{0,1}$ (since it can be defined by a single part with no edge-density constraints).

In what follows, for a property $P \in \mathcal{GPP}$ and a graph $G = (V, E)$ satisfying P, a partition $(V_1, \dots, V_k)$ of V is said to be a *witness partition* with respect to P, if in G, $(V_1, \dots, V_k)$ satisfies the constraints imposed by P. We say in such a case that the pair $(G, (V_1, \dots, V_k))$ satisfies P. We first prove that if P is easily testable, then either it is trivial, or for every graph G satisfying P and witness partition $(V_1, \dots, V_k)$, all parts are homogeneous. This is established by showing that if there exists a graph G in P with a witness partition that has some non-homogeneous part, then the premise that P is easily testable implies that P is trivial. The proof uses a type of "multiplying" operation on the graph G. Once we have only homogeneous parts, we can also establish the homogeneity of pairs (among those that are constrained in terms of edge-density). At this point it remains to show that there can be no size constraints on the parts.

To this end we prove a dichotomy claim. Let P' be the same property as P except that there are no size constraints. The claim is that either $P = P'$ or P' is in a certain sense far from P. We then show that the second case cannot hold if P is easily testable. In order to prove the dichotomy claim, we define a certain mathematical program, that, roughly speaking, is related to modifications of graph-partition pairs that satisfy P' to graph-partition pairs that satisfy P (by "fixing" the size constraints). In particular, the existence of a feasible solution corresponds to $P = P'$. On the other hand, if there is no feasible solution, then we show that P' is far from P. This proof involves a probabilistic construction of a graph that satisfies P but is sufficiently far from satisfying P'.

1.4 Related Work

Easily testable graph properties. Besides the class of graph partition properties, there are several results characterizing the set of easily testable graph properties among other classes. Alon [1] proved that the property of being H-free is easily testable if and only if H is bipartite. Alon and Shapira [8] proved that for any graph H besides P_2, P_3, P_4, C_4 and their complements, the property of being induced H-free is not easily testable. It was also shown in [8, 4] that induced H-freeness is easily-testable for P_2, P_3, P_4 and their complements, and the case of C_4 (and its complement) is the only one that remains open. In addition, the graph properties perfectness and comparability were shown to be not easily testable [4]. Gishboliner and Shapira [12] recently made significant progress by providing sufficient and necessary conditions for guaranteeing that a hereditary graph property is easily testable, implying all the positive and negative results mentioned above. It is worth noting, however, that their criteria do not apply to many properties in $\mathcal{GPP}_{0,1}$ (for example, (p,q)-colorability), that are shown to be easily testable in our work.

Testing properties in $\mathcal{GPP}_{NR}$ with one sided error. As mentioned previously, Goldreich and Trevisan [14] studied the one-sided error testability of $\mathcal{GPP}_{NR}$. They showed that every strongly testable property in $\mathcal{GPP}_{NR}$ belongs to a class of properties that generalizes k-colorability. Each property P in this class is defined by a set of pairs $A_P = \{(i,j)\,|0 \leq i,j \leq k\}$, where the property P is the set of k-colorable graphs with the additional constraint that if $(i,j) \in A_P$, then there are no edges between the vertices with color i and the vertices with color j. In addition, the property of being a clique and the property containing all graphs are both stongly testable graph properties in $\mathcal{GPP}_{NR}$.

We build on [14]'s technique of multiplying a graph-partition pair to derive the fact that all the easily testable properties in $\mathcal{GPP}$ only have homogeneous constraints on the edge density within and between parts. The idea of finding assignments to variables corresponding to moving vertices between parts also appears in [14], but they did not have to optimize over a mathematical program, and the assignments they defined could be used straightforwardly to establish the equivalence between a property and its relaxation. One of the main ideas used in [14] to derive the characterization was showing that strongly testable properties in $\mathcal{GPP}_{NR}$ are closed under removal of edges (except for the property of being a clique), and they rely on this fact heavily when deriving the implication regarding the assignments they define and when performing the multiplication. We could not use this idea as it does not hold in our case, because the existence of relative edge bounds in $\mathcal{GPP}$ enables easily testable properties to have lower bounds on the number of edges between or within parts. This is basically the main reason why the set of easily testable properties in $\mathcal{GPP}$ has a richer structure than in $\mathcal{GPP}_{NR}$. This is why we had to use notions that do not appear in their analysis such as weak and strong violations of assignments and rely on the probabilistic method to establish our result.

1.5 Organization

In Section 1.2 we give and analyze the one-sided error tester for properties in $\mathcal{GPP}_{0,1}$ (which proves the "if" direction of Theorem 2). In Section 3 we prove Theorem 1, providing a two-sided error tester for all the properties in $\mathcal{GPP}$. Then, in Section 4, we prove the result regarding the "only if" direction of Theorem 2, that all the easily testable properties in $\mathcal{GPP}$ are in $\mathcal{GPP}_{0,1}$. We've chosen to omit the proofs of some claims in places we believe that the details of the proof are not essential to the reader's understanding. For completeness, all the omitted proofs are given in Appendix B.

2 A One-Sided Error Tester for $\mathcal{GPP}_{0,1}$

In this section we assume P is in $\mathcal{GPP}_{0,1}$ as defined in the introduction. We describe a one-sided error tester for property P with query complexity $\text{poly}\left(\frac{1}{\epsilon}\right)$ and thus show that P is easily testable.

2.1 The Algorithm and its Analysis

The tester is very simple:

One-sided error tester for $P \in \mathcal{GGP}_{0,1}$

- Sample $36\frac{k\log(k)}{\epsilon^2}$ vertices uniformly, independently and at random.
- If the subgraph induced by the selected vertices satisfies P then ACCEPT, otherwise REJECT.

Since all the properties in $\mathcal{GGP}_{0,1}$ are hereditary, it clearly holds that if a graph G satisfies P, then every induced subgraph of G also does. Hence, if $G \in P$, then the above tester accepts with probability 1. From this point on we focus on the case that G is ϵ-far from satisfying P.

Let $G = (V, E)$ be a graph over n vertices such that G is ϵ-far from satisfying P. We first introduce several notions that we use in our proof that the tester rejects G with high constant probability and then provide the proof itself.

2.1.1 Partial partitions and their extensions

A *partial partition* of V is a partition of a subset of the vertices into k parts. Formally,

Definition 2. *A partial partition of V into k parts is a function $\pi : V \to [k] \cup \{\perp\}$. For a vertex $v \in V$, we say that π assigns v to part i if $\pi(v) = i$. If $\pi(v) = \perp$, then we say that v is not assigned by π.*

Given a partial partition π we denote by V_π the set of vertices assigned by π. That is,

$$V_\pi = \{v \in V : \pi(v) \neq \perp\} \ . \tag{2.1}$$

We also say in that case that π is *complete* with respect to V_π. Note that a partial partition is not necessarily strictly partial. That is, a partial partition that assigns all the vertices of G (i.e., that is complete with respect to V) is simply a partition.

Recall that $d : [k] \times [k] \to \{0, 1, \perp\}$ denotes the function defining the constraints of P.

Definition 3. *Given a partial partition π we say that a pair of vertices $\{u, v\}$ where $u, v \in V_\pi$ is a violating pair with respect to π and P if any of the following two cases holds:*
- $(u, v) \in E$ *and* $d_P(\pi(u), \pi(v)) = 0$
- $(u, v) \notin E$ *and* $d_P(\pi(u), \pi(v)) = 1$

We say that a partial partition π is valid with respect to P if there are no violating pairs with respect to π and P.

Let π be a valid partial partition of V.

Definition 4. *For each vertex $v \notin V_\pi$ we let $A_\pi(v)$ denote the set of indices of parts to which v can be assigned without introducing a violating pair (given that the vertices in V_π are assigned according to π). Namely,*

$$A_\pi(v) = \{i \in [k] \mid d_P(i, \pi(u)) \neq 0,\ \forall u \in V_\pi \cap \Gamma(v)\} \cup \{i \in [k] \mid d_P(i, \pi(u)) \neq 1,\ \forall u \in V_\pi \backslash \Gamma(v)\}\ .$$

We can extend π by creating a new partition π' that agrees with π on all the vertices belonging to V_π and also assigning the vertex v to some part $i \in A_\pi(v)$.

Definition 5. *We say that a partial partition π' is a v-extension of π for $v \notin V_\pi$ if π' is obtained from π by assigning each vertex $u \in V_\pi$ to $\pi(u)$ and in addition assigning v to a part $i \in [k]$.*

If π' is the v-extension of π in which $\pi'(v) = i$, then we say that we extend π by $v \to i$ to obtain π'. We also say that π' is the $v \to i$ extension of π.

We say that the $v \to i$ extension of π is valid with respect to P if $i \in A_\pi(v)$. Otherwise, we say that the $v \to i$ extension is an invalid v-extension of π.

Clearly, for every vertex $v \notin V_\pi$, the number of v-extensions of π is k and the number of valid v-extensions of π is $|A_\pi(v)|$.

Definition 6. *We let $|A_\pi|$ denote the total number of valid single-vertex extensions to the partial partition π. Formally,*

$$|A_\pi| = \sum_{v \notin S} |A_\pi(v)| \tag{2.2}$$

Intuitively, $|A_\pi|$ measures how much freedom we have in assigning more vertices given that the set of vertices in V_π have to be assigned according to π.

In addition, given a vertex $v \notin V_\pi$, we define the *price* of the $v \to i$ extension of π where $i \in A_\pi(v)$ as the amount of freedom lost by the extension. Formally,

Definition 7. *For $v \notin V_\pi$ and $i \in A_\pi(v)$, the price of extending π by $v \to i$, denoted by $\Delta_\pi(v \to i)$, is defined as follows:*

$$\Delta_\pi(v \to i) = |A_\pi| - |A_{\pi_{v \to i}}| \tag{2.3}$$

We define the price of extending π by v, denoted $\Delta_\pi(v)$, as the lowest price that can possibly be paid by doing so. Namely,

$$\Delta_\pi(v) = \min_{i \in A_\pi(v)} \{\Delta_\pi(v \to i)\} \tag{2.4}$$

If $A_\pi(v) = \emptyset$ then $\Delta_\pi(v)$ is undefined.

The interpretation of the price $\Delta_\pi(v \to i)$ is that if we extend π by $v \to i$ we lose $\Delta_\pi(v \to i)$ single-vertex extensions in the next extension round. That is, the partial partition π induces $\Delta_\pi(v \to i)$ more valid single-vertex extensions than the $v \to i$ extension of π. Clearly, $\Delta_\pi(v \to i)$ is non-negative.

Recall that all the definitions above assume that π is a valid partial partition with respect to P.

2.1.2 The tree of partial partitions

Let S be a sample of s vertices from the graph G, selected uniformly and independently (so that the same vertex may be selected more than once). We consider some arbitrary but fixed order over the s vertices in S. In this section we use S to define a k-ary tree T_S in which every node x in the tree is labeled by a particular partial partition π_x. The partial partition of a node is not necessarily valid with respect to P. Before giving the formal definition of T_S we present a general high level description of what the tree looks like.

Let x be a node in the tree. As aforesaid, the node x is labeled by a partial partition π_x. The partial partition π_x only assigns vertices of S (that is, $V_{\pi_x} \subseteq S$). If π_x is invalid with respect to P, then x has no children. For each internal node x, each child of x is labeled by a single-vertex extension of π_x where all the children are extended by the same vertex v_x. That is, each child of the node x is labeled by a partition that agrees with π_x on all the vertices assigned by π_x and in the addition, the i^{th} child partition assigns v_x to part $i \in [k]$. We refer to v_x as the *branching vertex* of the node x.

Now we formally define the tree T_S using structural induction.

1. As the basis of the structural induction we construct the root of T_S. The root of T_S is labeled by the the empty partial partition $\pi_\emptyset = (\emptyset, \dots, \emptyset)$ that does not assign any vertex.

2. Each step of the induction corresponds to a particular vertex in S. That is, to construct T_S we iterate over the vertices of S (according to the aforementioned fixed order) and extend the tree according to the instructions below.

3. Let $v \in S$ be the vertex considered in the current step. We use the following rules to extend the leaves of the tree built so far. For each leaf x where π_x is valid with respect to P (we do not extend leaves labeled by invalid partial partitions):

   (a) If $v \in V_{\pi_x}$ or $A_{\pi_x}(v) \neq \emptyset$ and $\Delta_{\pi_x}(v) < \frac{1}{2}\epsilon n$, then we do nothing with x.

   (b) Otherwise ($v \notin V_{\pi_x}$ and either $A_{\pi_x}(v) = \emptyset$ or $\Delta_{\pi_x}(v) \geq \frac{1}{2}\epsilon n$), we add the k children corresponding to the possible v-extensions of π_x, where child i is labeled by the $v \to i$ extension of π_x. In that case, the branching vertex $v_x = v$.

Observe that, by construction, every internal node of T_S is labeled by a partial partition that is valid with respect to P. We next establish some additional properties of the tree T_S.

