AN OPTIMAL ALGORITHM FOR CHECKING REGULARITY *

Y. KOHAYAKAWA†, V. RÖDL‡, AND L. THOMA§

Abstract. We present a deterministic algorithm $\mathcal{A}$ that, in $O(m^2)$ time, verifies whether a given m by m bipartite graph G is *regular*, in the sense of Szemerédi [E. Szemerédi, *Regular partitions of graphs*, Problèmes Combinatoires et Théorie des Graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976) (Paris), Colloques Internationaux CNRS n. 260, 1978, pp. 399–401]. In the case in which G is *not* regular enough, our algorithm outputs a *witness* to this irregularity. Algorithm $\mathcal{A}$ may be used as a subroutine in an algorithm that finds an ε-regular partition of a given n-vertex graph Γ in time $O(n^2)$. This time complexity is optimal, up to a constant factor, and improves upon the bound $O(M(n))$, proved by Alon, Duke, Lefmann, Rödl, and Yuster [N. Alon, R. A. Duke, H. Lefmann, V. Rödl, and R. Yuster, *The algorithmic aspects of the regularity lemma*, Journal of Algorithms, 16(1) (1994), pp. 80–109], where $M(n) = O(n^{2.376})$ is the time required to square a 0–1 matrix over the integers.

Our approach is elementary, except that it makes use of linear-sized expanders to accomplish a suitable form of deterministic sampling.

Key words. Szemerédi's regularity lemma, quasi-randomness, deterministic sampling, expander graphs, regular pairs

AMS subject classifications. 05C35, 05C85, 05C99, 68R10

1. Introduction and the main result.

Szemerédi's regularity lemma [31] is a fundamental result in graph theory (see [25] for an excellent survey). Roughly speaking, this lemma states that any graph admits a partition of its vertex set so that most pairs induce 'pseudorandom' or *regular* bipartite graphs. The original proof of the regularity lemma was non-constructive, but Alon, Duke, Lefmann, Rödl, and Yuster [1, 2] succeeded in developing a fast deterministic algorithm for finding such a partition. Many of the existential results based on the regularity lemma could then be turned into algorithmic results. The algorithm in [1, 2] finds a regular partition of an n-vertex graph in $O(M(n))$ deterministic time, where $M(n) = O(n^{2.376})$ (see [11]) is the time required to square a 0–1 matrix over the integers. More recently, Frieze and Kannan [18] (see also [19]) showed that *sampling* can be used to develop a $O(n)$ time *randomized* algorithm that, given an n-vertex graph G, outputs a partition for G that is regular with high probability.

In both algorithms above (and in all algorithms for variants of the regularity lemma), the main algorithmic problem is to decide whether a given m by m bipartite graph G is regular; if G is *not* regular, we are required to find a 'witness' for this irregularity. In this paper, we present a *deterministic* algorithm that solves this problem in $O(m^2)$ time. Given our algorithm, one can derive in a standard way an algorithm for Szemerédi's regularity lemma that finds a regular partition of an n-

*The first author was partially supported by MCT/CNPq through ProNEx Programme (Proc. CNPq 664107/1997–4), by CNPq (Proc. 300334/93–1 and 468516/2000–0), and by FAPESP (Proc. 96/04505–2). The second author was partially supported by NSF Grant 0071261. The third author was partially supported by NSF grant DMS–0301228. The collaboration of these authors is supported by a CNPq/NSF INT–0072064 cooperative grant.
Some of the results in this paper appeared in an extended abstract in the proceedings of SODA 2002.

†Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão 1010, 05508–090 São Paulo, SP, Brazil (yoshi@ime.usp.br).

‡Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, USA (rodl@mathcs.emory.edu).

§Department of Mathematics, University of Rhode Island, Kingston, RI 02881, USA (thoma@math.uri.edu).

vertex graph in time $O(n^2)$. A key feature of our approach lies in the use of linear-sized expanders for carrying out a certain procedure that may be thought of as *deterministic sampling*.

1.1. The main result. Let $G = (A, B; E)$ be a bipartite graph. If $\emptyset \neq U \subset A$, $\emptyset \neq V \subset B$, the *density* of (U, V) in G is $d(U, V) = e(U, V)/|U||V|$, where we write $e(U, V) = e_G(U, V)$ for the number of edges with one endpoint in U and the other endpoint in V. For $\varepsilon > 0$, we say that G is *ε-regular* if, for all $U \subset A$, $|U| \geq \varepsilon|A|$, and $V \subset B$, $|V| \geq \varepsilon|B|$, we have

$$|d(U, V) - d(A, B)| \leq \varepsilon \ . \tag{1.1}$$

In case G is not ε-regular and a certain pair (U, V) certifies this fact, then we say that (U, V) is a *witness* to the *ε-irregularity* of G.

Let $\Gamma = (V, E)$ be a graph. A partition $(V_i)_{i=0}^k$ of the vertex set V, $V = \bigcup_{i=0}^k V_i$, is said to be an *equitable partition* (with the exceptional class V_0) if $|V_1| = \cdots = |V_k|$. If $V = \bigcup_{i=0}^k V_i$ is an equitable partition of V such that the size of the exceptional class $|V_0| \leq \varepsilon n$ and at least $(1-\varepsilon)\binom{k}{2}$ pairs (V_i, V_j), where $1 \leq i < j \leq k$, are ε-regular, we say that the partition $(V_i)_{i=0}^k$ is an *ε-regular partition.* We say that a pair (U, W) is *ε-regular* if the bipartite graph induced by (U, W) is ε-regular.

Szemerédi's remarkable result may be stated as follows.

Theorem 1.1. *For any $\varepsilon > 0$ and any $k_0 \geq 1$, there is $K_0(\varepsilon, k_0)$ such that any graph Γ admits an ε-regular partition into k parts for some k satisfying $k_0 \leq k \leq K_0(\varepsilon, k_0)$.*

Alon, Duke, Lefmann, Rödl, and Yuster [1, 2] proved the following algorithmic version of Theorem 1.1.

Theorem 1.2. *There is a deterministic algorithm $\mathcal{A}_0$ that, given $\varepsilon > 0$, $k_0 \geq 1$, and Γ, produces an ε-regular partition for Γ into k parts for some k satisfying $k_0 \leq k \leq K_0'$, where $K_0' = K_0'(\varepsilon, k_0)$ depends only on ε and k_0. Moreover, algorithm $\mathcal{A}_0$ runs in time $O(M(n)) = O(n^{2.376})$ if Γ has n vertices.*

Consider now the following closely related decision problem.

Problem 1.3. Given a graph G, a pair (U, W) of non-empty, pairwise disjoint sets of vertices of G, and a positive ε, decide whether (U, W) is ε-regular with respect to G.

As it turns out, *the problem above is coNP-complete* [1, 2]. However, as observed already in [1, 2], to prove Theorem 1.2, it suffices to solve an *approximate* version of the decision problem above. For instance, the following result [15] suffices.

Theorem 1.4. *There exists an algorithm $\mathcal{A}_1$ for which the following holds. When $\mathcal{A}_1$ receives as input an $\varepsilon > 0$ and a bipartite graph $G = (A, B; E)$ with $|A| = |B| = m \geq (2/\varepsilon)^5$, it either correctly asserts that G is ε-regular, or else it returns a witness for the ε'-irregularity of G, where $\varepsilon' = \varepsilon'_{\mathcal{A}_1}(\varepsilon) = \varepsilon^5/16$. The running time of $\mathcal{A}_1$ is $O(M(m)) = O(m^{2.376})$.*

(See Frieze and Kannan [20] for a somewhat different approach to verifying regularity, based on singular values of matrices.) Our main result is an improvement of Theorem 1.4 above, and may be stated as follows.

Theorem 1.5 (The main result). *There exists an algorithm $\mathcal{A}$ for which the following holds. When $\mathcal{A}$ receives as input an $\varepsilon > 0$ and a bipartite graph $G = (A, B; E)$ with $|A| = |B| = m \geq m_0(\varepsilon)$, it either correctly asserts that G is ε-regular, or else it returns a witness for the ε'-irregularity of G, where $\varepsilon' = \varepsilon'_{\mathcal{A}}(\varepsilon) = \varepsilon^{20}/10^{24}$. The running time of $\mathcal{A}$ is $O(m^2)$.*

We describe our algorithm $\mathcal{A}$ in §3.1. Deriving an algorithm for the regularity lemma from Theorem 1.5 is standard (cf. §3.2).

COROLLARY 1.6. *There is a deterministic algorithm $\mathcal{A}'_0$ that, given $\varepsilon > 0$, $k_0 \geq 1$, and a graph Γ, produces an ε-regular partition for Γ into k parts for some k satisfying $k_0 \leq k \leq K''_0$, where $K''_0 = K''_0(\varepsilon, k_0)$ depends only on ε and k_0. Moreover, algorithm $\mathcal{A}'_0$ runs in time $O(n^2)$ if Γ has n vertices.*

Clearly, Algorithm $\mathcal{A}'_0$ above has optimal time complexity, up to the constant implicit in the big-O notation. In [1, 2], several algorithmic consequences are derived from Theorem 1.2. In the examples presented there, the time complexity of the algorithms is $O(M(n))$. *Using $\mathcal{A}'_0$ from Corollary 1.6, one obtains algorithms with optimal time complexity $O(n^2)$.* We also observe that a similar improvement may be obtained from Theorem 1.5 for the subgraph counting algorithm given in Duke, Lefmann, and Rödl [15].

Let us also mention that an important variant of the regularity lemma, suitable for finding *induced subgraphs*, was recently discovered by Alon, Fischer, Krivelevich, and Szegedy [3, 4], in the context of *property testing* (see, e.g., [21] and [22, 23]). In the applications of their regularity lemma in [3, 4] the authors do not need algorithms for finding their regular partitions; however, they observe that an algorithmic version of their lemma readily follows from results such as Theorem 1.4. Again, an $O(n^2)$ time algorithm follows immediately from Theorem 1.5.

