Deterministic vs Non-deterministic Graph Property Testing

Lior Gishboliner [*]   Asaf Shapira[†]

Abstract

A graph property $\mathcal{P}$ is said to be testable if one can check whether a graph is close or far from satisfying $\mathcal{P}$ using few random local inspections. Property $\mathcal{P}$ is said to be non-deterministically testable if one can supply a "certificate" to the fact that a graph satisfies $\mathcal{P}$ so that once the certificate is given its correctness can be tested. The notion of non-deterministic testing of graph properties was recently introduced by Lovász and Vesztergombi [5], who proved that (somewhat surprisingly) a graph property is testable if and only if it is non-deterministically testable. Their proof used graph limits, and so it did not supply any explicit bounds. They thus asked if one can obtain a proof of their result which will supply such bounds. We answer their question positively by proving their result using Szemerédi's regularity lemma.

An interesting aspect of our proof is that it highlights the fact that the regularity lemma can be interpreted as saying that all graphs can be approximated by finitely many "template" graphs.

1 Introduction

We consider properties of finite graph, where a property of graphs is simply a family of graphs closed under isomorphism. The main focus of our paper is the following notion of efficiently checking if a graph satisfies property $\mathcal{P}$ or is ϵ-far from satisfying it, where a graph G is said to be ϵ-far from satisfying $\mathcal{P}$ if one should add/delete at least ϵn^2 edges to turn G into a graph satisfying $\mathcal{P}$.

Definition 1.1. (Testable property) *A graph property $\mathcal{P}$ is called* testable *if there is an algorithm $\mathcal{T}_{\mathcal{P}}$, called a tester, that does the following: given $\epsilon > 0$ and a graph G, the tester $\mathcal{T}_{\mathcal{P}}$ samples a set S of $q_{\mathcal{P}}(\epsilon)$ vertices from G, checks for every $i, j \in S$ whether $(i, j) \in E(G)$ and then accepts/rejects deterministically based on the subgraph of G spanned by S. The success probability of $\mathcal{T}_{\mathcal{P}}$ should be at least $\frac{2}{3}$. In other words, if the input G satisfies $\mathcal{P}$ then $\mathcal{T}_{\mathcal{P}}$ accepts it with probability at least $\frac{2}{3}$, and if G is ϵ-far from satisfying $\mathcal{P}$ then $\mathcal{T}_{\mathcal{P}}$ rejects G with probability at least $\frac{2}{3}$. The function $q_{\mathcal{P}}(\epsilon)$ is called the* query complexity *of $\mathcal{T}_{\mathcal{P}}$, and does not depend on the size of the input graph.*

---

[*]School of Mathematics, Tel-Aviv University, Tel-Aviv, Israel 69978.

[†]School of Mathematics, Tel-Aviv University, Tel-Aviv, Israel 69978. Email: `asafico@tau.ac.il`.

The usual definition of a property $\mathcal{P}$ being testable, as introduced in [9], allows for the algorithm to be adaptive, but as proved in [1, 10] one can transform any tester into a tester of the form stated in Definition 1.1 with a very minor loss in the query complexity. Therefore we do not lose generality by restricting ourselves to Definition 1.1. The important point to observe about Definition 1.1 is that the algorithm makes its decision solely on the basis of the random inspections it makes into the input graph G. In other words, the decision of the algorithm is uniquely determined by the distribution of induced subgraphs of size $q_{\mathcal{P}}(\epsilon)$ in the input graph G.

Starting with [9], the problem of characterizing the testable graph properties received a lot of attention and by now there are several general results of this type, see [2, 6] and the recent surveys [8, 11] for more results and references on graph property testing. A drawback of these characterizations is that they are hard to state (and use). An alternative clean characterization was recently obtained by Lovász and Vesztergombi [5]. To state this characterization we need a bit of notation.

A k-colored graph on n vertices is a coloring of the edges of K_n (the complete graph on n vertices) using k colors. Thus, a graph can be thought of as a 2-colored graph. A property of k-colored graphs is again just a family of k-colored graphs closed under isomorphism, and it is said to be testable[1] if it is testable in the sense of Definition 1.1. A (k, m)-coloring of a graph G is a coloring of the edges and non-edges of G with the colors $\{1, \ldots, k\}$, so that edges are colored by $\{1, ..., m\}$ and non-edges are colored by $\{m + 1, ..., k\}$. The following is the notion of non-deterministic testing introduced in [5].

Definition 1.2. (Non-deterministically testable property) *A graph property $\mathcal{P}$ is called non-deterministically testable if there are integers k, m and a property $\mathcal{Q}$ of k-colored graphs so that:*

1. *A graph G satisfies $\mathcal{P}$ if and only if there is a (k, m)-coloring of G which satisfies $\mathcal{Q}$.*

2. *$\mathcal{Q}$ is testable.*

We are now ready to state the characterization of the testable graph properties that was obtained by Lovász and Vesztergombi [5].

Theorem 1. ([5]) *A property $\mathcal{P}$ is testable if and only if it is non-deterministically testable.*

Clearly any testable property is also non-deterministically testable, thus the interesting part of the above theorem is that given the fact that a property is non-deterministically testable, one can construct a standard tester for the property. Quoting [5], "one could say that this theorem shows that "P=NP" for property testing in dense graphs". We refer the reader to [5] for several nice illustrations showing how to apply Theorem 1.

---

[1] We define a k-colored graph to be ϵ-far from satisfying a property $\mathcal{Q}$ of k-colored graphs if one should modify the colors of at least ϵn^2 edges in order to turn G into a k-colored graph satisfying $\mathcal{P}$.

The proof of Theorem 1 in [5] used the machinery of graph limits. Hence, the proof was not explicit, that is, given the fact that a property $\mathcal{P}$ is non-deterministically testable (in the sense of Definition 1.2), it proved the existence of a standard testing algorithm for $\mathcal{P}$ (in the sense of Definition 1.1) but it did not supply any upper bound for the query complexity of the new tester (i.e. the function $q_{\mathcal{P}}(\epsilon)$ in Definition 1.1). Lovász and Vesztergombi [5] thus asked if Theorem 1 can be proved using explicit arguments that will give an effective bound. Our main result in this paper gives a positive answer to their question.

Our new proof of Theorem 1 uses several tools related to Szemerédi's regularity lemma [12]. In Section 2 we give the necessary background for applying this lemma, state some previous results as well as some preliminary lemmas that will be used in our new proof of Theorem 1. As the proofs of these technical lemmas are somewhat routine we differ them to Section 4. The proof of Theorem 1 appears in Section 3. As our proof applies the regularity lemma, although the bounds it supplies for $q_{\mathcal{P}}(\epsilon)$ are explicit, they are rather weak ones, given by Tower-type function of ϵ. Therefore, we will not keep track of the exact dependence of $q_{\mathcal{P}}(\epsilon)$ on ϵ. Finally, as we mentioned in the abstract, we believe that our proof gives a nice illustration of the fact that the regularity lemma implies that all graph can be "approximated" using only a finitely many template graphs. In fact, this intuition is the main idea behind the proof.