Claim 3. *Let G be a graph and let S be any (ordered) sample of the vertices of G. The depth of T_S is at most $\frac{2k}{\epsilon}$.*

Proof. Consider a path from the root of T_S to any leaf. Let x and y be a pair of consecutive nodes in the path such that x is the parent of y. Suppose y is not the last node on the path (that is, y is not a leaf). Hence, both π_x and π_y are valid partial partitions. By the definition of T_S, the partial partition π_y is a v_x-extension of π_x such that

$$\Delta_{\pi_x}(v_x) \geq \frac{1}{2}\epsilon n \; . \tag{2.5}$$

Hence,

$$|A_{\pi_x}| - |A_{\pi_y}| \geq \frac{1}{2}\epsilon n \; . \tag{2.6}$$

Therefore, in every step in which a prefix of the path is extended, the value of $|A_{\pi_x}|$ decreases by at least $\frac{1}{2}\epsilon n$. Since $|A_{\pi_x}|$ is non-negative (as it is the size of a set), the total number of nodes in the path is bounded from above by $\frac{|A_{\pi_\emptyset}|}{\frac{1}{2}\epsilon n}$ where $\pi_\emptyset$ is the partition of the root. Since $|A_{\pi_\emptyset}| = kn$ (as $|A_{\pi_\emptyset}(v)| = k$ for every vertex v, the length of the path is at most $\frac{nk}{\frac{1}{2}\epsilon n} = \frac{2k}{\epsilon}$. This holds for any path and therefore the depth of T_S is at most $\frac{2k}{\epsilon}$. □

Definition 8. *Given a complete partition ϕ of S, we define the path corresponding to ϕ in the tree T_S, denoted path(T_S, ϕ), as the following sequence of nodes. The first node in the sequence path(T_S, ϕ) is the root of T_S. Suppose we constructed a non-empty prefix of path(T_S, ϕ), where the last node in the prefix is x. If x is not a leaf of T_S, then the next node in the path is the child of x labeled by the $v_x \to \phi(v_x)$ extension of π_x. If x is a leaf, then x is the last node of the path.*

Claim 4. *Let G be a graph and let S be an (ordered) sample of the vertices of G. If each leaf of T_S is labeled by an invalid partition with respect to P, then every partition of S is invalid with respect to P.*

Proof. Let ϕ be a complete partition of S, and let x be the leaf at the end of path(T_S, ϕ). By the premise of the claim, x is labeled by an invalid partial partition. By Definition 8, $\phi(v) = \pi_x(v)$ for every $v \in V_{\pi_x}$. Hence, every violating pair with respect to π_x and P also serves as a violating pair with respect to ϕ and P. Therefore, since π_x is invalid with respect to P, so is ϕ. □

2.1.3 With high probability, every partition of S has a violation

Recall that we are working under the premise that the graph G is ϵ-far from P.

Claim 5. *Let G be a graph that is ϵ-far from P and let ϕ be a partial partition of V that is valid with respect to P. There are at least $\frac{1}{2}\epsilon n$ vertices $v \notin V_\phi$ satisfying either $A_\phi(v) = \emptyset$ or $\Delta_\phi(v) \geq \frac{1}{2}\epsilon n$.*

Proof. Assume, contrary to the claim, that the number of vertices $v \notin V_\phi$ satisfying either $A_\phi(v) = \emptyset$ or $\Delta_\phi(v) \geq \frac{1}{2}\epsilon n$ is less than $\frac{1}{2}\epsilon n$. We next construct a complete partition ϕ' of V that has at most ϵn^2 violating pairs with respect to the constraints imposed by P, in contradiction to G being ϵ-far from P.

The partition ϕ' agrees with ϕ on V_ϕ. For each vertex $v \notin V_\phi$, if $A_\phi(v) = \emptyset$, then we assign v to an arbitrary part. Otherwise ($A_\phi(v) \neq \emptyset$), we assign v to the part $i \in A_\phi(v)$ that minimizes $\Delta_\phi(v \to i)$. That is, we choose the part i satisfying $\Delta_\phi(v \to i) = \Delta_\phi(v)$. We next bound from above the number of violating vertex-pairs with respect to ϕ' and P.

First, let v be a vertex satisfying $A_\phi(v) \neq \emptyset$ and $\Delta_\phi(v) < \frac{1}{2}\epsilon n$. Suppose $\phi'(v) = i$. The price of assigning v to part i equals the maximum number of violating pairs v could be involved in. That is, the vertex v is involved in at most $\Delta_\phi(v \to i)$ violating pairs. By the definition of ϕ', the choice of i ensures that:

$$\Delta_\phi(v \to i) = \Delta_\phi(v) < \frac{1}{2}\epsilon n\ . \tag{2.7}$$

Hence, v is involved in at most $\frac{1}{2}\epsilon n$ violating pairs. The number of vertices satisfying the condition $A_\phi(v) \neq \emptyset$ and $\Delta_\phi(v) < \frac{1}{2}\epsilon n$ is at most n and therefore, these vertices contribute at most $\frac{1}{2}\epsilon n^2$ violating pairs to ϕ'.

Let v be a vertex that does not satisfy the above condition. That is, either $A_\phi(v) = \emptyset$ or $\Delta_\phi(v) \geq \frac{1}{2}\epsilon n$. Recall we assume that the number of these vertices is at most $\frac{1}{2}\epsilon n$. Each such vertex is involved in at most n violating pairs. Hence, the set of the vertices that do not satisfy the condition contribute at most $\frac{1}{2}\epsilon n^2$ violating pairs to ϕ'.

It follows that the total number of violating pairs with respect to ϕ' and P is bounded from above by $2 \cdot \frac{1}{2}\epsilon n^2 = \epsilon n^2$. □

The next claim concludes the analysis of the tester.

Claim 6. *Let G be a graph that is ϵ-far from P. If S is a sample of $36\frac{k\log(k)}{\epsilon^2}$ vertices that are selected uniformly, independently at random, then with a probability of at least $\frac{2}{3}$ over the choice of S, all the leaves of T_S are labeled by an invalid partition with respect to P.*

Proof. By construction, every internal node of T_S has exactly k children and by Claim 3 the depth of T_S is at most $\frac{2k}{\epsilon}$. Therefore, T_S can be embedded in the complete k-ary tree of depth $\frac{2k}{\epsilon}$. We denote that tree by $T_{k,\frac{2k}{\epsilon}}$. The number of nodes in $T_{k,\frac{2k}{\epsilon}}$ is:

$$\sum_{i=0}^{\frac{2k}{\epsilon}} k^i \le k^{\frac{2k}{\epsilon}+1} \tag{2.8}$$

Recall that during the construction of T_S, in each step we have a partial tree and we consider the next vertex selected in the sample. For each leaf x in the current partial tree, let $\mathcal{E}_x$ denote the event that x is extended in the current step. That is, $\mathcal{E}_x$ is the event that the next selected vertex v is such that $v \notin V_{\pi_x}$ and either $A_{\pi_x}(v) = \emptyset$ or $\Delta_{\pi_x}(v) \ge \frac{1}{2}\epsilon n$. By Claim 5, $\Pr[\mathcal{E}_x] \ge \frac{\epsilon}{2}$. By the definition of T_S, the occurrence of the event $\mathcal{E}_x$ for a leaf x results in extending x by k children.

Let x be a node in $T_{k,\frac{2k}{\epsilon}}$. If after $36\frac{k\log(k)}{\epsilon^2}$ rounds (of selecting random vertices), x is a leaf of T_S, then the total number of times the event has occurred for nodes on the path from the root to x equals the depth of x. Since in each round, $\Pr[\mathcal{E}_x]$ is at least $\frac{\epsilon}{2}$, the probability that x does not become a "dead-end" of T_S (that is, π_x is valid) after $36\frac{k\log(k)}{\epsilon^2}$ steps is bounded from above by the probability that the Binomial random variable $X \sim B\left(36\frac{k\log(k)}{\epsilon^2}, \frac{\epsilon}{2}\right)$ is less than $\frac{2k}{\epsilon}$.

$$\Pr\left[x \text{ is not a dead-end after } 36\frac{k\log(k)}{\epsilon^2} \text{ steps}\right] \le \Pr\left[X < \frac{2k}{\epsilon}\right] . \tag{2.9}$$

By Chernoff's inequality,

$$\Pr\left[X < \frac{2k}{\epsilon}\right] \le \exp\left(-\frac{\left(E[X] - \frac{2k}{\epsilon}\right)^2}{2E[X]}\right) . \tag{2.10}$$

Since X is a Binomial random variable, its expectation is:

$$E[X] = 36\frac{k\log(k)}{\epsilon^2} \cdot \frac{\epsilon}{2} = \frac{18k\log(k)}{\epsilon} . \tag{2.11}$$

Combining Equations (2.10) and (2.11) we obtain:

$$\Pr\left[X < \frac{2k}{\epsilon}\right] \le \exp\left(-\frac{\left(\frac{18k\log(k)}{\epsilon} - \frac{2k}{\epsilon}\right)^2}{\frac{24k\log(k)}{\epsilon}}\right) \le \exp\left(-\frac{\left(\frac{16k\log(k)}{\epsilon}\right)^2}{\frac{24k\log(k)}{\epsilon}}\right) \le k^{-\frac{3k}{\epsilon}} . \tag{2.12}$$

Hence, by the union bound over all at most $k^{\frac{2k}{\epsilon}+1}$ nodes of $T_{k,\frac{2k}{\epsilon}}$:

$$\Pr[\text{Not all leaves are dead-ends}] \le k^{\frac{2k}{\epsilon}+1} \cdot k^{-\frac{3k}{\epsilon}} = \frac{k}{k^{\frac{k}{\epsilon}}} \le \frac{1}{3} , \tag{2.13}$$

and the claim follows. ∎

The size of our sample S is $36\frac{k\log(k)}{\epsilon^2}$. Therefore, with a probability of at least $\frac{2}{3}$, all the leaves of T_S are labeled by invalid partitions with respect to P, and by Claim 4, S cannot be validly partitioned. Hence, the probability that our tester rejects is at least $\frac{2}{3}$.

3 A Two-Sided Error Tester for $\mathcal{GPP}$

3.1 The Class $\mathcal{GPP}_{NR}$

The class $\mathcal{GPP}_{NR}$ is the set of graph partition properties with no relative edge bounds. The family of these problems has a two-sided tester whose query complexity is polynomial in the proximity parameter [13]. The tester can be repeated several times to boost the probability of its correctness. We denote the tester for $\mathcal{GPP}_{NR}$ by $\mathsf{AbsTester}$ and set its correctness probability to at least $\left(\frac{2}{3}\right)^{\frac{1}{t}}$ where $t = \left(\frac{8k}{\epsilon}\right)^k \left(\frac{8k^2}{\epsilon}\right)^{k^2}$. We show how we can test properties in the generalized class $\mathcal{GPP}$, by invoking $\mathsf{AbsTester}$ several times. In order to describe the algorithm we first have to define the notion of a characterizing vector and the concept of structural equivalence.

3.2 The Characterizing Vector and Structural Equivalence

Definition 9. *Given an n-vertex graph G and a partition $(V_1, ..., V_k)$ we define the characterizing vector $\vec{\sigma}$ as follows. The vector can be indexed by either a single index $1 \le i \le k$ or by a pair i, j s.t. $i \le j$ and $1 \le i, j \le k$.*

$\vec{\sigma}_i = \frac{|V_i|}{n}$

$\vec{\sigma}_{ij} = \frac{e(V_i,V_j)}{n^2}$

We note that while a graph and a partition can be characterized by a unique vector, a given vector can characterize a variety of graph-partition pairs.

Definition 10. *We say that two graph-partition pairs are structurally equivalent if they're characterized by the same vector.*

We note that the structural equivalence is an equivalence relation of the graph-partition pairs.

Lemma 1 (The Structural Equivalence Lemma)**.** *Let G^* be a graph satisfying a GPP property P and let $(V_1^*, ..., V_k^*)$ be a witness partition. That is, a partition of $V[G]$ satisfying all the constraints specified by P. Denote by $\vec{\sigma}^*$ the characterizing vector of $(G^*, (V_1^*, ..., V_k^*))$. Let $(G, (V_1, ..., V_k))$ be a graph-partition pair whose characterizing vector is also $\vec{\sigma}^*$. Then G satisfies P as well.*

The correctness of *The Structural Equivalence Lemma* is directly derived from the way structural equivalence is defined.

3.3 Satisfying Interval Vectors

We define an *interval vector* $\vec{\rho}$ in a similar manner to how we defined a characterizing vector of a graph except that each element of $\vec{\rho}$ is an interval. That is, an interval vector $\vec{\rho}$ can be indexed by either a single index $1 \le i \le k$ or by a pair i, j s.t $i \le j$ and $1 \le i, j \le k$.

1. $\vec{\rho}_i = \left[\vec{\rho}_i^L, \vec{\rho}_i^U\right]$ is an interval representing a constraint saying $\vec{\rho}_i^L n \le |V_i| \le \vec{\rho}_i^U n$.
2. $\vec{\rho}_{ij} = \left[\vec{\rho}_{ij}^L, \vec{\rho}_{ij}^U\right]$ is an interval representing a constraint saying $\vec{\rho}_{ij}^L n^2 \le e\left(V_i, V_j\right) \le \vec{\rho}_{ij}^U n^2$.

Given a graph-partition pair $(G,(V_1,...,V_k))$ we can compute the corresponding characterizing vector $\vec{\sigma}$. We say that $\vec{\sigma}$ satisfies an interval vector $\vec{\rho}$ if every element of $\vec{\sigma}$ resides in the corresponding interval of $\vec{\rho}$, denoted $\vec{\sigma} \in \vec{\rho}$. Clearly, a graph $G-(V,E)$ satisfies the constraints induced by an interval vector $\vec{\rho}$ if and only if there exists a a partition of V such that the characterizing vector $\vec{\sigma}$ of the graph-partition pair satisfies $\vec{\rho}$. Every interval vector $\vec{\rho}$ corresponds to a property in $\mathcal{GPP}_{NR}$ denoted $GPP_{NR}(\vec{\rho})$.

We can split an interval vector $\vec{\rho}$ into two interval vectors, $\vec{a}$ and $\vec{b}$, where $\vec{a}$ is the part of $\vec{\rho}$ representing all the size constraints and $\vec{b}$ is the part of $\vec{\rho}$ representing the edge density constraints. We say that $\vec{\sigma} \in \vec{a}$ if $\vec{\sigma}$ satisfies the constraints in $\vec{a}$ and we say that $\vec{\sigma} \in \vec{b}$ if $\vec{\sigma}$ satisfies the constraints in $\vec{b}$.

3.4 Feasible and Compatible Interval Vectors

Clearly, not every interval vector attains a satisfying graph-partition pair. We say that such interval vectors are not feasible. Formally,

Definition 11. *An interval vector $\vec{\rho}$ is feasible if there exists an n-vertex graph $G=(V,E)$ and a partition $(V_1,...,V_k)$ of V s.t. the characterizing vector of $(G,(V_1,...,V_k))$ satisfies $\vec{\rho}$.*

A feasible interval vector $\vec{\rho}$ is compatible with a property P in $\mathcal{GPP}_{NR}$ if there exists an n-vertex graph G satisfying P and a partition $(V_1,...,V_k)$ of V s.t. the characterizing vector of $(G,(V_1,...,V_k))$ satisfies $\vec{\rho}$.