Finally, we mention that one may prove a 'non-bipartite version' of Theorem 1.5. This variant of our result implies that one may check in time $O(n^2)$ whether a given n-vertex graph Γ is *quasi-random*, in the sense of Chung, Graham, and Wilson [10]. Moreover, if Γ is not quasi-random, then our algorithm will produce a suitable witness proving this, i.e., an induced subgraph with $\Omega(n)$ vertices whose density deviates substantially from the density of Γ (see Section 1.3.5 for more details).

1.2. Local conditions for regularity. One may prove Theorem 1.4 by considering a certain 'local condition' on $G = (A, B; E)$ that is essentially equivalent to the regularity of G. For simplicity, let us suppose that G is degree-regular. The condition is simply that the following inequality should hold:

$$\sum_{x,\, y \in A} \left| d_G(x, y) - p(G)^2 m \right| \leq \delta p(G)^2 m^3, \tag{1.2}$$

where $d_G(x, y) = |N_G(x) \cap N_G(y)|$ is the so-called *codegree* of x and y, and $p(G) = |E|/|A||B| = |E|/m^2$, with $m = |A| = |B|$, is the density of G. Clearly, inequality (1.2) may be checked in $O(m^3)$ time, and, in fact, using fast matrix multiplication, one may verify (1.2) in $O(M(m)) = O(m^{2.376})$ time. The precise meaning of the equivalence of the ε-regularity of G and the validity of (1.2) is as follows: for all $\varepsilon > 0$ there is $\delta > 0$ such that if (1.2) holds, then G is ε-regular. Moreover, for all $\delta > 0$, there is $\varepsilon' > 0$ such that if (1.2) fails then G is not ε'-regular, and, in fact, a witness to this ε'-irregularity may be constructed explicitly in the same deterministic time. Some of the ideas described in this paragraph have appeared in the literature under many guises. (For a detailed discussion on the combinatorial aspects, see [24]; for applications of these ideas in theoretical computer science, see [28] and the references therein.) Basically, we are obtaining a somewhat surprising amount of information from 'pairwise independence'. We do not go into the details here.

The key idea in the proof of Theorem 1.5 is that we may restrict the sum in (1.2) to a *small*, *randomly selected* collection of pairs $\{x, y\}$ (and, naturally, scale down the right-hand side). This would not be so satisfactory, as we would have a randomized procedure: *we in fact show that we may achieve the same effect by 'deterministic sampling', using the edge set of a linear-sized expander J* (see the definition of property $\mathcal{P}(J,\delta)$ in §3.1.2).

1.3. Algorithmic applications. As mentioned above, Algorithms $\mathcal{A}$ and $\mathcal{A}_0'$ immediately imply improvements on deterministic algorithms that are based on Szemerédi's regularity lemma. Here we present a few typical examples of such algorithms. For more algorithmic applications of the regularity lemma see [2] and [15].

1.3.1. MAXCUT **in dense graphs.** There has been considerable interest in the following computational problem recently.

Problem 1.7 (MAXCUT). Given a graph G, find a partition (U,W) of the vertex set of G so that the number of edges $e(U,W)$ between U and W is maximum.

It follows from the algorithmic version of the regularity lemma that one may design a polynomial time approximation scheme for MAXCUT if the input graphs G are restricted to dense graphs. Let us be more precise.

Let α be a fixed positive real. In this section, we only consider graphs G with edge density $e(G)\binom{|V(G)|}{2}^{-1}\geq\alpha$. Theorem 1.2 implies the following result: for any ε and $\alpha>0$, there exist a constant $C(\varepsilon,\alpha)$ and a *deterministic* algorithm $\mathcal{A}_{\mathrm{MC}}$ so that, given an n-vertex graph G with edge density $\geq\alpha$, algorithm $\mathcal{A}_{\mathrm{MC}}$ returns a solution (U',W') for MAXCUT such that

$$e(U',W')\geq(1-\varepsilon)e(U^*,W^*), \tag{1.3}$$

where (U^*,W^*) is an optimal solution for G. Furthermore, the running time of $\mathcal{A}_{\mathrm{MC}}$ is $\leq C(\varepsilon,\alpha)M(n)$.

Algorithm $\mathcal{A}_{\mathrm{MC}}$ uses Algorithm $\mathcal{A}_0$ in Theorem 1.2 as a subroutine; we may use, instead, Algorithm $\mathcal{A}_0'$ in Corollary 1.6: let $\mathcal{A}_{\mathrm{MC}}'$ be the corresponding algorithm.

Theorem 1.8. *On input G as above, the deterministic algorithm $\mathcal{A}_{\mathrm{MC}}'$ produces a partition (U,W) satisfying (1.3) in time $\leq C'(\varepsilon,\alpha)n^2$, where $C'(\varepsilon,\alpha)$ is a constant that depends only on ε and α.*

We remark that a *randomized* algorithm with time complexity $O(n^2)$ was already given by de la Vega [14]. For related results concerning randomized algorithms, the reader is referred to Frieze and Kannan [18, 19].

1.3.2. The quasi-Ramsey number and maximum acyclic subgraphs. Let $f\colon E(K_n)\to\{-1,1\}$ be a function and set $f(S)=\sum_{e\in\binom{S}{2}}f(e)$ where $S\subseteq[n]$. Here, as usual, K_n stands for the complete graph on n vertices and $\binom{S}{2}$ denotes the set of all pairs on the set S.

The *quasi-Ramsey number* $g(n)$ is defined as

$$g(n)=\min_f\max_{S\subseteq[n]}|f(S)|\ .$$

Erdős and Spencer [17] showed that

$$c_1n^{3/2}\leq g(n)\leq c_2n^{3/2},$$

for some absolute constants c_1 and $c_2>0$.

Let T_n be a tournament and P_n a transitive tournament both on n vertices. Set $|T_n\cap P_n|$ to be the number of common oriented arcs of T_n and P_n. The tournament ranking function $h(n)$ is defined by

$$h(n)=\min_{T_n}\max_{P_n}|T_n\cap P_n|\ ,$$

i.e., $h(n)$ is the maximum number of edges one can choose in any tournament of order n without creating an oriented cycle. Spencer [29, 30] showed that

$$c_1 n^{3/2} \leq h(n) - \frac{1}{2}\binom{n}{2} \leq c_2 n^{3/2},$$

for some absolute constants c_1 and $c_2 > 0$.

A polynomial time approximation scheme (PTAS) for a maximization problem is a family of algorithms $\{\mathcal{S}_\varrho \colon 0 < \varrho < 1\}$ as follows. For any given $0 < \varrho < 1$, algorithm $\mathcal{S}_\varrho$ runs in polynomial time and finds a solution whose value is at least $(1-\varrho)\,\mathrm{OPT}$, where OPT in the optimal value. Using a constructive version of the regularity lemma Czygrinow, Poljak, and Rödl [13, Theorem 3] designed a PTAS for the 'dense' quasi-Ramsey problem and for tournament ranking.

For $f : E(K_n) \to \{-1, 1\}$ set $\mathrm{OPT}(f) = \max_{S \subseteq [n]} |f(S)|$. Our algorithm for the regularity lemma implies an improvement on the time complexity of the PTAS designed in [13].

THEOREM 1.9. *Let $c > 0$ be fixed. For every $0 < \varrho < 1$, there is a $O(n^2)$ time algorithm that constructs a set S such that*

$$|f(S)| \geq (1-\varrho)\,\mathrm{OPT}(f),$$

for any instance $f \colon E(K_n) \to \{-1, 1\}$ with $\mathrm{OPT}(f) \geq cn^2$.

Now, let $\mathrm{OPT}(T_n) = \max_{P_n} |T_n \cap P_n|$ where T_n is a tournament. Algorithm $\mathcal{A}'_0$ in Corollary 1.6 improves the time complexity of the PTAS designed in [13] to $O(n^2)$.

THEOREM 1.10. *Let $0 < \varrho < 1$. Then there is a $O(n^2)$ time algorithm that, given a tournament T_n, constructs an ordering σ of the vertices of T_n so that at least $(1-\varrho)\,\mathrm{OPT}(T_n)$ arcs agree with σ.*

1.3.3. Robustly high-chromatic graphs. Goldreich, Goldwasser, and Ron [22, 23] have recently initiated a systematic study of *property testing* for combinatorial structures. Roughly speaking, in property testing, one has a property P of interest, and one is given an object X and a real number $\varepsilon > 0$. The task is then to decide whether X has P or it is ε-far from any object Y having P (we suppose our objects are in some metric space). Furthermore, we wish to perform this test extremely quickly; typically, the tests examine a small random portion of X and distinguish between the two cases above with high probability of success. Thus, in an appropriate computational model, the tests have sublinear complexity (see [22, 23] for details).

A graph property P that has been proved to be testable [22, 23] is the property of having chromatic number at least k, for any fixed k. This result was in fact implicit in [16], where the regularity lemma is used to prove that 'robustly high-chromatic graphs' admit witnesses of bounded size. Indeed, the existential result in Theorem 1.11 below was proved in [16]. The algorithmic result in Theorem 1.11, but with time complexity $O(M(n))$, was proved in [1, 2].

THEOREM 1.11. *Let $k \geq 3$ be an integer and let $\varepsilon > 0$ be a real constant. Then there exist integers $n_0 = n_0(k, \varepsilon)$ and $f = f(k, \varepsilon)$ and a constant $\nu = \nu(k, \varepsilon) > 0$ such that if $G = (V, E)$ is a graph with $n \geq n_0$ vertices, then either*

(i) there exists a graph H on $h \leq f$ vertices with chromatic number $\chi(H) \geq k$ that occurs in G at least νn^h times as a subgraph, or else

(ii) there exists a set $E' \subseteq E$ with $|E'| \leq \varepsilon n^2$ such that the subgraph $G' = (V, E \setminus E')$ satisfies $\chi(G') < k$.

Furthermore, there is a deterministic algorithm that receives as input a graph $G = (V, E)$ as above and, in time $O(n^2)$, outputs either a graph H as in (i), or else it outputs a set of edges E' as in (ii), together with a proper coloring $\Delta\colon V \to \{1, \dots, k-1\}$ of the subgraph G'.