2 Tools and Preliminary Lemmas

As mentioned earlier, our proof of Theorem 1 will apply various tools related to Szemerédi's regularity lemma [12]. We will start with the basic definitions, then state some previous results that we will use (Theorem 2 and Lemmas 2.4 and 2.8) and then state some technical lemmas that we will need for the proof (Lemmas 2.13, 2.14 and 2.16). The proofs of these technical lemmas appear in Section 4. Here, and throughout the paper, when we write $x = y \pm z$ we mean that $y - z \leq x \leq y + z$.

Given two disjoint vertex sets U, V we use $E(U, V)$ to denote the set of edges connecting U to V and set $d(U, V) = |E(U, V)|/|U||V|$ to be the density of the bipartite graph between U and V. The basic notion of a regular pair is the following.

Definition 2.1. (Regular pair) *Suppose U, V are disjoint vertex sets in a graph and let $\gamma \in (0, 1)$. The pair (U, V) is said to be γ-regular if for every two subsets $U' \subseteq U$, $V' \subseteq V$ satisfying $|U'| \geq \gamma |U|$, $|V'| \geq \gamma |V|$ the inequality $|d(U, V) - d(U', V')| \leq \gamma$ holds.*

A γ-regular pair can/should be thought of as behaving almost like a random bipartite graph of the same density. A partition $V_1, ..., V_r$ of the vertex set of a graph is called an equipartition if $||V_i| - |V_j|| \leq 1$ for every $1 \leq i < j \leq r$. The order of a partition $V_1, ..., V_r$ is the number of parts in it (i.e. the integer r).

Definition 2.2. (Regular equipartition) *An equipartition $V_1, ..., V_r$ of the vertices of a graph is γ-regular if all but at most $\gamma \binom{r}{2}$ of the pairs (V_i, V_j) are γ-regular.*

We now define a graph property of having a γ-regular equipartition with a predefined set of densities.

Definition 2.3. (Regularity instance) *A graph* regularity instance R *is given by a regularity parameter γ, an integer r (the order of R), a set of densities $\eta_{i,j}$ where $1 \leq i < j \leq r$, and a set of non-regular pairs $\bar{R}$ of size at most $\gamma\binom{r}{2}$. A graph G is said to satisfy R if G has an equipartition $V_1, ..., V_r$ such that for every $(i,j) \notin \bar{R}$ the pair (V_i, V_j) is γ-regular and satisfies $d(V_i, V_j) = \eta_{i,j}$.*

A key element in the proof of Theorem 1 is the following result, which follows from the main results of [2, 3]. It allows us to test how close a graph is to satisfying a given regularity instance.

Theorem 2. ([2, 3]) *Let R be a graph regularity instance, let $\epsilon_2 > \epsilon_1 > 0$ and let $p < \frac{1}{2}$. Then there is a tester $\mathcal{T} = \mathcal{T}_2(R, \epsilon_1, \epsilon_2, p)$ that distinguishes graphs that are ϵ_1-close to satisfying R from graphs that are ϵ_2-far from satisfying R, with success probability at least $1 - p$.*

Furthermore, the query-complexity of T depends only on R, ϵ_1, ϵ_2 and p (and not on the input graph) and can be expressed as an explicit function of these parameters.

We note that the arguments used in [2, 3] to prove the above result all relied heavily on the regularity lemma. Therefore, the bounds they give have a very poor (yet explicit) Tower-type dependence on the input parameters.

The second result we will need is Corollary 3.8 from [2].

Lemma 2.4. ([2]) *Let R be a graph regularity instance of order r, regularity parameter γ, densities $\eta_{i,j}$ and a set of non-regular pairs $\bar{R}$. Suppose that a graph G has an equipartition $V_1, ..., V_r$ such that for every $(i,j) \notin \bar{R}$ the pair (V_i, V_j) is γ-regular[2] and satisfies $d(V_i, V_j) = \eta_{i,j} \pm \frac{\gamma^2 \epsilon}{50}$. Then G is ϵ-close to satisfying R.*

We now turn to consider k-colored graphs. We first generalize the definitions of a regular pair, regular equipartition and regularity instance, to the more general setting of k-colored graphs. We start with the following notation: Suppose U, V are two disjoint vertex sets in a k-colored graph. We use[3] $d^\ell(U, V)$ to denote the density of edges of color ℓ between U and V, that is $d^\ell(U, V)| = E^\ell(U, V)|/|A||B|$, where $E^\ell(U, V)$ is the set of edges with color ℓ that connect U to V. In case there is more than one graph, we use $d^\ell_G(U, V)$ to denote the density of edges colored by ℓ between U, V in the k-colored graph G.

Definition 2.5. (Regular pair in a k-colored graph) *Suppose U, V are disjoint vertex sets in a k-colored graph. The pair (U, V) is γ-regular if for every two subsets $U' \subseteq U$, $V' \subseteq V$ satisfying $|U'| \geq \gamma|U|$, $|V'| \geq \gamma|V|$, and for every $1 \leq \ell \leq k$, the inequality $\left| d^\ell(U, V) - d^\ell(U', V') \right| \leq \gamma$ holds.*

---

[2]Actually, Corollary 3.8 in [2] only needs to assume that (V_i, V_j) is $\left(\gamma + \frac{\gamma^2 \epsilon}{50} \right)$-regular.

[3]Here, and throughout the paper, we always use ℓ as a superscript and never as an exponent. So x^ℓ should read "x superscript ℓ" not "x to the power ℓ".

Definition 2.6. (Regular equipartition in a k-colored graph) *An equipartition $V_1, ..., V_r$ of the vertices of a k-colored graph is γ-regular if all but at most $\gamma\binom{r}{2}$ of the pairs (V_i, V_j) are γ-regular.*

Definition 2.7. (k-colored regularity instance) *A k-colored regularity instance R is given by a regularity parameter γ, an integer r (the order of R), a set of densities $\eta_{i,j}^\ell$ where $1 \le i < j \le r$ and $1 \le \ell \le k$, and a set of non-regular pairs $\bar{R}$ of size at most $\gamma\binom{r}{2}$. A k-colored graph G is said to satisfy R if G has an equipartition $V_1, ..., V_r$ such that for every $(i, j) \notin \bar{R}$ the pair (V_i, V_j) is γ-regular and satisfies $d^\ell(V_i, V_j) = \eta_{i,j}^\ell$ for every $1 \le \ell \le k$.*

The Regularity Lemma for k-colored graphs states that every k-colored graph has a γ-regular equipartition whose order can be bounded by a function of γ and k. It can be formulated in terms of regularity instances in the following way.

Lemma 2.8. (Regularity lemma for k-colored graphs) *For every $\gamma > 0$ and integers t and k, there exists $T = T_{2.8}(\gamma, t, k)$ so that every k-colored graph satisfies some k-colored regularity instance of order at least t and at most T, and regularity parameter γ.*

Note that the usual regularity lemma is the special case of the k-colored regularity lemma with $k = 2$. The proof of the k-colored version requires a minor adaptation of the proof of the standard regularity lemma. See [4] for the details.

Having described the known results that will be used in the proof of Theorem 1, we now turn to state the additional technical lemmas we shall rely on. We start with a lemma that allows one to approximate the number of copies of small k-colored graphs in a k-colored graph which satisfies a given regularity instance.

