A Combinatorial Characterization of the Testable Graph Properties: It's All About Regularity*

Noga Alon[†] Eldar Fischer[‡] Ilan Newman[§] Asaf Shapira[¶]

Abstract

A common thread in all the recent results concerning the testing of dense graphs is the use of Szemerédi's regularity lemma. In this paper we show that in some sense this is not a coincidence. Our first result is that the property defined by having any given Szemerédi-partition is testable with a constant number of queries. Our second and main result is a purely combinatorial characterization of the graph properties that are testable with a constant number of queries. This characterization (roughly) says that a graph property $\mathcal{P}$ can be tested with a constant number of queries **if and only if** testing $\mathcal{P}$ can be reduced to testing the property of satisfying one of finitely many Szemerédi-partitions. This means that in some sense, testing for Szemerédi-partitions is as hard as testing any testable graph property. We thus resolve one of the main open problems in the area of property-testing, which was first raised in the 1996 paper of Goldreich, Goldwasser and Ron [27] that initiated the study of graph property-testing. This characterization also gives an intuitive explanation as to what makes a graph property testable.

1 Background

1.1 Basic definitions

The meta-problem in the area of property testing is the following: given a combinatorial structure S, one should distinguish between the case that S satisfies some property $\mathcal{P}$ and the case that S is ϵ-*far* from satisfying $\mathcal{P}$. Roughly speaking, a combinatorial structure is said to be ϵ-far from satisfying some property $\mathcal{P}$ if an ϵ-fraction of its representation should be modified in order to make S satisfy $\mathcal{P}$. The main goal is to design randomized algorithms, which look at a very small portion of the input, and using this information distinguish with high probability between the above two cases.

*A preliminary version of this paper appeared in the Proc. of the 38^{th} Annual ACM Symposium on Theory of Computing (STOC), ACM Press (2006), 251-260.

[†]Schools of Mathematics and Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel, and IAS, Princeton, NJ. Email: nogaa@tau.ac.il. Research supported in part by a USA-Israeli BSF grant, by the Israel Science Foundation and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University.

[‡]Faculty of Computer Science, Technion – Israel institute of technology, Technion City, Haifa 32000, Israel. Email: eldar@cs.technion.ac.il. Research supported in part by an Israel Science Foundation grant number 55/03.

[§]Department of Computer Science, Haifa University, Haifa 31905, Israel. Email: ilan@cs.haifa.ac.il. Research supported in part by an Israel Science Foundation grant number 55/03.

[¶]School of Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel. Email: asafico@tau.ac.il. This work forms part of the author's Ph.D. thesis. Research supported in part by a Charles Clore Foundation Fellowship.

Such algorithms are called *property testers* or simply *testers* for the property $\mathcal{P}$. Preferably, a tester should look at a portion of the input whose size can be upper bounded by a function of ϵ only. Blum, Luby and Rubinfeld [12] were the first to formulate a question of this type, and the general notion of property testing was first formulated by Rubinfeld and Sudan [37], who were interested in studying various algebraic properties such as linearity of functions.

The main focus of this paper is the testing of properties of graphs. More specifically, we focus on property testing in the dense graph model as defined in [27]. In this case a graph G is said to be ϵ*-far* from satisfying a property $\mathcal{P}$, if one needs to add and/or delete at least ϵn^2 edges to G in order to turn it into a graph satisfying $\mathcal{P}$. A tester for $\mathcal{P}$ should distinguish with high probability, say 2/3, between the case that G satisfies $\mathcal{P}$ and the case that G is ϵ-far from satisfying $\mathcal{P}$. More precisely, it should accept G with probability at least 2/3 if it satisfies $\mathcal{P}$ and reject G with probability at least 2/3 if it is ϵ-far from $\mathcal{P}$. Here we assume that the tester can query some oracle whether a pair of vertices, i and j, are adjacent in the input graph G. We also assume that the tester has access to the size of the input graph[1]. In what follows we will say that a tester for a graph property $\mathcal{P}$ has *one-sided error* if it accepts every graph satisfying $\mathcal{P}$ with probability 1 (and still rejects those that are ϵ-far from $\mathcal{P}$ with probability at least 2/3). If the tester may reject graphs satisfying $\mathcal{P}$ with non-zero probability then it is said to have *two-sided error.* The following notion of efficient testing will be the main focus of this paper:

Definition 1.1 (Testable) *A graph property $\mathcal{P}$ is* testable *if there is a randomized algorithm $\mathcal{T}$, that can distinguish for every $\epsilon > 0$ and with probability 2/3, between graphs satisfying $\mathcal{P}$ and graphs that are ϵ-far from satisfying $\mathcal{P}$, while making a number of edge queries[2] which is bounded by some function $q(\epsilon)$ that is independent of the size of the input.*

The study of the notion of testability for combinatorial structures, and mainly the dense graph model, was introduced in the seminal paper of Goldreich, Goldwasser and Ron [27]. Graph property testing has also been studied in the *bounded-degree* model [29], and the newer *general density* model [33]. We note that in these models a property is usually said to be testable if the number of queries is $o(n)$. Following [12, 27, 37] property testing was studied in various other contexts such as boolean functions [4, 19, 21, 22, 32, 34], geometric objects [2, 14] and algebraic structures [10, 12, 23, 24]. See the surveys [17, 36] for additional results and references.

1.2 Background on the characterization project

With this abundance of results on property testing, a natural question is what makes a combinatorial property testable. In particular, characterizing the testable graph properties was considered one of the main open problems in the area of property testing, and was raised already in the 1996 paper of Goldreich, Goldwasser and Ron [27], see also [11, 25, 28]. In this paper we obtain for the first time a characterization of the testable graph properties. We next discuss some results related to this problem.

A natural strategy toward obtaining a characterization of the testable graphs was to either prove the testability/non-testability of general families of graph properties or to obtain characterizations

[1]We need this extra assumption to avoid certain issues related to the computability of the query complexity as a function of ϵ. See [8] for a thorough discussion of these issues.

[2]We allow the tester to be adaptive, although one can assume with a slight loss of generality that it is non-adaptive. See Lemma 4.2.

for special cases of testers. The main result of [27] was that a general family of so called "partition-problems" are all testable. These include the properties of being k-colorable, having a large cut and having a large clique. [28] gave a characterization of the partition-problems that can be tested with 1-sided error. They also proved that not all graph properties that are closed under edge-removal are testable. [15] studied property testing via the framework of *abstract combinatorial programs* and gave certain characterizations within this framework. [3] tried to obtain a *logical* characterization of the testable graph properties. More specifically, it was shown that every first order graph-property of type $\exists\forall$ (see [3]) is testable, while there are first-order graph properties of type $\forall\exists$ that are not testable. The main technical result of [3] was that certain abstract colorability properties are all testable. These results were generalized in [16]. In [6] it was shown that every graph property that is closed under removal of edges and vertices is testable. This result was extended in [7], where it was shown that in fact, being closed under vertex removal is already sufficient for being testable (see also [13] for an alternative proof). [7] also contains a characterization of the graph properties that can be tested with *one-sided* error by certain restricted testers. Finally, [28] following [3], proved that a tester may be assumed to be non-adaptive (see Lemma 4.2 below), and [20] proved that if a graph property is testable then it is also possible to estimate how far is a given graph from satisfying the property (see Theorem 3 below). These last two results are key ingredients in the present paper.

2 The Main Result

2.1 Background on Szemerédi's regularity lemma

Our main result in this paper gives a purely combinatorial characterization of the testable graph properties. As we have previously mentioned, the first properties that were shown to be testable in [27] were certain graph partition properties. As it turns out, our characterization relies on certain "enhanced" partition properties, whose existence is guaranteed by the celebrated regularity lemma of Szemerédi [38]. We start by introducing some standard definitions related to the regularity lemma. For a comprehensive survey about the regularity lemma the reader is referred to [31].

For every two nonempty disjoint vertex sets A and B of a graph G, we define $e(A, B)$ to be the number of edges of G between A and B. The *edge density* of the pair is defined by $d(A, B) = e(A, B)/(|A||B|)$.

Definition 2.1 (γ-regular pair) *A pair (A, B) is* γ-regular*, if for any two subsets $A' \subseteq A$ and $B' \subseteq B$, satisfying $|A'| \geq \gamma|A|$ and $|B'| \geq \gamma|B|$, the inequality $|d(A', B') - d(A, B)| \leq \gamma$ holds.*

Throughout the paper it will be useful to observe that in the above definition it is enough to require that $|d(A', B')-d(A, B)| \leq \gamma$ for sets $A' \subseteq A$ and $B' \subseteq B$ of sizes $|A'| = \gamma|A|$ and $|B'| = \gamma|B|$. A partition $\mathcal{A} = \{V_i \mid 1 \leq i \leq k\}$ of the vertex set of a graph is called an *equipartition* if $|V_i|$ and $|V_j|$ differ by no more than 1 for all $1 \leq i < j \leq k$ (so in particular every V_i has one of two possible sizes). The *order* of an equipartition denotes the number of partition classes (k above).

Definition 2.2 (γ-regular equipartition) *An equipartition $\mathcal{B} = \{V_i \mid 1 \leq i \leq k\}$ of the vertex set of a graph is called* γ-regular *if all but at most $\gamma\binom{k}{2}$ of the pairs (V_i, V_j) are γ-regular.*

In what follows an equipartition is said to *refine* another if every set of the former is contained in one of the sets of the latter. Szemerédi's regularity lemma can be formulated as follows.

Lemma 2.3 ([38]) *For every m and $\gamma > 0$ there exists $T = T_{2.3}(m, \gamma)$ with the following property: If G is a graph with $n \geq T$ vertices, and $\mathcal{A}$ is any equipartition of the vertex set of G of order at most m, then there exists a refinement $\mathcal{B}$ of $\mathcal{A}$ of order k, where $m \leq k \leq T$ and $\mathcal{B}$ is γ-regular. In particular, for every m and $\gamma > 0$ there exists $T = T_{2.3}(m, \gamma)$, such that any graph with $n \geq T$ vertices has a γ-regular equipartition of order k, where $m \leq k \leq T$.*

The regularity lemma guarantees that every graph has a γ-regular equipartition of (relatively) small order. As it turns out in many applications of the regularity lemma, one is usually interested in the densities of the bipartite graphs connecting the sets V_i of the regular partitions. In fact, one important consequence of the regularity lemma is that in many cases knowing the densities connecting the sets V_i (approximately) tells us all we need to know about a graph. Roughly speaking, if a graph G has a regular partition of order k and we define a weighted complete graph $R(G)$, of size k, where the weight of edge (i, j) is $d(V_i, V_j)$, then by considering an appropriate property of $R(G)$ one can infer many properties of G. As the order of the equipartition is guaranteed to be bounded by a function of γ, this means that for many applications, every graph has an approximate description of *constant-complexity* (we will return to this aspect in a moment). As it turns out, this interpretation of the regularity lemma is the key to our characterization. We believe that our characterization of the testable graph properties is an interesting application of this aspect of the regularity lemma.

Given the above discussion it seems natural to define a graph property, which states that a graph has a given γ-regular partition, that is, an equipartition which is γ-regular and such that the densities between the sets V_i belong to some predefined set of densities.

Definition 2.4 (Regularity-Instance) *A* regularity-instance *R is given by an error-parameter $0 < \gamma \leq 1$, an integer k, a set of $\binom{k}{2}$ densities $0 \leq \eta_{ij} \leq 1$ indexed by $1 \leq i < j \leq k$, and a set $\overline{R}$ of pairs (i, j) of size at most $\gamma\binom{k}{2}$. A graph is said to* satisfy *the regularity-instance if it has an equipartition $\{V_i \mid 1 \leq i \leq k\}$ such that for all $(i, j) \notin \overline{R}$ the pair (V_i, V_j) is γ-regular and satisfies $d(V_i, V_j) = \eta_{i,j}$. The* complexity *of the regularity-instance is* $\max(k, 1/\gamma)$.

Comment 2.5 *In the above definition, as well as throughout the paper, when we say that $d(V_i, V_j) = \eta_{i,j}$ we mean that the number of edges between V_i and V_j is $\lfloor \eta_{i,j}|V_i||V_j| \rfloor$. This will allow us to disregard divisibility issues that will have no real difference in any of our proofs.*

Note, that in the above definition the set $\overline{R}$ corresponds to the set of pairs (i, j) for which (V_i, V_j) is not necessarily a γ-regular pair (note that there may be up to $\gamma\binom{k}{2}$ such pairs). Also, note that the definition of a regularity-instance does not impose any restriction on the graphs spanned by any single set V_i. By Lemma 2.3, for any $0 < \gamma \leq 1$ any graph satisfies some regularity instance with an error parameter γ and with an order bounded by a function of γ. The first step needed in order to obtain our characterization of the testable properties, is that the property of satisfying any given regularity-instance is testable. This is also the main technical result of this paper.