Given an interval vector $\vec{\rho}$ we can determine whether it is both feasible and compatible with P by checking if the feasible region defined by a specific set of quadratic constraints is empty. In particular, the vector $\vec{\rho}$ is feasible and compatible with P if and only if some quadratic program is feasible. The program is defined over the following set of $k+k^2$ real decision variables: $\{x_i\}_{i=1}^{k} \cup \{y_{ij}\}_{(i,j)\in[k]\times[k]}$. The variables $\{x_i\}_{i=1}^{k}$ correspond to sizes of parts (relative to n) and the variables $\{y_{ij}\}_{(i,j)\in[k]\times[k]}$ correspond to number of edges between parts (or within a part for $i=j$) relative to n^2. To simplify the description of the program we assume that each entry of the vector $\vec{\rho}$ holds an interval that is contained in the corresponding interval specified by P. That is, we denote by $\left[\rho_i^L,\rho_i^U\right]$ an interval interval in the vector $\vec{\rho}$ and assume that the constraint imposed by P on the size of part i is an interval that contains any point belonging to $\left[\rho_i^L,\rho_i^U\right]$. We assume the same for the ρ_{ij} entries of the interval vector. Hence, in the following description of the program, the entries $\left[\rho_i^L,\rho_i^U\right]$ and $\left[\rho_{ij}^L,\rho_{ij}^U\right]$ are entries of the interval vector $\vec{\rho}$, and the entries $\left[\alpha_{ij}^L,\alpha_{ij}^U\right]$ come from the description of the property P. The full quadratic program is defined specified in Appendix A. We can simplify the program defined in Appendix A and eliminate the variables $\{y_{ij}\}_{(i,j)\in[k]\times[k]}$ by expressing the constraints on those variables in terms of $\{x_i\}_{i=1}^{k}$. That is, the program has a feasible solution if and only if the following one does.

$$\sum_{i=1}^{k} x_i = 1 \tag{3.1}$$

$$\forall i \in [k] : \rho_i^L \le x_i \le \rho_i^U \tag{3.2}$$

$$\forall i \in [k] : \max\left\{\alpha_{ii}^L,\rho_{ii}^L\right\} \le x_i^2 \le \min\left\{\alpha_{ii}^U,\rho_{ii}^U\right\} \tag{3.3}$$

$$\forall i \ne j \in [k]\times[k] : \max\left\{2\alpha_{ij}^L,2\rho_{ij}^L\right\} \le x_i \cdot x_j \le \min\left\{2\alpha_{ij}^U,2\rho_{ij}^U\right\} \tag{3.4}$$

The problem of deciding whether the above program has a feasible solution is equivalent to deciding if it belongs to a set known as The Existential Theory of the Reals. This decision is decidable in time exponential in k [15, 16]. Hence, given an interval vector we can decide whether or not it is both feasible and compatible with the property P.

3.5 The Two-Sided Error Tester

We denote the parameters of the original problem by $k, \left\{\widetilde{\rho}_i^L, \widetilde{\rho}_i^U, \widetilde{\rho}_{ij}^L, \widetilde{\rho}_{ij}^U, \widetilde{\alpha}_{ij}^L, \widetilde{\alpha}_{ij}^U\right\}_{1 \leq i,j \leq k}$ and reserve the Tilde-free symbols for the sequence of sub-problems to be defined soon. The first stage of the algorithm is to generate interval vectors on which we're going to test the input graph using the AbsTester. We compute the set of such interval vectors by splitting the intervals defining the property $\mathcal{P}$.

3.5.1 Splitting the intervals

We describe a process in which we split the property's intervals into smaller sub-intervals. The algorithm is going to combine those sub-intervals into interval vectors corresponding to easily solvable sub-problems: ones without relative formulations.

1. **Size Intervals:** For every part i we take the interval $\left[\widetilde{\rho}_i^L, \widetilde{\rho}_i^U\right]$ and split it into non-overlapping sub-intervals, each of size $\epsilon' = \frac{1}{8k}\epsilon$ (the size of the last sub-interval is allowed to be smaller). We denote the set of the sub-intervals by A_i.

2. **Edge Density Intervals:** Let $\prod_{i=1}^k A_i$ be the Cartesian product of the k sets of intervals. That is, a vector $\vec{\rho} \in \prod_{i=1}^k A_i$ holds k intervals, one for each part, each of size at most ϵ'. The cardinality of $\prod_{i=1}^k A_i$ is at most $\left(\frac{1}{\epsilon'}\right)^k$. Let $\vec{a} = \left(\left[\rho_1^L, \rho_1^U\right], ..., \left[\rho_k^L, \rho_k^U\right]\right) \in \prod_{i=1}^k A_i$. For every $1 \leq i, j \leq k$ where $i < j$ we define:

$$\rho_{ij}^L = \max\left\{\widetilde{\rho}_{ij}^L, \widetilde{\alpha}_{ij}^L \cdot \rho_i^L \cdot \rho_j^L\right\} \tag{3.5}$$

$$\rho_{ij}^U = \min\left\{\widetilde{\rho}_{ij}^U, \widetilde{\alpha}_{ij}^U \cdot \rho_i^U \cdot \rho_j^U\right\} \tag{3.6}$$

For $i = j$:

$$\rho_{ii}^U = \min\left\{\rho_{ii}^U, \widetilde{\alpha}_{ii}^U \cdot \left(\rho_i^U\right)^2\right\} \tag{3.7}$$

$$\rho_{ii}^L = \max\left\{\rho_{ii}^L, \widetilde{\alpha}_{ii}^L \cdot \left(\rho_i^L\right)^2\right\} \tag{3.8}$$

Now we take the interval $\left[\rho_{ij}^L, \rho_{ij}^U\right]$ and split it into non-overlapping sub-intervals, each of size $\epsilon'' = \frac{1}{8k^2}\epsilon$ (again, the size of the last sub-interval is allowed to be smaller). We denote the set of the sub-intervals by $B_{ij}^{\vec{a}}$.

For every vector $\vec{a} \in \prod_{i=1}^k A_i$ we iterate over the vectors $\vec{b} \in \prod_{i \leq j} B_{ij}^{\vec{a}}$ and combine $\vec{a}$ and $\vec{b}$ into a single interval vector $\vec{\rho}$ in such a way that:

$$\vec{\rho}_i = \vec{a}_i \tag{3.9}$$

$$\vec{\rho}_{ij} = \vec{b}_{ij} \tag{3.10}$$

For every vector $\vec{a} \in \prod_{i=1}^k A_i$ the number of interval vectors we have is at most $\left(\frac{1}{\epsilon''}\right)^{k^2}$. Hence, the total number of interval vectors we consider is at most $\left(\frac{1}{\epsilon'}\right)^k \cdot \left(\frac{1}{\epsilon''}\right)^{k^2}$. We only consider feasible vectors that are compatible with the original problem.

3.5.2 Running AbsTester on the interval vectors

For each feasible compatible interval vector $\vec{\rho}$ we run AbsTester on $\vec{\rho}$ with proximity parameter $\frac{1}{2}\epsilon$. If any of them accepts, we accept too. If they all reject, so do we.

3.6 Analysis of the Two-Sided Error Tester

We have to show that the algorithm outputs the right answer with high probability. That is, we have to show that if a graph G satisfies P then the tester accepts with high probability and that if G is far from satisfying P then the tester rejects with high probability.

3.6.1 If G is a YES instance, we accept with high probability

Let G be a YES instance and let $(V_1, ..., V_k)$ be a witness partition of $V[G]$. That is, $(V_1, ..., V_k)$ satisfies the constraints of the original problem. Let $\vec{\sigma}$ be the characterizing vector of $(G, (V_1, ..., V_k))$. For every part $1 \leq i \leq k$:

$$\vec{\sigma}_i \in \left[\widetilde{\rho}_i^L, \widetilde{\rho}_i^U\right] \tag{3.11}$$

Therefore, there exists an interval $\left[\rho_i^L, \rho_i^U\right] \in A_i$ s.t.:

$$\vec{\sigma}_i \in \left[\rho_i^L, \rho_i^U\right] \tag{3.12}$$

Hence, there exists $\vec{a} \in \prod_{i=1}^k A_i$ s.t.

$$\vec{\sigma} \in \vec{a} \tag{3.13}$$

Claim 7. *There exists $\vec{b} \in \prod_{i \leq j} B_{ij}^{\vec{a}}$ s.t. $\vec{\sigma} \in \vec{b}$.*

The proof of Claim 7 is given in Appendix B.1. It follows that there exists a valid interval vector $\vec{\rho}$ such that $\vec{\sigma} \in \vec{\rho}$ and therefore G will be accepted in the $\vec{\rho}$ iteration with high probability (as AbsTester is a 2-sided error tester for the property).

3.6.2 If G is far from satisfying the property, we reject with high probability

Suppose G is ϵ-far from satisfying P. We show that G is $\frac{1}{2}\epsilon$-far from $GPP_{NR}(\vec{\rho})$ for every feasible compatible interval vector $\vec{\rho}$ (among the ones we've constructed).

Suppose for the sake of contradiction that G is not $\frac{1}{2}\epsilon$-far from $GPP_{NR}(\vec{\rho})$ for some feasible $\vec{\rho}$. We show how by modifying at most ϵn^2 vertex pairs we make G satisfy the original property.

First we make $\frac{1}{2}\epsilon n^2$ edge modifications to make G satisfy $GPP_{NR}(\vec{\rho})$. Denote by G' the graph resulting from G by applying the above edge modifications. Let $(V_1', ..., V_k')$ be a witness partition demonstrating G' satisfies $GPP_{NR}(\vec{\rho})$.

Since $\vec{\rho}$ is feasible there exists a graph $G^* \in \mathcal{P}$ and a witness partition $(V_1^*, ..., V_k^*)$ satisfying the constraints induced by $GPP_{NR}(\vec{\rho})$. Let $\vec{\sigma}^*$ be the characterizing vector of $(G^*, (V_1^*, ..., V_k^*))$. We show how to modify $(G', (V_1', ..., V_k'))$ in such a way that they'll be characterized by the vector $\vec{\sigma}^*$. We do this in four steps.

1. We move vertices between the various parts in $(V_1', ..., V_k')$ so that for every $1 \leq i \leq k$ the number of vertices in V_i becomes $\vec{\sigma}_i^* n$. We choose a sequence of movements which is minimal in the sense that it involves the smallest possible number of vertices being assigned to a new part. We denote the obtained partition by $(V_1, ..., V_k)$.

2. We remove all the edges from every vertex we've moved in the previous step. Since both $(G', (V'_1, ..., V'_k))$ and $(G^*, (V^*_1, ...V^*_k))$ satisfy the constraints induced by $\vec{\rho}$, for every i we have $||V'_i| - |V^*_i|| \leq \epsilon' n$. Therefore the procedure of moving the vertices so that our graph agrees with G^* on each part's size requires moving less than $k\epsilon' n$ vertices. Hence, the cost of the current step (number of edge modifications) is at most $k\epsilon' n^2$.

3. For every i, j s.t $i \leq j$ if $e'(V_i, V_j) > \vec{\sigma}^*_{ij} n^2$ we remove edges between V_i and V_j until we have exactly $\vec{\sigma}^*_{ij} n^2$ edges between the two parts. Since both $(G', (V'_1, ..., V'_k))$ and $(G^*, (V^*_1, ...V^*_k))$ satisfy the constraints induced by $\vec{\rho}$, for every such i, j we have $e(V_i, V_j) - e^*\left(V^*_i, V^*_j\right) \leq \epsilon'' n^2$. Therefore the cost of the current step (number of edge modifications) is at most $k^2 \epsilon'' n^2$.

4. For every i, j s.t $i \leq j$ if $e'(V_i, V_j) < \sigma^*_{ij} n^2$ we add edges between V_i and V_j until we have exactly $\sigma^*_{ij} n^2$ edges between the two parts. The cost of the step in the worst case involves both making up for the edges we've removed in step 2 (by adding them all) and then adding at most $e^*\left(V^*_i, V^*_j\right) - e(V_i, V_j) \leq \epsilon'' n^2$ edges to narrow the gap. Thus, the total cost of this step is at most $k\epsilon' n^2 + k^2 \epsilon'' n^2$.

Clearly, the graph resulting from applying the above four steps together with the partition $(V_1, ..., V_k)$ are characterized by the vector $\vec{\sigma}^*$. Since G^* satisfies P the structural equivalence lemma implies that the resulting graph satisfies $\mathcal{P}$. The number of edge modifications we had to make in order to make G' satisfy $\mathcal{P}$ is at most:

$$k\epsilon' n^2 + k^2 \epsilon'' n^2 + k\epsilon' n^2 + k^2 \epsilon'' n^2 = 2k\epsilon' n^2 + 2k^2 \epsilon'' n^2$$

$$= \left(\frac{2k}{8k}\epsilon + \frac{2k^2}{8k^2}\epsilon\right) n^2 = \frac{1}{2}\epsilon n^2$$

Together with the (at most) $\frac{1}{2}\epsilon n^2$ edge modifications we've made in order to transform G into G' we get a total of less than ϵn^2 edge modifications for making G satisfying the original property P. This contradicts G being ϵ-far from satisfying the property. Hence, we conclude that G must be $\frac{1}{2}\epsilon$-far from satisfying $GPP_{NR}(\vec{\rho})$ for every feasible $\vec{\rho}$.

Let $\Pr[Reject]$ be the probability of our algorithm rejecting G when G is ϵ-far from satisfying $\mathcal{P}$. Remember we set AbsTester to succeed with probability at least $\left(\frac{2}{3}\right)^{\frac{1}{t}}$ where $t = \left(\frac{8k}{\epsilon}\right)^k \left(\frac{8k^2}{\epsilon}\right)^{k^2}$. We observe t is an upper bound on the number of interval vectors $\vec{\rho}$ we consider.

$$\Pr[REJECT] \geq \left(\left(\frac{2}{3}\right)^{\frac{1}{t}}\right)^t = \frac{2}{3} \tag{3.14}$$

4 Easily Testable Graph Partition Properties Must be in $\mathcal{GPP}_{0,1}$

As discussed in Section 2, all the graph partition properties in $\mathcal{GPP}_{0,1}$ are easily testable. In this section we provide the proof for the claim that if a graph partition property is easily testable, then $P \in \mathcal{GPP}_{0,1}$. We first define the concept of a t-multiplier, which is used several times throughout the proof. In what follows, given a graph $G = (V, E)$ and a set of vertices $U \subseteq V$ we denote by $G[U]$ the subgraph induced by U.