The approach in [22, 23] does not use the regularity lemma, and implies the existential part of Theorem 1.11. Moreover that approach also gives a randomized, polynomial time algorithm for the constructive part of Theorem 1.11.

Finally, we mention that Czumaj and Sohler [12] have recently proved that the property of having chromatic number at least k is also testable for hypergraphs.

1.3.4. Counting subgraphs. In this section, we describe an algorithm for approximately counting small subgraphs in large graphs. This algorithm will also be an application of Algorithm $\mathcal{A}_0'$ from Corollary 1.6.

We need to introduce some notation. We shall follow [15]. Let $G = (V, E)$ be a graph on n vertices whose vertex set $V = \{v_1, \dots, v_n\}$ is ordered by $v_1 < \dots < v_n$. Let the set $W = \{w_1, \dots, w_k\}$ be ordered by $w_1 < \dots < w_k$. We say that a graph H with vertex set W is *order isomorphic* to an induced subgraph H' of G if there exists an isomorphism $\phi : H \to H'$ with the property that for each i and j, if $w_i < w_j$, then $\phi(w_i) < \phi(w_j)$. Let $H_1, \dots, H_t$, where $t = 2^{\binom{k}{2}}$, be the list of all graphs on the set W and let $\sigma_k(G) = (h_1, \dots, h_t)$ be the t-dimensional vector in which each h_i is the number of induced subgraphs of G to which H_i is order isomorphic.

The following proposition asserts the existence of a certain type of approximation algorithm for the vector $\sigma_k(G)$. For more details, see [15].

Theorem 1.12. *Let $k \geq 3$ be a fixed integer and suppose $\delta > 0$ is a fixed real. There is an algorithm that, on input G, a labeled, ordered graph on n vertices, produces an approximation $\overline{\sigma}_k(G) = (\overline{h}_1, \dots, \overline{h}_t)$ to the vector $\sigma_k(G) = (h_1, \dots, h_t)$ with the property that*

$$|h_i - \overline{h}_i| \leq \delta \binom{n}{k}$$

for all $1 \leq i \leq t$. This algorithm runs in time $O(n^2)$.

In [15], the authors consider the problem of approximating $\sigma_k(G)$ for $k = k(n)$ slowly increasing functions of n. Our results may be used to improve on the time complexity of the algorithms given in [15] for such $k = k(n)$, but we shall not go into the details.

1.3.5. Checking quasi-randomness. Thomason [32] and Chung, Graham, and Wilson [10] initiated a systematic study of *quasi-random* properties of graphs: these are properties that are shared by almost all graphs, and are in fact *deterministically* asymptotically equivalent, i.e., if a large graph has one of these properties, then it in fact has all of them.

The investigation of quasi-randomness in combinatorics turned out to be a very rich line of research, as shows the series of papers by Chung and Graham on the subject (see [9] for recent developments, and the references therein). Besides graphs, other combinatorial structures such as tournaments, set-systems, and subsets of $\mathbb{Z}/n\mathbb{Z}$, have been studied from this perspective (see [6, 7, 8].) Finally, we mention that applications of some of the underlying ideas in this area have occurred in the literature in different contexts; the interested reader is referred to [5, Chapter 9] and [24].

In this section, we shall consider the computational problem of determining whether or not a given graph is quasi-random. We are also interested in an additional requirement: in the case in which the input graph is *not* quasi-random, a 'witness' to certify this fact should be efficiently produced.

We shall use the following definition.

DEFINITION 1.13. *Let reals $0 < \varepsilon \le 1$ and $0 < \delta \le 1$ be given. We shall say that a graph G is $(1/2, \varepsilon, \delta)$-quasi-random if, for all U, $W \subset V(G)$ with $U \cap W = \emptyset$ and $|U|$, $|W| \ge \delta n$, we have*

$$\left| e_G(U, W) - \frac{1}{2}|U||W| \right| \le \frac{1}{2}\varepsilon |U||W|.$$

In [24] the authors consider a new quasi-random property to develop an algorithm for testing quasi-randomness. Let G be a graph on n vertices and let J be a (ϱ, L)-uniform graph on the same vertex set (for the definition, see Section 2). To state the results we need to introduce some notation. For a vertex i of G we set $N(i)$ to be its neighborhood. Further, we write $N(i) \triangle N(j)$ for the symmetric difference of the sets $N(i)$ and $N(j)$.

To decide about the quasi-randomness of G we introduce the following couple of properties. Let $0 < \varepsilon, \delta \le 1$ be real numbers. We say that G satisfies property $\mathcal{T}_\triangle(J, \varepsilon)$ if we have

$$\sum_{\{i,j\} \in E(J)} \left| |N(i) \triangle N(j)| - \frac{1}{2}n \right| \le \frac{1}{2}\varepsilon n e(J) \ .$$

Note that this property is closely related to our property $\mathcal{P}$ introduced below. Similarly, we say that G satisfies property $\mathcal{T}'_\triangle(J, \gamma, \varepsilon)$ if the inequality

$$\left| |N(i) \triangle N(j)| - \frac{1}{2}n \right| \le \frac{1}{2}\varepsilon n$$

fails for at most $\gamma e(J)$ edges $\{i, j\} \in E(J)$.

The following two characterization theorems are proved in [24, Theorems 56, 57].

THEOREM 1.14. *For any $0 < \varepsilon, \delta \le 1$ and any L, there exist $\varepsilon_0 = \varepsilon_0(\varepsilon, \delta, L) > 0$ and $r_0 = r_0(\varepsilon, \delta, L) \ge 1$ for which the following holds. Let G and J be two graphs on the same vertex set of n vertices. Assume further that J is a (ϱ, L)-uniform graph with the average degree $r = \varrho n \ge r_0$. Then, if G satisfies the property $\mathcal{T}_\triangle(J, \varepsilon')$ for some $0 < \varepsilon' \le \varepsilon_0$, then G is $(1/2, \varepsilon, \delta)$-quasi-random.*

THEOREM 1.15. *For any $0 < \gamma, \varepsilon \le 1$ and any L, there exist $\varepsilon_1 = \varepsilon_1(\gamma, \varepsilon, L) > 0$, $\delta_1 = \delta_1(\gamma, \varepsilon, L) > 0$, $r_1 = r_1(\gamma, \varepsilon, L) \ge 1$, and $N_1 = N_1(\gamma, \varepsilon, L) \ge 1$, for which the following holds. Let G and J be two graphs on the same vertex set of $n \ge N_1$ vertices. Assume further that J is a (ϱ, L)-uniform graph with the average degree $r = \varrho n \ge r_1$. Then, if G is $(1/2, \varepsilon', \delta')$-quasi-random for some $0 < \varepsilon' \le \varepsilon_1$ and $0 < \delta' \le \delta_1$, then property $\mathcal{T}'_\triangle(J, \gamma, \varepsilon)$ holds for G.*

It is straightforward to see that the properties $\mathcal{T}$ and $\mathcal{T}'$ can be checked in $O(n^2)$ deterministic time. Moreover, if a graph $\mathcal{G}$ does not satisfy property $\mathcal{T}'_\triangle(J, \gamma, \varepsilon)$, then one can, using the ideas from our present paper, construct a witness for the non-quasi-randomness of G in $O(n^2)$ time.

2. Preliminaries. In this section, we discuss some basic properties and the algorithmic construction of certain very well known random looking graphs.

2.1. (ϱ, L)**-uniformity.** Let $0 < \varrho \le 1$ and $L > 0$ be fixed. We say that a graph J on m vertices is *(ϱ, L)-uniform* if, for any U, $W \subset V(J)$ with $U \cap W = \emptyset$, we have

$$\big|e_J(U,W) - \varrho|U||W|\big| \le L\sqrt{r|U||W|}, \tag{2.1}$$

where $r = \varrho m$. The following lemma is immediate.

LEMMA 2.1. *Let $R = (V,E)$ be a (ϱ, L)-uniform, m-vertex graph and let $\emptyset \neq A \subset V$ be given. Put $J = R[A]$. Then J is an (ϱ, L')-uniform graph with $L' = L\sqrt{m/|A|}$.*

NOTATION 2.2. We use the following non-standard notation: we write $O_1(x)$ for any term y such that $|y| \le x$.

We shall need estimates on the number of edges induced on subsets of (ϱ, L)-uniform graphs. Below, if Γ is a graph, we write $e(\Gamma)$ for the number of edges in Γ.

LEMMA 2.3. *Let $J = (V,E)$ be a (ϱ, L)-uniform graph and let $S \subseteq V$ be a nonempty subset of vertices of J. Then*

$$\begin{aligned} e(J[S]) &= \varrho\binom{|S|}{2} + O_1\big(Lr^{1/2}(|S|+1)\big) \\ &= \varrho\frac{|S|^2}{2} + O_1\big(2Lr^{1/2}|S|\big), \end{aligned} \tag{2.2}$$

where $r = \varrho|V|$.

Proof. Put $s = |S|$. Note that, for any $1 \le t < s$, we have $2e(S)\binom{s-2}{t-1} = \sum_T e(T, S\setminus T)$, where the sum is extended over all $T \subset S$ with $|T| = t$. Thus

$$e(S) = \frac{1}{2}\binom{s}{t}\binom{s-2}{t-1}^{-1}\Big\{\varrho|T||S\setminus T| + O_1\big(L\{rt(s-t)\}^{1/2}\big)\Big\}$$

for any $1 \le t < s$. We use this relation with $t = \lfloor s/2 \rfloor$. Note that

$$\binom{s}{\lfloor s/2\rfloor}\binom{s-2}{\lfloor s/2\rfloor - 1}^{-1} = \frac{s(s-1)}{\lfloor s/2\rfloor\lceil s/2\rceil} \le 4,$$

and so

$$e(S) = \varrho\binom{s}{2} + O_1\big(2L\{r\lfloor s/2\rfloor\lceil s/2\rceil\}^{1/2}\big) = \varrho\binom{s}{2} + O_1\big(Lr^{1/2}(s+1)\big),$$

and the result follows. ∎

In what follows, the following simple consequences of Lemma 2.3 will be useful.