Theorem 1 *For any regularity-instance R, the property of satisfying R is testable.*

2.2 The characterization

Many of the recent results on testing graph properties in the dense graph model relied on the Regularity Lemma. Our main result shows that this is not a coincidence. Previous results which applied the regularity lemma to test a graph property used it to infer a property of the graph that could actually be inferred by looking at a regular partition of the graph. These results however, use the properties of the regular partition in an *implicit* way. For example, the main observation needed in order to infer that triangle-freeness is testable, is that if the regular partition has three sets V_i, V_j, V_k, which are connected by regular and dense bipartite graphs, then the graph contains many triangles. However, to *test* triangle freeness we do not need to know the regular partition, we just need to find a triangle in the graph. As Theorem 1 allows us to test for having a certain regular partition, it seems natural to try and test properties by *explicitly* checking for properties of the regular partition of the input. Returning to the previous discussion on viewing the regularity lemma as a constant complexity description of a graph, being able to explicitly test for having a given regular partition should allow us to test more complex properties as we can obtain all the information about the regular partition and not just certain *consequences* of having some regular partition. The next definition tries to capture the graph properties $\mathcal{P}$ that can be tested via testing a certain set of regularity instances.

Definition 2.6 (Regular-Reducible) *A Graph property $\mathcal{P}$ is* regular-reducible *if for any $\delta > 0$ there exists an $r = r_{\mathcal{P}}(\delta)$ such that for*[3] *any n there is a family $\mathcal{R}$ of at most r regularity-instances each of complexity at most r, such that the following holds for every $\epsilon > 0$ and every n-vertex graph G:*

1. *If G satisfies $\mathcal{P}$ then for some $R \in \mathcal{R}$, G is δ-close to satisfying R.*
2. *If G is ϵ-far from satisfying $\mathcal{P}$, then for any $R \in \mathcal{R}$, G is $(\epsilon - \delta)$-far from satisfying R.*

The reader may observe that in the above definition the value of δ may be arbitrarily close to 0. If we think of $\delta = 0$ then we get that a graph satisfies $\mathcal{P}$ if and only if it satisfies one of the regularity instances of $\mathcal{R}$. With this interpretation in mind, in order to test $\mathcal{P}$ one can test the property of satisfying any one of the instances of $\mathcal{R}$. Therefore, in some sense we "reduce" the testing of the property $\mathcal{P}$ to the testing of regularity-instances. We are now ready to state our characterization of the testable graph properties.

Theorem 2 (Main Result) *A graph property is testable* **if and only if** *it is regular-reducible.*

If we have to summarize the moral of our characterization in one simple sentence, then it says that a graph property $\mathcal{P}$ is testable if and only if $\mathcal{P}$ is such that knowing a regular partition of a graph G is sufficient for telling whether G is far or close to satisfying $\mathcal{P}$. In other words, there is a short "proof" that G is either close or far from satisfying $\mathcal{P}$. Thus, in a more "computational complexity" jargon, we could say that a graph property is testable if and only if it has the following "interactive proof": A prover gives a verifier the description of a regularity-instance R, which the input G is (supposedly) close to satisfying. The verifier, using Theorem 1, then verifies if G is indeed close[4] to satisfying R. The way to turn this interactive proof into a testing algorithm is to apply the constant-complexity

[3] We note that we allow different values of n to have different regularity instances in order to handle cases where the property we are testing depends on the size of the graphs. Some papers (e.g. [7, 13]) consider only properties whose definition does not involve the size of the input.

[4] Actually, Theorem 1 only allows us to test if a graph satisfies $\mathcal{R}$. To be able to tell if a graph is close to $\mathcal{R}$ we will apply a result from [20]

properties of the regularity lemma that we have previously discussed; as the order of the regular partition is bounded by a function of ϵ, there are only *finitely* many regularity-instances that the prover may potentially send to the verifier. Therefore, the verifier does not need to get an alleged regularity-instance, it can simply try them all! Theorem 2 thus states that in some sense testing regularity-instances is the "hardest" property to test, because by Theorem 2 any testing algorithm can be turned into a testing algorithm for regularity-instances. However, we stress that this is true only on the *qualitative* level. The reason is that applying Theorem 2 to a testable property $\mathcal{P}$, thus obtaining a tester that tests for regularity-instances, may result in a tester whose query complexity is much larger than the query complexity of the optimal tester for $\mathcal{P}$. The main reason is that the proofs of Theorems 1 and 2 apply Lemma 2.3 and thus only give weak upper bounds (it is known that the constants in Lemma 2.3 have a tower-type dependence on γ, see [30]). We also note that the reason for choosing the term regular-reducible in Definition 2.6 is because in order to prove one of the directions of Theorem 2 we actually test a property $\mathcal{P}$, which is regular-reducible to a set $\mathcal{R}$, by testing the regularity-instances of $\mathcal{R}$. Theorem 2 also gives further convincing evidence to the "combinatorial" nature of property testing in the dense graph model as was recently advocated by Goldreich [26].

As is evident from Definition 2.6, the characterization given in Theorem 2 is not a "quick recipe" for inferring whether a given property is testable. Still, we can use Theorem 2 in order to obtain unified proofs for several previous results. As we have alluded to before, these results can be inferred by showing that it is possible (or impossible) to reduce the testing of the property to testing if a graph satisfies certain regularity-instances. We believe that these proofs give some (non-explicit) structural explanation as to what makes a graph property testable. See Section 7 for more details. It is thus natural to ask if one can come up with more "handy" characterizations. We doubt that such a characterization exists, mainly because it should (obviously) be equivalent to Theorem 2. One supporting evidence is a recent related (and independent) study of graph homomorphism [13] that led to a different characterization of the testable graph *parameters*, which is also somewhat complicated to apply. See [13] for more details.

2.3 Organization and overview of the paper

The first main technical step of the proof of Theorem 1 is taken in Section 3. In this section we prove that if the densities of pairs of subsets of vertices of a bipartite graph are close to the density of the bipartite graph itself, then the bipartite graph can be turned into a regular-pair using relatively few edge modifications. Rephrasing this gives that we can increase the regularity measure of a bipartite pair by making relatively few edge modifications. The second main step is taken in Section 5. In this section, we show that sampling a constant number of vertices guarantees that the sample and the graph will have (roughly) the *same* set of regular partitions. We believe that this result may be of independent interest. By applying the results of Sections 3 and 5 we prove Theorem 1 in Section 6. In this section we also prove one of the directions of Theorem 2, asserting that if a graph property is regular-reducible then it is testable. Along with Theorem 1, a second tool that we need in order to prove this direction is the main result of [20]. We apply this result in order to infer that for any regularity-instance R, one can not only test the property of satisfying R, but can also estimate how far is a given graph from satisfying R. This *estimation* of the distance to satisfying regularity-instances is key to *testing* a property via a regularity-reduction.

The proof of the second direction of Theorem 2 appears in Section 4. To prove this direction we first show that knowing that a graph G satisfies a regularity instance enables us to estimate the number of copies of certain graphs in G. We then apply the main result of [28] about canonical testers along with the main result of Section 3 in order to "pick" those regularity-instances that can constitute the family $\mathcal{R}$ in Definition 2.6. In Section 7 we use Theorem 2 in order to reprove some previously known results in property-testing. The main interest of these proofs is that they apply Theorem 2 in order to prove in a unified manner results that had distinct proofs. Section 8 contains some concluding remarks.

Throughout the paper, whenever we use a notation like $\alpha_{3.2}$ we refer to the constant α defined in Lemma/Theorem/Claim 3.2. To avoid using floor/ceiling signs, in most parts of the paper we will assume that the number of vertices of a graph is divisible by some small integer k. This will allow us to assume, for example, that the sets of an equipartition are all of the same size. This will not change any of the asymptotic results of the paper.

3 Enhancing Regularity with Few Edge Modifications

The definition of a γ-regular pair of density η requires a pair of sets of vertices to satisfy several density requirements. The main goal in this section is to show that if a pair of vertex sets "almost" satisfy these requirements, then it is indeed close to a γ-regular pair of density η. For example, consider the property of being a 0.1-regular pair with edge density 0.4. Intuitively, it seems that if the edge density of a bipartite graph G on vertex sets A and B of size m each is close to 0.4, and the density of any pair $A' \subseteq A$ and $B' \subseteq B$ of sizes $0.1m$ is close to 0.4 ± 0.1, then G should be close to satisfying the property. However, note that it may be the case that there are pairs (A', B'), whose density is smaller than 0.3, and other pairs, whose density is larger than 0.5. Thus, only removing or only adding edges (even randomly) will most likely not turn G into a 0.1-regular pair of density 0.4. In order to show that G is indeed close to satisfying the property, we take a "convex combination" of G with a random graph, whose density is 0.4. The intuition is that the random graph will not change the density of G much, but, because a random graph is highly regular, it will increase the regularity of G. The main result of this section is formalized in the following lemma, which is an important ingredient in the proofs of both directions of Theorem 2.

In this lemma, as well as throughout the rest of the paper, when we write $x = a \pm b$ we mean that $a - b \leq x \leq a + b$.

Lemma 3.1 *The following holds for any $0 < \delta \leq \gamma \leq 1$: Suppose that (A, B) is a $(\gamma + \delta)$-regular pair with density $\eta \pm \delta$, where $|A| = |B| = m \geq m_{3.1}(\eta, \delta)$. Then, it is possible to make at most $50\frac{\delta}{\gamma^2}m^2$ edge modifications and turn (A, B) into a γ-regular pair with density precisely η.*

The proof of Lemma 3.1 has two main steps, which are captured in Lemmas 3.2 and 3.3 below. The first step, given in Lemma 3.2 below, enables us to make relatively few edge modifications and thus make sure that the density of a pair is exactly what it should be, while at the same time not decreasing its regularity by much.

Lemma 3.2 *Suppose that (A, B) is a $(\gamma + \delta)$-regular pair satisfying $d(A, B) = \eta \pm \delta$, where $|A| = |B| = m \geq m_{3.2}(\eta, \delta)$. Then, it is possible to make at most $2\delta m^2$ modifications, and thus turn (A, B) into a $(\gamma + 2\delta)$-regular pair with density precisely η.*

The second and main step, which implements the main idea presented at the beginning of this section, takes a bipartite graph, whose density is precisely η, and returns a bipartite graph, whose density is still η but with a better regularity measure.

Lemma 3.3 *The following holds for any $0 < \delta \leq \gamma \leq 1$. Let A and B be two vertex sets of size $m \geq m_{3.3}(\delta, \gamma)$, satisfying $d(A, B) = \eta$. Suppose further that for any pair of subsets $A' \subseteq A$ and $B' \subseteq B$ of size γm we have $d(A', B') = \eta \pm (\gamma + \delta)$. Then, it is possible to make at most $\frac{3\delta}{\gamma}m^2$ edge modifications that turn (A, B) into a γ-regular pair with density precisely η.*

We now turn to prove the above three lemmas. Following them is a corollary of Lemma 3.1, which will be used in the proof of Theorem 2. For the proofs of this section we need the following standard Chernoff-type large deviation inequality (see, e.g., the appendix of [9]).

Lemma 3.4 *Suppose $X_1, \dots, X_n$ are n independent Boolean random variables, where $Prob[X_i = 1] = p_i$. Let $E = \sum_{i=1}^{n} p_i$. Then, $Prob[|\sum_{i=1}^{n} X_i - E| \geq \delta n] \leq 2e^{-2\delta^2 n}$.*

Proof of Lemma 3.2: Suppose that $d(A, B) = \eta + p$, where $|p| \leq \delta$, and assume for now that $0 \leq p \leq \delta$. Suppose first that $p \leq \delta(\gamma + 2\delta)^2$. In this case we just remove any $pm^2 (\leq \delta m^2)$ edges and thus make sure that $d(A, B) = \eta$. Furthermore, as for any pair (A', B') of size $(\gamma + 2\delta)m$ we initially had $d(A', B') = \eta + p \pm (\gamma + \delta)$, it is easy to see that because we remove $pm^2 \leq \delta(\gamma + 2\delta)^2 m^2$ edges, we now have $\eta - \gamma - 2\delta \leq d(A', B') \leq \eta + \gamma + \delta$, which satisfies $d(A', B') = \eta \pm (\gamma + 2\delta)$. Thus, in this case we turned (A, B) into a $(\gamma + 2\delta)$-regular pair of density η.

Suppose now that $p \geq \delta(\gamma + 2\delta)^2$. Our way for turning (A, B) into a $(\gamma + 2\delta)$-regular pair with density η will consist of two stages. In the first we will randomly remove some of the edges between A and B. We will then deterministically make some additional modifications. To get that after these two stages (A, B) has the required properties we show that with probability $3/4$ the pair (A, B) is $(\gamma + 2\delta)$-regular and with the same probability $d(A, B) = \eta$. By the union bound we will get that with probability at least $1/2$ the pair (A, B) has the required two properties.