4.1 t-Multipliers

Definition 12. *Let $G = (V, E)$ be a graph over n vertices and let $(V_1, ..., V_k)$ be a partition of V. For an integer t we say that a graph-partition pair $(G', (V'_1, ..., V'_k)$ where $G' = (V', E')$ is a* t-multiplier *of $(G, (V_1, ..., V_k))$, if the following holds.*
Vertices: *$|V'| = t \cdot n$.*
Partition: *For each $1 \leq i \leq k$, $|V'_i| = t \cdot |V_i|$.*
Within Edges: *Suppose $G[V_i]$ has $\alpha_{ii} |V_i|^2$ edges. Then $G'[V'_i]$ has $\alpha_{ii} t^2 |V_i|^2$ edges.*
Between Edges: *Suppose G has $2\alpha_{ij} |V_i| \cdot |V_j|$ edges between V_i and V_j. Then G' has $2\alpha_{ij} \cdot t^2 |V_i| \cdot |V_j|$ edges between V'_i and V'_j.*

Recall that given a graph $G = (V, E)$ and a partition $(V_1, ..., V_k)$ of V that satisfies the constraints imposed by property P in $\mathcal{GPP}$, we say that $(V_1, ..., V_k)$ is a *witness* partition to the fact that G satisfies P. In short, we say that the graph-partition pair $(G, (V_1, ..., V_k))$ *satisfies* P. We prove the following claim regarding t-multipliers.

Claim 8. *If $(G, (V_1, ..., V_k))$ is a graph-partition pair satisfying P and $(G', (V'_1, ..., V'_k))$ is a t-multiplier of $(G, (V_1, ..., V_k))$, then the pair $(G', (V'_1, ..., V'_k))$ also satisfies P.*

The proof of Claim 8 is given in Appendix B.2.

4.2 Easily Testable Graph Partition Properties are Homogeneous

Let P be a graph partition property. We note that even if there are no explicit bounds on a part's size or on the edge density within a part or between a pair of parts, such constraints may be implicitly induced by the combination of other constraints. This leads to the following definition.

Definition 13. *Given an integer $1 \leq i \leq k$, we say that P has* no constraints on the edge density within part *i if for every graph-partition pair $(G, (V_1, ..., V_k))$ satisfying P, any graph G', obtained from G by performing arbitrary vertex-pair modifications within $G[V_i]$, satisfies P and $(V_1, ..., V_k)$ serves as a witness partition. Otherwise, we say that P* constrains the edge density within part *i. Similarly, given a pair of integers $(i, j) \in [k] \times [k]$, we say that P has* no constraints on the edge density between the parts *(i, j) if for every graph-partition pair $(G, (V_1, ..., V_k))$ satisfying P, any graph G', obtained from G by performing arbitrary vertex-pair modifications between V_i and V_j in G, satisfies P and $(V_1, ..., V_k)$ serves as a witness partition. Otherwise, we say that P* constrains the edge density between parts *(i, j).*

We say that a graph is *homogeneous* if it is either an independent set or a clique. We can classify the properties in $\mathcal{GPP}$ into two sets, corresponding to the following complementary two cases.

- **Case (a):** There exists a graph-partition pair $(G, (V_1, ..., V_k))$ satisfying P and an integer $1 \leq i \leq k$ such that $G[V_i]$ is heterogeneous.
- **Case (b):** For every graph-partition pair $(G, (V_1, ..., V_k))$ that satisfies P it holds that $G[V_i]$ is homogeneous for every $1 \leq i \leq k$.

We note that Case (b) in fact implies a stronger statement. If Case (b) holds, then not only every part is homogeneous, but rather the following holds: For each $1 \leq i \leq k$, either for every graph-partition pair $(G, (V_1, ..., V_k))$ satisfying the property, $G[V_i]$ is an independent set, or for every graph-partition pair $(G, (V_1, ..., V_k))$ satisfying the property, $G[V_i]$ is a clique. The reason is that otherwise it can be shown that Case (a) holds, but Case (a) cannot hold if Case (b) does. We next establish the following implication of Case (a).

Claim 9. *Let P be an easily testable graph partition property for which Case (a) holds. Then for every proximity parameter ϵ, every graph is ϵ-close to satisfying P.*

Proof. Since Case (a) holds, there exists a graph-partition pair $(G, (V_1, ..., V_k))$ satisfying P and an integer $1 \leq i \leq k$ such that $G[V_i]$ is non-homogeneous. Suppose by way of contradiction that there exists a one-sided error tester for P that is input-size oblivious. Denote the tester by $\mathcal{T}$. By [2, 14], we can assume without loss of generality that the algorithm $\mathcal{T}$ makes its decision based on an inspection of the subgraph induced by a random sample of s_ϵ vertices chosen independently and uniformly at random, where s_ϵ is a function of ϵ and is independent of n.

By Claim 8, for every $(G', (V_1', ..., V_k'))$ that is an s_ϵ-multiplier of $(G, (V_1, ..., V_k))$, we have that G' satisfies P (with the witness partition $(V_1', ..., V_k')$). Therefore, $\mathcal{T}$ must accept each such G' with probability 1. We next show that for every graph H over at most s_ϵ vertices, there exists at least one such graph G' for which $G'[V_i']$ contains H as an induced subgraph. The claim will then follow since the tester must accept given any induced subgraph that it observes, implying that it accepts all graphs with probability 1.

Let $|V_i| = n_i$ and let $v_i^1, \ldots, v_i^{s_\epsilon n_i}$ denote the vertices in V_i'. Observe that since $G[V_i]$ has at least one edge and at least one non-edge, for every $(G', (V_1', ..., V_k'))$ that is an s_ϵ-multiplier of $(G, (V_1, ..., V_k))$, it holds that $G'[V_i']$ has $m_i' \geq s_\epsilon^2$ edges and $(t \cdot n_i)^2 - m_i' \geq s_\epsilon^2$ non-edges. Let H be some fixed graph over $s \leq s_\epsilon$ vertices with m_H edges (and $s^2 - m_H$ non-edges). Since $m_H \leq s_\epsilon^2 \leq m_i'$ and $s^2 - m_H \leq s_\epsilon^2 \leq (t \cdot n_i)^2 - m_i'$, the definition of an s_ϵ-multiplier allows to let the subgraph of $G'[V_i']$ induced by the vertices $v_i^1, \ldots, v_i^s$ be H.

That is, the tester $\mathcal{T}$ accepts every graph with probability 1, and hence, for every ϵ, every graph is ϵ-close to satisfying P. □

We emphasize that Claim 9 holds for every graph and not only for sufficiently large graphs. It follows that if P is an easily testable graph partition property that satisfies Case (a) then P is in fact the trivial graph partition property that contains all graphs. This property clearly belongs to $\mathcal{GPP}_{0,1}$. Hence, from now on we can assume that Case (b) holds. In other words, we consider properties $P \in \mathcal{GPP}$ for which every graph satisfying P can only be validly partitioned in such a way that every part of the partition is homogeneous. That is, if G is a graph satisfying P and $(V_1, ..., V_k)$ is a witness partition, then for every pair $(i, j) \in [k] \times [k]$, the subgraph $G[V_i \cup V_j]$ is either a split graph, or a bipartite graph or a cobipartite graph. We can use this fact together with an application of an appropriate multiplier to establish the following claim.

Claim 10. *Let P be an easily testable property in $\mathcal{GPP}$. If there exists a graph-partition pair $(G, (V_1, ..., V_k))$ that satisfies P such that the edge density between a pair of parts is neither 0 nor 1, then P has no constraints on the edge density between the two parts.*

Proof. Let V_i and V_j be a pair of parts having an edge density α_{ij} where $0 < \alpha_{ij} < 1$. That is, $e_G(V_i, V_j) = 2\alpha_{ij} |V_i| \cdot |V_j|$. Let $(G', (V_1', ..., V_k'))$ be an $\frac{s^2}{2\alpha_{ij}(1-\alpha_{ij})}$-multiplier of $(G, (V_1, ..., V_k))$. Since $(G', (V_1', ..., V_k'))$ is a multiplier of a graph-partition pair satisfying P, then $\mathcal{T}$ must accept G' with probability 1. We first show that both the number of edges and the number of nonedges between V_i' and V_j' in G' is at least s^2.

$$e_{G'}\left(V_i', V_j'\right) = 2\alpha_{ij} \frac{s^2}{2\alpha_{ij}\left(1 - \alpha_{ij}\right)} |V_i| \, |V_j| \geq 2\alpha_{ij} \frac{s^2}{2\alpha_{ij}} = s^2 \tag{4.1}$$

$$\bar{e}_{ij} = \left(1 - 2\alpha_{ij}\right) \frac{s^2}{\alpha_{ij}\left(1 - 2\alpha_{ij}\right)} |V_i| \, |V_j| \geq \left(1 - 2\alpha_{ij}\right) \frac{s^2}{\left(1 - 2\alpha_{ij}\right)} = s^2 \tag{4.2}$$

Assume without loss of generality that $G[V_i]$ is a clique and $G[V_j]$ is an independent set (the proof is similar for the case of two cliques or two independent sets). Let H be an s-vertex split graph with clique size s_1 and independent set size s_2. We obtain a graph G'_H satisfying P by modifying G' as follows: we choose s_1 vertices from V_i and s_2 vertices from V_j. Then we modify the subgraph induced by the chosen $s = s_1 + s_2$ vertices so that it becomes H. We don't have to modify edges in $G'[V'_i]$ or $G'\left[V'_j\right]$ as they already have all or none of the edges respectively. That is, $G'[V'_i]$ is a clique of the right size and $G'\left[V'_j\right]$ is an independent set of the right size. Therefore, in order to perform the modification we had to make at most $2s_1 \cdot s_2 < 2s^2$ vertex pair modifications in the subgraph induced by the s vertices. Since the number of edges between V'_i and V'_j is at least $2s^2$ and the number of nonedges is also at least s^2 we can make up for the modification by adding an edge for every deletion and deleting an edge for every addition in such a way that the subgraph induced by the s vertices becomes H and the total number of edges between V'_i and V'_j in is maintained. The procedure clearly results in a graph satisfying P that has H as an induced subgraph. This implies that the one-sided tester must accept upon seeing any split graph as the induced subgraph it happens to sample. Hence, the property cannot have any restrictions on the edge density between part i and part j. The argument works similarly if we assume both parts are independent sets (in which case we replace the induced split graph with an induced bipartite graph) or both parts are cliques (in which case we replace the induced split graph with an induced co-bipartite graph). □

Claim 11. *For every split graph H of size s there exists a graph G'_H satisfying P on the same set of vertices as G' where $(V'_1, ..., V'_k)$ serves as a witness partition and H is an induced subgraph of G'_H.*

Proof. Let H be an s-vertex split graph with clique size s_1 and independent set size s_2. We obtain a graph G'_H satisfying P by modifying G' as follows: we choose s_1 vertices from V_i and s_2 vertices from V_j. Then we modify the subgraph induced by the chosen $s = s_1 + s_2$ vertices so that it becomes H. We don't have to modify edges in $G'[V'_i]$ or $G'\left[V'_j\right]$ as they already have all or none of the edges respectively. That is, $G'[V'_i]$ is a clique of the right size and $G'\left[V'_j\right]$ is an independent set of the right size. Therefore, in order to perform the modification we had to make at most $2s_1 \cdot s_2 < 2s^2$ vertex pair modifications in the subgraph induced by the s vertices. Since the number of edges between V'_i and V'_j is at least $2s^2$ and the number of nonedges is also at least s^2 we can make up for the modification by adding an edge for every deletion and deleting an edge for every addition in such a way that the subgraph induced by the s vertices becomes H and the total number of edges between V'_i and V'_j in is maintained. The procedure clearly results in a graph satisfying P that has H as an induced subgraph. □

In this subsection we showed that if a graph partition property P is easily testable and P constrains the edge density within a particular part, then the part must be homogeneous. (To be precise, P either forces the part to be an independent set or it forces it to be a clique.) Similarly, if P constrains the edge density between a pair of parts then it either forces the edge density within the pair to be 0 or it forces it to be 1. We call such properties *homogeneous graph partition properties.* That is, a homogeneous graph partition property is defined similarly to a property in $\mathcal{GPP}_{0,1}$ except that unlike $\mathcal{GPP}_{0,1}$, a homogeneous graph partition property possibly has size constraints on the parts. In the next subsection we show that if such a property is easily testable, then it has no size constraints.

4.3 No Constraints on the Sizes of Parts

Suppose P is a homogeneous property in $\mathcal{GPP}$ that possibly has size constraints in its specification and is easily testable. We show that if this is the case, then there exists an equivalent formulation of P, which we denote by P', where P' has no size constraints. Namely, we show that the simple relaxation of P whose specification contains the same homogeneity constraints as those specified by P but excludes its size constraints, serves as an equivalent property (under the assumption that P is easily testable). From now on, given a graph partition property P, we denote the relaxation of P (obtained by deleting the size constraints) by P'.

In order to obtain the above we prove a *dichotomy of properties*. In particular, we show that every homogeneous graph partition property P falls into one of two disjoint categories. The first category contains all those properties P for which the removal of their size constraints leaves them unchanged. That is, all the properties satisfying $P = P'$. The second category contains the properties P for which there exists an $\epsilon > 0$ such that P is ϵ-far from P'. That is, there are infinitely many graphs in P' that are ϵ-far from P.

Lemma 2 (The Dichotomy of Properties)**.** *Every homogeneous property P in the class $\mathcal{GPP}$ satisfies one of the following:*

1. $P = P'$.

2. There exists $\epsilon > 0$ such that for every n_0 there exists a graph $G' \in P'$ of size $n > n_0$ where G' is ϵ-far from P.