Definition 2.9. ($IC(B, W, \sigma)$) *Suppose B is a k-colored graph on vertices $[q]$ in which (i, j) is colored by $c(i, j)$. Suppose $W = \{\eta_{i,j}^\ell : 1 \le i < j \le q, \ 1 \le \ell \le k\}$ are densities and $\sigma : [q] \to [q]$ is a permutation. Define:*

$$IC(B, W, \sigma) = \prod_{i<j} \eta_{i,j}^{c(\sigma(i),\sigma(j))}$$

Definition 2.10. ($IC(B, W)$) *Suppose B and W are as in the previous definition. Define:*

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

where $Aut(B)$ is the number of automorphisms of B, that is, the number of injections $\phi : V(B) \mapsto V(B)$ that preserve the color of the edges.

Definition 2.11. ($IC(B, R)$) *Let R be a k-colored regularity instance of order r and densities $\{\eta_{i,j}^\ell : 1 \le i < j \le r, \ 1 \le \ell \le k\}$. Let B be a k-colored graph on the vertex set $[q]$. For every $A \subseteq [r]$ of size q put $W(A) = \{\eta_{i,j}^\ell : i, j \in A, \ 1 \le \ell \le k\}$. Define:*

$$IC(B, R) = \frac{1}{\binom{r}{q}} \sum_{A \subseteq [r], |A|=q} IC(B, W(A))$$

Remark 2.12. *It is easy to see that $IC(B, W, \sigma)$, $IC(B, W)$ and $IC(B, R)$ are quantities in $[0, 1]$.*

To understand Definition 2.9, consider a random k-colored graph whose vertices are $V_1 \cup ... \cup V_q$. Suppose that the probability that the color of (v_i, v_j) is ℓ is $\eta^\ell_{i,j}$ ($v_i \in V_i, v_j \in V_j$). Suppose also that $|V_1| = ...|V_q| = n$. Let B be a fixed k-colored graph on the vertices $[q]$ and let σ be a permutation of $[q]$. What is the expected number of q-tuples $v_1 \in V_1, ..., v_q \in V_q$ which span a copy of B such that v_i plays the role of $\sigma(i)$? It is easy to see that this number is $IC(B, W, \sigma)n^q$ where we set $W = \{\eta^\ell_{i,j} : 1 \leq i < j \leq q, \ 1 \leq \ell \leq k\}$. We show (in Lemma 4.2) that for every δ, if all pairs (V_i, V_j) are γ-regular for some small enough γ, then the number of such q-tuples $v_1, ..., v_q$ is $(IC(B, W, \sigma) \pm \delta)n^q$. This fact demonstrates the almost random behavior of regular partitions. The expression $IC(B, W)$ (in Definition 2.10) is used to approximate the total number of q-tuples $v_1 \in V_1, ..., v_q \in V_q$ which span a copy of B. The expression $IC(B, R)$ (in Definition 2.11) is used to approximate the number of copies of B in a graph that satisfies the regularity instance R. The most general result of this sort is the following lemma.

Lemma 2.13. *For any $\delta > 0$ and integers k and q there are $\gamma = \gamma_{2.13}(\delta, q, k)$ and $t = t_{2.13}(\delta, q, k)$ with the following property: For any k-colored regularity instance R of order at least t and regularity parameter at most γ, and for any family $\mathcal{B}$ of k-colored graphs on q vertices, the number of copies of k-colored graphs $B \in \mathcal{B}$ in any k-colored graph on n vertices satisfying R is*

$$\left(\sum_{B \in \mathcal{B}} IC(B, R) \pm \delta \right) \binom{n}{q}$$

The proof of Lemma 2.13 appears in Subsection 4.1. The second lemma we will need is the following.

Lemma 2.14. *For every δ and integers q and k there is $\lambda = \lambda_{2.14}(\delta, q, k)$ such that the following holds: Let R and M be k-colored regularity instances of order r, and densities*

$$\{\eta^\ell_{i,j} : 1 \leq i < j \leq r, \ 1 \leq \ell \leq k\}$$

and

$$\{\mu^\ell_{i,j} : 1 \leq i < j \leq r, \ 1 \leq \ell \leq k\}$$

respectively. Let $\mathcal{B}$ be a family of k-colored graphs of order q. Suppose that $\mu^\ell_{i,j} = \eta^\ell_{i,j} \pm \lambda$. Then

$$\left| \sum_{B \in \mathcal{B}} IC(B, R) - \sum_{B \in \mathcal{B}} IC(B, M) \right| \leq \delta . \tag{1}$$

The proof of Lemma 2.14 appears in Subsection 4.2. The last ingredient we will need is the following lemma whose proof appears in Subsection 4.3.

Definition 2.15. (Chopping) *Let R be a graph regularity instance of order r, regularity parameter γ, densities $\mu_{i,j}$ and a set $\bar{R}$ of non-regular pairs. A (k,m)-**chopping** of R is any k-colored regularity instance R' of order r, regularity parameter 2γ, non-regular set $\bar{R}' = \bar{R}$ and densities $\eta_{i,j}^{\ell}$ that satisfy*

$$\sum_{\ell=1}^{m} \eta_{i,j}^{\ell} = \mu_{i,j} \quad and \quad \sum_{\ell=m+1}^{k} \eta_{i,j}^{\ell} = 1 - \mu_{i,j}$$

Lemma 2.16. *For every $\gamma > 0$ and integers t and k, there is $n_{2.16}(\gamma, t, k)$ such that the following holds: Suppose R is a graph regularity instance of order at most t and regularity parameter γ, that R' is a (k,m)-chopping of R and that G is a graph satisfying R with at least $n_{2.16}(\gamma, t, k)$ vertices. Then G has a (k,m)-coloring that satisfies R'.*

3   The New Proof of Theorem 1

Consider any $\epsilon > 0$. Let property $\mathcal{Q}$ and integers k and m be those from Definition 1.2, that is, so that $\mathcal{Q}$ is a property of k-colored graphs and so that a graph satisfies $\mathcal{P}$ if and only if it has a (k,m)-coloring satisfying $\mathcal{Q}$. Suppose $\mathcal{Q}$ can be tested by a tester $\mathcal{T}_\mathcal{Q}$ as in Definition 1.1. Let $q = q_\mathcal{Q}(\frac{\epsilon}{2})$ be the query-complexity of $\mathcal{T}_\mathcal{Q}$, i.e. the number of vertices that $\mathcal{T}_\mathcal{Q}$ samples when testing if a k-colored graph satisfies $\mathcal{Q}$ or is $\frac{\epsilon}{2}$-far from satisfying it. Let $\mathcal{B}$ be the set of all k-colored graphs B on q vertices, such that when $\mathcal{T}_\mathcal{Q}$ samples a k-colored graph isomorphic to B, it accepts the input. Put

$$t = t_{2.13}(1/12, q, k), \quad \gamma = \gamma_{2.13}(1/12, q, k), \quad T = T_{2.8}(\gamma/2k, t, k)$$

and

$$\eta = \min\left\{\lambda_{2.14}(1/12, q, k), \frac{\epsilon\left(\frac{\gamma}{2}\right)^2}{200m}\right\}.$$

Let I be the set of all k-colored regularity instances of order at least t and at most T, regularity parameter γ and densities from the set $\{0, \eta, 2\eta, 3\eta, ..., 1\}$. Observe that all the above constants, as well as $|I|$, depend only on ϵ, k and the properties $\mathcal{P}$ and $\mathcal{Q}$. We now arrive at the critical definition of the proof:

Definition 3.1. (Good regularity instance) *A graph regularity instance R with regularity measure $\gamma/2$ is considered **good** if it has a (k,m)-chopping R' that satisfies:*

1. *$R' \in I$.*

2. *$\sum_{B \in \mathcal{B}} IC(B, R') \geq \frac{1}{2}$.*

*We say that R' is a **witness** to the fact that R is good. We set GOOD to be the family of good regularity instances.*

Suppose first that the input graph has less than $n_{2.16}(\frac{\gamma}{2}, T, k)$ vertices. In this case the algorithm can just ask about all edges of G and then check if G satisfies property $\mathcal{P}$. Since γ, T and k are all functions of ϵ, $\mathcal{P}$ and $\mathcal{Q}$, we get that so is the query complexity in this case. Hence from this point on we will assume that $n \geq n_{2.16}(\frac{\gamma}{2}, T, k)$.