LEMMA 2.4. *Let $\eta > 0$ and $L > 0$ be given. Then there is an $\overline{r} = \overline{r}(\eta, L)$ such that any m-vertex (ϱ, L)-uniform graph J with $\varrho m \ge \overline{r}$ has the two properties below.*

(a) If $S \subset V(J)$ is such that $|S| = \nu m \ge \eta m$, then

$$e(J[S]) = (1 + O_1(\eta))\nu^2 e(J)\ . \tag{2.3}$$

(b) If $S \subset V(J)$ is such that $|S| < \eta m$, then

$$e(J[S]) < 2\eta^2 e(J)\ . \tag{2.4}$$

2.2. Auxiliary results on expander graphs. The celebrated Ramanujan graphs of Lubotzky, Phillips, and Sarnak [26, 27] are explicitly constructible examples of linear-sized $(\varrho, 2)$-uniform graphs. We shall make crucial use of their construction.

The Ramanujan graphs $X^{p,q}$ constructed in [26, 27] depend on certain primes p and q, which have to satisfy certain simple arithmetical conditions. The graphs $X^{p,q}$ that we shall be interested in are $(p+1)$-regular and have $q(q^2-1)/2$ vertices. However, we shall need to construct linear-sized $(\varrho, O(1))$-uniform graphs with m vertices and average degree around r, where m and r are arbitrary integers (which we may assume to be large). The main result of this section, Lemma 2.5 below, asserts that this can be done efficiently. As the reader will see, we shall simply check that, given m and r, we may find suitable primes p and q so that an induced subgraph of $X^{p,q}$ will do.

LEMMA 2.5. *There exists an Algorithm $\mathcal{E}$ satisfying the following properties. There is an absolute constant r_1 such that for all $r_0 \geq r_1$ there are constants $m_0 = m_0(r_0)$ and $C_0 = C_0(r_0)$ for which the following holds. Algorithm $\mathcal{E}$ receives as input integers $r_0 \geq r_1$ and $m \geq m_0 = m_0(r_0)$, and returns an adjacency list representation of a particular $(\varrho, 3)$-uniform graph J on m vertices with $r = \varrho m$ satisfying $r_0 \leq r \leq 2r_0$. Furthermore, Algorithm $\mathcal{E}$ runs in time $\leq C_0 m(\log m)^2$.*

In the remainder of this section, we prove Lemma 2.5 above for completeness.

2.2.1. Ramanujan graphs. Before we start with the proof of Lemma 2.5, we recall the construction of Lubotzky, Phillips, and Sarnak [26, 27].

As usual, in what follows, if a is an integer and p is a prime with a not divisible by p, the Legendre symbol $(\frac{a}{p})$ is defined as 1 if a is a quadratic residue modulo p and as -1 if a is a quadratic non-residue modulo p. To describe the construction in [26, 27], let p and q be two unequal primes satisfying

$$p,\, q \equiv 1 \pmod 4, \tag{2.5}$$

and

$$\left(\frac{p}{q}\right) = 1. \tag{2.6}$$

We now let S and T be the following sets. Below, i is an arbitrary fixed integer such that $i^2 \equiv -1 \pmod q$. We let

$$\begin{aligned} S = \big\{(\alpha_0, \alpha_1, \alpha_2, \alpha_3) \in \mathbb{Z}^4 \colon & \alpha_0^2 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2 = p \\ & \text{with } \alpha_0 > 0, \text{ odd, and } \alpha_1, \alpha_2, \alpha_3 \text{ even}\big\} \end{aligned}$$
$$T = \left\{ \begin{pmatrix} \alpha_0 + i\alpha_1 & \alpha_2 + i\alpha_3 \\ -\alpha_2 + i\alpha_3 & \alpha_0 - i\alpha_1 \end{pmatrix} : (\alpha_0, \alpha_1, \alpha_2, \alpha_3) \in S \right\}. \tag{2.7}$$

We now consider $\mathrm{PSL}(2, \mathbb{Z}/q\mathbb{Z})$ (the projective special linear group), which consists of the 2×2 matrices over $\mathbb{Z}/q\mathbb{Z}$ whose determinants are non-zero quadratic residues mod p, quotiented out by the equivalence relation that makes two such matrices equivalent if one is a non-zero scalar multiple of the other.

It will be convenient to observe that each element of $\mathrm{PSL}(2, \mathbb{Z}/q\mathbb{Z})$ (i.e., each equivalence class) may be represented by a matrix whose second row is either $(0, 1)$, or else the second row is $(1, x)$, where x is some arbitrary element of $\mathbb{Z}/q\mathbb{Z}$. The existence of this simple 'canonical representation' for the elements of $\mathrm{PSL}(2, \mathbb{Z}/q\mathbb{Z})$ will be helpful below.

Observe that, if we consider the entries of the matrices in T modulo q, we get 2×2 matrices over $\mathbb{Z}/q\mathbb{Z}$, with determinant $p \pmod q$, which is a non-zero quadratic residue modulo q (cf. (2.6)). By a well known result of Jacobi and some simple arguments, one may check that there are $p+1$ elements in T and that they are all distinct in $\mathrm{PSL}(2,\mathbb{Z}/q\mathbb{Z})$.

The graph $X^{p,q}$ constructed in [26, 27] is the Cayley graph of $\mathrm{PSL}(2,\mathbb{Z}/q\mathbb{Z})$ relative to the set T. The vertices of $X^{p,q}$ are the elements of $\mathrm{PSL}(2,\mathbb{Z}/q\mathbb{Z})$ and the edge set of $X^{p,q}$ is so that $\{x,y\}$ is an edge of $X^{p,q}$ if and only if there is a $t\in T$ such that $x=ty$ (one may check that this is a symmetric relation). A key result concerning the graphs $X^{p,q}$ is the following.

THEOREM 2.6. *The graph $X^{p,q}$ is a non-bipartite $(p+1)$-regular graph on $n=q(q^2-1)/2$ vertices. Moreover, if the eigenvalues of $X^{p,q}$ are $|\lambda_1|\geq\cdots\geq|\lambda_n|$, then $\lambda_1=p+1$ and*

$$|\lambda_j|\leq 2\sqrt{p}\quad\textit{for all}\quad j>1. \tag{2.8}$$

Because of (2.8), the graphs $X^{p,q}$ are called *Ramanujan graphs.* We now state the following well known pseudorandom property of the graphs $X^{p,q}$, which follows from (2.8) (see, e.g., Corollary 9.2.5 in [5]).

COROLLARY 2.7. *The graph $X^{p,q}$ is $(\varrho,2)$-uniform, where $\varrho=(p+1)/n$.*

Having covered the basics of the Lubotzky, Phillips, and Sarnak construction, we turn to the proof of Lemma 2.5.

2.2.2. Proof of Lemma 2.5. We start with a simple lemma asserting the existence of appropriate primes p and q.

LEMMA 2.8. *There exists an absolute constant r_1 such that, for any $r_0\geq r_1$, there exists an integer $m_0=m_0(r_0)$ for which the following holds. There is an Algorithm $\mathcal{P}$ that, on input (r_0,m), where $r_0\geq r_1$ and $m\geq m_0=m_0(r_0)$, produces a pair of primes p and q which satisfy*

$$p\neq q,\quad p,\,q\equiv 1 \pmod 4,\quad\text{and}\quad \left(\frac{p}{q}\right)=1, \tag{2.9}$$

$$1.4r_0\leq p+1\leq 2r_0, \tag{2.10}$$

and

$$\sqrt[3]{2.1m}\leq q\leq 1.1\sqrt[3]{2.1m}. \tag{2.11}$$

Algorithm $\mathcal{P}$ runs in time $\leq C_1m^{1/2}(\log m)^2$, where $C_1=C_1(r_0)$ depends only on r_0.

Proof. Let us start recalling Dirichlet's theorem on primes in arithmetic progressions. In particular, the quantitative version of Dirichlet's theorem implies that for integers a and b with $(a,b)=1$, there is an integer $t_{a,b}$ such that for all $t\geq t_{a,b}$ there is a prime $p\equiv a \pmod b$ in the interval $[t,11t/10]=\{x\colon t\leq x\leq 11t/10\}$.

We let $r_1=t_{1,8}+1$ and proceed to show that this choice of r_1 will do. Thus, let an arbitrary integer $r_0\geq r_1$ be given, and let us define $m_0=m_0(r_0)$ as required in our lemma. To that end, first observe that, by the choice of r_1, there is a prime $p\equiv 1 \pmod 8$ satisfying (2.10). We fix such a prime p. Observe that we have $p\equiv 1 \pmod 4$. Moreover, since $p\equiv 1 \pmod 8$, we have that

$$2\text{ is a quadratic residue modulo }p. \tag{2.12}$$

Since $(4, p) = 1$, by the Chinese remainder theorem, there is a unique integer s with $1 \leq s \leq 4p$ satisfying

$$s \equiv 2 \pmod{p} \qquad \text{and} \qquad s \equiv 1 \pmod{4}. \tag{2.13}$$

We are finally ready to define $m_0 = m_0(r_0)$. We let

$$m_0 = m_0(r_0) = \max\left\{\frac{10}{21}t_{s,4p}^3,\ \frac{8}{2.1}r_0^3\right\}. \tag{2.14}$$

Our aim now is to show that the choice for $m_0 = m_0(r_0)$ in (2.14) will do. Thus, let $m \geq m_0(r_0)$ be given. We shall now describe a procedure $\mathcal{P}$ to find the primes p and q as required. Our description will be quite informal.

The prime p with $p \equiv 1 \pmod{8}$ satisfying (2.10) may be found easily. We now need to determine a suitable value for q. We choose q among the integers in the arithmetic progression $\{4pk + s\colon k = 0, 1, 2, \dots\}$, where s is the integer satisfying $1 \leq s \leq 4p$ and (2.13). By Dirichlet's theorem and our choice of $m_0 \geq (10/21)t_{s,4p}^3$, there is a prime $q \equiv s \pmod{4p}$ satisfying (2.11). We claim that our choice of s implies all properties promised for q. Indeed, $q \equiv s \equiv 1 \pmod{4}$. Furthermore, the quadratic reciprocity law implies $\left(\frac{p}{q}\right) = \left(\frac{q}{p}\right)$, since both $p, q \equiv 1 \pmod{4}$. Using that $q \equiv s \equiv 2 \pmod{p}$ and recalling (2.12), we have

$$\left(\frac{p}{q}\right) = \left(\frac{q}{p}\right) = \left(\frac{s}{p}\right) = \left(\frac{2}{p}\right) = 1.$$

Finally, note that $m_0 \geq 8r_0^3/2.1$ guarantees $q > p$ and consequently condition (2.9) is satisfied. Therefore, the primes p and q as required do exist.