In the first (random) step, we remove each of the edges connecting A and B randomly and independently with probability $\frac{p}{\eta + p}$. Then the expected number of edges removed is

$$\frac{p}{\eta + p}(\eta + p)|A||B| = p|A||B| \leq \delta|A||B| \ ,$$

and the expected value of $d(A, B)$ is thus η. As we assumed that $p \geq \delta(\gamma + 2\delta)^2$ we have $d(A, B) \geq p \geq \delta(\gamma + 2\delta)^2$. Therefore, by Lemma 3.4, for large enough $m \geq m_{3.2}(\delta, \gamma)$, the probability that $d(A, B)$ deviates from η by more than $1/m^{0.5}$ is at most $1/4$ [5]. In particular, the number of edge modifications made is at most $\frac{3}{2}\delta m^2$ with probability at least $3/4$. Now (this is the second, deterministic step) we can add or remove at most $m^{1.5}$ edges arbitrarily and thus make sure that $d(A, B) = \eta$. The total number of edge modifications is also at most $\frac{3}{2}\delta m^2 + m^{1.5} \leq 2\delta m^2$, for large enough $m \geq m_{3.2}(\delta, \gamma)$. Note that we have thus established that with probability at least $3/4$ after the above two stages $d(A, B) = \eta$.

[5] Note that we need a lower bound on the density as otherwise we cannot apply Lemma 3.4 to obtain a tail bound that goes to zero fast enough.

As (A, B) was assumed to be $(\gamma + \delta)$-regular, we initially had $d(A', B') = \eta + p \pm (\gamma + \delta)$ for any pair of subsets $A' \subseteq A$ and $B' \subseteq B$ of size $(\gamma + 2\delta)m$. As in the first step we removed each edge with probability $\frac{p}{\eta+p}$, the expected value of $d(A'B')$ after the first step is between

$$(\eta + p + \gamma + \delta)(1 - \frac{p}{\eta + p}) \leq \eta + \gamma + \delta$$

and

$$(\eta + p - \gamma - \delta)(1 - \frac{p}{\eta + p}) \geq \eta - \gamma - \delta.$$

Recall that we have already established that with probability at least 3/4 we have $d(A, B) = \eta$ and that for any pair (A', B') of size $(\gamma + 2\delta)m$ the expected value of $d(A', B')$ is $\eta \pm (\gamma + \delta)$. Hence, to show that after the two steps (A, B) is a $(\gamma + 2\delta)$-regular pair with probability at least 1/2, it suffices to show that with probability at least 3/4, the densities of all pairs (A', B') do not deviate from their expectation by more than δ.

Suppose first that $d(A', B')$ was originally at most $\frac{1}{2}\delta$. This means that when we randomly remove edges from (A, B) we can change $d(A', B')$ by at most $\frac{1}{2}\delta$. Thus in this case $d(A', B')$ can deviate from its expectation by at most $\frac{1}{2}\delta$. Also, when adding or removing $m^{1.5}$ edges to (A, B) in the second step we can change $d(A', B')$ by at most $1/m^{0.5}(\gamma+2\delta)^2 \leq \frac{1}{2}\delta$ for large enough $m \geq m_{3.2}(\delta, \gamma)$. Thus, for such pairs we are guaranteed that $d(A', B') = \eta \pm (\gamma + 2\delta)$.

Suppose now that $d(A', B')$ was at least $\frac{1}{2}\delta$. Thus the number of edges, which were considered for removal between A' and B' in the first step was at least $\frac{1}{2}\delta(\gamma+2\delta)^2m^2$. Hence, by Lemma 3.4 the probability that $d(A', B')$ deviates from its expectation by more than $\frac{1}{2}\delta$ is at most $2e^{-2(\frac{1}{2}\delta)^2\frac{1}{2}\delta(\gamma+2\delta)^2m^2}$. Thus, as there are at most 2^{2m} pairs of such sets (A', B'), we conclude by the union-bound that for large enough $m \geq m_{3.2}(\delta, \gamma)$, with probability at least 3/4 all sets (A', B') of size $(\gamma + 2\delta)m$ satisfy $d(A', B') = \eta \pm (\gamma + \frac{3}{2}\delta)$. As in the previous paragraph, adding or removing $m^{1.5}$ edges in the second step can change $d(A', B')$ by at most $\frac{1}{2}\delta$, so in this case we also have $d(A', B') = \eta \pm (\gamma + 2\delta)$.

Finally, in the case that p above satisfies $-\delta \leq p \leq 0$ we can use essentially the same argument. The only modification is that we add edges instead of remove them. ■

Proof of Lemma 3.3: For any pair of vertices $a \in A$ and $b \in B$ we do the following: we flip a coin that comes up heads with probability $\frac{2\delta}{(\delta+\gamma)}$ and tails with probability $1 - \frac{2\delta}{(\delta+\gamma)}$. If the coin comes up tails we make no modification between the vertices a and b. If the coin comes up heads then we disregard the adjacency relation between a and b and do the following: we flip another coin that comes up heads with probability η and tails with probability $1 - \eta$. If the coin comes up heads then we put an edge connecting a and b, and otherwise we do not put such an edge. In what follows we call the coins flipped in the first step the *first* coins, and those flipped in the second step the *second* coins.

Claim 3.5 *With probability at least* $3/4$*, we make at most* $\frac{3\delta}{\gamma}m^2$ *edge modifications.*

Proof: Note that the number of edge modifications is at most the number of first coins that came up heads. The distribution of these m^2 coins is given by the Binomial distribution $B(m^2, \frac{2\delta}{(\delta+\gamma)})$, whose expectation is $\frac{2\delta}{(\delta+\gamma)}m^2$, and by Lemma 3.4 the probability of deviating by more than $\frac{1}{2}\delta m^2$

from this expectation is at most $2e^{-2(\delta/2)^2 m^2}$. For large enough $m \geq m_{3.3}(\delta,\gamma)$ we get that with probability at least 3/4 we make at most $\frac{2\delta}{(\delta+\gamma)}m^2 + \frac{1}{2}\delta m^2 \leq \frac{2.5\delta}{\gamma}m^2$ modifications. ■

The following observation will be useful for the next two claims: Fix a pair of adjacent vertices $a \in A$ and $b \in B$. For them to become non-adjacent the first coin should come up heads (this happens with probability $\frac{2\delta}{(\delta+\gamma)}$) and the second tails (this happens with probability $(1-\eta)$), thus the probability of them staying adjacent is $(1-\frac{2\delta}{(\delta+\gamma)}+\frac{2\eta\delta}{(\delta+\gamma)})$. Now, fix a pair of non-adjacent vertices $a \in A$ and $b \in B$. For them to become adjacent, both coins must come up heads, so the probability of them becoming adjacent is $\frac{2\eta\delta}{(\delta+\gamma)}$.

Claim 3.6 *With probability at least* 3/4*, we have* $d(A,B) = \eta \pm 1/m^{0.5}$.

Proof: Recall that by assumption the number of adjacent vertices was ηm^2. Thus, by the above observation the expected number of adjacent vertices is

$$\eta m^2(1-\frac{2\delta}{(\delta+\gamma)}+\frac{2\eta\delta}{(\delta+\gamma)}) + (1-\eta)m^2\frac{2\eta\delta}{(\delta+\gamma)} = \eta m^2.$$

By Lemma 3.4 we get that for large enough $m \geq m_{3.3}(\delta,\gamma)$ the probability of deviating from this expectation by more than $m^{1.5}$ is at most 1/4. Normalizing by m^2 we get the required bound on $d(A,B)$. ■

Claim 3.7 *With probability at least* 3/4*, all sets* $A' \subseteq A$ *and* $B' \subseteq B$*, whose size is* γm*, satisfy* $d(A',B') = \eta \pm (\gamma - \frac{1}{2}\delta)$.

Proof: Fix any pair of such sets. Let e denote the number of edges originally spanned by these sets. As in the previous claim we get that the expected number of edges spanned by (A',B') is

$$e(1-\frac{2\delta}{(\delta+\gamma)}+\frac{2\eta\delta}{(\delta+\gamma)}) + (|A'||B'|-e)\frac{2\eta\delta}{(\delta+\gamma)} = e(1-\frac{2\delta}{(\delta+\gamma)}) + |A'||B'|\frac{2\eta\delta}{(\delta+\gamma)}.$$

Recall that by assumption $e = |A'||B'|(\eta \pm (\gamma+\delta))$. Thus, the expected number of edges spanned by (A',B') is at most

$$\begin{aligned} |A'||B'|(\eta+\gamma+\delta)(1-\frac{2\delta}{(\delta+\gamma)}) + |A'||B'|\frac{2\eta\delta}{(\delta+\gamma)} &= \\ |A'||B'|(\eta+\gamma+\delta-\frac{2\delta\gamma}{\delta+\gamma}-\frac{2\delta^2}{\delta+\gamma}) &= \\ |A'||B'|(\eta+\gamma-\delta)\ . \end{aligned}$$

Similarly, we infer that the expected number of edges spanned by (A',B') is at least

$$\begin{aligned} |A'||B'|(\eta-\gamma-\delta)(1-\frac{2\delta}{(\delta+\gamma)}) + |A'||B'|\frac{2\eta\delta}{(\delta+\gamma)} &= \\ |A'||B'|(\eta-\gamma-\delta+\frac{2\delta\gamma}{\delta+\gamma}+\frac{2\delta^2}{\delta+\gamma}) &= \\ |A'||B'|(\eta-\gamma+\delta)\ . \end{aligned}$$

By Lemma 3.4 the probability that the number of edges between A' and B' will deviate from its expectation by more than $\frac{1}{2}\delta|A'||B'|$ is at most $2e^{-2(\delta/2)^2|A'||B'|} = 2e^{-2(\delta/2)^2(\gamma m)^2}$. As the number of pairs (A', B') is at most 2^{2m} we get by the union bound, that if $m \geq m_{3.3}(\delta, \gamma)$ is large enough then with probability at least 3/4 all the pairs (A', B') of size γm satisfy this property. Thus for all pairs (A', B') of size γm we have $d(A', B') = \eta \pm (\gamma - \frac{1}{2}\delta)$. ■

Combining the above three claims we get that with constant probability we make at most $\frac{2.5\delta}{\gamma}m^2$ modifications and thus make sure that $d(A, B) = \eta \pm 1/m^{0.5}$ and furthermore that for any pair of sets (A', B') of size γm we have $d(A', B') = \eta \pm (\gamma - \frac{1}{2}\delta)$. Now we can add or remove at most $m^{1.5}$ edges to make sure that $d(A, B) = \eta$. For any pair of sets (A', B') of size γm this will change $d(A', B')$ by at most $1/m^{0.5}\gamma^2 \leq \frac{1}{2}\delta$ for large enough m. This means that we will have $d(A', B') = \eta \pm \gamma$, implying that (A, B) is γ-regular with density η, completing the proof of the lemma. ■

Proof of Lemma 3.1: By Lemma 3.2 we can make at most $2\delta m^2$ edge modifications and thus turn (A, B) into a $(\gamma + 2\delta)$-regular pair with density η. Thus, every pair of subsets $A'' \subseteq A$ and $B'' \subseteq B$ of size γm has density at most

$$(\eta + \gamma + 2\delta)(\gamma + 2\delta)^2 m^2/\gamma^2 m^2 \leq (\eta + \gamma + 2\delta)(1 + 8\delta/\gamma) \leq \eta + \gamma + 14\delta/\gamma.$$

Similarly, the density of such a pair is at least $\eta - \gamma - 14\delta/\gamma$. We thus conclude that (A, B) has density precisely η, and every pair of subsets (A'', B'') of size γm has density $\eta \pm (\gamma + 14\delta/\gamma)$. Now we can use Lemma 3.3 to make at most $3\frac{14\delta/\gamma}{\gamma}m^2 = 42\frac{\delta}{\gamma^2}m^2$ additional edge modifications and thus turn (A, B) into a γ-regular pair with density precisely η. The total number of modifications is $42\frac{\delta}{\gamma^2}m^2 + 2\delta m^2 \leq 50\frac{\delta}{\gamma^2}m^2$ as needed. ■

We finish this section with the following application of Lemma 3.1 that will be useful later in the paper.

Corollary 3.8 *Let R be a regularity-instance of order k, error-parameter γ, $\binom{k}{2}$ edge densities $\eta_{i,j}$ and a set of non-regular pairs $\overline{R}$. Suppose a graph G has an equipartition $\mathcal{V} = \{V_1, \ldots, V_k\}$ of order k such that*

1. *$d(V_i, V_j) = \eta_{i,j} \pm \frac{\gamma^2\epsilon}{50}$ for all $i < j$.*
2. *Whenever $(i, j) \notin \overline{R}$, the pair (V_i, V_j) is $(\gamma + \frac{\gamma^2\epsilon}{50})$-regular.*

Then G is ϵ-close to satisfying R.

Proof: For any $(i, j) \notin \overline{R}$ we can use Lemma 3.1 and make at most $50\frac{\gamma^2\epsilon/50}{\gamma^2}(n/k)^2 \leq \epsilon n^2/k^2$ edge modifications to turn (V_i, V_j) into a γ-regular pair with density $\eta_{i,j}$. As there are at most $\binom{k}{2}$ pairs this is a total of at most ϵn^2 modifications. We have thus turned G into a graph satisfying R by making at most ϵn^2 edge modifications, as needed. ■

4 Any Testable Property is Regular-Reducible

In this section we prove the first direction of Theorem 2.