The following are the key observations we will need for the proof.

Claim 3.2. *If G satisfies $\mathcal{P}$, then G is $\frac{\epsilon}{4}$-close to satisfying some $R \in GOOD$.*

Claim 3.3. *If G is ϵ-far from satisfying $\mathcal{P}$, then G is $\frac{\epsilon}{2}$-far from satisfying any $R \in GOOD$.*

Let us first complete the proof based on these claims. We describe a randomized algorithm that distinguishes between graphs satisfying $\mathcal{P}$ and graphs that are ϵ-far from satisfying $\mathcal{P}$, with success probability at least $\frac{2}{3}$, and by making a number of queries that can be bounded by a function of ϵ. Put $p = \frac{1}{3|GOOD|}$. Let G be a graph on at least $n_{2.16}(\frac{\gamma}{2}, T, k)$ vertices. In order to test G for property $\mathcal{P}$ we do the following: For every $R \in GOOD$ use $\mathcal{T}_{13}(R, \frac{\epsilon}{4}, \frac{\epsilon}{2}, p)$ (recall Theorem 2) to test whether G is $\frac{\epsilon}{4}$-close to satisfying R or $\frac{\epsilon}{2}$-far from satisfying it. If one of these tests accepts, then accept the input G, otherwise reject it. If G satisfies $\mathcal{P}$ then by Claim 3.2 it is $\frac{\epsilon}{4}$-close to some $R \in GOOD$, and our tester accepts it with probability at least $1 - p \geq \frac{2}{3}$. If G is ϵ-far from satisfying $\mathcal{P}$ then by Claim 3.3 it is $\frac{\epsilon}{2}$-far from satisfying any $R \in GOOD$. Our tester accepts G with probability at most $|GOOD| \cdot p = \frac{1}{3}$ and so it rejects with probability at least $\frac{2}{3}$, as required. Finally, since all the parameters involved are given by explicit functions of ϵ and the properties $\mathcal{P}$ and $\mathcal{Q}$, we get via Theorem 2 that the number of queries made by the tester can be bounded by an explicit function of ϵ.

We now complete the proof of Theorem 1 by proving Claims 3.2 and 3.3.

Proof (of Claim 3.2): Suppose G satisfies $\mathcal{P}$. Then there exists some (k, m)-coloring of G that satisfies $\mathcal{Q}$. Denote this k-colored graph by H. By Lemma 2.8, H satisfies some k-colored regularity instance R'_1 of order $t \leq r \leq T$, regularity parameter $\frac{\gamma}{2k}$ and densities

$$\{\eta^\ell_{i,j} : 1 \leq i < j \leq r,\ 1 \leq \ell \leq k\}.$$

Since H satisfies $\mathcal{Q}$, we infer that $\mathcal{T}_\mathcal{Q}$ must accept H with probability at least $\frac{2}{3}$. This means that when sampling q vertices from H, the probability to get a k-colored graph isomorphic to one of the elements of $\mathcal{B}$ is at least $\frac{2}{3}$. By the choice of γ and t via Lemma 2.13 we get that this probability is $\sum_{B \in \mathcal{B}} IC(B, R'_1) \pm \frac{1}{12}$. Therefore

$$\sum_{B \in \mathcal{B}} IC(B, R'_1) \geq \frac{7}{12}\ . \tag{2}$$

Let $V_1, ..., V_r$ be an equipartition of H which corresponds to R'_1. We claim that $V_1, ..., V_r$ is also a $\frac{\gamma}{2}$-regular equipartition of G. To see this let $(i, j) \notin \bar{R}'_1$. For every $x \in V_i, y \in V_j$, the edge (x, y) is in G if and only if (x, y) is colored in H by a color $\ell \in \{1, ..., m\}$. Therefore

$d_G(V_i, V_j) = \sum_{\ell=1}^m d_H^\ell(V_i, V_j)$. Let $U' \subseteq V_i, V' \subseteq V_j$ such that $|U'| \geq \frac{\gamma}{2}|V_i|, |V'| \geq \frac{\gamma}{2}|V_j|$. As we assume that (V_i, V_j) is a $\frac{\gamma}{2k}$-regular pair in H, we have $|d_H^\ell(U', V') - d_H^\ell(V_i, V_j)| \leq \frac{\gamma}{2k}$ for every $1 \leq \ell \leq k$. By the triangle inequality, we have

$$|d_G(U', V') - d_G(V_i, V_j)| \leq \sum_{\ell=1}^m |d_H^\ell(U', V') - d_H^\ell(V_i, V_j)| \leq \frac{m\gamma}{2k} \leq \frac{\gamma}{2} \ .$$

We thus infer that G satisfies a regularity instance R_1 with order r, regularity parameter $\frac{\gamma}{2}$, a set of irregular pairs $\bar{R}_1'$ and densities $\{\eta_{i,j} : 1 \leq i < j \leq r\}$ where $\eta_{i,j} = \sum_{l=1}^m \eta_{i,j}^\ell$.

Let R_2' be the k-colored regularity instance that is obtained from R_1' by replacing each of the densities $\eta_{i,j}^\ell$ with the closest integer multiples of η. Observe that we thus change each density by at most η. As we chose $\eta \leq \lambda_{2.14}(\frac{1}{12}, q, k)$, we get from Lemma 2.14 and (2) that

$$\sum_{B \in \mathcal{B}} IC(B, R_2') \geq \sum_{B \in \mathcal{B}} IC(B, R_1') - \frac{1}{12} \geq \frac{1}{2} \ . \tag{3}$$

Denote the densities of R_2' by $\mu_{i,j}^\ell$. Let R_2 be the graph regularity instance of order r, regularity parameter $\frac{\gamma}{2}$, densities $\{\mu_{i,j} : 1 \leq i < j \leq r\}$, where $\mu_{i,j} = \sum_{\ell=1}^m \mu_{i,j}^\ell$, and a set of irregular pairs $\bar{R}_2'$. By Definition 2.15 R_2' is a (k, m)-chopping of R_2. Furthermore $R_2' \in I$ and we get from (3) that $\sum_{B \in \mathcal{B}} IC(B, R_2') \geq \frac{1}{2}$. By Definition 3.1 R_2' is a witness to the fact that R_2 is good. Finally, recalling that $\eta \leq \frac{\epsilon(\frac{\gamma}{2})^2}{200m}$, we get that for every $i < j$ we have

$$|\eta_{i,j} - \mu_{i,j}| = \left| \sum_{l=1}^m \eta_{i,j}^\ell - \sum_{l=1}^m \mu_{i,j}^\ell \right| \leq \sum_{l=1}^m \left| \eta_{i,j}^\ell - \mu_{i,j}^\ell \right| \leq m\eta \leq \frac{\epsilon(\frac{\gamma}{2})^2}{200} \ .$$