Let us now consider the time complexity of our procedure $\mathcal{P}$ above. We first observe that the search for $p < 2r_0$ takes a quantity of steps that depends only on r_0. To find q, we have enough time to check all integers in the interval $[\sqrt[3]{2.1m}, 1.1\sqrt[3]{2.1m}]$. Since this interval is of length $O(m^{1/3})$, this will take $O(m^{1/3} \cdot m^{1/6}(\log m)^2) = O(m^{1/2}(\log m)^2)$ steps. The $(\log m)^2$ term accounts for the time complexity of arithmetic operations with integers having $O(\log m)$ digits. □

We now describe Algorithm $\mathcal{E}$, the existence of which is asserted in Lemma 2.5. Consider the integer r_1 and the function $m_0(r_0)$ whose existences are guaranteed by Lemma 2.8. On input (r_0, m), where $r_0 \geq r_1$ and $m \geq m_0(r_0)$, Algorithm $\mathcal{E}$ performs the following steps.

1. Run Algorithm $\mathcal{P}$ on input (r_0, m) to obtain primes p and q as in the statement of Lemma 2.8.
2. List all elements of $\mathrm{PSL}(2, \mathbb{Z}/q\mathbb{Z})$, i.e., the vertex set of $X^{p,q}$, by enumerating all the canonical representatives of the elements in $\mathrm{PSL}(2, \mathbb{Z}/q\mathbb{Z})$.
3. Find all solutions to $\alpha_0^2 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2 = p$ that belong to S and construct T (see (2.7)).
4. For each vertex x of $X^{p,q}$, construct its adjacency list.
5. Set J to be any induced subgraph of $X^{p,q}$ on m vertices.

The following claim will finish the proof of Lemma 2.5.

Claim 2.9. *Algorithm $\mathcal{E}$ produces a graph J that is $(\varrho, 3)$-uniform in time $\leq Cm(\log m)^2$, where $r = \varrho m$ satisfies $r_0 \leq r \leq 2r_0$, and C is a constant that depends only on r_0.*

Proof. We start with the correctness of $\mathcal{E}$. We already know that $\mathcal{P}$ produces suitable primes p and q. Hence, we only need to argue that the graph J obtained in Step 5 has the required properties.

By Theorem 2.6 and Corollary 2.7, the graph $X^{p,q}$ constructed in Steps 2–4 has $n = q(q^2-1)/2$ vertices, is $(p+1)$-regular, and is $(\varrho, 2)$-uniform with $\varrho = (p+1)/n$. Furthermore, note that (2.11) implies $1 \le n/m \le 1.05(1.1)^3$. Lemma 2.1 implies J is a (ϱ, L)-uniform graph with $\varrho = (p+1)/n$ and $L = 2\sqrt{n/m} \le 2.1\sqrt{231/200} = 2.256\cdots < 3$. Thus J is indeed a $(\varrho, 3)$-uniform graph.

Since $1 \le n/m \le 1.05(1.1)^3$, we deduce that $r = \varrho m = (p+1)m/n$ is such that

$$r \le 2r_0$$

and

$$r \ge 1.4r_0 \cdot \frac{20}{21}\left(\frac{10}{11}\right)^3 \ge r_0 \ .$$

Finally, we argue about the time complexity of each of the steps in Algorithm $\mathcal{E}$. By Lemma 2.8, we already know that Step 1 takes time $\le C_1(r_0)m(\log m)^2$, where $C_1(r_0)$ is a constant that depends only on r_0.

Recalling the form of the canonical representatives of the elements in $\mathrm{PSL}(2, \mathbb{Z}/q\mathbb{Z})$, we see that Step 2 may be performed in time $O(m(\log m)^2)$. The time complexity of Step 3 depends only on r_0.

In Step 4, we take one by one the vertices of $X^{p,q}$, i.e., the elements of $\mathrm{PSL}(2, \mathbb{Z}/q\mathbb{Z})$, and generate their adjacency lists. Since $|T| = p+1 \le 2r_0$, to generate the adjacency list of a particular vertex takes only $O((\log q)^2)$ steps.

Finally, taking m vertices of $X^{p,q}$ arbitrarily and adjusting their adjacency lists to create an adjacency list representation for the corresponding induced subgraph takes $O(m \log m)$ time. Therefore, the time complexity of Algorithm $\mathcal{E}$ is $\le C_0(r_0)m(\log m)^2$, as promised. □

3. Algorithms. Before we describe the algorithm, let us introduce some notation. Let $\Gamma = (V, E)$ be a graph and $v \in V$ a vertex. We write $\Gamma(v)$ for the neighborhood of v, i.e., for the set of all vertices adjacent to v in Γ, and $d(v)$ for the degree of v, i.e. $d(v) = |\Gamma(v)|$. To shorten our notation we will use that same letter to denote a graph and the set of its edges. For example, Γ will also stand for $E(\Gamma)$ and, hence, $e(\Gamma) = |\Gamma|$.

3.1. Regularity of bipartite graphs. In this section we describe algorithm $\mathcal{A}$ which takes as an input a bipartite graph $G = (A, B; E), |A| = |B| = m$, and $0 < \varepsilon < 1$. The algorithm in time $O(m^2)$ either confirms that G is ε-regular or finds sets $A' \subseteq A, B' \subseteq B, |A'| \ge \varepsilon' m, |B'| \ge \varepsilon' m$ such that

$$|d(A', B') - d(A, B)| \ge \varepsilon' \ .$$

Our algorithm $\mathcal{A}$ consists of the preprocessing stage $\mathcal{A}_P$ and the main procedure $\mathcal{A}_M$.

3.1.1. The preprocessing stage. In order to describe the preprocessing stage we need to define

$$\varepsilon' = \left(\frac{\varepsilon}{10}\right)^{20} \frac{1}{10^4} \ . \tag{3.1}$$

To describe $\mathcal{A}_P$ we need the constants ε and ε' only. Note that other constants are used for describing the other part $\mathcal{A}_M$; these other constants will be defined later in §3.1.2 and will be related to ε and ε' above.

Algorithm $\mathcal{A}_P$ is based on the following standard observation and Lemma 3.1. We observe that if the bipartite graph G on vertex set $A \cup B, |A| = |B| = m$, is such that $p(G) := d(A, B) = |G|/m^2 \leq \varepsilon^3$, then G is ε-regular (we omit the proof of this standard observation).

Lemma 3.1 quantifies a further observation that we may remove some vertices of our graph G so that we either obtain a subgraph $H \subseteq G$ that is essentially degree-regular (all degrees are about the same), or else in the process of removing these vertices we locate a witness to the ε'-irregularity of G. This is formalized as follows.

LEMMA 3.1. *Suppose G is a bipartite graph with vertex set $A \cup B$, $|A| = |B| = m$, and suppose that $p(G) > \varepsilon^3$ holds. There is a procedure that runs in time $O(m^2)$ that either (i) produces a witness to the ε'-irregularity of G or (ii) produces a bipartite subgraph $H \subseteq G$, say $H = (U, V; F)$, such that*

(a)

$$(1 - 2\varepsilon')m < |U|, |V| \leq m\ , \tag{3.2}$$

(b)

$$\big|p(H) - p(G)\big| \leq 5\varepsilon', \tag{3.3}$$

where $p(H) = |H|/|U||V|$, and

(c) *for all $u \in U$ and $v \in V$, we have*

$$\begin{aligned} d(u) &= (p(H) + O_1(10\varepsilon')) \cdot |V|\ , \\ d(v) &= (p(H) + O_1(10\varepsilon')) \cdot |U|\ . \end{aligned} \tag{3.4}$$

Proof. We first omit the vertices v in V for which the condition

$$d(v) = (p(G) + O_1(\varepsilon'))m \tag{3.5}$$

fails. If the number of such vertices is $\geq 2\varepsilon' m$, we may easily produce a witness to the ε'-irregularity of G as in (i) in the statement of our lemma. Suppose therefore that the number of such vertices is $< 2\varepsilon' m$. Let $V \subseteq B$ be the resulting subset of B. Hence $|V| > (1 - 2\varepsilon')m$. We now omit the vertices $u \in A$ for which the condition

$$d(u) = (p(G) + O_1(\varepsilon')) \cdot |V| \tag{3.6}$$

fails. Again, if the number of such vertices is $\geq 2\varepsilon' m$, we may easily produce a witness to the ε'-irregularity of G. If the number of such vertices is $< 2\varepsilon' m$, the resulting graph H is as in (ii) in the statement of our lemma.

The time complexity assertion will be verified in the proof of Lemma 3.4 (cf. algorithm $\mathcal{A}_P$ below). □

For convenience, we let $\Psi(m, \varepsilon')$ be the family of subgraphs H of G that satisfy (a)–(c) in (ii) of Lemma 3.1 above.

Now we are ready to describe *algorithm $\mathcal{A}_P$*:

1. Given G and ε, decide if $p(G) < \varepsilon^3$.
2. If $p(G) \leq \varepsilon^3$, then G is ε-regular and $\mathcal{A}_P$ halts.
3. If $p(G) > \varepsilon^3$, apply Lemma 3.1 to construct a subgraph H of G which satisfies (a)–(c). (Algorithm $\mathcal{A}_M$ will be applied to H.)

3.1.2. The main procedure. In view of step 3 in algorithm $\mathcal{A}_P$ (cf. Lemma 3.1) we will now assume that $H \in \Psi(m, \varepsilon')$.