Lemma 4.1 *If a graph property is testable then it is regular-reducible.*

Our starting point in the proof of Lemma 4.1 is the following result of [28] (extending a result of [3]) about *canonical testers*:

Lemma 4.2 ([3, 28]) *If a graph property $\mathcal{P}$ can be tested on n-vertex graphs with $q = q(\epsilon, n)$ edge queries, then it can also be tested by a one that makes its queries by uniformly and randomly choosing a set of $2q$ vertices, querying all the pairs and then accepting or rejecting (deterministically) according to the (isomorphism class of the) graph induced by the sample, the value of ϵ and the value of n. In particular, it is a non-adaptive tester making $\binom{2q}{2}$ queries.*

Restating the above, by (at most) squaring the query complexity, we can assume without loss of generality that a property-tester works by sampling a set of vertices of size $q(\epsilon, n)$ and accepting or rejecting according to some graph property of the sample. Such a testing algorithm is said to be canonical. As noted in [28], the graph property that the algorithm may search for in the sample may be different from the property which is being tested. In fact, the property the algorithm checks for in the sample may depend on ϵ and on the size of the input graph[6]. We will use Lemma 4.2 in order to "pick" the graphs of size q that cause a tester for $\mathcal{P}$ to accept. The first technical step that we take towards proving Lemma 4.1 is proving some technical results about induced copies of (small) graphs spanned by graphs satisfying a given regularity-instance. These results enable us to deduce from the fact that a graph satisfies some regularity-instance, what is the (approximate) probability that a given tester accepts the graph. We then use these results along with Lemma 4.2, Corollary 3.8 and some additional arguments in order to prove that any testable property is regular reducible. The details follow.

Definition 4.3 *Let H be a graph on h vertices, let W be a weighted complete graph on h vertices, where the weight of an edge (i, j) is $\eta_{i,j}$. For a permutation $\sigma : [h] \to [h]$ define*

$$IC(H, W, \sigma) = \prod_{(i,j)\in E(H)} \eta_{\sigma(i),\sigma(j)} \prod_{(i,j)\notin E(H)} (1 - \eta_{\sigma(i),\sigma(j)})$$

Suppose $V_1, \ldots, V_k$ are k vertex sets, each of size m, and suppose the bipartite graph spanned by V_i and V_j is a bipartite random graph with edge density $\eta_{i,j}$. Let H be a graph of size k, and let $\sigma : [k] \to [k]$ be some permutation. What is the expected number of k-tuples of vertices $v_1 \in V_1, \ldots, v_k \in V_k$ that span an induced copy of H, such that for every i we have v_i playing the role of $\sigma(i)$? It is easy to see that the answer is $IC(H, W, \sigma)m^k$, where W is the weighted complete graph with weights $\eta_{i,j}$. The following claim shows that this is approximately the case when instead of random bipartite graphs we take regular enough bipartite graphs. The proof is a standard application of the definition of a regular pair and is thus omitted. See Lemma 4.2 in [18] for a version of the proof.

[6]For most "natural" graph properties there will not be any dependence on the size of the input, but as we are dealing here with arbitrary properties we must take this possibility into account.

Claim 4.4 *For any δ and h, there exists a $\gamma = \gamma_{4.4}(\delta, h)$ such that the following holds: Suppose $V_1, \ldots, V_h$ are h sets of vertices of size m each, and that all the pairs (V_i, V_j) are γ-regular. Define W to be the weighted complete graph on h vertices, whose weights are $\eta_{i,j} = d(V_i, V_j)$. Then, for any graph H on h vertices and for any $\sigma : [k] \to [k]$, the number of h-tuples $v_1 \in V_1, \ldots, v_h \in V_h$, which span an induced copy of H with each v_i playing the role of the vertex $\sigma(i)$ is*

$$(IC(H, W, \sigma) \pm \delta)m^h$$

Definition 4.5 *Let H be a graph on h vertices, let W be a weighted complete graph on h vertices, where the weight of edge (i, j) is $\eta_{i,j}$. Let $Aut(H)$ denote the number of automorphisms of H. Define*

$$IC(H, W) = \frac{1}{Aut(H)} \sum_{\sigma} IC(H, W, \sigma).$$

Continuing the discussion before Claim 4.4, suppose now we want to estimate the total number of induced copies of H having one vertex in each of the sets V_i. It is easy to see that in this case the expected number of such copies of H is $IC(H, W)$. Again, we can show that the same is approximately true when we replace random bipartite graphs with regular enough bipartite graphs.

Claim 4.6 *For any δ and k, there exists a $\gamma = \gamma_{4.6}(\delta, k)$ such that the following holds: Suppose that $V_1, \ldots, V_k$ are sets of vertices of size m each, and that all the pairs (V_i, V_j) are γ-regular. Define K to be the weighted complete graph on k vertices, whose weights are $\eta_{i,j} = d(V_i, V_j)$. Then, for any graph H of size k, the number of induced copies of H, which have precisely one vertex in each of the sets $V_1, \ldots, V_k$ is*

$$(IC(H, W) \pm \delta)m^k$$

Proof: Set $\gamma_{4.6}(\delta, k) = \gamma_{4.4}(\delta/k!, k)$. Suppose $V_1, \ldots, V_k$ are as in the statement of the claim and let H be any graph on k vertices. By Claim 4.4 for every permutation $\sigma : [k] \to [k]$, the number of induced copies of H which have precisely one vertex v_i in each set V_i such that v_i plays the role of vertex $\sigma(i)$ is $IC(H, W, \sigma) \pm \delta m^k/k!$. If we sum over all permutations $\sigma : [k] \to [k]$ we get $\sum_\sigma (IC(H, W, \sigma) \pm \delta/k!)m^k$. This summation, however, counts copies of H several times. More precisely, each copy is thus counted $Aut(H)$ times. Dividing by $Aut(H)$ gives that the number of such induced copies is

$$\begin{aligned} \frac{1}{Aut(H)} \left(\sum_{\sigma} (IC(H, W, \sigma) \pm \delta/k!) \right) m^k &= \\ \left(\frac{1}{Aut(H)} \sum_{\sigma} IC(H, W, \sigma) \pm \delta \right) m^k &= \\ (IC(H, W) \pm \delta)m^k &. \end{aligned}$$

■

We would now want to consider the number of induced copies of a graph H, when the number of sets V_i is larger than the size of H.

Definition 4.7 *Let H be a graph on h vertices, let R be a weighted complete graph of size at least h, where the weight of an edge (i,j) is $\eta_{i,j}$, and let $\mathcal{W}$ denote all the subsets of $V(W)$ of size h. Define*

$$IC(H,R) = \sum_{W \in \mathcal{W}} IC(H,W).$$

The following lemma shows that knowing that a graph satisfies some regularity-instance R, enables us to estimate the number of induced copies spanned by any graph satisfying R.

Lemma 4.8 *For any δ and q, there are $k = k_{4.8}(\delta,q)$ and $\gamma = \gamma_{4.8}(\delta,q)$ with the following properties: For any regularity-instance R of order at least k and with error parameter at most γ, and for every graph H of size $h \le q$, the number of induced copies of H in any n-vertex graph satisfying R is*

$$(IC(H,R) \pm \delta)\binom{n}{h}$$

Proof: Put

$$k = k_{4.8}(\delta,q) = \frac{\delta}{10q^2} ,$$

and

$$\gamma = \gamma_{4.8}(\delta,q) = \min\{\frac{\delta}{3q^2}, \gamma_{4.6}(\frac{1}{3}\delta,q)\} .$$

Let R be any regularity instance as in the statement, let G be any graph satisfying R, and let H be any graph of size $h \le q$. Let $V_1,\ldots,V_\ell$ be an equipartition of G satisfying R. For the proof of the lemma it will be simpler to consider an equivalent statement of the lemma, stating that if one samples an h-tuple of vertices from G, then the probability that it spans an induced copy of H is $IC(H,R) \pm \delta$.

First, note that by our choice of k we get from a simple birthday-paradox argument, that the probability that the h-tuple of vertices has more than one vertex in any one of the sets V_i is at most $\frac{1}{3}\delta$. Second, observe that as the equipartition of R is γ-regular and $\gamma \le \delta$, we get that the probability that the h-tuple of vertices contains a pair $v_i \in V_i$ and $v_j \in V_j$ such that (V_i,V_j) is not γ-regular is at most $\binom{h}{2}\gamma \le \binom{q}{2}\gamma \le \frac{1}{3}\delta$. Consider the events (i) the h vertices $v_1,\ldots,v_h$ belong to distinct sets V_i (ii) if the tuple $v_1,\ldots,v_h$ is such that $v_i \in V_i$ then for every $i<j$ the pair (V_i,V_j) is γ-regular. As each of these events holds with probability $1-\frac{1}{3}\delta$ it is enough for us to show that conditioned on them that the probability that $v_1,\ldots,v_h$ span an induced copy of H is $IC(H,R) \pm \frac{1}{3}\delta$. Assuming events (i) and (ii) hold let us compute the probability that the h-tuple of vertices spans an induced copy of H, while conditioning on the h sets from $V_1,\ldots,V_\ell$ which contain the h vertices. For every possible set W of h sets V_i we get from the choice of γ via Claim 4.6 that the probability that they span an induced copy of H is $IC(H,W) \pm \frac{1}{3}\delta$. This means that the conditional probability that the h-tuple of vertices spans an induced copy of H is $IC(H,R) \pm \frac{1}{3}\delta$, as needed. ■

The following corollary strengthens the above lemma by allowing us to consider families of graphs.

Corollary 4.9 *For any δ and q, there are $k = k_{4.9}(\delta,q)$ and $\gamma = \gamma_{4.9}(\delta,q)$ with the following properties: For any regularity-instance R of order at least k and with error parameter at most γ, and for* *every family $\mathcal{A}$ of graphs of size q, the number of induced copies of graphs $H \in \mathcal{A}$ in any n-vertex graph satisfying R is*

$$(\sum_{H \in \mathcal{A}} IC(H,R) \pm \delta)\binom{n}{q}$$

Proof: We just take $k = k_{4.9}(\delta, q) = k_{4.8}(\delta/2^{\binom{q}{2}}, q)$ and $\gamma = \gamma_{4.9}(\delta, q) = \gamma_{4.8}(\delta/2^{\binom{q}{2}}, q)$. By Lemma 4.8, the number of induced copies of a graph H or size q, in any graph satisfying a regularity instance R of order at least k and with error parameter at most γ, is $(IC(H,R) \pm \delta/2^{\binom{q}{2}})\binom{n}{h}$. As $\mathcal{A}$ contains at most $2^{\binom{q}{2}}$ graphs the total error after summing over all graphs $H \in \mathcal{A}$ is bounded by $\delta\binom{n}{h}$. ■

Proof of Lemma 4.1: Suppose $\mathcal{P}$ is testable by a tester $\mathcal{T}$, and assume without loss of generality that $\mathcal{T}$ is canonical. This assumption is possible by Lemma 4.2. Let $q(\epsilon)$ be the upper bound guarantee for the query complexity of $\mathcal{T}$. Fix any n and δ and assume that $\delta < 1/12$ (otherwise, replace δ with 1/13). Let $q = q(\delta, n) \leq q(\delta)$ be the query complexity, which is sufficient for $\mathcal{T}$ to distinguish between n-vertex graphs satisfying $\mathcal{P}$ and those that are δ-far from satisfying it, with success probability at least 2/3. As $\mathcal{T}$ is canonical, if it samples a set of vertices and gets a graph of size q, it either rejects or accepts deterministically. Hence, we can define a set $\mathcal{A}$, of all the graphs Q of size q, such that if the sample of vertices spans a graph isomorphic to Q, then $\mathcal{T}$ accepts the input. We finally put

$$k = k_{4.9}(\delta, q) \ ,$$

$$\gamma = \gamma_{4.9}(\delta, q) \ ,$$

and

$$T = T_{2.3}(k, \gamma) \ .$$

For any $k \leq t \leq T$ consider all the (finitely many) regularity-instances of order t, where for the edge densities $\eta_{i,j}$ we choose a real from the set $\{0, \frac{\delta\gamma^2}{50q^2}, 2\frac{\delta\gamma^2}{50q^2}, 3\frac{\delta\gamma^2}{50q^2}, \ldots, 1\}$. Let $\mathcal{I}$ be the union of all these regularity-instances. Note, that all the above constants, as well as the size of $\mathcal{I}$ and the complexity of the regularity-instances in $\mathcal{I}$, are determined as a function of δ only (and the property $\mathcal{P}$).

We claim that we can take $\mathcal{R}$ in Definition 2.6 to be

$$\mathcal{R} = \{R \in \mathcal{I} : \sum_{H \in \mathcal{A}} IC(R,H) \geq 1/2\} \ ,$$

where $IC(R,H)$ was defined in Definition 4.7. Indeed, first note that the expression $\sum_{H \in \mathcal{A}} IC(R,H)$ is an estimation of the fraction of induced copies of graphs from $\mathcal{A}$ in a graph satisfying R. As we chose k and γ via Corollary 4.9 we infer that the expression $\sum_{H \in \mathcal{A}} IC(R,H)$ is an estimate of the number of induced copies of graphs from $\mathcal{A}$ in a graph satisfying R, up to an additive error of at most $\delta\binom{n}{q}$.