The implication of the *dichotomy of properties* is that a homogeneous graph partition property P with size constraints cannot be close to its relaxation P'. Either P is equivalent to P' or P is far from P' (for an appropriate distance measure). We are going to prove that if the latter case holds then P is not easily-testable. To do so we use the following auxiliary claim.

Claim 12. *Let P be any graph partition property. Suppose that for every $1 \le i \le k$ there exists a size vector $\vec{\rho}$ where $\rho_i \neq 0$ and $\vec{\rho}$ satisfies the size constraints induced by P. Then there exists a size vector $\vec{\rho}^*$ that satisfies the size constraints induced by P and for every $1 \le i \le k$ it holds that $\rho^*_i \neq 0$.*

The proof of Claim 12 is given in Appendix B.3. We use Claim 12 to prove that if Case 2 in the dichotomy of properties holds then P is not easily testable.

Claim 13. *Suppose there exists $\epsilon > 0$ such that for every n_0 there exists a graph $G' \in P'$ of size $n > n_0$ where G' is ϵ-far from P. Then P is not easily testable.*

Proof. Suppose by way of contradiction that P is easily testable. We set the proximity parameter to ϵ, the constant whose existence is assumed by the claim. We can also assume that the tester $\mathcal{T}$ is the canonical tester which accepts if and only if a random induced subgraph has a certain property (not necessarily P). That is, the tester chooses uniformly at random a set of vertices and operates by querying all the vertex pairs involving these vertices. We denote by s the number of vertices the tester selects (its sample size). The query complexity of the tester for P doesn't depend on the size of the input graph, but is a function of the proximity parameter ϵ and the parameters defining P. Hence, the value of the sample size s doesn't grow with n.

Consider the set of size vectors that are valid with respect to P (the vectors satisfying the size constraints induced by P). We can assume that for every coordinate i there exists a valid size vector $\vec{\rho}$ in which $\rho_i \neq 0$ because the existence of a coordinate i in which ρ_i is always null implies that P could be reformulated as a partition property with $k-1$ parts without the parameters referring to part i. Hence, by the previous claim, there exists a size vector $\vec{\rho}^*$ that is valid with respect to P and for every coordinate i, $\rho^*_i > 0$. We set n_0 in such a way that $\rho^*_i n_0 \ge s$ for every coordinate i.

Let $G' = (V', E')$ be a graph of size $n > n_0$ that satisfies P' and is ϵ-far from P. Let $(V_1, ..., V_k)$ be a witness partition to the fact that G' satisfies P'. We show that for every set of vertices $S \subseteq V'$ of size s there exists a graph $G \in P$ such that $G'[S]$ is isomorphic to an induced subgraph of G.

Let $S \subseteq V'$ be a set of vertices such that $|S| = s$. For every $1 \leq i \leq k$ we denote $S_i = S \cap V_i$.

We construct a graph G^* of size n. We set the vertices of G^* as $V^* = \{1, ..., n\}$. We partition V^* into k parts $(V_1^*, ..., V_k^*)$ by assigning every vertex to one of the parts arbitrarily in such a way that $|V_i^*| = \rho_i^* n$ for every $1 \leq i \leq k$. In order to set the edges of G^* we create a mapping from S to a subset of the vertices in G^* as follows. For every $1 \leq i \leq k$ and for every $u \in S_i$ we arbitrarily assign a distinct vertex $v \in V_i^*$. The number of vertices in V_i^* is sufficient to construct the mapping because $|S_i| \leq |S| \leq \rho_i^* n = |V_i^*|$. We denote the mapping by f. Particularly, if $v \in V_i^*$ is assigned by the mapping to $u \in S_i$ we denote $f(u) = v$ or equivalently $f^{-1}(v) = u$. For every $1 \leq i \leq k$ we denote by S_i^* the set of vertices in V_i^* that participate in the mapping. That is,

$$S_i^* = \{v \in V_i^* | \exists u \in V_i : f(u) = v\} \tag{4.3}$$

Additionally, we denote by S^* all the vertices in V^* that participate in the mapping. That is,

$$S^* = \bigcup_{i=1}^{k} S_i^* \tag{4.4}$$

Clearly, $|S_i^*| = |S_i|$ for every $1 \leq i \leq k$ and therefore $|S^*| = |S| = s$. With the mapping f at hand, we define E^*, the set of edges in G^*. Let $v_1, v_2 \in V^*$ be a pair of vertices.

Case 1. There exists a pair of vertices $u_1, u_2 \in S$ such that $u_1 \in S_i, u_2 \in S_j$ and $f(v_1) = u_1, f(v_2) = v_2$. In this case, $(v_1, v_2) \in E^*$ if and only if $(u_1, u_2) \in E'$.

Case 2. At least one of v_1 or v_2 is not assigned by f. Suppose $v_1 \in V_i^*, v_2 \in V_j^*$. In this case, $(v_1, v_2) \in E^*$ if and only if $d_P(i, j) = 1$.

Clearly, the mapping f serves as an isomorphism from $G'[S]$ to $G^*[S^*]$ because a vertex $u_1 \in S$ is connected to another vertex $u_2 \in S$ if and only if $f(u_1)$ is connected to $f(u_2)$. We prove that G^* satisfies P by showing that $(V_1^*, ..., V_k^*)$ serves as a witness partition. The size constraints are clearly satisfied since the definition of the partition explicitly indicates that for every $1 \leq i \leq k$, $|V_i| = \rho_i^* n$ where $\vec{\rho}^*$ is a valid size vector with respect to P. Consider a pair of vertices $v_1, v_2 \in V^*$ such that $v_1 \in V_i^*$ and $v_2 \in V_j^*$. Assume without loss of generality that $d_P(i, j) = 0$. We have to prove that v_1 is disconnected from v_2 in G^* (if $d_P(i, j) = 1$ we would've had to prove that v_1 is connected to v_2 and the proof is similar to the one we detail next). If either v_1 or v_2 (or both) is not assigned by f then by Case 2 of the definition of E^* it holds that v_1 is disconnected from v_2 because $d_P(i, j) \neq 1$. Suppose both v_1 and v_2 are assigned by f and let $u_1 = f^{-1}(v_1), u_2 = f^{-1}(v_2)$. Since f maps a vertex u to V_i^* only if $u \in V_i$ then $u_1 \in V_i$ and similarly $u_2 \in V_j$. Therefore, as $(V_1, ..., V_k)$ serves as a witness to $G' \in P'$ and $d_P(i, j) = 0$ it holds that $(u_1, u_2) \notin E'$. Therefore, from Case 1 of the definition of E^*, it must hold that $(v_1, v_2) \notin E^*$. This concludes the proof that G^* satisfies P.

We've shown that every induced s-sized subgraph of G' is isomorphic to an induced subgraph of a graph in P. Since $\mathcal{T}$ must accept with probability 1 upon seeing an induced s-sized subgraph of a graph belonging to P, it must also accept G' with probability 1. However, since G' is far from P, this is a contradiction to $\mathcal{T}$ being a one-sided error tester for P. Therefore, P is not easily testable. $\square$

Combining Claim 13 with *the dichotomy of properties* establishes the result we were aiming for: If a graph partition property P enforces size constraints (that cannot be ignored without changing the property) then P is not easily testable. It still remains to prove that *the dichotomy of properties* indeed holds. In order to do so, we first define the notion of a property's set of assignments. Then we show the existence of another dichotomy, *the trivial dichotomy*, which, as we prove below, implies *the dichotomy of properties* defined above.

Definition 14. *Given a property P in $\mathcal{GPP}$ we define a set of variables $X = \{x_{ij} | (i,j) \in [k] \times [k]\}$. An assignment φ is a function from X to $[0,1]$.*

We interpret an assignment φ as a transformation from a size vector $\vec{\rho}$ that violates the size constraints of P to a size vector that satisfies the size constraints. In particular, we interpret $\varphi(x_{ij})$ as the fraction of vertices (relative to n) that should be transferred from part i to part j (or stay in part i if $i = j$) in order to satisfy the size constraints imposed by P. Naturally, we restrict our attention to assignments that are *valid sizewise* as defined below.

Definition 15. *Given a property P in $\mathcal{GPP}$ and a size vector $\vec{\rho}$ we say that an assignment φ is valid sizewise if:*

$$\forall i' \in [k] : \sum_{i=1}^{k} \varphi(x_{i'i}) = \rho_{i'} \tag{4.5}$$

$$\forall i \in [k] : \rho_i^L \leq \sum_{i'=1}^{k} \varphi(x_{i'i}) \leq \rho_i^U \tag{4.6}$$

The first requirement of being valid sizewise can be interpreted as *the number of vertices being transferred from part i' equals the number of vertices assigned to part i' in the first place.* The second requirement can be interpreted as *the number of vertices being transferred to part i satisfies the size constraints imposed by the property P.*

Clearly, applying the transformation induced by a sizewise valid assignment φ to a graph $G' \in P'$ results in a graph G that satisfies the size constraints imposed by P. However, the assignment being valid sizewise doesn't necessarily induce a transformation resulting in a graph that satisfies the edge constraints. This leads to the definition of a violation in an assignment.

Definition 16. *Given a homogeneous property P in $\mathcal{GPP}$ and an assignment φ we say that a pair of variables $\{x_{i'i}, x_{j'j}\}$ constitutes a* violation *in the assignment φ with respect to P if:*

$$\varphi(x_{i'i}) \neq 0, \varphi(x_{j'j}) \neq 0 \tag{4.7}$$

$$d_P(i,j) \neq \perp \tag{4.8}$$

$$d_P(i',j') \neq d_P(i,j) \tag{4.9}$$

Definition 17. *If $d_P(i',j') = \perp$, then we say that the violation is* weak*. Otherwise, we say the violation is* strong.

The situation stated in Definition 16 is considered to be a violation of the edge constraints because it can be interpreted as vertices being transferred to a pair of parts whose edge density differs from that of the pair of parts those vertices originated from. Before describing *the trivial dichotomy* we have one more definition regarding violations.

Definition 18. *Given a violation* $\{x_{i'i}, x_{j'j}\}$ *of an assignment* φ *we define the violation's size as* $\min\{x_{i'i}, x_{j'j}\}$. *We call a violation of size at least* δ *a* δ*-violation.*

The following claim describes *the trivial dichotomy.*

Proposition 1 (The Trivial Dichotomy)**.** *Let* P *be a homogeneous graph partition property. Exactly one of the following holds:*

1'. For every size vector $\vec{\rho}$ *there exists a sizewise valid assignment* φ *which is free of violations with respect to* P*.*

2'. There exists a size vector $\vec{\rho}$ *for which every sizewise valid assignment* φ *contains a violation with respect to* P*.*

The trivial dichotomy (Proposition 1) trivially holds as the second case is the complement of the first. In order to prove that *the trivial dichotomy* implies *the dichotomy of properties* (Claim 2) we have to prove that Case 1' in *the trivial dichotomy* implies Case 1 in *the dichotomy of properties* and that Case 2' in *the trivial dichotomy* implies Case 2 in *the dichotomy of properties.*

4.3.1 Case 1' implies Case 1

Suppose Case 1' holds. Let G' be a graph satisfying P' over n vertices and let $(U_1, ..., U_k)$ be a witness partition. We have to show that $G' \in P$.

Let $\vec{\rho}$ be the size vector induced by $(U_1, ..., U_k)$. That is, $\vec{\rho} = \left(\frac{|U_1|}{n}, ..., \frac{|U_k|}{n}\right)$. Let φ be the assignment whose existence is promised by Case 1'.

For every $1 \leq i' \leq k$ we partition $U_{i'}$ into disjoint sets $V_{i'1}, ..., V_{i'k}$ in such a way that $|V_{i'i}| = \varphi(x_{i'i})\, n$. The choice of the vertices to be put in each part is arbitrary. For every $1 \leq i' \leq k$ it is possible to partition $U_{i'}$ in such a way since the number of vertices in $U_{i'}$ equals the number of vertices in $\bigcup_{i=1}^{k} V_{i'i}$ because φ is valid sizewise:

$$|U_{i'}| = \rho_{i'} n = n \sum_{i=1}^{k} \varphi(x_{i'i}) = \sum_{i=1}^{k} \varphi(x_{i'i}) \tag{4.10}$$

We use the partition defined above to devise a witness partition $(V_1, ..., V_k)$ for $G' \in P$:

$$\forall i \in [k] : V_i = \bigcup_{i'=1}^{k} V_{i'i} \tag{4.11}$$

Claim 14. *The partition* $(V_1, ..., V_k)$ *serves as a witness partition for* $G' \in P$*.*

The proof of Claim 14 is given in Appendix B.4. The graph G' satisfies property P and hence Case 1' implies Case 1.

4.3.2 Case 2' implies Case 2

Let $\vec{\rho}$ be a size vector for which every assignment φ that is valid sizewise contains a violation.

We define a set of decision variables Y:

$$Y = \{y_{i'i} | 1 \leq i \leq k, 1 \leq i' \leq k\} \tag{4.12}$$

Definition 19. *We say that a pair of decision variables $y_{i'i}, y_{j'j} \in Y$ are in potential conflict if $d_P(i,j) \neq \perp$ and $d_P(i',j') \neq d_P(i,j)$.*

That is, $y_{i'i}$ is in potential conflict with $y_{j'j}$ if the pair $\{x_{i'i}, x_{j'j}\}$ constitutes a violation with respect to P in an assignment φ which assigns $\varphi(x_{i'i}) > 0$ and $\varphi(x_{j'j}) > 0$.