(a) Definition of constants. Before describing algorithm $\mathcal{A}_M$ we will define constants needed for it. Recall that algorithm $\mathcal{A}$, and so $\mathcal{A}_M$, is given $0 < \varepsilon < 1$. In §3.1.1 we already defined $\varepsilon' = (\varepsilon/10)^{20}/10^4$.

We first let

$$\delta = \frac{1}{4}\left(\frac{\varepsilon}{2}\right)^5, \qquad L = 5, \quad \text{and} \quad r_A = \frac{10^6}{\varepsilon^8}\,. \tag{3.7}$$

We now let

$$\mu = \left(\frac{\varepsilon}{10}\right)^{10} \frac{1}{100} \tag{3.8}$$

and put

$$\eta = \frac{\mu}{3}\,. \tag{3.9}$$

We also let $r_B = \overline{r}(\eta, L)$ be as given in Lemma 2.4 and let

$$r_0 = \max\{r_A, r_B\}. \tag{3.10}$$

Finally, we set

$$\varepsilon_1 = \frac{1}{4}\left(\frac{\varepsilon}{2}\right)^{16}\,. \tag{3.11}$$

This will be used only later in a proof. However, it might be helpful to see the relation of ε_1 to the other constants introduced here.

The reader may find it useful to keep in mind the following hierarchy of the constants for ε small:

$$\varepsilon' < \varepsilon^{20} \ll \varepsilon_1 < \varepsilon^{16} \ll \eta = \frac{\mu}{3} < \varepsilon^{10} \ll \delta \ll \varepsilon^3 \ll \varepsilon,\, p\,. \tag{3.12}$$

(Here inequalities $<$ are used to compare two quantities which differ by an absolute constant.) Note that for the description of $\mathcal{A}_M$ we only need to know r_0 and δ defined above. The other constants are needed in the proofs below.

(b) Property $\mathcal{P}(J, \delta)$. We introduce some notation. Let H be a bipartite graph with vertex set $U \cup V$. Let J be a (ϱ, L)-uniform graph with vertex set U. We say that H *has property* $\mathcal{P}(J, \delta)$ if

$$\sum_{\{u,u'\}\in J} \left| d_H(u, u') - p(H)^2 \cdot |V| \right| \le \delta p(H)^2 |V| \cdot |J| \tag{3.13}$$

holds. Recall that due to our notation $\{u, u'\} \in J$ means that $\{u, u'\}$ is an edge of J. Moreover, let us write J_v $(v \in V)$ for the graph $J[H(v)]$ induced by the neighbourhood $H(v)$ of v in H. We define a 0–1 matrix $M = (m(e, v))_{e,v}$ indexed by $J \times V$ as follows:

$$m(e, v) = \begin{cases} 1 & \text{if } e \in J_v \\ 0 & \text{otherwise.} \end{cases} \tag{3.14}$$

Therefore, clearly, $m(e, v) = 1$ if and only if both endpoints of e are adjacent to v.

(c) Description of $\mathcal{A}_M$. We assume $\mathcal{A}_M$ is given a bipartite graph $H \in \Psi(m, \varepsilon')$ having vertex set $U \cup V$. *Algorithm $\mathcal{A}_M$* now proceeds as follows:

1. Apply procedure $\mathcal{E}$ to construct a (ϱ, L)-uniform graph J with vertex set U and average degree $r = \varrho \cdot |U|$ satisfying
$$r_0 = r_0(\eta) \le r \le 2r_0 \tag{3.15}$$
(cf. Lemma 2.5).
2. Verify whether H has $\mathcal{P}(J, \delta)$. If it has, then then G is ε-regular (see Lemma 3.2 below) and $\mathcal{A}_M$ halts.
3. If $\mathcal{P}(J, \delta)$ fails for H, construct matrix $M = (m(e, v))_{e,v}$ defined above. Find and fix a vertex $v_0 \in V$ such that
$$\sum_{v\in V} \sum_{e\in J_{v_0}} m(e, v) \ge \left(1 + \frac{\delta^2}{2}\right) p(H)^4 |V| \cdot |J|. \tag{3.16}$$
[The existence of such a vertex v_0 is proved later, cf. Lemma 3.5.]
4. Set $U' = H(v_0)$ and
$$V' = \left\{ v' \in V \ : \ \sum\{m(e, v') \colon e \in J_{v_0}\} \ge \left(1 + \frac{\delta^2}{4}\right) p(H)^4 \cdot |J| \right\}. \tag{3.17}$$
5. $\mathcal{A}_M$ outputs (U', V') and claims that this is a witness to the ε'-irregularity of G.

3.1.3. Correctness and analysis of Algorithm $\mathcal{A}$. The correctness of $\mathcal{A}$ follows from Lemmas 3.2 and 3.3 below. The first lemma says that if Algorithm $\mathcal{A}_M$, and hence $\mathcal{A}$, claims that G is ε-regular in Step 2, then this is indeed the case.

LEMMA 3.2. *If H enjoys property $\mathcal{P}(J, \delta)$, then G is ε-regular.*

LEMMA 3.3. *If property $\mathcal{P}(J, \delta)$ fails for H, then G is not ε'-regular, and the pair (U', V') produced by Algorithm $\mathcal{A}_M$ is indeed a witness for the ε'-irregularity of G.*

We shall prove the two lemmas above in §4. The next two lemmas immediately imply that $\mathcal{A}$ has time complexity $O(m^2)$.

LEMMA 3.4. *Algorithm $\mathcal{A}_P$ described in §3.1.1 has time complexity $O(m^2)$.*

Proof. The steps of $\mathcal{A}_P$ have the following time complexity. The only computations are in steps 1 and 3:

Step 1. Since $p(G) = |G|/m^2$, it takes at most $O(m^2)$ steps to compute $p(G)$ and decide whether $p(G) < \varepsilon^3$.

Step 3. Based on the proof of Lemma 3.1, a vertex of G is put into H iff it satisfies (3.4). Hence, we proceed through all $2m$ vertices of G each time checking (3.4) which takes $O(m)$ steps. Thus, this step takes $O(m^2)$ steps, too.

The overall time complexity of $\mathcal{A}_P$ is $O(m^2)$. □

LEMMA 3.5. *Algorithm $\mathcal{A}_M$ described in §3.1.2 runs in time $O(m^2)$.*

Proof. The steps of $\mathcal{A}_M$ have the following time complexity:

Step 1. To perform procedure $\mathcal{E}$ we need $O(m(\log m)^2)$ steps, cf. Lemma 2.5.

Step 2. To verify $\mathcal{P}(J, \delta)$ we need to add $|J| = O(m)$ summands. Computing each of these summands takes $O(m)$ steps. Consequently, one can decide $\mathcal{P}(J, \delta)$ in $O(m^2)$ time.

Step 3. Deciding if $m(e, v) = 1$ or 0 can be performed in constant time. Thus, constructing M takes $O(|J||V|) = O(m^2)$ time. To check (3.16), we first write it in the equivalent form
$$\sum_{e\in J_{v_0}} \sum_{v\in V} m(e, v) \ge \left(1 + \frac{\delta^2}{2}\right) p(H)^4 |V| \cdot |J| \,. \tag{3.18}$$

Using matrix M we compute the column sums $\sum_{v\in V} m(e,v)$ for each $e \in J$. This takes $O(m^2)$ steps.
Now to check (3.18) for a fixed vertex $v_0 \in V$, we first find J_{v_0}, which takes $|J_{v_0}| = O(m)$ steps. Then we add together the column sums $\sum_{v\in V} m(e,v)$ for all $e \in J_{v_0}$ and decide about the truth of (3.18). This takes another $O(m)$ steps. In the worst case we need to check all $v_0 \in V$. Since $|V| = m$, to find v_0 that satisfies (3.18) will take at most $O(m^2)$ steps.

Step 4. Since for each $v' \in V$ the condition $\sum_{e\in J_{v_0}} m(e,v') \geq (1+\delta^2/4)p(H)^4|J|$ can be verified in $O(m)$ time, the set V' can be constructed in $O(m^2)$ time.

Step 5. This step takes a constant time.

Therefore, the time complexity of $\mathcal{A}_M$ is $O(m^2)$. Finally let us point out that the fact that J has $O(m)$ edges was crucial. □

3.2. The regularity lemma. For the sake of completeness, we include an algorithm necessary for proving Corollary 1.6. Given Theorem 1.5 we can derive the necessary algorithm in a standard way.

Let $V_0, V_1, \ldots, V_k$ be an equitable partition P of the set of vertices of a graph. We define the *index* of P, cf. [31], by

$$\mathrm{ind}(P) = \frac{1}{k^2} \sum_{1\leq r<s\leq k} d(V_r, V_s)^2 .$$

To present a proof of Corollary 1.6 we will use the following lemma, which was proved in [31]. Note that no comment is made in [31] on the running time. However, the proof of the lemma implies an algorithm of time complexity $O(n)$.

LEMMA 3.6. *Fix k and γ and let $G = (V,E)$ be a graph on n vertices. Let P be an equitable partition of V into classes $V_0, V_1, \ldots, V_k$. Assume $|V_1| > 4^{2k}$ and $4^k > 600\gamma^{-5}$. Given proofs that more than γk^2 pairs (V_r, V_s) are not γ-regular (where by proofs we mean subsets $X = X(r,s) \subseteq V_r$, $Y = Y(r,s) \subseteq V_s$ that violate the condition of γ-regularity of (V_r,V_s)), then one can find in $O(n)$ time an equitable partition P' (which is a refinement of P) into $1 + k4^k$ classes, with the exceptional class of cardinality at most $|V_0| + n/4^k$ and such that*

$$\mathrm{ind}(P') \geq \mathrm{ind}(P) + \frac{\gamma^5}{20} .$$

Proof of Corollary 1.6. Theorem 1.5 and Lemma 3.6 already imply Corollary 1.6. Let $\varepsilon > 0$ and k_0 be a positive integer. Let $\varepsilon' = \varepsilon^{20}/10^{24}$. We set $N = N(\varepsilon, k_0)$ and $T = T(\varepsilon, k_0)$ as follows: Let a be the least positive integer such that

$$4^a > 600\left(\frac{\varepsilon'}{4}\right)^{-5}, \quad a \geq k_0 . \tag{3.19}$$

Let k_i be a sequence of integers defined inductively as

$$k_0 = a, \quad k_{i+1} = k_i 4^{k_i} .$$

Set $T = k_{\lceil 10(\varepsilon'/4)^{-5}\rceil}$ and $N = \max\{T4^{2T}, 2T/\varepsilon'^2\}$. Finally we set $K_0'' = N \geq T$.