In other words, the densities of R_1 and R_2 differ by at most $\frac{\epsilon(\frac{\gamma}{2})^2}{200}$. We now get via Lemma 2.4 that G is $\frac{\epsilon}{4}$-close to satisfying R_2. $\blacksquare$

Proof (of Claim 3.3): We will prove that if G is $\frac{\epsilon}{2}$-close to satisfying some $R \in GOOD$ then G is ϵ-close to satisfying $\mathcal{P}$. Suppose that G is $\frac{\epsilon}{2}$-close to a graph G' that satisfies some $R \in GOOD$. By Definition 3.1 R has a (k, m)-chopping R' such that $\sum_{B \in \mathcal{B}} IC(B, R') \geq \frac{1}{2}$. By Lemma 2.16 G' has a (k, m)-coloring satisfying R'. Call this k-colored graph H'. By Lemma 2.13 the probability to get a k-colored graph isomorphic to an element of $\mathcal{B}$ when sampling q vertices from H' is $\sum_{B \in \mathcal{B}} IC(B, R') \pm \frac{1}{12}$. Therefore this probability is at least $\frac{5}{12}$. If H' was $\frac{\epsilon}{2}$-far from satisfying $\mathcal{Q}$ this probability would have to be at most $\frac{1}{3}$, because $\mathcal{T}_Q$ would have to reject H' with probability at least $\frac{2}{3}$. So we infer that H' is $\frac{\epsilon}{2}$-close to satisfying $\mathcal{Q}$. This means that H' can be turned into a k-colored graph H'' that satisfies $\mathcal{Q}$ by changing the colors of at most $\frac{\epsilon}{2}n^2$ edges.

Construct a graph G'' by doing the following: For every $x, y \in V(H'')$, put an edge between x and y if (x, y) is colored by a color $\ell \in \{1, ..., m\}$ in H''. First, G'' satisfies $\mathcal{P}$ because H'' is a (k, m)-coloring of G'' which satisfies $\mathcal{Q}$. Furthermore, we claim that G'' is $\frac{\epsilon}{2}$-close to G'. Indeed, observe that the number of edge modifications we performed is exactly the number of pairs (x, y)

so that in one of the graphs H', H'' the color of (x, y) belonged to the set $\{1, \ldots, m\}$ while in the other it belonged to $\{m+1, \ldots, k\}$. This number is clearly bounded from above by the number of modifications made when turning H' to H''. Since H'' and H' differ in at most $\frac{\epsilon}{2}n^2$ edges the same thus holds for G'' and G', implying that G' is $\frac{\epsilon}{2}$-close to G''. Since G is assumed to be $\frac{\epsilon}{2}$-close to G', we infer that G is ϵ-close to G''. Since G'' satisfies $\mathcal{P}$ the proof is complete. ∎

4   Proofs of Auxiliary Lemmas

4.1   Proof of Lemma 2.13

We will need the following folklore result stating the a q-tuple of vertex sets that are pairwise regular have the correct number of copies of K_q (the complete graph on q vertices). A detailed proof can be found in [7].

Lemma 4.1. *For every $\delta > 0$ and q there exists $\gamma' = \gamma'_{4.1}(\delta, q)$ such that the following holds: Suppose $V_1, \ldots, V_q$ are disjoint vertex sets in a graph, $|V_1| = \ldots = |V_q| = n$, and all pairs (V_i, V_j) are γ'-regular. Put $IC(K_q; V_1, \ldots, V_q) = \prod_{i<j} d(V_i, V_j)$. Then the number of q-tuples $v_1 \in V_1, \ldots, v_q \in V_q$ that span a copy of K_q is*

$$(IC(K_q; V_1, \ldots, V_q) \pm \delta)n^q$$

As a first step towards proving Lemma 2.13 we prove a variant of Lemma 4.1 for k-colored graphs with respect to $IC(B, W, \sigma)$. We will then obtain similar lemmas with respect to $IC(B, W)$ and $IC(B, R)$ (recall Definitions 2.9, 2.10 and 2.11) and then derive from them the proof of Lemma 2.13.

Lemma 4.2. *For every $\delta > 0$ and q there exists $\gamma = \gamma_{4.2}(\delta, q)$ such that the following holds: Suppose $V_1, \ldots, V_q$ are disjoint vertex sets in a k-colored graph, $|V_1| = \ldots = |V_q| = n$, and all pairs (V_i, V_j) are γ-regular. Put $W = \{d^\ell(V_i, V_j) : 1 \leq i < j \leq q, \ 1 \leq \ell \leq k\}$. Then for every k-colored graph B on the vertices $[q]$, and for any permutation $\sigma : [q] \to [q]$, the number of q-tuples $v_1 \in V_1, \ldots, v_q \in V_q$ which span copy a of B with v_i playing the role of $\sigma(i)$ is*

$$(IC(B, W, \sigma) \pm \delta)n^q$$

Proof: While the proof of Lemma 4.1 can be adapted to the more general setting of Lemma 4.2 it will be easier to reduce Lemma 4.2 to Lemma 4.1. Set $\gamma = \gamma_{4.2}(\delta, q) = \gamma'_{4.1}(\delta, q)$ and suppose (V_i, V_j) is γ-regular for every $1 \leq i < j \leq q$. We call a q-tuple $v_1 \in V_1, \ldots, v_q \in V_q$ **proper**, if $v_1, \ldots, v_q$ span a copy of B with v_i playing the role of $\sigma(i)$.

We denote by $c(i, j)$ the color of the edge (i, j) in B. Let $E_{i,j}$ be the set of edges connecting a vertex in V_i to a vertex in V_j whose color is $c(\sigma(i), \sigma(j))$. If $v_1 \in V_1, \ldots, v_q \in V_q$ is proper, then the color of (v_i, v_j) is $c(\sigma(i), \sigma(j))$. We see that the edges in $E_{i,j}$ are the only edges between V_i and V_j

that can "participate" in a proper q-tuple. Define a q-partite graph S with vertex sets $V_1, ..., V_q$, in which the edges between V_i and V_j are $E_{i,j}$. A q-tuple $v_1 \in V_1, ..., v_q \in V_q$ is proper if and only if it spans a copy of K_q in S. So in order to prove Lemma 4.2 it is enough to show that the number of copies of K_q in S is $(IC(B, W, \sigma) \pm \delta)n^q$. By Lemma 4.1, the number of copies of K_q in S is $(IC(K_q; V_1, ..., V_q) \pm \delta)n^q$ where $IC(K_q; V_1, ..., V_q) = \prod_{i<j} d_S(V_i, V_j)$. So to complete the proof it is enough for us to show that $IC(B, W, \sigma) = IC(K_q; V_1, ..., V_q)$. Indeed, we have

$$IC(B, W, \sigma) = \prod_{i<j} d^{c(\sigma(i), \sigma(j))}(V_i, V_j) = \prod_{i<j} \frac{|E_{i,j}|}{|V_i||V_j|} = \prod_{i<j} d_S(V_i, V_j) = IC(K_q; V_1, ..., V_q)$$