Suppose a graph G satisfies $\mathcal{P}$. This means that $\mathcal{T}$ accepts G with probability at least 2/3. In other words, this means that at least $\frac{2}{3}\binom{n}{q}$ of the subsets of q vertices of G span a graph isomorphic to one of the members of $\mathcal{A}$. By Lemma 2.3 G has some γ-regular partition of size at least k and at most T. By construction of $\mathcal{I}$ we get that the densities of the regular partition of G differ by at most $\frac{\delta\gamma^2}{50q^2}$ from the densities of one[7] of the regularity-instances $R \in \mathcal{I}$. Corollary 3.8 implies that G is δ/q^2-close to satisfying one of the regularity-instances of $\mathcal{I}$. Note that adding and/or removing an edge can decrease the number of induced copies of members of $\mathcal{A}$ in G by at most $\binom{n-2}{q-2}$. Thus adding and/or removing $\delta n^2/q^2$ edges can decrease the number of induced copies of members of $\mathcal{A}$ in G by at most $\delta \frac{n^2}{q^2}\binom{n-2}{q-2} \leq \delta\binom{n}{q}$. Thus, after these at most $\delta n^2/q^2$ edge modifications we get a graph that satisfies one of the regularity-instances $R \in \mathcal{I}$ where at least $(\frac{2}{3} - \delta)\binom{n}{q} > (\frac{1}{2} + \delta)\binom{n}{q}$ of the subsets of q vertices of the new graph span a member of $\mathcal{A}$ (here we use the assumption that $\delta < 1/12$). As explained in the previous paragraph, by our choice of k and γ via Corollary 4.9, this means that $\sum_{H \in \mathcal{A}} IC(R, H) \geq 1/2$. By the definition of $\mathcal{R}$ this means that $R \in \mathcal{R}$, so G is indeed δ-close to satisfying one of the regularity-instances of $\mathcal{R}$.

Suppose now that a graph G is ϵ-far from satisfying $\mathcal{P}$. If $\delta \geq \epsilon$ then there is nothing to prove, so assume that $\delta < \epsilon$. If G is $(\epsilon - \delta)$-close to satisfying a regularity-instance $R \in \mathcal{R}$, then by the definition of $\mathcal{R}$ and our choice of k and γ via Corollary 4.9, it is $(\epsilon - \delta)$-close to a graph G', such that at least $(\frac{1}{2} - \delta)\binom{n}{q} > (\frac{1}{3} + \delta)\binom{n}{q}$ of the subsets of q vertices of G' span an induced copy of a graph from $\mathcal{A}$. In other words, this means that $\mathcal{T}$ accepts G' with probability at least $\frac{1}{3} + \delta$. This means that G' cannot be δ-far from satisfying $\mathcal{P}$ as we assume that q is enough for $\mathcal{T}$ to reject graphs that are δ-far from satisfying $\mathcal{P}$ with probability at least $2/3$. However, as G is ϵ-far from satisfying $\mathcal{P}$ any graph that is $(\epsilon - \delta)$-close to G must be δ-far from satisfying $\mathcal{P}$, a contradiction. ■

5 Sampling Regular Partitions

The main result of this section asserts (roughly) that for every fixed γ, if we sample a constant number of vertices from a graph G, then with high probability the graph induced by the sample and the graph G will have the same set of γ-regular partitions. To formally state this result we introduce the following definition:

Definition 5.1 (δ-similar regular-partition) *An equipartition $\mathcal{U} = \{U_i \mid 1 \leq i \leq k\}$ is δ-*similar *to a γ-regular equipartition $\mathcal{V} = \{V_i \mid 1 \leq i \leq k\}$, of the same order k (where $0 < \gamma \leq 1$), if:*

1. *$d(U_i, U_j) = d(V_i, V_j) \pm \delta$ for all $i < j$.*
2. *Whenever (V_i, V_j) is γ-regular, (U_i, U_j) is $(\gamma + \delta)$-regular.*

Observe that in the above definition, the two equipartitions $\mathcal{V}$ and $\mathcal{U}$ may be equipartitions of different graphs. In what follows, if $G = (V, E)$ is a graph and $Q \subseteq V(G)$, then $G[Q]$ denotes the subgraph induced by G on Q. Our main result in this section is the following lemma that roughly asserts that a large enough sample Q from a graph G will be such that G and the graph spanned by Q are close to satisfying the same regularity instances.

Lemma 5.2 *For every k and δ there exists a $q = q_{5.2}(k, \delta)$ such that the following holds for every $\gamma \geq \delta$ and $k' \leq k$: A sample Q of q vertices from a graph G, satisfies the following with probability at least $2/3$:*

[7]In other words, this follows from the fact that the elements of $\mathcal{I}$ form a $\frac{\delta\gamma^2}{50q^2}$-net in the ℓ_∞-norm.

1. *If G has a γ-regular equipartition $\mathcal{V}$ of order k', then $G[Q]$ has an equipartition $\mathcal{U}$ of the same order which is δ-similar to $\mathcal{V}$.*

2. *If $G[Q]$ has a γ-regular equipartition $\mathcal{U}$ of order at most k', then G has an equipartition $\mathcal{V}$ of the same order which is δ-similar to $\mathcal{U}$.*

We stress that in the above lemma we do not refer to some fixed regular partition of G or $G[Q]$ but to any possible regular partition of either one of them.

The proof of Lemma 5.2 has two main stages. For the first one we need a weaker result, which says that a sample of vertices has a regular partition, but with a *weaker* regularity measure.

Lemma 5.3 ([16]) *For every k and δ there exists $q = q_{5.3}(k, \delta)$ such that the following holds for every $\gamma \geq \delta$ and $k' \leq k$: if a graph G has a γ-regular equipartition $\mathcal{V} = \{V_1, \ldots, V_{k'}\}$, then with probability at least $2/3$, a sample of q vertices will have an equipartition $\mathcal{U} = \{U_1, \ldots, U_{k'}\}$ satisfying:*

1. *$d(U_i, U_j) = d(V_i, V_j) \pm \delta$ for all $i < j$.*

2. *Whenever (V_i, V_j) is γ-regular, (U_i, U_j) is $50\gamma^{1/5}$-regular.*

For our purposes however, we cannot allow a weaker regularity as in the above lemma. Our main tool in the proof of Lemma 5.2 is Lemma 5.5 below, which establishes that if two graphs share *one* γ-regular equipartition, then they share *all* the γ'-regular-partitions where γ' is larger than γ. This will allow us to strengthen Lemma 5.3 and thus obtain Lemma 5.2. For the statement of this lemma we need the following definition:

Definition 5.4 ((δ, γ)-similar regular-partitions) *Two equipartitions $\mathcal{V} = \{V_i \mid 1 \leq i \leq k\}$ and $\mathcal{U} = \{U_i \mid 1 \leq i \leq k\}$ of the same order k, are said to be* (δ, γ)-similar *if:*

1. *$d(U_i, U_j) = d(V_i, V_j) \pm \delta$ for all $i < j$.*

2. *For all but at most $\gamma\binom{k}{2}$ of the pairs $i < j$, both (V_i, V_j) and (U_i, U_j) are γ-regular.*

The following is the main technical lemma of this section. This lemma allows us to infer from the fact that two graphs are close to satisfying a *single* common regular partition, that they are close to satisfying *all* the regularity-instances (but with a larger error parameter).

Lemma 5.5 *For every k and δ there exists $\zeta = \zeta_{5.5}(k, \delta)$ such that the following holds for every $k' \leq k$: suppose that two graphs $G = (V, E)$ and $\overline{G} = (\overline{V}, \overline{E})$ have (ζ, ζ)-similar regular-equipartitions $\mathcal{V} = \{V_1, \ldots, V_\ell\}$ and $\overline{\mathcal{V}} = \{\overline{V}_1, \ldots, \overline{V}_\ell\}$ with $\ell \geq 1/\zeta$. Then, if $\overline{G}$ has a γ-regular equipartition $\overline{\mathcal{A}} = \{\overline{A}_1, \ldots, \overline{A}_{k'}\}$ then G has an equipartition $\mathcal{A} = \{A_1, \ldots, A_{k'}\}$, which is δ-similar to $\overline{\mathcal{A}}$.*

We turn to prove Lemma 5.5, and then use it to prove Lemma 5.2. But we first state the following simple fact.

Claim 5.6 *Let $a_1, \ldots, a_\ell$ and $b_1, \ldots, b_\ell$ satisfy $\sum_{1\leq i\leq \ell} a_i = \sum_{1\leq i\leq \ell} b_i = 1$ and $0 \leq a_i, b_i \leq k/\ell$, where $k \leq \ell$. Then $\sum_{1\leq i\leq \ell} a_i b_i \leq k/\ell$.*

Proof: Observe that $\sum_{1\leq i\leq \ell} a_i b_i \leq \max_{1\leq i\leq \ell}\{a_i\} \sum_{1\leq i\leq \ell} b_i \leq k/\ell$. ■

Proof of Lemma 5.5: We consider the case $k' = k$ as the cases $k' < k$ are identical and in fact follow from the case $k' = k$ by monotonicity. Let $\overline{A}_1, \ldots, \overline{A}_k$ be any equipartition of $\overline{G}$. Recall that ℓ denotes the order of the equipartition $\overline{\mathcal{V}}$, which is also the order of $\mathcal{V}$. For every $1 \leq p \leq \ell$ and $1 \leq q \leq k$ set $\overline{AV}_{p,q} = \overline{V}_p \cap \overline{A}_q$ and $\alpha_{p,q} = |\overline{AV}_{p,q}|/|\overline{V}_p|$. For every $1 \leq p \leq \ell$ we partition V_p arbitrarily into k disjoint subsets $AV_{p,1}, \ldots, AV_{p,k}$ in a way that for every $1 \leq q \leq k$ we have $|AV_{p,q}| = \alpha_{p,q}|V_p|$. Finally for every $1 \leq q \leq k$ define $A_q = \bigcup_{p=1}^{\ell} AV_{p,q}$. Instead of stating $\zeta_{5.5}(k, \delta)$ explicitly, we state along the way different upper bounds on $\zeta_{5.5}(k, \delta)$ that will depend only on k and δ. One can then take the minimum of all these values as $\zeta_{5.5}(k, \delta)$.

Claim 5.7 *If $(\overline{A}_q, \overline{A}_{q'})$ is γ-regular then $(A_q, A_{q'})$ is $(\gamma + \delta)$-regular.*

Proof: To simplify the notation we assume that $(\overline{A}_1, \overline{A}_2)$ is γ-regular and prove that (A_1, A_2) is $(\gamma + 2\delta)$-regular. Set $\eta = d(\overline{A}_1, \overline{A}_2)$. Claim 5.8 below asserts $d(A_1, A_2) = \eta \pm \delta$, so we need to show that $d(A_1', A_2') = \eta \pm (\gamma + \delta)$ for every $A_1' \subseteq A_1$ and $A_2' \subseteq A_2$ of sizes $(\gamma + \delta)|A_1|$ and $(\gamma + \delta)|A_2|$ respectively. For simplicity we show that $d(A_1', A_2') \leq \eta + \gamma + \delta$, as showing that $d(A_1', A_2') \geq \eta - \gamma - \delta$ is similar. Recall that every set A_q is the union of ℓ sets $AV_{1,q}, \ldots, AV_{\ell,q}$. For every $1 \leq i \leq \ell$ and $1 \leq j \leq \ell$ put $AV_{i,1}' = AV_{i,1} \cap A_1'$ and $AV_{j,2}' = AV_{j,2} \cap A_2'$. We can rephrase our goal in terms of the number of edges as follows: we wish to show that

$$\sum_{1 \leq i,j \leq \ell} e(AV_{i,1}', AV_{j,2}') \leq (\eta + \gamma + \delta)|A_1'||A_2'| = (\eta + \gamma + \delta)(\gamma + \delta)^2|A_1||A_2|. \tag{1}$$

Let n denote the number of vertices of G. Observe that every set V_i is of size n/ℓ and every set A_i is of size n/k. Recall that by assumption all but $\zeta\binom{\ell}{2}$ of the pairs $1 \leq i < j \leq \ell$ are such that both (V_i, V_j) and $(\overline{V}_i, \overline{V}_j)$ are ζ-regular. Let us then denote by $\mathcal{M}$ the pairs i, j for which both (V_i, V_j) and $(\overline{V}_i, \overline{V}_j)$ are ζ-regular. To prove (1) we turn to bound the contribution to the LHS (Left Hand Side) of (1) of three types of pairs of (i, j):

- **Pairs (i, j) for which $i = j$:** Observe that the maximum possible number of edges connecting a pair $(AV_{i,1}', AV_{i,2}')$ is at most

  $$|AV_{i,1}'||AV_{i,2}'| \leq |AV_{i,1}||AV_{i,2}| = \alpha_{i,1}\alpha_{i,2}|V_i||V_i| = \alpha_{i,1}\alpha_{i,2}\frac{k}{\ell}|A_1|\frac{k}{\ell}|A_2| \,,$$

  where in the last equality we use the fact that the sets V_i have size n/ℓ and the sets A_i have size n/k. Therefore, the maximum contribution of such pairs is given by $|A_1||A_2| \sum_i \frac{k}{\ell}\alpha_{i,1}\frac{k}{\ell}\alpha_{i,2}$. By Claim 5.6 we get that $|A_1||A_2| \sum_i \frac{k}{\ell}\alpha_{i,1}\frac{k}{\ell}\alpha_{i,2} \leq \frac{k}{\ell}|A_1||A_2|$ and if we choose a ζ satisfying $\ell \geq 1/\zeta \geq 6k/\delta^3 \geq 6k/\delta(\gamma + \delta)^2$ we can infer that the contribution of the pairs (i, i) to the LHS of (1) is at most $\frac{1}{6}\delta(\gamma + \delta)^2|A_1||A_2|$. Note that the fact that $\ell \geq 1/\zeta$ is guaranteed by the assertion of the lemma.