We define a mathematical program on the set of decision variables Y. The objective of the program is to minimize the function

$$\max_{(y_{i'i}, y_{j'j}) \in Y} \{\min\{y_{i'i}, y_{j'j}\}\} \tag{4.13}$$

The feasible region is defined by the following linear constraints:

$$\forall i' \in [k] : \sum_{i=1}^{k} y_{i'i} = \rho_{i'} \tag{4.14}$$

$$\forall i \in [k] : \rho_i^L \leq \sum_{i'=1}^{k} y_{i'i} \leq \rho_i^U \tag{4.15}$$

$$\forall (i', i) \in [k] : 0 \leq y_{i'i} \leq 1 \tag{4.16}$$

There is a one to one correspondence between feasible solutions to the program and sizewise valid assignments. This is because the feasibility constraints in fact force $\vec{y}$ to be valid sizewise with respect to P. Hence, given an assignment φ that is valid sizewise we can obtain a feasible solution $\vec{y}$ by assigning $y_{i'i} = \varphi(x_{i'i})$. Similarly, given a feasible solution $\vec{y}$ we can obtain a sizewise valid assignment φ by assigning $\varphi(x_{i'i}) = y_{i'i}$. Moreover, the correspondence between assignments and feasible solutions has the property that an assignment is free of violations if and only if the corresponding feasible solution attains a value of 0 to the objective function. The reason is simple: A pair $\{x_{i'i}, x_{j'j}\}$ constitutes a violation if and only if $y_{i'i}$ is in conflict with $y_{j'j}$ and $\min\{\varphi(x_{i'i}), \varphi(x_{j'j})\} > 0$.

The feasibility constraints of the program define a compact set (the set defined by the constraints is both closed and bounded, and therefore compact). Also, the objective function is continuous as it is defined as a composition of continuous functions (the max function is continuous and so is the min function). Hence, according to the Extreme Value Theorem, the objective function must have both a maximum and a minimum on the set. That is, there exists a feasible solution $\vec{y}$ (and a corresponding assignment φ) that minimizes $\max_{(y_{i'i}, y_{j'j}) \in Y} \{\min\{y_{i'i}, y_{j'j}\}\}$. Denote by δ the value of the objective function under that solution. Since we've assumed that every sizewise valid assignment φ contains a violation, there is no feasible solution attaining an objective function value of 0. In other words $\delta > 0$. Therefore, every sizewise valid assignment contains a violation of size at least δ.

Let $\epsilon = \frac{1}{16}\delta^2$. We have to show that for every n_0 there exists a graph of size $n > n_0$ which is ϵ-far from P. We first show the existence of such a graph under the assumption that every assignment contains a *strong* δ-violation. Then we show that even if we only have weak δ-violations in some of the assignments there still exists a graph $G' \in P'$ which is ϵ-far from P.

Handling strong violations

We suppose for now that every sizewise valid assignment φ contains a strong δ-violation. We construct a graph $G' \in P'$ of size n and prove that G' is ϵ-far from P.

Let $G' = (V', E')$. We set $V' = \{1, ..., n\}$. We arbitrarily partition V' into k disjoint sets $U_1, ..., U_k$ in such a way that the number of vertices in U_i is $\rho_i n$. Clearly,

$$V' = \bigcup_{i=1}^{k} U_i \tag{4.17}$$

Now we define the set of edges E'. For every pair of vertices u, v in V' where $u \in U_{i'}, v \in V_{j'}$ (possibly $i' = j'$) we connect u and v if and only if $d_P(i', j') = 1$. By this construction, $G' \in P'$.

We are going to prove that G' is ϵ-far from P. Let $G \in P$ be a graph obtained from G' by any number of vertex-pair modifications. We show that the number of modifications must be at least ϵn^2.

Since $G \in P$, it possesses witness partitions. Let $(V_1, ..., V_k)$ be one of those witness partitions demonstrating $G \in P$. We use this witness partition to construct an assignment φ. For every $(i', i) \in [k] \times [k]$ we define:

$$\varphi(x_{i'i}) = \frac{|U_{i'} \cap V_i|}{n} \tag{4.18}$$

Claim 15. *The assignment φ is valid sizewise with respect to $\vec{\rho}$ and P.*

The proof of Claim 15 is given in Appendix B.5. Recall that we assume that every assignment that is valid sizewise with respect to P and $\vec{\rho}$ contains a strong δ-violation. Hence, the assignment φ defined above contains such a violation which we denote by $\{x_{i'i}, x_{j'j}\}$. That is,

$$\varphi(x_{i'i}) \geq \delta, \varphi(x_{j'j}) \geq \delta \tag{4.19}$$

and also

$$d_P(i,j) \neq \perp, d_P(i', j') \neq \perp, d_P(i,j) \neq d_P(i', j') \tag{4.20}$$

We assume without loss of generality that $d_P(i,j) = 0$ and $d_P(i', j') = 1$. Since $\varphi(x_{i'i}) \geq \delta$ and $\varphi(x_{j'j}) \geq \delta$ we have

$$|U_{i'} \cap V_i| \geq \delta n \text{ and } |U_{j'} \cap V_j| \geq \delta n \tag{4.21}$$

Since $d_P(i', j') = 1$, in the graph G', every vertex in $U_{i'}$ is connected to every vertex in $V_{j'}$.

Claim 16. $e_{G'}(V_i, V_j) \geq \frac{1}{4}\delta^2 n$

The proof of Claim 16 is given in Appendix B.6.

We've seen that if $d_P(i,j) = 0$ then $e_{G'}(V_i, V_j) \geq \frac{1}{4}\delta^2 n^2$. Similarly, if $d_P(i,j) = 1$ then $\overline{e}_{G'}(V_i, V_j) \geq \frac{1}{4}\delta^2 n^2$. Therefore, in order for $(V_1, ..., V_k)$ to become a witness partition for belonging to P, at least $\frac{1}{4}\delta^2 n^2$ vertex-pairs must be modified. That is, G' is $\frac{1}{4}\delta^2$-far from P. In particular, G' is also ϵ-far from P.

We've seen that if there exists a size vector $\vec{\rho}$ such that every sizewise valid assignment φ contains a strong violation with respect to P then for every n_0 there exists a graph $G' \in P'$ which is ϵ-far from P. Now we have to prove that this is the case even if every sizewise valid assignment contains a δ-violation that is not necessarily strong.

Handling Weak Violations

We construct a random graph $G' = (V', E')$ of size n and prove that with positive probability G' is ϵ-far from every n-size graph $G \in P$.

We define $V' = \{1, ..., n\}$. We arbitrarily partition V' into k disjoint sets $U_1, ..., U_k$ in such a way that the number of vertices in U_i is $\rho_i n$.

We next define E'. For each pair of vertices $u \in U_i, v \in U_j$:

- If $d_P(i,j) = 0$ we do not connect u and v.
- If $d_P(i,j) = 1$ we do connect u and v.
- If $d_P(i,j) = \perp$ we connect u and v with probability $\frac{1}{2}$.

For every $1 \le i \le k$ where $d_P(i,j) = \perp$ we define the event $\mathcal{E}_i$ as follows:

> For every set $A \subseteq U_i$ of size $|A| \ge \delta n$ it holds that $e_{G'}(A,A) \ge \epsilon n^2$ and $\overline{e}_{G'}(A,A) \ge \epsilon n^2$. We denote the $e_{G'}(A,A)$ sub-event by $\mathcal{E}_i^1$ and the $\overline{e}_{G'}(A,A)$ sub-event by $\mathcal{E}_i^0$. That is, $\mathcal{E}_i = \mathcal{E}_i^0 \cap \mathcal{E}_i^1$.

For every $(i,j) \in [k] \times [k]$ where $d_P(i,j) = \perp$ we define the event $\mathcal{E}_{ij}$ as follows (possibly $i = j$):

> For every pair of disjoint sets $(A,B) \subseteq U_i \times U_j$ where $|A| \ge \delta n$ and $|B| \ge \delta n$ it holds that $e_{G'}(A,B) \ge \epsilon n^2$ and $\overline{e}_{G'}(A,B) \ge \epsilon n^2$. As before, we denote the two sub-events by $\mathcal{E}_{ij}^0$ and $\mathcal{E}_{ij}^1$ so that $\mathcal{E}_{ij} = \mathcal{E}_{ij}^0 \cap \mathcal{E}_{ij}^1$.

Observe that in the above definition the requirement that the sets A and B are disjoint is redundant if i≠j but should be stated explicitly if $i = j$.

Definition 20. *We denote by $\mathcal{E}$ the conjunction of all the events defined above. That is,*

$$\mathcal{E} = \left(\bigcap_{i=1}^{k} \mathcal{E}_i\right) \cap \left(\bigcap_{(i,j)\in[k]^2} \mathcal{E}_{ij}\right) \tag{4.22}$$

Claim 17. $\Pr[\mathcal{E}] > 0$

The proof of Claim 17 is given in Appendix B.7.

Claim 17 implies the existence of a graph $G^* \in P'$ for which $\mathcal{E}$ holds. Let $G \in P$ be a graph of size n and let $(V_1, ..., V_k)$ be a witness partition for $G \in P$. We define the assignment φ as done before ($\varphi(x_{i'i}) = \frac{|U_{i'} \cap V_i|}{n}$). As shown above, φ is valid sizewise. Hence, φ contains a δ-violation $\{x_{i'i}, x_{j'j}\}$. If the violation is strong, we're done (by adapting the same analysis as before). Suppose the violation is weak. That is,

$$d_P(i', j') = \perp, d_P(i,j) \neq \perp \tag{4.23}$$

As before, $\varphi(x_{i'i}) \ge \delta$ and $\varphi(x_{j'j}) \ge \delta$ imply that

$$|U_{i'} \cap V_i| \ge \delta n \text{ and } |U_{j'} \cap V_j| \ge \delta n \tag{4.24}$$

Let $A = U_{i'} \cap V_i$ and $B = U_{j'} \cap V_j$. Clearly, $A \subseteq U_{i'}, B \subseteq U_{j'}, |A| \ge \delta n, |B| \ge \delta n$.
The event $\mathcal{E}_{i'j'}$ holds and so do $\mathcal{E}_{i'}$ and $\mathcal{E}_{j'}$. That is,

$$e_{G'}(A,B) \ge \epsilon n^2 \text{ and } \overline{e}_{G'}(A,B) \ge \epsilon n^2 \tag{4.25}$$

Therefore,

$$e_{G'}(V_i, V_j) \ge \epsilon n^2 \text{ and } \overline{e}_{G'}(V_i, V_j) \ge \epsilon n^2 \tag{4.26}$$

However, since $d_P(i,j) \neq \perp$, the number of vertex-pair modifications required in order to obtain G from G' is at least ϵn^2. In other words, the graph G', although in P', is ϵ-far from every graph $G \in P$.

That is to say, easily-testable homogeneous graph partition properties cannot have size constraints, and if they do then these size constraints are in fact redundant.

4.4 Wrapping Things Up

In Section 4.2 we've shown that if a graph partition property is easily-testable then it must be homogeneous. In Section 4.3 we've shown that if a homogeneous graph partition property is easily-testable then it has no size constraints. Hence, by combining the results of Sections 4.2 and 4.3 we obtain the following: If a graph partition property P is then $P \in \mathcal{GPP}_{0,1}$. This proves the "only if" direction of Theorem 2.

5 References

[1] Noga Alon. Testing subgraphs in large graphs. *Random Structures & Algorithms*, 21(3-4):359–370, 2002.

[2] Noga Alon, Eldar Fischer, Michael Krivelevich, and Mario Szegedy. Efficient testing of large graphs. *Combinatorica*, 20(4):451–476, 2000.

[3] Noga Alon, Eldar Fischer, Ilan Newman, and Asaf Shapira. A combinatorial characterization of the testable graph properties: it's all about regularity. *SIAM Journal on Computing*, 39(1):143–167, 2009.

[4] Noga Alon and Jacob Fox. Easily testable graph properties. *Combinatorics, Probability and Computing*, 24(4):646–657, 2015.

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

[6] Noga Alon and Asaf Shapira. Testing satisfiability. *Journal of Algorithms*, 47(2):87–103, 2003.

[7] Noga Alon and Asaf Shapira. Testing subgraphs in directed graphs. *Journal of Computer and System Sciences*, 69(3):354–382, 2004.

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

[9] Noga Alon and Asaf Shapira. A characterization of the (natural) graph properties testable with one-sided error. *SIAM Journal on Computing*, 37(6):1703–1727, 2008.

[10] Christian Borgs, Jennifer Chayes, László Lovász, Vera T Sós, Balázs Szegedy, and Katalin Vesztergombi. Graph limits and parameter testing. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 261–270. ACM, 2006.

[11] David Conlon and Jacob Fox. Graph removal lemmas. *Surveys in combinatorics*, 1(2):3, 2013.

[12] Lior Gishboliner and Asaf Shapira. Removal lemmas with polynomial bounds. pages 510–522, 2017.

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

[14] Oded Goldreich and Luca Trevisan. Three theorems regarding testing graph properties. *Random Structures & Algorithms*, 23(1):23–57, 2003.

[15] James Renegar. On the computational complexity and geometry of the first-order theory of the reals. part i: Introduction. preliminaries. the geometry of semi-algebraic sets. the decision problem for the existential theory of the reals. *Journal of symbolic computation*, 13(3):255–299, 1992.

[16] James Renegar. On the computational complexity and geometry of the first-order theory of the reals. part ii: The general decision problem. preliminaries for quantifier elimination. *Journal of Symbolic Computation*, 13(3):301–327, 1992.