Let $\Gamma = (V,E)$ be a graph on n vertices, $n \geq N$. The following algorithm constructs an ε-regular partition of Γ into $k+1$ classes with $k_0 \leq k \leq T \leq K_0''$.

Algorithm $\mathcal{A}_0'$ proceeds as follows:

1. Arbitrarily divide the vertices of Γ into an equitable partition P_1 with classes V_0, $V_1, \dots, V_a$ where $|V_1| = \lfloor n/a \rfloor$ and $|V_0| < a$. Set $k_1 = a$.
2. For every pair (V_r, V_s) of P_i, verify if it is ε-regular or find $V_r' \subseteq V_r$, $V_s' \subseteq V_s$, $|V_r'| \geq (\varepsilon'/4)|V_r|$, $|V_s'| \geq (\varepsilon'/4)|V_s|$, such that $|d(V_r', V_s') - d(V_r, V_s)| \geq \varepsilon'$.
3. If there are at most $\varepsilon\binom{k_i}{2}$ pairs that are not verified as ε-regular, then stop. The partition P_i is an ε-regular partition.
4. Apply Lemma 3.6 where $P = P_i$, $k = k_i$, $\gamma = \varepsilon'/4$ and obtain a partition P' with $1 + k_i 4^{k_i}$ classes.
5. Let $k_{i+1} = k_i 4^{k_i}$, $P_{i+1} = P'$, $i = i+1$, and go to step 2.

CLAIM 3.7. *Algorithm $\mathcal{A}_0'$ described is correct and runs in $O(n^2)$ time.*

Proof. To prove correctness of algorithm $\mathcal{A}_0'$ is quite standard. Since the index of partitions P_i constructed by $\mathcal{A}_0'$ strictly increases and at the same time is bounded by 1 from above, algorithm $\mathcal{A}_0'$ uses Lemma 3.6, and thus Theorem 1.5, only finitely many times. Each such use takes $O(n^2)$ time. □

4. Proofs of the main lemmas. We use the notation introduced in Section 3.1. Before the proofs let us point out a straightforward estimate on the size of neighborhoods used throughout the proofs.

REMARK 4.1. *Let $H \in \Psi(m, \varepsilon')$ be the graph in the statement of Lemma 3.1, and let and $u \in U$ and $v \in V$ be vertices of H. As an immediate consequence of (3.4) and the definition of ε' and μ, we get*

$$\begin{aligned} |H(u)| &= (p(H) + O_1(10\varepsilon'))|V| = (1 + O_1(\mu))p(H)|V|\ , \\ |H(v)| &= (p(H) + O_1(10\varepsilon'))|U| = (1 + O_1(\mu))p(H)|U|\ . \end{aligned} \tag{4.1}$$

4.1. Proof of Lemma 3.2. Let $H \in \Psi(m, \varepsilon')$ be the input graph of algorithm $\mathcal{A}_M$. Let $U \cup V$ be the vertex set of H. Set $m_U = |U|, m_V = |V|$, and $p := p(H) = |H|/m_U m_V$.

We say that the graph H has *property* $\mathcal{Q}(J, \delta)$ if the inequality

$$\sum_{\{u,u'\} \in J} \left| d_H(u,u') + p^2 m_V - (d_H(u) + d_H(u'))p \right| \leq \delta p^2 m_V \cdot |J| \tag{4.2}$$

holds true.

CLAIM 4.2. *Let $\delta > 0$ be fixed. If a graph $H \in \Psi(m, \varepsilon')$ has property $\mathcal{P}(J, \delta)$, then it has property $\mathcal{Q}(J, 2\delta)$.*

Proof. Since H enjoys $\mathcal{P}(J, \delta)$ and $H \in \Psi(m, \varepsilon')$, we have

$$\begin{aligned} &\sum_{\{u,u'\} \in J} \left| d_H(u,u') + p^2 m_V - (d_H(u) + d_H(u'))p \right| \\ &\qquad \leq \sum_{\{u,u'\} \in J} \left| d_H(u,u') - p^2 m_V \right| \\ &\qquad\qquad + 2p \sum_{u \in U} \left| p m_V - d_H(u) \right| d_J(u) \\ &\qquad \leq \delta p^2 m_V |J| + 20 p \varepsilon' m_V \sum_{u \in U} d_J(u) \\ &\qquad \leq 2\delta p^2 m_V |J|\ , \end{aligned}$$

since $\varepsilon' \leq \delta \varepsilon^3/40 \leq \delta p/40$ (see (3.1, 3.7)). □

In view of Claim 4.2, we are going to prove

Claim 4.3. *Every $H \in \Psi(m, \varepsilon')$ that has property $\mathcal{Q}(J, 2\delta)$ is $(\varepsilon/2)$-regular.*

The proof is presented below. First, let us observe that the $\varepsilon/2$-regularity of H implies the ε-regularity of G. Indeed, let $X \subseteq A$, $|X| \geq \varepsilon|A|$, and $Y \subseteq B$, $|Y| \geq \varepsilon|B|$. Set $X' = X \cap U$ and $Y' = Y \cap V$. By Lemma 3.1,

$$|X'| \geq \varepsilon|A| - 2\varepsilon'|A| \geq (\varepsilon - 2\varepsilon')|U| \geq \frac{\varepsilon}{2}|U|$$

and, similarly,

$$|Y'| \geq \frac{\varepsilon}{2}|V| .$$

A standard argument based on the fact that X' and Y' are almost equal to X and Y (recall Lemma 3.1) shows that $|d(X,Y) - d(X',Y')| \leq \varepsilon/4$. Thus, using the $\varepsilon/2$-regularity of H and Lemma 3.1, we get

$$\begin{aligned}|d(X,Y) - d(A,B)| &\leq |d(X,Y) - d(X',Y')| + |d(X',Y') - d(U,V)| \\ &\quad + |d(U,V) - d(A,B)| \leq \frac{\varepsilon}{4} + \frac{\varepsilon}{2} + 5\varepsilon' < \varepsilon .\end{aligned}$$

Hence, G is ε-regular. □

Proof of Claim 4.3. Let $\mathbb{A} = (a_{u,v})_{u,v}$ be a matrix indexed by $U \times V$, with entries

$$a_{u,v} = \begin{cases} -(1-p) & \text{if } \{u,v\} \in H \\ p & \text{otherwise.} \end{cases}$$

Moreover, for $u \in U$, let $\xi_u = (a_{u,1}, \dots, a_{u,m_V})$ be the u-th row of $\mathbb{A}$. The following claim follows easily from the definition of property $\mathcal{Q}$ (see (4.2)).

Claim 4.4. *For every $H \in \Psi(m, \varepsilon')$ that has property $\mathcal{Q}(J, 2\delta)$, the row-vectors of $\mathbb{A}$ satisfy the following inequality*

$$\sum_{\{u,u'\} \in J} |\langle \xi_u, \xi_{u'} \rangle| \leq 2\delta p^2 m_V |J| . \tag{4.3}$$

Proof. Since

$$\begin{aligned}\langle \xi_u, \xi_{u'} \rangle &= d_H(u,u')(1-p)^2 - (d_H(u) + d_H(u') - 2d_H(u,u'))p(1-p) \\ &\quad + (m_V - (d_H(u) + d_H(u') - d_H(u,u'))p^2 \\ &= d_H(u,u')[(1-p)^2 + 2p(1-p) + p^2] + m_V p^2 - (d_H(u) + d_H(u'))p \\ &= d_H(u,u') + m_V p^2 - (d_H(u) + d_H(u'))p\end{aligned}$$

for any pair of vertices $u, u' \in U$, we have that $\sum_{\{u,u'\} \in J} |\langle \xi_u, \xi_{u'} \rangle|$ equals the sum on the left hand side of (4.2). Thus the claim follows. □

Let $U' \subseteq U$ and $V' \subseteq V$. To shorten our notation $\sum_{u,u' \in J}^{U'}$ will denote summation over $\{u, u'\} \in J$ such that u, $u' \in U'$. Furthermore, for $u \in U'$ let ψ_u be the restriction of the vector ξ_u to V', i.e., $\psi_u = (a_{u,v})_{v \in V'}$. We clearly have $\sum_{u,u' \in J}^{U'} |\langle \xi_u, \xi_{u'} \rangle| \leq \sum_{\{u,u'\} \in J} |\langle \xi_u, \xi_{u'} \rangle|$. We now compare $\sum_{u,u' \in J}^{U'} \langle \xi_u, \xi_{u'} \rangle$ with $\sum_{u,u' \in J}^{U'} \langle \psi_u, \psi_{u'} \rangle$. We have

$$\sum_{u,u' \in J}^{U'} \langle \xi_u, \xi_{u'} \rangle = \sum_{u,u' \in J}^{U'} \langle \psi_u, \psi_{u'} \rangle + \sum_{v \notin V'} \sum_{u,u' \in J}^{U'} a_{u,v} a_{u',v}. \tag{4.4}$$

For $v \in V$ we set $S_v^{U'} = \sum_{u,u' \in J}^{U'} a_{u,v} a_{u',v}$, and proceed to estimate this quantity. Let $S := H(v) \cap U' \subseteq U'$ be the set of neighbours of the vertex v in U' and $T := (U' \setminus S) \subseteq U'$ be the set of non-neighbours of the vertex v in U'.

Set $\alpha = |S|/|U'|$ and $\beta = |T|/|U'|$. Note $\alpha + \beta = 1$. Thus we can write

$$S_v^{U'} = e_J(S)(1-p)^2 + e_J(T)p^2 - e_J(S,T)p(1-p). \tag{4.5}$$

The (ϱ, L)-uniformity of J implies the following claim. Recall that for two numbers a, b we write $a = O_1(b)$ if and only if $|a| \leq b$.