- **Pairs (i, j) for which either $|AV_{i,1}'| < \zeta|V_i|$ or $|AV_{j,2}'| < \zeta|V_j|$:** Consider the $1 \leq i \leq \ell$ in (1) for which $|AV_{i,1}'| < \zeta|V_i| = \zeta n/\ell$. The total number of vertices of G that belong to such sets is clearly at most ζn, therefore the total number of such vertices in A_1 is at most $k\zeta|A_1|$ (because $|A_1| = n/k$). Similarly, the total number of vertices of A_2 which belong to sets $|AV_{j,2}'|$ for which $|AV_{j,2}'| < \zeta|V_j|$ is at most $k\zeta|A_2|$. Therefore the contribution of pairs (i, j) to the LHS of (1) for which either $|AV'_{i,1}| < \zeta|V_i|$ or $|AV'_{j,2}| < \zeta|V_j|$ is at most $2k\zeta|A_1||A_2|$. If we choose ζ so that it satisfies $\zeta \le \frac{\delta^3}{12k} \le \frac{\delta(\gamma+\delta)^2}{12k}$, such pairs (i,j) can contribute to the LHS of (1) a total of at most $\frac{1}{6}\delta(\gamma+\delta)^2|A_1||A_2|$.

- **Pairs (i,j) that do not belong to $\mathcal{M}$:** As all but at most $\zeta\binom{\ell}{2}$ of the pairs (V_i,V_j) belong to $\mathcal{M}$ (defined above), we may deduce that at most ζn^2 edges of G connect pairs of clusters (V_i,V_j) that are not ζ-regular. As $|A_1| = |A_2| = n/k$, this means that the number of edges connecting A_1 and A_2 that belong to pairs (V_i,V_j) that are not ζ-regular is at most $k^2\zeta(n/k)^2 = k^2\zeta|A_1||A_2|$. If we choose ζ so that $\zeta \le \frac{1}{6}\delta^3/k^2 \le \frac{1}{6}\delta(\gamma+\delta)^2/k^2$, such pairs can contribute at most $\frac{1}{6}\delta(\gamma+\delta)^2|A_1||A_2|$ to the sum in (1).

We have thus accounted for all pairs (i,j) in (1) for which either $i=j$, (V_i,V_j) is not ζ-regular, $|AV'_{i,1}| < \zeta|V_i|$ or $|AV'_{j,2}| < \zeta|V_j|$. Specifically, we have shown that they can contribute at most $\frac{1}{2}\delta(\gamma+\delta)^2|A_1||A_2| = \frac{1}{2}\delta|A'_1||A'_2|$ to the LHS of (1). Therefore, we can now reduce proving (1) to showing that

$$\sum_{(i,j)\in B} e(AV'_{i,1},AV'_{j,2}) = \sum_{(i,j)\in B} d(AV'_{i,1},AV'_{j,2})|AV'_{i,1}||AV'_{j,2}| \le (\eta+\gamma+\frac{1}{2}\delta)|A'_1||A'_2| \;, \qquad (2)$$

where B is the set of pairs (i,j) that satisfy $i \ne j$, $|AV'_{i,1}| \ge \zeta|V_i|$, $|AV'_{j,2}| \ge \zeta|V_j|$ as well as that (V_i,V_j) and $(\overline{V}_i,\overline{V}_j)$ are ζ-regular. Therefore, all $(i,j)\in B$ are such that

$$d(AV'_{i,1},AV'_{j,2}) = d(V_i,V_j) \pm \zeta \;. \qquad (3)$$

and

$$d(\overline{AV'}_{i,1},\overline{AV'}_{j,2}) = d(\overline{V}_i,\overline{V}_j) \pm \zeta \;. \qquad (4)$$

If we choose ζ so that $\zeta \le \frac{1}{6}\delta$ we can use (3) to reduce (2) to showing

$$\sum_{(i,j)\in B} d(V_i,V_j)|AV'_{i,1}||AV'_{j,2}| \le (\eta+\gamma+\frac{1}{3}\delta)|A'_1||A'_2| \;. \qquad (5)$$

As we assume that $\mathcal{V}$ and $\overline{\mathcal{V}}$ are (ζ,ζ)-similar we have $d(V_i,V_j) = d(\overline{V}_i,\overline{V}_j) \pm \zeta$ for every $i<j$. If we choose ζ so that $\zeta \le \frac{1}{6}\delta$, we can reduce (5) to showing that

$$\sum_{(i,j)\in B} d(\overline{V_i},\overline{V_j})|AV'_{i,1}||AV'_{j,2}| \le (\eta+\gamma+\frac{1}{6}\delta)|A'_1||A'_2| \;. \qquad (6)$$

By (4) we can further reduce (6) to showing that

$$\sum_{(i,j)\in B} d(\overline{AV'_{i,1}},\overline{AV'_{j,2}})|AV'_{i,1}||AV'_{j,2}| \le (\eta+\gamma)|A'_1||A'_2| \;. \qquad (7)$$

It will be more convenient to prove (7) by deriving the following stronger assertion [8]:

$$\sum_{1\le i,j\le \ell} d(\overline{AV'_{i,1}},\overline{AV'_{j,2}})|AV'_{i,1}||AV'_{j,2}| \le (\eta+\gamma)|A'_1||A'_2| \;. \qquad (8)$$

[8] The assertion is stronger since we are summing over all pairs (i,j) and not only over the pairs in the set B.

Put $\beta_{i,1} = |AV'_{i,1}|/|A'_1|$ and $\beta_{j,2} = |AV'_{j,2}|/|A'_2|$. For every $1 \leq i \leq \ell$ let $\overline{AV}'_{i,1}$ be any subset of $\overline{AV}_{i,1}$ of size $\beta_{i,1}|\overline{AV}_{i,1}|$. Similarly, for every $1 \leq j \leq \ell$ let $\overline{AV}'_{j,2}$ be any subset of $\overline{AV}_{j,2}$ of size $\beta_{j,2}|\overline{AV}_{j,2}|$. Put $\overline{A}'_1 = \bigcup_{1\leq i\leq \ell} \overline{AV}_{i,1}$ and $\overline{A}'_2 = \bigcup_{1\leq j\leq \ell} \overline{AV}_{j,2}$ and note that just as $|A'_1| \geq \gamma|A_1|$ and $|A'_2| \geq \gamma|A_2|$ we also have $|\overline{A}'_1| \geq \gamma|\overline{A}_1|$ and $|\overline{A}'_2| \geq \gamma|\overline{A}_2|$. Dividing by $|A'_1||A'_2|$ we can restate (8) as

$$\sum_{1\leq i,j\leq \ell} d(\overline{AV'_{i,1}}, \overline{AV'_{j,2}})\beta_{i,1}\beta_{j,2} \leq \eta + \gamma \ .$$

Finally, note that the above holds because

$$\sum_{1\leq i,j\leq \ell} d(\overline{AV'_{i,1}}, \overline{AV'_{j,2}})\beta_{i,1}\beta_{j,2} = d(\overline{A}'_1, \overline{A}'_2) \leq \eta + \gamma \ , \tag{9}$$

where the inequality follows from the fact that $(\overline{A}_1, \overline{A}_2)$ is by assumption γ-regular, $d(\overline{A}_1, \overline{A}_2) = \eta$, $|\overline{A}'_1| \geq \gamma|\overline{A}_1|$ and $|\overline{A}'_2| \geq \gamma|\overline{A}_2|$. This completes the proof of the claim. ■

Claim 5.8 *For all $q < q'$ we have $d(A_q, A_{q'}) = d(\overline{A}_q, \overline{A}_{q'}) \pm \delta$*

Proof: The proof is essentially identical to the above proof. Instead of working with two subsets of A_1 and A_2 we work with the sets themselves. There is only one place in the proof that needs to be changed; note that in equation (9) we use the density of a pair of subsets of $\overline{A_1}$ and $\overline{A_2}$ in order to bound the density of the subsets of A_1 and A_2. To this end we have used the fact that the pair $(\overline{A_1}, \overline{A_2})$ is γ-regular. The only change we now have to make is that as we work with the sets A_1, A_2 and not subsets of them then we also work with the sets $\overline{A_1}, \overline{A_2}$ and not subsets of them. Since the density of $(\overline{A_1}, \overline{A_2})$ is precisely η we do not lose an additional γ in the estimation of $d(A_1, A_2)$ as we did in Claim 5.7. ■

The proof of Lemma 5.5 now follows from the above two claims. ■

Proof of Lemma 5.2: Set $\zeta = (\zeta_{5.5}(k,\delta)/50)^5$ and $\zeta' = \zeta_{5.5}(k,\delta) = 50\zeta^{1/5}$, and note that $\zeta, \zeta' \leq \zeta_{5.5}(k,\delta)$. Let $\mathcal{V} = \{V_1, \ldots, V_\ell\}$ be a ζ-regular partition of G of order $\ell \geq 1/\zeta$. Such an equipartition of order at most $T_{2.3}(1/\zeta, \zeta)$ exists by Lemma 2.3. By Lemma 5.3 we get that if we sample a set Q of at least $q_{5.3}(\ell, \zeta)$ vertices from G then with probability at least $2/3$ the graph induced on Q, which we denote by $G[Q]$, will have an equipartition $\mathcal{U} = \{U_1, \ldots, U_\ell\}$, such that $d(V_i, V_j) = d(U_i, U_j) \pm \zeta'$ and such that if (V_i, V_j) is ζ-regular then (U_i, U_j) is ζ'-regular. This means that with probability at least $2/3$, the graph $G[Q]$ is such that G and $G[Q]$ have equipartitions, which are $(\zeta_{5.5}(k,\delta), \zeta_{5.5}(k,\delta))$-similar. Indeed, as these equipartitions we can take $\mathcal{V}$ and $\mathcal{U}$, because $\zeta' = \zeta_{5.5}(k,\delta)$ then $d(V_i, V_j) = d(U_i, U_j) \pm \zeta_{5.5}(k,\delta)$. Also, as $\zeta \leq \zeta' = \zeta_{5.5}(k,\delta)$, we have for all but at most $\zeta_{5.5}(k,\delta)\binom{k}{2}$ of the pairs $i < j$, that both (V_i, V_j) and (U_i, U_j) are $\zeta_{5.5}(k,\delta)$-regular.

Thus, Lemma 5.5 implies for any γ-regular partition in G of order at most k, that $G[Q]$ has an equipartition that is δ-similar to it. Similarly, Lemma 5.5 implies for any γ-regular partition in $G[Q]$ of size at most k, that G has an equipartition that is δ-similar to it. We can thus set $q_{5.2}(k,\delta) = q_{5.3}(\ell,\zeta)$ in the statement of the lemma because ℓ and ζ depend only on k and δ. ■

6 Testing Regular Partitions and Proof of the Main Result

In this section we apply the results of Sections 3 and 5 to prove Theorem 2. We start by proving the main technical result of this paper by showing that the property of satisfying a regularity-instance is testable with a constant number of queries. We then use this result to prove the main result of this paper.

Proof of Theorem 1: Suppose the regularity-instance R has error parameter γ, $\binom{k}{2}$ edge densities $\eta_{i,j}$ and a set of non-regular pairs $\overline{R}$. Given $G=(V,E)$ and ϵ, the algorithm for testing the property of satisfying R, samples a set of vertices Q, of size q, where q will be chosen later, and accepts G if and only if the graph spanned by Q is $\frac{\gamma^4\epsilon}{200k^2}$-close to satisfying R. In what follows we denote by $G[Q]$ the graph spanned by Q.

Claim 6.1 *If G satisfies R, and $q \geq q_1(\epsilon,k,\gamma)$, then $G[Q]$ is $\frac{\gamma^4\epsilon}{200k^2}$-close to satisfying R with probability at least* $2/3$.