A Appendix: The Quadratic Program

The quadratic program is defined by the following set of constraints:

$$\sum_{i=1}^{k} x_i = 1 \tag{A.1}$$

$$\forall i \in [k] : \rho_i^L \le x_i \le \rho_i^U \tag{A.2}$$

$$\forall i \in [k] : y_{ii} \le x_i^2 \tag{A.3}$$

$$\forall (i,j) \in [k] \times [k] : y_{ij} \le 2x_i \cdot x_j \tag{A.4}$$

$$\forall i \in [k] : \rho_{ii}^L \le y_{ii} \le \rho_{ii}^U \tag{A.5}$$

$$\forall i \ne j \in [k] \times [k] : 2\rho_{ij}^L \le y_{ij} \le 2\rho_{ij}^U \tag{A.6}$$

$$\forall i \in [k] : \alpha_{ii}^L \le y_{ii} \le \alpha_{ii}^U \tag{A.7}$$

$$\forall i \ne j \in [k] \times [k] : 2\alpha_{ij}^L \le y_{ij} \le 2\alpha_{ij}^U \tag{A.8}$$

B Appendix: Proofs of Auxiliary Claims

B.1 Proof of Claim 7

For every pair of parts $1 \le i, j \le k$ s.t. $i \ne j$ we have:

$$\begin{aligned} e(V_i, V_j) &\in \left[\widetilde{\rho}_{ij}^L n^2, \widetilde{\rho}_{ij}^U n^2\right] \cap \left[\widetilde{\alpha}_{ij}^L |V_i| |V_j|, \widetilde{\alpha}_{ij}^U |V_i| |V_j|\right] \\ &\subseteq \left[\widetilde{\rho}_{ij}^L n^2, \widetilde{\rho}_{ij}^U n^2\right] \cap \left[\widetilde{\alpha}_{ij}^L \rho_i^L \rho_j^L n^2, \widetilde{\alpha}_{ij}^U \rho_i^U \rho_j^U n^2\right] \end{aligned}$$

Hence,

$$\begin{aligned} \vec{\sigma}_{ij} &\in \left[\widetilde{\rho}_{ij}^L, \widetilde{\rho}_{ij}^U\right] \cap \left[\widetilde{\alpha}_{ij}^L \rho_i^L \rho_j^L, \widetilde{\alpha}_{ij}^U \rho_i^U \rho_j^U\right] \\ &= \left[\max\left\{\widetilde{\rho}_{ij}^L, \widetilde{\alpha}_{ij}^L \cdot \rho_i^L \cdot \rho_j^L\right\}, \min\left\{\widetilde{\rho}_{ij}^U, \widetilde{\alpha}_{ij}^U \cdot \rho_i^U \cdot \rho_j^U\right\}\right] \end{aligned}$$

If $i = j$:

$$
\begin{aligned}
e(V_i) &\in \left[\widetilde{\rho}_{ii}^L n^2, \widetilde{\rho}_{ii}^U n^2\right] \cap \left[\widetilde{\alpha}_{ii}^L |V_i|^2 n^2, \widetilde{\alpha}_{ii}^U |V_i|^2 n^2\right] \\
&\subseteq \left[\widetilde{\rho}_{ii}^L n^2, \widetilde{\rho}_{ii}^U n^2\right] \cap \left[\widetilde{\alpha}_{ii}^L \left(\rho_i^L\right)^2 n^2, \widetilde{\alpha}_{ii}^U \left(\rho_i^U\right)^2 n^2\right]
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\vec{\sigma}_{ij} &\in \left[\widetilde{\rho}_{ii}^L, \widetilde{\rho}_{ii}^U\right] \cap \left[\widetilde{\alpha}_{ii}^L \left(\rho_i^L\right)^2, \widetilde{\alpha}_{ii}^U \left(\rho_i^U\right)^2\right] \\
&= \left[\max\left\{\rho_{ii}^L, \widetilde{\alpha}_{ii}^L \left(\rho_i^L\right)^2\right\}, \min\left\{\rho_{ii}^U, \widetilde{\alpha}_{ii}^U \left(\rho_i^U\right)^2\right\}\right]
\end{aligned}
$$

Therefore, there exists $\vec{b} \in \prod_{i \le j} B_{ij}^{\vec{a}}$ s.t. $\vec{\sigma} \in \vec{b}$. □

B.2 Proof of Claim 8

We show that $(G', (V_1', ..., V_k'))$ satisfies all of $\mathcal{P}$'s constraints.

Size constraints: Denote by ρ_i the relative size of V_i in $(G, (V_1, ..., V_k))$. That is, $|V_i| = \rho_i n$.

$$
\left|V_i'\right| = t\,|V_i| = t\rho_i n = \rho_i n'^2 \tag{B.1}
$$

Between constraints: For every $1 \le i, j \le k$ where $i \ne j$ we denote by α_{ij} the relative relative edge density of the pair (V_i, V_j) in $(G, (V_1, ..., V_k))$ and we denote by ρ_{ij} the absolute edge density of the pair. That is,

$$
e_G(V_i, V_j) = 2\alpha_{ij}\,|V_i| \cdot |V_j| = \rho_{ij} n^2
$$

In $(G', (V_1', ..., V_k'))$ we have

$$
e_{G'}\left(V_i', V_j'\right) = 2\alpha_{ij} t^2\,|V_i| \cdot |V_j| = 2\alpha_{ij}\left|V_i'\right| \cdot \left|V_j'\right|
$$

That is, the relative edge density of $\left(V_i', V_j'\right)$ in $(G', (V_1', ..., V_k'))$ is also α_{ij}.

As for the absolute edge densities in $(G', (V_1', ..., V_k'))$,

$$
e_{G'}\left(V_i', V_j'\right) = 2\alpha_{ij} t^2\,|V_i| \cdot |V_j| = e(V_i, V_j) \cdot t^2 = \rho_{ij} n^2 t^2 = \rho_{ij} n'^2
$$

That is, the absolute edge density of $\left(V_i', V_j'\right)$ in $(G', (V_1', ..., V_k'))$ is also ρ_{ij}.

Within constraints: For every $1 \le i \le k$ where we denote by α_{ii} the relative relative edge density of V_i in $(G, (V_1, ..., V_k))$ and we denote by ρ_{ii} the absolute edge density of the V_i. That is,

$$
e_G(V_i, V_i) = \alpha_{ii}\,|V_i|^2 = \rho_{ii} n^2
$$

In $(G', (V_1', ..., V_k'))$ we have

$$
e_{G'}\left(V_i', V_i'\right) = \alpha_{ii} t^2\,|V_i|^2 = 2\alpha_{ij}\left|V_i'\right| \cdot \left|V_j'\right|
$$

That is, the relative edge density of $\left(V_i', V_j'\right)$ in $(G', (V_1', ..., V_k'))$ is also α_{ij}.

As for the absolute edge densities in $(G', (V_1', ..., V_k'))$,

$$
e_{G'}\left(V_i', V_j'\right) = \alpha_{ij} t^2\,|V_i| \cdot |V_j| = e(V_i, V_j) \cdot t^2 = \rho_{ij} n^2 t^2 = \rho_{ij} n'^2
$$

That is, the absolute edge density of $\left(V_i', V_j'\right)$ in $(G', (V_1', ..., V_k'))$ is also ρ_{ij}.

All the relative sizes and the edge densities are the same for both the original graph-partition pair and its t-multiplier and therefore if the original pair satisfies $\mathcal{P}$ then so is its multiplier. □

B.3 Proof of Claim 12

Let $\vec{\rho} = (\rho_1, ..., \rho_k)$ be a size vector satisfying the size constraints induced by P. If $\vec{\rho}$ is null-free (that is, $\rho_i \neq 0$ for every $1 \leq i \leq k$) then we're done. Suppose there are $\ell > 0$ nulls in $\vec{\rho}$. We claim that there exists a coordinate j where $1 \leq j \leq k$ such that ρ_j is strictly greater than ρ_j^L. Suppose by way of contradiction that this is not the case. That is, for every $1 \leq j \leq k$, $\rho_j = \rho_j^L$. If this is the case, then for every size vector $\vec{\rho'}$ that satisfies the size constraints but differs from $\vec{\rho}$ it holds that $\rho'_j \geq \rho_j$ for every $1 \leq j \leq k$ and there exists j^* such that $\rho'_{j^*} > \rho_{j^*}$. Hence, if $\vec{\rho'}$ satisfies the size constraints and differs from $\vec{\rho}$ then

$$\sum_{i=1}^{k} \rho'_i > \sum_{i=1}^{k} \rho_i = 1 \tag{B.2}$$

But the sum of the coordinates of a size vector must equal exactly 1. Hence, either there exists a coordinate in $\vec{\rho}$ strictly greater than its lower bound or $\vec{\rho}$ is the only size vector satisfying the size constraints. However, $\vec{\rho}$ cannot be the only such vector because $\vec{\rho}$ contains nulls, and we've assumed that for every i there exists a valid size vector whose ith coordinate is not null. Therefore we can assume there exists a coordinate in $\vec{\rho}$ whose value is strictly greater than the lower bound imposed on it by P. We denote this coordinate by j. That is,

$$\rho_j > \rho_j^L \tag{B.3}$$

We use the existence of this coordinate to construct a size vector $\vec{\rho^*}$ that is null-free:

$$\rho_i^* = \begin{cases} \rho_j - \Delta & \text{if } i = j \\ \frac{\Delta}{\ell} & \text{if } \rho_i = 0 \\ \rho_i & \text{otherwise} \end{cases} \tag{B.4}$$

where

$$\Delta = \min \left\{ \min_{i:\rho_i=0} \left\{ \rho_i^U \right\}, \frac{\rho_j - \rho_j^L}{2} \right\} \tag{B.5}$$

We have to prove that $\vec{\rho^*}$ is a valid size vector that satisfies the size constraints imposed by P and that it is free of nulls. That is, we need to prove that $\vec{\rho^*}$ admits three properties:

1. $\sum_{i=1}^{k} \rho_i^* = 1$
2. for every i, $\rho_i^L \leq \rho_i^* \leq \rho_i^U$
3. for every i, $\rho_i^* > 0$

We begin by proving $\sum_{i=1}^{k} \rho_i^* = 1$.

$$\begin{aligned} \sum_{i=1}^{k} \rho_i^* &= \rho_j^* + \sum_{i:\rho_i=0} \rho_i^* + \sum_{i:\rho_i\neq 0, i\neq j} \rho_i^* \\ &= \rho_j - \Delta + \ell \cdot \frac{\Delta}{\ell} + \sum_{i:\rho_i\neq 0, i\neq j} \rho_i \\ &= \rho_j + \sum_{i:i\neq j} \rho_i = \sum_{i=1}^{k} \rho_i \\ &= 1 \end{aligned}$$

Now we prove that for every i, $\rho_i^L \leq \rho_i^* \leq \rho_i^U$. There are three cases:

- Case 1: $i = j$. In this case,

$$\rho_i^* = \rho_j^* = \rho_j - \Delta < \rho_j \leq \rho_j^U \tag{B.6}$$

and

$$\rho_i^* = \rho_j^* = \rho_j - \Delta \geq \rho_j - \frac{\rho_j - \rho_j^L}{2} = \frac{\rho_j + \rho_j^L}{2} \geq \frac{2\rho_j^L}{2} = \rho_j^L \tag{B.7}$$

- Case 2: $\rho_i = 0$. In this case,

$$\rho_i^* = \frac{\Delta}{\ell} \leq \frac{\rho_i^U}{\ell} \leq \rho_i^U \tag{B.8}$$

and

$$\rho_i^* = \frac{\Delta}{\ell} \geq 0 = \rho_i = \rho_i^L \tag{B.9}$$

- Case 3: $\rho_i \neq 0$ and $i \neq j$. In this case,

$$\rho_i^* = \rho_i \in \left[\rho_i^L, \rho_i^U\right] \tag{B.10}$$

This concludes the three cases. Now we prove that for every i, $\rho_i^* > 0$. Again, there are three cases.

- Case 1: $i = j$. In this case,

$$\rho_i^* = \rho_j^* = \rho_j - \Delta \geq \rho_j - \frac{\rho_j - \rho_j^L}{2} = \frac{\rho_j + \rho_j^L}{2} > 0 \tag{B.11}$$

- Case 2: $\rho_i = 0$. In this case,

$$\rho_i^* = \frac{\Delta}{\ell} > 0 \tag{B.12}$$

- Case 3: $\rho_i \neq 0$ and $i \neq j$. In this case,

$$\rho_i^* = \rho_i > 0 \tag{B.13}$$

In conclusion, the vector $\vec{\rho}^*$ admits the three properties. The claim follows.

□

B.4 Proof of Claim 14

We prove that $(V_1, ..., V_k)$ indeed serves as a witness partition for $G' \in P$. In order to do so we demonstrate that $(V_1, ..., V_k)$ satisfies both the size constraints imposed by P and the edge density constraints.

Claim 18. *The partition $(V_1, ..., V_k)$ satisfies the size constraints imposed by P.*

Proof. For every $1 \leq i \leq k$:

$$|V_i| = \left|\bigcup_{i'=1}^{k} V_{i'i}\right| = \sum_{i'=1}^{k} |V_{i'i}| = \sum_{i'=1}^{k} \varphi\left(x_{i'i}\right) n = n \cdot \sum_{i'=1}^{k} \varphi\left(x_{i'i}\right) \tag{B.14}$$

The second equality holds since the sets $\{V_{i'i}\}_{i'=1}^{k}$ are disjoint. The third equality results directly from the definition of $V_{i'i}$.

Since the assignment φ is valid sizewise we have:

$$\rho_i^L n \leq \sum_{i'=1}^{k} \varphi\left(x_{i'i}\right) \leq \rho_i^U n \tag{B.15}$$

Hence, $\rho_i^L n \leq |V_i| \leq \rho_i^U n$ as required. □

Claim 19. *The partition $(V_1, ..., V_k)$ satisfies the edge density constraints imposed by P.*

Proof. We have to show that for every $1 \leq i \leq k$, if $d_P(i,j) = 0$ then $e_G(V_i, V_j) = 0$ and that if $d_P(i,j) = 1$ then $\bar{e}_G(V_i, V_j) = 0$. Let V_i, V_j be a pair of parts in $(V_1, ..., V_k)$ (possibly $i = j$). Assume without loss of generality that $d_P(i,j) = 0$. Suppose by way of contradiction that $e_G(V_i, V_j) \neq 0$. That is, there are two vertices $u \in V_i, v \in V_j$ such that $(u,v) \in E'$. Since $u \in V_i$ there exists $1 \leq i' \leq k$ for which $u \in V_{i'i}$. Similarly, since $v \in V_j$ there exists $1 \leq j' \leq k$ for which $v \in V_{j'j}$. Hence, $\varphi\left(x_{i'i}\right) = \frac{|V_{i'i}|}{n} > 0$ and $\varphi\left(x_{j'j}\right) = \frac{\left|V_{j'j}\right|}{n} > 0$.

We show that $\left\{x_{i'i}, x_{j'j}\right\}$ serves as a violation of φ with respect to P, in contradiction to the assumption that φ is free of violations.