■

Lemma 4.3. *For every $\delta > 0$ and every q there exists $\gamma = \gamma_{4.3}(\delta, q)$ such that the following holds: Suppose that $V_1, ..., V_q$ are disjoint vertex sets of size n each, and all pairs (V_i, V_j) are γ-regular. Put $W = \{d^\ell(V_i, V_j) : 1 \le i < j \le q, \ 1 \le \ell \le k\}$. Then for every k-colored graph B on the vertices $[q]$, the number of copies of B which have precisely one vertex in each of the sets $V_1, ..., V_q$ is*

$$(IC(B, W) \pm \delta)n^q$$

Proof: Set $\gamma_{4.3}(\delta, q) = \gamma_{4.2}(\frac{\delta}{q!}, q)$. Let $V_1, ..., V_q$ be as in the statement, and let B be any k-colored graph. By Claim 4.2 for any permutation $\sigma : [q] \to [q]$, the number of copies of B spanned by $v_1 \in V_1, ..., v_q \in V_q$ such that v_i plays the role of $\sigma(i)$ is $\left(IC(B, W, \sigma) \pm \frac{\delta}{q!}\right)n^q$. If we sum over all permutations $\sigma : [q] \to [q]$, we get $\sum_\sigma \left(IC(B, W, \sigma) \pm \frac{\delta}{q!}\right)n^q$. In this summation, we count every copy of B exactly $Aut(B)$ times. Thus, by dividing by $Aut(B)$, we get that the number of copies of B is

$$\frac{1}{Aut(B)}\left(\sum_\sigma \left(IC(B, W, \sigma) \pm \frac{\delta}{q!}\right)n^q\right) = \left(\frac{1}{Aut(B)}\sum_\sigma IC(B, W, \sigma) \pm \delta\right)n^q$$
$$= (IC(B, W) \pm \delta)n^q$$

■

Lemma 4.4. *For every $\delta > 0$ and q there are $\gamma = \gamma_{4.4}(\delta, q)$ and $t = t_{4.4}(\delta, q)$ such that the following holds: Suppose that R is a k-colored regularity instance of order at least t and regularity parameter at most γ. Then for every k-colored graph B on q vertices, the number of copies of B in any n-vertex k-colored graph satisfying R is*

$$(IC(B, R) \pm \delta)\binom{n}{q}$$

Proof: Put

$$t = t_{4.4}(\delta, q) = \left\lceil \frac{4q^2}{\delta} \right\rceil$$

and

$$\gamma = \gamma_{4.4}(\delta, q) = \min \left\{ \frac{\delta}{4q^2}, \gamma_{4.3}\left(\frac{\delta}{4}, q\right) \right\}.$$

Let R be a k-colored regularity instance as in the statement, let G be an n-vertex k-colored graph satisfying R and let B be any k-colored graph on q vertices. Let $V_1, ..., V_r$ be an equipartition of $V(G)$ satisfying R. Let $\mathcal{C}$ be the collection of all q-tuples that have at most one vertex in each of the sets V_i. By a union bound, the number of q-tuples that have more than one vertex in one of the sets V_i is at most

$$r \left(\frac{n}{r}\right)^2 \binom{n-2}{q-2} \leq \frac{q^2}{r} \binom{n}{q} \leq \frac{q^2}{t} \binom{n}{q} \leq \frac{1}{4}\delta \binom{n}{q}.$$

So $|\mathcal{C}| \geq \left(1 - \frac{\delta}{4}\right) \binom{n}{q}$. Therefore the lemma will follow from showing that the number of q-tuples belonging to $\mathcal{C}$ which span a copy of B is $\left(IC(B, R) \pm \frac{3}{4}\delta\right) |\mathcal{C}|$.

Given $A = \{x_1, ..., x_q\} \subseteq \{1, ..., r\}$ let $N(A)$ denote the number of q-tuples $v_1 \in V_{x_1}, ..., v_q \in V_{x_q}$ which span a copy of B. We say that A is **good** if all the pairs (V_{x_i}, V_{x_j}) $(1 \leq i < j \leq q)$ are γ-regular. Otherwise A is called **bad**. If A is good we get from our choice of γ via Lemma 4.3 that

$$N(A) = \left(IC(B, W(A)) \pm \frac{1}{4}\delta\right) \left(\frac{n}{r}\right)^q$$

where we set

$$W(A) = \{d^\ell(V_i, V_j) : i, j \in A, \ 1 \leq \ell \leq k\}.$$

We can thus estimate the number of q-tuples belonging to $\mathcal{C}$ which span a copy of B by

$$\sum_{A \subseteq [r], |A| = q} N(A) = \sum_{A \ is \ good} \left(\left(IC(B, W(A)) \pm \frac{1}{4}\delta\right) \left(\frac{n}{r}\right)^q\right) + \sum_{A \ is \ bad} N(A) =$$

$$\sum_{A \subseteq [r], |A| = q} \left(\left(IC(B, W(A)) \pm \frac{1}{4}\delta\right) \left(\frac{n}{r}\right)^q\right) + \sum_{A \ is \ bad} \left(N(A) - IC(B, W(A)) \left(\frac{n}{r}\right)^q\right) =$$

$$\left(IC(B, R) \pm \frac{1}{4}\delta\right) \binom{r}{q} \left(\frac{n}{r}\right)^q + \sum_{A \ is \ bad} \left(N(A) - IC(B, W(A)) \left(\frac{n}{r}\right)^q\right) \tag{4}$$

Since (V_i, V_j) is γ-regular for every $(i, j) \notin \bar{R}$ there are at most $\gamma \binom{r}{2}$ pairs (V_i, V_j) which are not γ-regular. Therefore the number of bad sets $A \subseteq \{1, ..., r\}$ is at most

$$\gamma \binom{r}{2} \binom{r-2}{q-2} \leq \gamma q^2 \binom{r}{q} \leq \frac{1}{4}\delta \binom{r}{q}$$

Using the facts that $0 \leq IC(B, W(A)) \leq 1$ and $0 \leq N(A) \leq \left(\frac{n}{r}\right)^q$ for every $A \subseteq [r]$, and the bound on the number of bad sets, we get that

$$\left| \sum_{A \text{ is bad}} \left(N(A) - IC(B, W(A)) \left(\frac{n}{r}\right)^q \right) \right| \leq \frac{1}{2} \delta \binom{r}{q} \left(\frac{n}{r}\right)^q$$

By plugging the above inequality in (4) we get that the number of q-tuples belonging to $\mathcal{C}$ which span a copy of B is $\left(IC(B, R) \pm \frac{3}{4}\delta \right) \binom{r}{q} \left(\frac{n}{r}\right)^q$. Observe that $|\mathcal{C}| = \binom{r}{q} \left(\frac{n}{r}\right)^q$. Therefore the number of those q-tuples is $\left(IC(B, R) \pm \frac{3}{4}\delta \right) |\mathcal{C}|$, as required. ∎

Proof (of Lemma 2.13): Put $t = t_{2.13}(\delta, q, k) = t_{4.4}(k^{-\binom{q}{2}}\delta, q)$, $\gamma = \gamma_{2.13}(\delta, q, k) = \gamma_{4.4}(k^{-\binom{q}{2}}\delta, q)$. Let R be a regularity instance of order at least t and regularity parameter at most γ and let G be a k-colored graph satisfying R. Let $B \in \mathcal{B}$. By our choice of γ and t via Lemma 4.4, the number of copies of B in G is $(IC(B, R) \pm k^{-\binom{q}{2}}\delta)\binom{n}{q}$. Clearly $|\mathcal{B}| \leq k^{\binom{q}{2}}$, so the number of copies of graphs $B \in \mathcal{B}$ in G is

$$\sum_{B \in \mathcal{B}} \left(\left(IC(B, R) \pm k^{-\binom{q}{2}}\delta \right) \binom{n}{q} \right) = \left(\sum_{B \in \mathcal{B}} IC(B, R) \pm \delta \right) \binom{n}{q}$$

∎

4.2 Proof of Lemma 2.14

We will derive Lemma 2.14 from the following lemma.