Proof: If $G=(V,E)$ satisfies R, then V has an equipartition into $V_1,\ldots,V_k$ such that for all $(i,j)\notin\overline{R}$ the pair (V_i,V_j) is γ-regular. If we take $q_1(\epsilon,k,\gamma)=q_{5.2}(k,\frac{\gamma^6\epsilon}{10000k^2})$, then by Lemma 5.2, with probability at least 2/3 the graph $G[Q]$ will have an equipartition into k sets $A_1,\ldots,A_k$, such that $d(A_i,A_j)=\eta_{i,j}\pm\frac{\gamma^6\epsilon}{10000k^2}$ for all $i<j$, and if (V_i,V_j) is γ-regular then (A_i,A_j) is $(\gamma+\frac{\gamma^6\epsilon}{10000k^2})$-regular. By Corollary 3.8, this means that $G[Q]$ is $\frac{\gamma^4\epsilon}{200k^2}$-close to satisfying R. ■

Claim 6.2 *If G is ϵ-far from satisfying R, and $q\geq q_2(\epsilon,k,\gamma)$, then $G[Q]$ is $\frac{\gamma^4\epsilon}{200k^2}$-far from satisfying R with probability at least* $2/3$.

Proof: We take $q_2(\epsilon,k,\delta)=q_{5.2}(k,\frac{\gamma^4\epsilon}{200k^2})$. By Lemma 5.2 we get that with probability at least 2/3 the graph $G[Q]$ is such that if it has a γ'-regular equipartition of order k, then G has an equipartition which is $\frac{\gamma^4\epsilon}{200k^2}$-similar to it. We claim that if this event occurs then $G[Q]$ is $\frac{\gamma^4\epsilon}{200k^2}$-far from satisfying R, which is what we want to show. Suppose $G[Q]$ satisfies the above property and assume on the contrary that it is $\frac{\gamma^4\epsilon}{200k^2}$-close to satisfying R. Consider the $\frac{\gamma^4\epsilon}{200k^2}q^2$ edge modifications that make $G[Q]$ satisfy R and consider an equipartition $\mathcal{U}=\{U_1,\ldots,U_k\}$ of $G[Q]$, which satisfies R after performing these modifications. As we made at most $\frac{\gamma^4\epsilon}{200k^2}q^2$ edge modifications, we initially had $d(U_i,U_j)=\eta_{i,j}\pm\frac{\gamma^4\epsilon}{200}$. Consider now any $(i,j)\notin\overline{R}$. After these modifications (U_i,U_j) must be γ-regular with density $\eta_{i,j}$. Therefore, after these modifications every pair $U'_i\subseteq U_i, U'_j\subseteq U_j$ satisfying $|U'_i|\geq\gamma|U_i|$ and $|U'_j|\geq\gamma|U_j|$ satisfies $d(U'_i,U'_j)=\eta_{i,j}\pm\gamma$. Hence, before the modifications every such pair satisfied $d(U'_i,U'_j)=\eta_{i,j}\pm(\gamma+\frac{\gamma^2\epsilon}{200})$. Note that this means that every such pair was originally $(\gamma+\frac{\gamma^2\epsilon}{100})$-regular. By our assumption on $G[Q]$ this means that G has an equipartition in $V_1,\ldots,V_k$ such that $d(V_i,V_j)=\eta_{i,j}\pm\frac{\gamma^2\epsilon}{50}$ holds for all $i<j$, and for all $(i,j)\notin\overline{R}$ the pair (V_i,V_j) is $(\gamma+\frac{\gamma^2\epsilon}{50})$-regular. By Corollary 3.8, this means that G is ϵ-close to satisfying R, contradicting our assumption. ■

Combining the above two claims we infer that if $q=\max\{q_1(\epsilon,k,\gamma),q_2(\epsilon,k,\gamma)\}$ then with probability at least 2/3 the algorithm distinguishes between the required two cases. Furthermore, the number of queries performed by the algorithm depends only on ϵ, k and γ, and is thus bounded from above by a function of ϵ. This completes the proof of the theorem. ■

Having established the testability of any given regularity-instance we can prove Theorem 2. The last tool we need for the proof is the main result of [20] about estimating graph properties.

Theorem 3 ([20]) *Suppose that a graph property $\mathcal{P}$ is testable. Then for every $0 \le \epsilon_1 < \epsilon_2 \le 1$ there is a randomized algorithm for distinguishing (with success probability at least $2/3$) between graphs that are ϵ_1-close to satisfying $\mathcal{P}$ and graphs that are ϵ_2-far from satisfying it. Furthermore, the query complexity of the algorithm can be bounded from above by a function of ϵ_1 and ϵ_2, which is independent of the size of the input.*

Proof of Theorem 2: The first direction is given in Lemma 4.1. For the other direction, suppose that a graph property $\mathcal{P}$ is regular-reducible as per Definition 2.6. Let us fix n and ϵ. Put $r = r(\frac{1}{4}\epsilon)$ and let $\mathcal{R}$ be the corresponding set of regularity instances for $\delta = \frac{1}{4}\epsilon$ as in Definition 2.6. Recall that Definition 2.6 guarantees that the number and the complexity of the regularity-instances of $\mathcal{R}$ are bounded by a function of $\delta = \frac{1}{4}\epsilon$. By Theorem 1 for any regularity-instance $R \in \mathcal{R}$, the property of satisfying R is testable. Thus, by Theorem 3 for any such R, we can distinguish graphs that are $\frac{1}{4}\epsilon$-close to satisfying R from those that are $\frac{3}{4}\epsilon$-far from satisfying it, while making a number of queries, which is bounded by a function of ϵ. In particular, by repeating the algorithm of Theorem 3 an appropriate number of times (that depends only on $r = r(\frac{1}{4}\epsilon)$), and taking the majority vote, we get an algorithm for distinguishing between the above two cases, whose query complexity is a function of ϵ and r, which succeeds with probability at least $1 - \frac{1}{3r}$. As r itself is bounded by a function of ϵ, the number of queries of this algorithm can be bounded by a function of ϵ only.

We are now ready to describe our tester for $\mathcal{P}$: given a graph G of size n and $\epsilon > 0$, the algorithm uses for every $R \in \mathcal{R}$ the version of Theorem 3 described in the previous paragraph, which succeeds with probability at least $1 - \frac{1}{3r}$ in distinguishing between the case that G is $\frac{1}{4}\epsilon$-close to satisfying R from the case that it is $\frac{3}{4}\epsilon$-far from satisfying it. If it finds that G is $\frac{1}{4}\epsilon$-close to satisfying some $R \in \mathcal{R}$, then the algorithm accepts, and otherwise it rejects. Observe that as there are at most r regularity-instances in $\mathcal{R}$, we get by the union-bound that with probability at least $2/3$ the subroutine for estimating how far is G from satisfying some $R \in \mathcal{R}$ never errs. We now prove that the above algorithm is indeed a tester for $\mathcal{P}$. Suppose first that G satisfies $\mathcal{P}$. As we set $\delta = \frac{1}{4}\epsilon$ and $\mathcal{P}$ is regular-reducible to $\mathcal{R}$, the graph G must be $\frac{1}{4}\epsilon$-close to satisfying some regularity-instance $R' \in \mathcal{R}$. Suppose now that G is ϵ-far from satisfying $\mathcal{P}$. Again, as we assume that $\mathcal{P}$ is regular-reducible to $\mathcal{R}$, we conclude that G must be $\frac{3}{4}\epsilon$-far from satisfying all of the regularity-instances $R \in \mathcal{R}$. As with probability at least $2/3$ the algorithm correctly decides for any $R \in \mathcal{R}$ if G is $\frac{1}{4}\epsilon$-close to satisfying R or $\frac{3}{4}\epsilon$-far from satisfying it, we get that if G satisfies $\mathcal{P}$ then with probability at least $2/3$ the algorithm will find that G is $\frac{1}{4}\epsilon$-close to satisfying some $R \in \mathcal{R}$, while if G is ϵ-far from satisfying $\mathcal{P}$ then with probability at least $2/3$ the algorithm will find that G is $\frac{3}{4}\epsilon$-far from all $R \in \mathcal{R}$. By the definition of the algorithm, we get that with probability at least $2/3$ it distinguishes between graphs satisfying $\mathcal{P}$ from those that are ϵ-far from satisfying it. This means that the algorithm is indeed a tester for $\mathcal{P}$. ■

7 Applications of the Main Result

In this section we show that Theorem 2 can be used in order to derive some positive and negative results on testing graph properties. We would like to stress that all these proofs implicitly apply the main intuition behind our characterization, which was explained after the statement of Theorem 2, that a graph property is testable if and only if knowing a regular enough partition of the graph is sufficient for inferring if a graph is far from satisfying the property. Our first application of Theorem 2 concerns testing for H-freeness. A graph is said to be H-free if it contains no (not necessarily induced) copy of H. It was implicitly proved in [1] that for any H, the property of being H-free is testable. The main idea of the proof in [1] is that if G is ϵ-far from being H-free then a large enough sample of vertices will contain a copy of H with high probability. Here we derive this result from Theorem 2 by giving an alternative proof, which checks if the input satisfies some regularity-instance. For simplicity, we only consider testing triangle-freeness. We briefly mention that an argument similar to the one we use to test triangle-freeness can be used to test any monotone graph property. However, to carry out the proof one needs one additional non-trivial argument, which was proved in [6], so we refrain from including the proof.

Corollary 7.1 *Triangle-freeness is testable.*

Proof: By Theorem 2 it is enough to show that triangle-freeness is regular-reducible. Fix any $\delta > 0$ and set $\gamma' = \gamma_{4.6}(\delta, 3)$. Define $\gamma = \min\{\gamma', \delta\}$. We define $\mathcal{R}$ to be all the regularity-instances R satisfying the following:

(i) They have regularity parameter γ.

(ii) They have order at least $1/\gamma$ and at most $T_{2.3}(1/\gamma, \gamma)$.

(iii) Their densities $\eta_{i,j}$ are taken from $\{0, \gamma, 2\gamma, \ldots, 1\}$.

(iv) They do not contain three clusters V_i, V_j, V_k such that $\eta_{i,j}, \eta_{j,k}, \eta_{i,k}$ are all positive.

To show that this is a valid reduction (in the sense of Definition 2.6), assume first that G is ϵ-far from being triangle-free. Assume G is $(\epsilon - \delta)$-close to satisfying a regularity instance $R \in \mathcal{R}$. We can thus make $(\epsilon - \delta)n^2$ edge modifications and get a graph satisfying R. We also remove all edges inside the sets V_i. As by item (ii) each set has size at most $\gamma n \leq \delta n$ we remove less than δn^2 edges. The total number of edges removed is thus less than ϵn^2. By property (iv) of the regularity instances of $\mathcal{R}$ this means that the new graph is triangle-free, which is impossible because we made less than ϵn^2 edge modifications and G was assumed to be ϵ-far from being triangle-free.

Assume now that G is triangle-free. By Lemma 2.3 G has a γ-regular equipartition $V_1, \ldots, V_k$ of order $1/\gamma \leq k \leq T_{2.3}(1/\gamma, \gamma)$. Note that by our choice of γ' via Claim 4.6, and because $\gamma \leq \gamma'$, there are no i, j, k such that (V_i, V_j), (V_j, V_k), (V_i, V_k) are γ-regular and $d(V_i, V_j), d(V_j, V_k), d(V_i, V_k) \geq \delta$ because such sets span at least one triangle (in fact, many). As by item (iii) the densities of the instances in $\mathcal{R}$ are taken from $\{0, \gamma, 2\gamma, \ldots, 1\}$ we can make at most $\gamma n^2 \leq \delta n^2$ changes and "round down" the densities between the sets into a multiple of γ, while maintaining the regularity of the regular-pairs (we can use Lemma 3.1 here). This means that the new graph satisfies a regularity-instance $R \in \mathcal{R}$, which means that G was δ-close to satisfying R. ∎

Our second application of Theorem 2 is concerned with testing k-colorability. This property was first implicitly proved to be testable in [35]. Much better upper bounds were obtained in [27], and further improved by [5]. As in the case of H-freeness, the main ideas of the proofs in [35, 27, 5] is that if G is ϵ-far from being k-colorable then a large enough sample of vertices will not be k-colorable with high probability. Here we derive this result by applying Theorem 2. Though we derive here only the testability of k-colorability, simple variants of the argument can be used to show that all the partition-problems studied in [27] are testable[9].

Corollary 7.2 *k-colorability is testable.*

Proof: By Theorem 2 it is enough to show that k-colorability is regular-reducible. Fix any $\delta > 0$ and define $\mathcal{R}$ to be all the regularity-instances R satisfying the following:

(i) They have regularity measure δ

(ii) They have order at least $1/\delta$ and at most $T_{2.3}(2k/\delta, \delta)$.

(iii) Their densities $\eta_{i,j}$ are taken from $\{0, \delta, 2\delta, \ldots, 1\}$.

(iv) The following graph $T = T(R)$ is k-colorable: if R has order t then T has t vertices, and $(i, j) \in E(T)$ iff $\eta_{i,j} > 0$.