From the definition of $V_{i'i}$ and $V_{j'j}$ we have that $V_{i'i} \subseteq U_{i'}$ and $V_{j'j} \subseteq U_{j'}$. This, together with the fact that $u \in V_{i'i}$ and $v \in V_{j'j}$, implies

$$u \in U_{i'} \text{ and } v \in U_{j'}$$

That is, G' contains an edge connecting a vertex in $U_{i'}$ to a vertex in $U_{j'}$. Since $(U_1, ..., U_k)$ serves as a witness partition for $G' \in P'$ and there are edges between $U_{i'}$ and $U_{j'}$, it must hold that $d_{P'}(i', j') \neq 0$. Since P doesn't differ from P' in the edge density constraints, $d_P(i', j') \neq 0$.

Hence, we found a pair of variables $x_{i'i}, x_{j'j}$ where $\varphi\left(x_{i'i}\right) \neq 0, \varphi\left(x_{j'j}\right) \neq 0$ in the assignment φ such that:

1. $d_P(i,j) \neq \perp$ (as we've assumed $d_P(i,j) = 0$)
2. $d_P(i', j') \neq P(i,j)$ (as we've shown $d_P(i', j') \neq 0$)

Therefore, the pair $\left\{x_{i'i}, x_{j'j}\right\}$ serves as a violation in the assignment φ with respect to P, in contradiction to φ being free of violations. Hence, if $d_P(i,j) = 0$ then $e_G(V_i, V_j) = 0$. Similarly, if $d_P(i,j) = 1$ then $\bar{e}_G(V_i, V_j) = 0$. □

We've shown that the partition $(V_1, ..., V_k)$ satisfies both the size constraints and the edge density constraints imposed by P. Hence, $(V_1, ..., V_k)$ serves as a witness partition for $G' \in P$.

B.5 Proof of Claim 15

1. For every $1 \leq i' \leq k$

$$n \cdot \sum_{i=1}^{k} \varphi\left(x_{i'i}\right) = \sum_{i=1}^{k} |U_{i'} \cap V_i| = \left|\bigcup_{i=1}^{k} (U_{i'} \cap V_i)\right| = |U_{i'}| = \rho_{i'} n \tag{B.16}$$

That is, $\sum_{i=1}^{k} \varphi\left(x_{i'i}\right) = \rho_{i'}$

2. For every $1 \leq i \leq k$

$$n \cdot \sum_{i'=1}^{k} \varphi\left(x_{i'i}\right) = \sum_{i'=1}^{k} \left|U_{i'} \cap V_i\right| = \left|\bigcup_{i'=1}^{k} \left(U_{i'} \cap V_i\right)\right| = \left|V_i\right| \tag{B.17}$$

The partition $(V_1, ..., V_k)$ serves as a witness partition to $G \in P$ and therefore satisfies the size constraints. That is, for every $i \in [k]$, $\rho_i^L n \leq |V_i| \leq \rho_i^U n$. Therefore,

$$\rho_i^L \leq \sum_{i'=1}^{k} \varphi\left(x_{i'i}\right) \leq \rho_i^U \tag{B.18}$$

Items 1 and 2 imply that φ is valid sizewise. □

B.6 Proof of Claim 16

We first prove the claim for the case where $i = j$ and $i' = j'$. In this case $G'\left[U_{i'}\right]$ is a clique (because $d_P\left(i', i'\right) = 1$). Since $U_{i'} \cap V_i \subseteq U_{i'}$ then so is $G'\left[U_{i'} \cap V_i\right]$. Since $\left|U_{i'} \cap V_i\right| \geq \delta n$ there are at least $\binom{\delta n}{2} \geq \frac{1}{4}\delta^2 n^2$ edges in the clique $G'\left[U_{i'} \cap V_i\right]$. Additionally, $U_{i'} \cap V_i \subseteq V_i$. Therefore, the number of edges in $G'\left[V_i\right]$ is also at least $\frac{1}{4}\delta^2 n^2$. Hence if $i = j$ and $i' = j'$ then:

$$e_{G'}\left(V_i, V_j\right) = e_{G'}\left(V_i, V_i\right) \geq \frac{1}{4}\delta^2 n^2 \tag{B.19}$$

Now suppose this is not the case. That is, either $i \neq j$ or $i' \neq j'$ (or both). In the graph G', every vertex in $U_{i'}$ is connected to every vertex in $U_{j'}$ (because $d_P\left(i', j'\right) = 1$). Therefore, in G', every vertex in $U_{i'} \cap V_i$ is connected to every vertex in $U_{j'} \cap V_j$. The set $U_{i'} \cap V_i$ is disjoint from the set $U_{j'} \cap V_j$ because either $i \neq j$ or $i' \neq j'$. Hence,

$$e_{G'}\left(U_{i'} \cap V_i, U_{j'} \cap V_j\right) = \left|U_{i'} \cap V_i\right| \cdot \left|U_{j'} \cap V_j\right| \geq \delta^2 n^2 \tag{B.20}$$

Therefore,

$$e_{G'}\left(V_i, V_j\right) \geq \delta^2 n^2 \geq \frac{1}{4}\delta^2 n^2 \tag{B.21}$$

The claim follows. □

B.7 Proof of Claim 17

To prove Claim 17 we first prove the following two claims.

Claim 20. *For every* $i \in [k]$, $\Pr\left[\overline{\mathcal{E}_i^1}\right] \leq e^{-n\left(\frac{\delta^2}{64}n - 2\right)}$ *and* $\Pr\left(\overline{\mathcal{E}_i^0}\right) \leq e^{-n\left(\frac{\delta^2}{64}n - 2\right)}$.

Proof. We prove $\Pr\left[\overline{\mathcal{E}_i^1}\right] \leq e^{-n\left(\frac{\delta^2}{64}n - 2\right)}$ in detail as the proof of the bound of $\Pr\left[\overline{\mathcal{E}_i^0}\right]$ is similar. Let $A \subseteq U_i$ be a set of at least δn vertices. We bound from above the probability that the number of edges in $G'[A]$ is at most ϵn^2. In order to obtain that upper bound we consider a set $A' \subseteq A$ of size exactly δn. Since every edge in $G'\left[A'\right]$ is also in $G'[A]$,

$$\Pr\left[e_{G'}(A, A) < \epsilon n^2\right] \leq \Pr\left[e_{G'}\left(A', A'\right) < \epsilon n^2\right] \tag{B.22}$$

Therefore, we proceed by finding an upper bound on $\Pr\left[e_{G'}\left(A', A'\right) < \epsilon n^2\right]$.

For every pair of vertices $u, v \in A'$ we define an indicator random variable χ_{uv} as follows.

$$\chi_{uv} = \begin{cases} 0 & (u,v) \notin E[G'] \\ 1 & (u,v) \in E[G'] \end{cases}$$

We define a random variable χ as the sum of the indicators.

$$\chi = \sum_{u,v \in A'} \chi_{uv} \tag{B.23}$$

Clearly, $e_{G'}(A', A') = \chi$. Hence, we have to bound from above the probability that $\chi < \epsilon n^2$. Since $\epsilon n^2 \leq \frac{1}{2} \cdot \frac{1}{2} \cdot \binom{\delta n}{2}$,

$$\Pr\left[\chi < \epsilon n^2\right] \leq \Pr\left[\chi < \frac{1}{2} \cdot \frac{1}{2} \cdot \binom{\delta n}{2}\right] \tag{B.24}$$

To bound the probability from above by applying Chernoff's bound, we first find the expectation of χ using the linearity of expectation:

$$E[\chi] = \sum_{u,v \in A'} \chi_{uv} = \sum_{u,v \in A} \frac{1}{2} = \frac{1}{2} \cdot \binom{\delta n}{2} \tag{B.25}$$

We apply Chernoff's bound:

$$\Pr\left[\chi < \frac{1}{2} \cdot \frac{1}{2} \cdot \binom{\delta n}{2}\right] = \Pr\left[\chi < \frac{1}{2} E[\chi]\right] \leq e^{-\left(\frac{1}{2}\right)^2 \cdot \frac{E[\chi]}{2}} = e^{-\frac{1}{16} \cdot \binom{\delta n}{2}} \leq e^{-\frac{1}{64}\delta^2 n^2} \tag{B.26}$$

Hence,

$$\Pr\left[e_{G'}(A, A) < \epsilon n^2\right] \leq e^{-\frac{1}{64}\delta^2 n^2} \tag{B.27}$$

The number of ways to choose a subset A of size at least δn from U_i is at most 2^n. Therefore, by applying the union bound, the probability that there exists a set $A \subseteq U_i$ of size at least δn in which the number of edges is less than ϵn^2 is at most $2^n \cdot e^{-\frac{1}{64}\delta^2 n^2}$.

That is,

$$\Pr\left[\overline{\mathcal{E}_i^1}\right] \leq 2^n \cdot e^{-\frac{1}{64}\delta^2 n^2} \leq e^{2n} \cdot e^{-\frac{1}{64}\delta^2 n^2} \leq e^{-n\left(\frac{\delta^2}{64}n - 2\right)} \tag{B.28}$$

Similarly, $\Pr\left[\overline{\mathcal{E}_i^0}\right] \leq e^{-n\left(\frac{\delta^2}{64}n-2\right)}$. The claim follows. $\square$

Claim 21. *For every $i \in [k]$, $\Pr\left[\overline{\mathcal{E}_{ij}^1}\right] \leq e^{-n\left(\frac{\delta^2}{64}n-2\right)}$ and $\Pr\left[\overline{\mathcal{E}_{ij}^0}\right] \leq e^{-n\left(\frac{\delta^2}{64}n-2\right)}$.*

Proof. As in the proof of the previous claim, we only prove $\Pr\left[\overline{\mathcal{E}_{ij}^1}\right] \leq e^{-n\left(\frac{\delta^2}{64}n-2\right)}$, because proving the bound on $\Pr\left[\overline{\mathcal{E}_i^0}\right]$ is very similar. Let $(A, B) \subseteq U_i \times U_j$ be two disjoint sets, each of size at least δn. We explicitly require the two subsets to be disjoint because if $i = j$ then an arbitrary pair of subsets are not necessarily disjoint. We bound from above the probability that the number of edges between A and B in the graph G' is at most ϵn^2. In order to obtain that upper bound we consider a pair of subsets $(A', B') \subseteq A \times B$ where $|A'| = |B'| = \delta n$. Clearly,

$$\Pr\left[e_{G'}(A, B) < \epsilon n^2\right] \leq \Pr\left[e_{G'}\left(A', B'\right) < \epsilon n^2\right] \tag{B.29}$$

As in the previous proof, we define a random variable χ_{uv} for every pair of vertices $(u,v) \in A' \times B'$ indicating whether or not there exists an edge between u and v. We define the random variable χ as the sum of the indicators. That is, $e_{G'}(A', B') = \chi$. Hence, we have to bound from above the probability that $\chi < \epsilon n^2$. Again, in order to apply Chernoff's bound, we compute the expectation of χ.

$$E[\chi] = \sum_{(u,v)\in A'\times B'} \chi_{uv} = \sum_{u,v\in A} \frac{1}{2} = \frac{1}{2} \cdot \delta^2 n^2 \tag{B.30}$$

We apply Chernoff's bound:

$$\Pr\left[\chi < \epsilon n^2\right] \leq \Pr\left[\chi < \frac{1}{2} E[\chi]\right] \leq e^{-\left(\frac{1}{2}\right)^2 \cdot \frac{E[\chi]}{2}} = e^{-\frac{1}{16}\cdot\delta^2 n^2} \leq e^{-\frac{1}{64}\delta^2 n^2} \tag{B.31}$$

Hence,

$$\Pr\left[e_{G'}(A,A) < \epsilon n^2\right] \leq e^{-\frac{1}{64}\delta^2 n^2} \tag{B.32}$$

The number of ways to choose a pair of disjoint subsets (A,B) each of size at least δn from $U_i \times U_j$ is at most $2^n \cdot 2^n = 4^n$ (in fact, if $i = j$ then the number of ways is at most $3^n < 4^n$). Therefore, by applying the union bound, the probability that there exists a pair of subsets $(A,B) \subseteq U_i \times U_j$ where each subset is of size at least δn and the number of edges between them is less than ϵn^2 is at most $4^n \cdot e^{-\frac{1}{64}\delta^2 n^2}$.

That is,

$$\Pr\left[\overline{\mathcal{E}^1_{ij}}\right] \leq 4^n \cdot e^{-\frac{1}{64}\delta^2 n^2} \leq e^{2n} \cdot e^{-\frac{1}{64}\delta^2 n^2} \leq e^{-n\left(\frac{\delta^2}{64}n - 2\right)} \tag{B.33}$$

Similarly, $\Pr\left[\overline{\mathcal{E}^0_{ij}}\right] \leq e^{-n\left(\frac{\delta^2}{64}n-2\right)}$. The claim follows. $\square$

Now we can prove claim 17. That is, we prove that $\Pr[\mathcal{E}] > 0$.

Proof of Claim 17. We use the union bound to bound from above the probability of the event $\overline{\mathcal{E}}$.

For every i where $d_P(i,i) = \perp$:

$$\Pr\left[\overline{\mathcal{E}_i}\right] \leq 2 \cdot \Pr\left[\overline{\mathcal{E}^0_i}\right] \leq 2 \cdot e^{-n\left(\frac{\delta^2}{64}n-2\right)} \tag{B.34}$$

For every i, j where $d_P(i,j) = \perp$:

$$\Pr\left[\overline{\mathcal{E}_{ij}}\right] \leq 2 \cdot \Pr\left(\overline{\mathcal{E}^0_{ij}}\right) \leq 2 \cdot e^{-n\left(\frac{\delta^2}{64}n-2\right)} \tag{B.35}$$

The number of events $\mathcal{E}_i$ is at most k and the number of events $\mathcal{E}_{ij}$ is at most k^2. Hence the number of events we use in the union bound is at most $k + k^2 \leq 2k^2$

$$\Pr\left[\overline{\mathcal{E}}\right] = 2k^2 \cdot 2 \cdot e^{-n\left(\frac{\delta^2}{64}n-2\right)} = 4k^2 \cdot e^{-n\left(\frac{\delta^2}{64}n-2\right)} < 1 \tag{B.36}$$

where the last inequality holds since we can choose n to be sufficiently large.

Hence, $\Pr[\mathcal{E}] > 0$ $\square$

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il