Lemma 4.5. *For every δ and q there is $\lambda = \lambda_{4.5}(\delta, q)$ such that the following holds: Let R and M be k-colored regularity instances of order r, and densities $\{\eta_{i,j}^\ell : 1 \leq i < j \leq r, \ 1 \leq \ell \leq k\}$ and $\{\mu_{i,j}^\ell : 1 \leq i < j \leq r, \ 1 \leq \ell \leq k\}$ respectively. Let B be a k-colored graph on the vertices $[q]$. Suppose that $\mu_{i,j}^\ell = \eta_{i,j}^\ell \pm \lambda$. Then*

$$|IC(B, R) - IC(B, M)| \leq \delta$$

Proof: Put

$$\lambda = \lambda_{4.5} = (\delta, q) = \frac{\delta}{2^{\binom{q}{2}} q!}.$$

Let R, M be k-colored regularity instances as in the statement. Let $A = \{x_1, ..., x_q\} \subseteq \{1, ..., r\}$, and put

$$W_R(A) = \{\eta_{x_i, x_j}^\ell \ : \ 1 \leq i < j \leq q, 1 \leq \ell \leq k\}$$

and

$$W_M(A) = \{\mu_{x_i, x_j}^\ell \ : \ 1 \leq i < j \leq q, 1 \leq \ell \leq k\} .$$

Denote the color of (i,j) in B by $c(i,j)$. Let $\sigma : [q] \to [q]$ be a permutation. By Definition 2.9 we have

$$|IC(B, W_R(A), \sigma) - IC(B, W_M(A), \sigma)| = \left| \prod_{1 \leq i < j \leq q} \eta_{x_i, x_j}^{c(\sigma(i), \sigma(j))} - \prod_{1 \leq i < j \leq q} \mu_{x_i, x_j}^{c(\sigma(i), \sigma(j))} \right| =$$

$$\left| \prod_{1 \leq i < j \leq q} \eta_{x_i, x_j}^{c(\sigma(i), \sigma(j))} - \prod_{1 \leq i < j \leq q} \left(\eta_{x_i, x_j}^{c(\sigma(i), \sigma(j))} \pm \lambda \right) \right|$$

Opening the parentheses in the above product gives $2^{\binom{q}{2}} - 1$ summands, all of which are multiples of $\pm \lambda$. Therefore $|IC(B, W_R(A), \sigma) - IC(B, W_M(A), \sigma)| \leq \lambda 2^{\binom{q}{2}}$. By Definition 2.10, the triangle inequality and our choice of γ we have

$$
\begin{aligned}
|IC(B, W_R(A)) - IC(B, W_M(A))| &= \left| \frac{1}{Aut(B)} \sum_\sigma (IC(B, W_R(A), \sigma) - IC(B, W_M(A), \sigma)) \right| \\
&\leq q! \lambda 2^{\binom{q}{2}} = \delta \ .
\end{aligned}
$$

By Definition 2.11 and the triangle inequality we have

$$|IC(B, R) - IC(B, M)| = \left| \frac{1}{\binom{r}{q}} \sum_{A \subseteq \{1, \ldots, r\}, \ |A| = q} (IC(B, W_R(A)) - IC(B, W_M(A))) \right| \leq \delta \ ,$$

as needed. ∎

Proof (of Lemma 2.14): Put $\lambda = \lambda_{2.14}(\delta, q, k) = \lambda_{4.5}(k^{-\binom{q}{2}} \delta, q)$. By the choice of λ via Lemma 4.5 we get that $|IC(B, R) - IC(B, M)| \leq k^{-\binom{q}{2}} \delta$ for every $B \in \mathcal{B}$. Since $|\mathcal{B}| \leq k^{\binom{q}{2}}$ the triangle inequality thus gives (1). ∎

4.3 Proof of Lemma 2.16

We will derive Lemma 2.16 from the following lemma.

Lemma 4.6. *Let U, V be disjoint vertex sets in a graph satisfying $|U| = |V| = n \geq n_{4.6}(\gamma, k)$. Suppose the bipartite graph (U, V) is γ-regular with $d(U, V) = \mu$. Let $\{\eta^\ell : 1 \leq \ell \leq k\}$ be nonnegative reals satisfying $\sum_{\ell=1}^m \eta^\ell = \mu$ and $\sum_{\ell=m+1}^k \eta^\ell = 1 - \mu$. Then there is a (k, m)-coloring of (U, V) such that the resulting k-colored graph is 2γ-regular and satisfies $d^\ell(U, V) = \eta^\ell$ for every $1 \leq \ell \leq k$.*

For the proof of Lemma 4.6 we need the following standard Chernoff-type inequality:

Lemma 4.7. *Suppose $X_1, \ldots, X_m$ are independent Boolean random variables and $\mathbb{P}(X_i = 1) = p_i$. Let $E = \sum_{i=1}^m p_i$. Then $\mathbb{P}(|\sum_{i=1}^m X_i - E| \geq \delta m) \leq 2e^{-2\delta^2 m}$.*

Proof (of Lemma 4.6): We will show that the edges between U and V can be colored with colors $1, ..., m$ in a way that satisfies the requirements. The same argument can be used to color the non-edges with colors $m+1, ..., k$. First assume that $\mu \leq \gamma$. If this is the case, just color any $\eta^\ell n^2$ of the edges between U and V with color ℓ, for every $1 \leq \ell \leq m$. This way we made sure that $d^\ell(U, V) = \eta^\ell$. Let $U' \subseteq U, V' \subseteq V$ with $|U'|, |V'| \geq 2\gamma n$. Before the coloring we had $d(U', V') = \mu \pm \gamma$. Therefore after the coloring we have $0 \leq d^\ell(U', V') \leq \mu + \gamma$, so $|d^\ell(U', V') - \eta^\ell| \leq \mu + \gamma \leq 2\gamma$.