To show that this is a valid reduction, assume first that G is ϵ-far from being k-colorable. Assume G is $(\epsilon - \delta)$-close to satisfying a regularity instance $R \in \mathcal{R}$. We can thus make $(\epsilon - \delta)n^2$ edge modifications and get a graph satisfying R. We also remove all edges inside the sets V_i. As by item (ii) each set has size at most δn we remove less than δn^2 edges. The total number of edges removed is thus less than ϵn^2. By property (iv) of the regularity instances of $\mathcal{R}$ this means that the new graph is k-colorable, which is impossible because we made less than ϵn^2 edge modifications and G was assumed to be ϵ-far from being k-colorable.

Assume now that G is k-colorable and let $V_1, \ldots, V_k$ be the partition of $V(G)$, which is determined by a legal k-coloring of G. Break every set V_i into sets $U_{i,1}, \ldots, U_{i,j_i}$ of size $\frac{\delta}{2k}n$ each. In case $\frac{\delta}{2k}n$ does not divide $|V_i|$, put the remaining vertices in a "garbage set" L. Note that the size of L is at most $\frac{1}{2}\delta n$. By Lemma 2.3, starting from the equipartition into the sets $U_{i,j}$, whose order is at most $2k/\delta$ we can get a δ-regular equipartition of G of order at most $T_{2.3}(2k/\delta, \delta)$. Note that disregarding the refinement of L the new equipartition must satisfy item (iv) in the definition of $\mathcal{R}$. As by item (iii) the densities of the instances in $\mathcal{R}$ are taken from $\{0, \delta, 2\delta, \ldots, 1\}$ we can make at most δn^2 edge modifications and thus "round down" the densities between the sets into a multiple of δ, while maintaining the regularity of the regular-pairs (we can use Lemma 3.1 here). This means that the new graph satisfies a regularity-instance $R \in \mathcal{R}$, which means that G was δ-close to satisfying R. ■

The examples that were discussed above apply Theorem 2 to obtain positive results. Our third application of Theorem 2 derives a negative result. The main focus of [18] is testing for isomorphism

[9] An alert reader may note that our proof of Theorem 2 applies the result of [20], which relies on the result of [27]. Thus, in the strict sense it is wrong to say that we infer the result of [27] from ours. However, it is not difficult to see that the result of [27] also follows from our (self-contained) proof of Lemma 5.2. Also, while k-colorability can actually be tested with 1-sided error, many other partition problems that were studied in [27] can only be tested with 2-sided error.

to a given fixed graph[10]. It shows that the query complexity of testing for isomorphism grows with a certain parameter, which measures the "complexity" of the graph. Without going into too much detail we just mention that under this measure random graphs are complex. Here we prove that testing for being isomorphic to a graph generated by $G(n, 0.5)$ requires a super-constant number of queries.

Corollary 7.3 *Let I be a graph generated by $G(n, 0.5)$. Then, with probability $1 - o(1)$ the property of being isomorphic to I is not testable.*

Proof (skecth): By Theorem 2 it is enough to show that with probability $1 - o(1)$ the property of being isomorphic to I is not regular-reducible. Note, that now there is only one value of n to consider in Definition 2.6 because the property we consider is a property of n-vertex graphs. Consider a graph generated by $G(n, 0.5)$. Clearly, by Lemma 3.4 the bipartite graph induced on any pair of sets of vertices of size $\sqrt{n}$ has density ≈ 0.5. We claim that if I satisfies this condition then the graph property $\mathcal{P}_I$ of being isomorphic to I is not regular-reducible. Suppose it is regular-reducible and consider a small δ, say $\delta = 0.01$. Let $\mathcal{R}$ be the set of regularity-instances, which corresponds to this value of δ. Let G be a graph isomorphic to I. By Definition 2.6 it must be the case that G is δ-close to satisfying some $R \in \mathcal{R}$. By the properties of I this means that most densities of R must be close to 0.5. Let k denote the order of R and let $\eta_{i,j}$ denote its densities. Consider a random k-partite graph on sets of vertices $V_1, \dots, V_k$ each of size n/k, where the bipartite graph connecting V_i and V_j is a random bipartite graph with edge density $\eta_{i,j}$. Clearly this graph is δ-close to satisfying R. On the other hand, it is not difficult to see that as most of the densities $\eta_{i,j}$ should be close to 0.5, then with high probability such a graph must be α-far from being isomorphic to I, for some fixed $\alpha > 0$, say $\alpha = 0.03$. This means that we have a graph that is 0.03-far from satisfying the property and is yet 0.01-close to satisfying one of the regularity-instances of $\mathcal{R}$. As we chose $\delta = 0.01$, this violates the second condition of Definition 2.6. ■

8 Concluding Remarks and Open Problems

The main result of this paper gives a combinatorial characterization of the graph properties, which can be tested with a constant number of edge queries in the dense graph model, possibly with a two-sided error. Together with the (near) characterization of [7] of the graph properties that can be tested with one-sided error, and the result of [20] showing that any testable property is also estimable, we get a more or less complete answer to many of the *qualitative* questions on testing graph properties in the dense model. An intriguing open question is to address the above problems quantitatively, and specifically to characterize the graph properties that can be tested with query complexity that is bounded by a polynomial of $1/\epsilon$.

While property testing in the dense model is relatively well understood, there are no general positive or negative results on testing graph properties in the bounded-degree model [29] or the general density model [33]. In these models the query complexity of the tester usually depends on the size of the input. It seems interesting and challenging to obtain general results in these models. One interesting problem is which of the partition problems which were studied in [27] can be tested

[10]Of course, we refer to isomorphism of un-labeled graphs.

using a sublinear number of queries. It will also be very interesting to give general positive and negative results concerning the testing of boolean functions.

Acknowledgements: The authors would like to thank the organizers of the Dagstuhl Workshop on Sublinear Algorithms, where this research has emerged. Asaf Shapira would also like to thank his hosts at the Theory Group of Microsoft Research, where parts of this research were conducted.

References

[1] N. Alon, R. A. Duke, H. Lefmann, V. Rödl and R. Yuster, The algorithmic aspects of the Regularity Lemma, Proc. of 33^{rd} IEEE FOCS, Pittsburgh, IEEE (1992), 473-481. Also: J. of Algorithms 16 (1994), 80-109.

[2] N. Alon, S. Dar, M. Parnas and D. Ron, Testing of clustering, Proc. 41^{st} IEEE FOCS, IEEE (2000), 240-250. Also: SIAM J. Discrete Math. 16 (2003), 393-417.

[3] N. Alon, E. Fischer, M. Krivelevich and M. Szegedy, Efficient testing of large graphs, Proc. of 40^{th} FOCS, New York, NY, IEEE (1999), 656–666. Also: Combinatorica 20 (2000), 451-476.

[4] N. Alon, M. Krivelevich, I. Newman and M. Szegedy, Regular languages are testable with a constant number of queries, Proc. 40^{th} FOCS, New York, NY, IEEE (1999), 645–655. Also: SIAM J. on Computing 30 (2001), 1842-1862.

[5] N. Alon and M. Krivelevich, Testing k-colorability, SIAM J. Discrete Math., 15 (2002), 211-227.

[6] N. Alon and A. Shapira, Every monotone graph property is testable, Proc. of STOC 2005, 128-137. Also, SIAM J. on Computing, to appear.

[7] N. Alon and A. Shapira, A characterization of the (natural) graph properties testable with one-sided error, Proc. of FOCS 2005, 429-438. Also, SIAM J. on Computing, to appear.

[8] N. Alon and A. Shapira, A separation theorem in property testing, Combinatorica, to appear.

[9] N. Alon and J. H. Spencer, **The Probabilistic Method**, Second Edition, Wiley, New York, 2000.

[10] M. Ben-Or, D. Coppersmith, M. Luby and R. Rubinfeld, Non-abelian homomorphism testing, and distributions close to their self-convolutions, Proc. of APPROX-RANDOM 2004, 273-285.

[11] E. Ben-Sasson, P. Harsha and S. Raskhodnikova, Some 3-CNF properties are hard to test, Proc. of STOC 2003, 345-354. Also, SIAM J. on Computing, 35 (2005), 1-21.

[12] M. Blum, M. Luby and R. Rubinfeld, Self-testing/correcting with applications to numerical problems, JCSS 47 (1993), 549-595.

[13] C. Borgs, J. Chayes, L. Lovász, V.T. Sós, B. Szegedy and K. Vesztergombi, Graph limits and parameter testing, Proc. of STOC 2006, 261-270.

[14] A. Czumaj, C. Sohler and M. Ziegler, Property testing in computational geometry, Proc. of ESA 2000, 155–166.

[15] A. Czumaj and C. Sohler, Abstract combinatorial programs and efficient property testers, SIAM Journal on Computing 34 (2005), 580-615.

[16] E. Fischer, Testing graphs for colorability properties, Proc. of the 12^{th} SODA (2001), 873-882. Also, Random Struct. and Algorithms 26 (2005), 289-309.

[17] E. Fischer, The art of uninformed decisions: A primer to property testing, The Computational Complexity Column of The Bulletin of the European Association for Theoretical Computer Science 75 (2001), 97-126. Also, Current Trends in Theoretical Computer Science: The Challenge of the New Century, G. Paun, G. Rozenberg and A. Salomaa (editors), World Scientific Publishing (2004), Vol. I 229-264.

[18] E. Fischer, The difficulty of testing for isomorphism against a graph that is given in advance, SIAM Journal on Computing, 34, 1147-1158.

[19] E. Fischer, G. Kindler, D. Ron, S. Safra, and A. Samorodnitsky, Testing juntas, JCSS 68 (2004), 753-787. Also, Proc. of FOCS (2002), 103-112.

[20] E. Fischer and I. Newman, Testing versus estimation of graph properties, Proc. of 37^{th} STOC, 2005, 138-146. SIAM J. on Computing, to appear.

[21] E. Fischer, I. Newman and J. Sgall, Functions that have read-twice constant width branching programs are not necessarily testable, Random Structures and Algorithms, in press.

[22] E. Fischer, E. Lehman, I. Newman, S. Raskhodnikova, R. Rubinfeld and A. Samorodnitsky, Monotonicity testing over general poset domains, Proc. of STOC (2002), 474-483.

[23] K. Friedl, G. Ivanyos and M. Santha, Efficient testing of groups, Proc. of STOC (2005), 157-166.

[24] A. Frieze and R. Kannan, Quick approximation to matrices and applications, *Combinatorica* 19 (1999), 175-220.

[25] O. Goldreich, Combinatorial property testing - a survey, In: Randomization Methods in Algorithm Design (P. Pardalos, S. Rajasekaran and J. Rolim eds.), AMS-DIMACS (1998), 45-60.

[26] O. Goldreich, Contemplations on testing graph properties, manuscript, 2005. Available at http://www.wisdom.weizmann.ac.il/∼oded.

[27] O. Goldreich, S. Goldwasser and D. Ron, Property testing and its connection to learning and approximation, Proc. of 37^{th} Annual IEEE FOCS, (1996), 339–348. Also, JACM 45(4): 653-750 (1998).

[28] O. Goldreich and L. Trevisan, Three theorems regarding testing graph properties, Proc. 42^{nd} IEEE FOCS, IEEE (2001), 460-469. Also, Random Structures and Algorithms, 23(1):23-57, 2003.

[29] O. Goldreich and D. Ron, Property Testing in Bounded-Degree Graphs, Proc. of STOC 1997, 406-415. Also, Algorithmica 32 (2002), 302-343.

[30] T. Gowers, Lower bounds of tower type for Szemerédi's uniformity lemma, GAFA 7 (1997), 322-337.

[31] J. Komlós and M. Simonovits, Szemerédi's Regularity Lemma and its applications in graph theory. In: *Combinatorics, Paul Erdős is Eighty*, Vol II (D. Miklós, V. T. Sós, T. Szönyi eds.), János Bolyai Math. Soc., Budapest (1996), 295–352.

[32] I. Newman, Testing of functions that have small width branching programs, Proc. of 41^{th} FOCS (2000), 251-258.

[33] M. Parnas and D. Ron, Testing the diameter of graphs, Random structures and algorithms, 20 (2002), 165-183. Also, Random Struct. and Algorithms 20 (2002), 165-183.

[34] M. Parnas, D. Ron and R. Rubinfeld, Testing membership in parenthesis languages, Random Struct. and Algorithms 22 (2003), 98-138.

[35] V. Rödl and R. Duke, On graphs with small subgraphs of large chromatic number, *Graphs and Combinatorics* 1 (1985), 91–96.

[36] D. Ron, Property testing, in: P. M. Pardalos, S. Rajasekaran, J. Reif and J. D. P. Rolim, editors, *Handbook of Randomized Computing*, Vol. II, Kluwer Academic Publishers, 2001, 597–649.

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

[38] E. Szemerédi, Regular partitions of graphs, In: *Proc. Colloque Inter. CNRS* (J. C. Bermond, J. C. Fournier, M. Las Vergnas and D. Sotteau, eds.), 1978, 399–401.