Assume from this point on that $\mu \geq \gamma$. For every edge $e \in E(U, V)$ roll a die with sides $1, ..., m$ so that probability of side ℓ is $\frac{\eta^\ell}{\mu}$. If the die falls on side ℓ then color e with color ℓ. Then the expected number of edges of color ℓ is $\eta^\ell n^2$. By Lemma 4.7, the probability that the number of edges colored by ℓ deviates from its expectation by more than $n^{\frac{3}{2}}$ is at most $2e^{-2n/\mu} \leq 2e^{-2n}$. If $n \geq n_{4.6}(\gamma, k)$ then this probability is less than $\frac{1}{4k}$. This means, by a union bound, that with probability at least $\frac{3}{4}$ the number of edges colored by ℓ is $\eta^\ell n^2 \pm n^{\frac{3}{2}}$ for every $1 \leq \ell \leq m$.

Claim 4.8. *With probability at least $\frac{3}{4}$ all sets $U' \subseteq U, V' \subseteq V$ such that $|U'|, |V'| \geq 2\gamma n$ satisfy $d^\ell(U', V') = \eta^\ell \pm \frac{3\gamma}{2}$*

Proof: Let $U' \subseteq U, V' \subseteq V$ such that $|U'|, |V'| \geq 2\gamma n$. The density of edges between U' and V' before the coloring is $\mu \pm \gamma$. Therefore the expected density of edges with color ℓ is $\frac{\eta^\ell}{\mu}(\mu \pm \gamma) = \eta^\ell \pm \gamma$. So it is enough to show that with probability at least $\frac{3}{4}$, there are no sets U', V' and color $1 \leq \ell \leq m$ such that the density of edges of color ℓ between U' and V' deviates from its expectation by more than $\frac{\gamma}{2}$.

By Lemma 4.7, the probability that the density of edges of color ℓ between U' and V' deviates from its expected value by more than $\frac{\gamma}{2}$ is at most $2e^{-2(\frac{\gamma}{2})^2|U'||V'|/d(U', V')} \leq 2e^{-\gamma^2|U'||V'|/2}$. We assumed that $|U'|, |V'| \geq 2\gamma n$, so this probability is at most $2e^{-\gamma^4 n^2/2}$. The number of choices of sets U', V' as above is at most 2^{2n}, and the number of colors is at most k, so by a union bound we get: The probability that there are sets U', V' and a color ℓ, such that the density of edges between U', V' with color ℓ deviates from its expectation by more than $\frac{\gamma}{2}$ is at most $k2^{2n}2e^{-\gamma^4 n^2/2}$. This expression is less than $\frac{1}{4}$ if n is large enough, namely $n \geq n_{4.6}(\gamma, k)$. ∎

Getting back to the proof of Lemma 4.6 we see that so far we proved that with probability at least $\frac{1}{2}$ the following two conditions hold:

1. $d^\ell(U, V) = \eta^\ell \pm n^{-\frac{1}{2}}$ for every $1 \leq \ell \leq m$.

2. $d^\ell(U', V') = \eta^\ell \pm \frac{3\gamma}{2}$ for every $1 \leq \ell \leq m$ and every two sets $U' \subseteq U, V' \subseteq V$ of size at least $2\gamma n$.

Now take a coloring that satisfies conditions 1 and 2. Let us write $d^\ell(U, V) = \eta^\ell + \epsilon^\ell$ where $|\epsilon^\ell| \leq n^{-\frac{1}{2}}$. Observe that

$$\sum_{\ell=1}^{m} \left(\eta^\ell + \epsilon^\ell\right) = \sum_{\ell=1}^{m} d^\ell(U, V) = \mu = \sum_{\ell=1}^{m} \eta^\ell$$

Therefore $\sum_{\ell=1}^{m} \epsilon^\ell = 0$. We can change the colors of at most $mn^{\frac{3}{2}} \leq kn^{\frac{3}{2}}$ edges to make sure that $d^\ell(U,V)$ is exactly η^ℓ. For every $U' \subseteq U, V' \subseteq V$ of size at least $2\gamma n$ this final change changes $d^\ell(U',V')$ by at most $\frac{k}{(2\gamma)^2 n^{\frac{1}{2}}}$ which is less than $\frac{\gamma}{2}$ if $n \geq n_{4.6}(\gamma,k)$. So in the end we have $d^\ell(U',V') = \eta^\ell \pm 2\gamma$ as required. ∎

Remark 4.9. *In fact, we could have proved the following stronger lemma: Let U,V be disjoint sets in a graph, $|U| = |V| = n \geq n_{4.6}(\gamma,\epsilon,k)$. Suppose U,V is γ-regular and $d(A,B) = \mu$, and let $\{\eta^\ell : 1 \leq \ell \leq k\}$ be nonnegative numbers satisfying $\sum_{l=1}^{m} \eta^\ell = \mu$ and $\sum_{l=m+1}^{k} \eta^\ell = 1 - \mu$. Then there is a (k,m)-coloring of U,V such that the resulting k-colored graph is $\gamma(1+\epsilon)$-regular and satisfies $d^\ell(U,V) = \eta^\ell$. The choice of 2γ in Lemma 4.6 and in Definition 2.15 is for convenience.*

Proof (of Lemma 2.16): Put $n_{2.16}(\gamma,t,k) = t \cdot n_{4.6}(\gamma,k)$. Let R be a graph regularity instance of order $r \leq t$ and regularity parameter γ. Let R' be a (k,m)-chopping of R' and let $\{\eta^\ell_{i,j} : 1 \leq i < j \leq r, \ 1 \leq \ell \leq k\}$ be the densities of R'. Let G be a graph with at least $n_{2.16}(\gamma,t,k)$ vertices that satisfies R. Let $V_1,...,V_r$ be a γ-regular equipartition of $V(G)$ that corresponds to R. For every $1 \leq i \leq r$ we have $|V_i| \geq \frac{n_{2.16}(\gamma,t,k)}{r} \geq n_{4.6}(\gamma,k)$. If $(i,j) \notin \bar{R}$ apply Lemma 4.6 for V_i,V_j and $\{\eta^1_{i,j},...,\eta^k_{i,j}\}$. Color the rest of the edges and non-edges arbitrarily. The resulting k-colored graph satisfies R'. ∎

References

[1] N. Alon, E. Fischer, M. Krivelevich and M. Szegedy, Efficient testing of large graphs, Combinatorica 20 (2000), 451-476.

[2] N. Alon, E. Fischer, I. Newman and A. Shapira, A combinatorial charactarization of testable graph properties - it is all about regularity, SIAM J. on Computing 39 (2009), 143-167.

[3] E. Fischer and I. Newman, Testing versus estimation of graph properties, SIAM J. on Computing 37 (2007), 482-501.

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

[5] L. Lovász and K. Vesztergombi, Nondeterministic graph property testing, manuscript, 2012.

[6] L. Lovász, B. Szegedy: Testing properties of graphs and functions, Israel J. Math. 178 (2010), 113156.

[7] E. Fischer, The difficulty of testing for isomorphism against a graph that is given in advance, SIAM Journal on Computing 34 (2005), 1147-1158.

[8] O. Goldreich (ed), Property Testing: Current Research and Suerveys, LNCS 6390, Springer, 2010.

[9] O. Goldreich, S. Goldwasser and D. Ron, Property testing and its connection to learning and approximation, J. ACM 45 (1998), 653-750.

[10] O. Goldreich and L. Trevisan, Three theorems regarding testing graph properties, Random Structures and Algorithms 23 (2003), 23-57.

[11] R. Rubinfeld and A. Shapira, Sublinear time algorithms, SIAM J. on Discrete Math 25 (2011), 1562-1588.

[12] 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.