Claim 4.5. *For all $v \in V'$ and $U' \subseteq U$, we have*

$$S_v^{U'} = \frac{\varrho}{2}|U'|^2 \cdot \big[\alpha(1-p) - \beta p\big]^2 + O_1\big(3L\sqrt{r} \cdot |U'|\big)\ . \tag{4.6}$$

Proof. Set $m' = |U'|$. Since J is a (ϱ, L)-uniform graph, Lemma 2.3 implies

$$\begin{aligned}
e_J(S) &= \frac{\varrho}{2}|S|^2 + O_1\big(2L\sqrt{r} \cdot |S|\big) \\
&= \frac{\varrho}{2}(\alpha m')^2 + O_1\big(2L\sqrt{r} \cdot \alpha m'\big)\ , \\
e_J(T) &= \frac{\varrho}{2}|T|^2 + O_1\big(2L\sqrt{r} \cdot |T|\big) \\
&= \frac{\varrho}{2}(\beta m')^2 + O_1\big(2L\sqrt{r} \cdot \beta m'\big)\ , \\
e_J(S,T) &= \varrho|S| \cdot |T| + O_1\big(L\sqrt{r|S| \cdot |T|}\big) \\
&= \varrho\alpha\beta(m')^2 + O_1\big(L\sqrt{r\alpha\beta} \cdot m'\big)\ .
\end{aligned}$$

Using (4.5) we get

$$\begin{aligned}
S_v^{U'} &= \big[\frac{\varrho}{2}(\alpha m')^2 + O_1\big(2L\sqrt{r} \cdot \alpha m'\big)\big](1-p)^2 \\
&\qquad + \big[\frac{\varrho}{2}(\beta m')^2 + O_1\big(2L\sqrt{r} \cdot \beta m'\big)\big]p^2 \\
&\qquad - \big[\varrho\alpha\beta(m')^2 + O_1\big(L\sqrt{r\alpha\beta} \cdot m'\big)\big]p(1-p) \\
&= \frac{\varrho}{2}(m')^2\big[\alpha(1-p) - \beta p\big]^2 + \Delta,
\end{aligned} \tag{4.7}$$

where

$$\Delta = O_1\big(2L\sqrt{r} \cdot \alpha m'(1-p)^2 + 2L\sqrt{r} \cdot \beta m' p^2 + L\sqrt{r\alpha\beta} \cdot m' p(1-p)\big)\ . \tag{4.8}$$

To bound Δ we are going to use the inequalities $\alpha(1-p)^2 + \beta p^2 \leq \alpha + \beta$ and $\sqrt{\alpha\beta} \leq (\alpha + \beta)$. Thus,

$$\begin{aligned}
|\Delta| &\leq 2L\sqrt{r} \cdot \alpha m'(1-p)^2 + 2L\sqrt{r} \cdot \beta m' p^2 + L\sqrt{r\alpha\beta} \cdot m' p(1-p) \\
&\leq L\sqrt{r}m'\big(2\alpha + 2\beta + \sqrt{\alpha\beta}\big) \\
&\leq L\sqrt{r}m' \cdot 3(\alpha + \beta) \\
&= 3L\sqrt{r}m'\ .
\end{aligned} \tag{4.9}$$

Expressions (4.7) and (4.9) already imply the claim. We only note that we got a bound on Δ linear in $m' = |U'|$ since L and r are constants. □

Next we proceed with an upper and lower bounds on $\sum_{u,u'\in J}^{U'}\langle\psi_u,\psi_{u'}\rangle$. To derive an upper bound for this quantity we use (4.4) to relate $\sum_{u,u'\in J}^{U'}\langle\psi_u,\psi_{u'}\rangle$ and $S_v^{U'}$ as follows:

$$\sum\nolimits_{u,u'\in J}^{U'}\langle\psi_u,\psi_{u'}\rangle=\sum\nolimits_{u,u'\in J}^{U'}\langle\xi_u,\xi_{u'}\rangle-\sum_{v\notin V'}S_v^{U'}.$$

Note that (4.6) implies $S_v^{U'}\geq -3L\sqrt{r}|U'|$. This lower bound and Claim 4.4 imply

$$\sum\nolimits_{u,u'\in J}^{U'}\langle\psi_u,\psi_{u'}\rangle\leq 2\delta p^2 m_V|J|+3L\sqrt{r}|U'|(m_V-|V'|)\ . \tag{4.10}$$

To estimate $\sum_{u,u'\in J}^{U'}\langle\psi_u,\psi_{u'}\rangle$ from below, we first write

$$\sum\nolimits_{u,u'\in J}^{U'}\langle\psi_u,\psi_{u'}\rangle=\sum_{v\in V'}\sum\nolimits_{u,u'\in J}^{U'}a_{u,v}a_{u',v}=\sum_{v\in V'}S_v^{U'}\ . \tag{4.11}$$

Using (4.6) we get a lower bound on the expression in our last equation

$$\begin{aligned}\sum\nolimits_{u,u'\in J}^{U'}\langle\psi_u,\psi_{u'}\rangle&=\sum_{v\in V'}S_v^{U'}\\ &\geq\sum_{v\in V'}\bigl(\frac{\varrho}{2}|U'|^2\bigl[\alpha(1-p)-\beta p\bigr]^2-3L\sqrt{r}\cdot|U'|\bigr)\\ &\geq\frac{\varrho}{2}\cdot\frac{|U'|^2}{|V'|}\bigl[\sum_{v\in V'}(\alpha(1-p)-\beta p)\bigr]^2-3L\sqrt{r}|U'|\cdot|V'|\ .\end{aligned}\tag{4.12}$$

We used the Cauchy–Schwarz inequality to get the last line of our bound. Comparing the lower and upper bounds on $\sum_{u,u'\in J}^{U'}\langle\psi_u,\psi_{u'}\rangle$, cf. (4.12) and (4.10), we infer

$$\Bigl[\sum_{v\in V'}(\alpha(1-p)-\beta p)\Bigr]^2\leq\frac{2|V'|}{\varrho|U'|^2}\cdot\bigl(2\delta p^2 m_V|J|+3L\sqrt{r}|U'|m_V\bigr)\ . \tag{4.13}$$

Now we are ready to show that H is $\frac{\varepsilon}{2}$-regular. Fix $U'\subseteq U$, $|U'|\geq\frac{\varepsilon}{2}|U|=\frac{\varepsilon}{2}m_U$, and $V'\subseteq V$, $|V'|\geq\frac{\varepsilon}{2}|V|=\frac{\varepsilon}{2}m_V$. Recall that $d(U,V)=p=p(H)$ in our notation. First we relate the difference of densities to the left-hand side of (4.13) as follows:

$$\begin{aligned}|d(U',V')-d(U,V)|^2&=\left|\frac{e(U',V')}{|U'||V'|}-p\right|^2\\ &=\frac{1}{|U'|^2|V'|^2}\Bigl[e(U',V')-p|U'|\cdot|V'|\Bigr]^2\\ &=\frac{1}{|U'|^2|V'|^2}\Bigl[\sum_{v\in V'}(|H(v)\cap U'|-p|U'|)\Bigr]^2\\ &=\frac{1}{|U'|^2|V'|^2}\Bigl[\sum_{v\in V'}\alpha|U'|-p|U'|\Bigr]^2\\ &=\frac{1}{|V'|^2}\Bigl[\sum_{v\in V'}(\alpha(1-p)-\beta p)\Bigr]^2\ .\end{aligned}\tag{4.14}$$

Finally, we bound the expression for the difference of densities in (4.14) using (4.13). We get

$$\begin{aligned}\left|d(U',V')-d(U,V)\right|^2 &\le \frac{2}{\varrho|U'|^2|V'|}\left(2\delta p^2 m_V|J| \;+3L\sqrt{r}\cdot|U'|m_V\right)\\ &\le \frac{2}{\varrho|U'|^2|V'|}\times 2\delta p^2 m_V\times\frac{rm_U}{2}+\frac{6L\sqrt{r}\cdot|U'|m_V}{\varrho|U'|^2|V'|}\\ &\le \frac{2\delta p^2 {m_U}^2 m_V}{|U'|^2|V'|}+\frac{6Lm_Vm_U\sqrt{r}}{r|U'|\cdot|V'|}\\ &\le \frac{16\delta}{\varepsilon^3}+\frac{24L}{\sqrt{r}\cdot\varepsilon^2}\\ &\le \frac{1}{2}\cdot\left(\frac{\varepsilon}{2}\right)^2+\frac{1}{2}\cdot\left(\frac{\varepsilon}{2}\right)^2\\ &=\left(\frac{\varepsilon}{2}\right)^2.\end{aligned}$$

The last inequality follows because of our choice of δ and $r \ge r_0 \ge r_A$. □

4.2. Proof of Lemma 3.3. For this proof, the reader may find it convenient to recall the hierarchy of the constants given in (3.12). Recall we have

$$p = p(H) = p(G) + O_1(5\varepsilon') \ge \frac{1}{2}\varepsilon^3. \tag{4.15}$$

Note that (4.15) is guaranteed to hold for the graph H that we obtain after preprocessing H as described in §3.1.1 since $p(G) \ge \varepsilon^3$ in this case. As in §3.1.2 we assume $U \cup V$ is the vertex set of H and set $m_U = |U|$ and $m_V = |V|$.

Suppose that property $\mathcal{P}(J, \delta)$ fails for H. Thus

$$\sum_{\{u,u'\}\in J} |d_H(u,u') - p^2 m_V| > \delta p^2 m_V |J|\,. \tag{4.16}$$

Let us first observe that

$$\begin{aligned}&\sum_{\{u,u'\}\in J}\left(d_H(u,u')-p^2m_V\right)^2\\ &\quad=\sum_{\{u,u'\}\in J}\left(d_H(u,u')^2-2d_H(u,u')p^2m_V+p^4{m_V}^2\right)\\ &\quad=\sum_{\{u,u'\}\in J}d_H(u,u')^2-2p^2m_V\sum_{\{u,u'\}\in J}d_H(u,u')+p^4{m_V}^2|J|\,.\end{aligned}\tag{4.17}$$

However,

$$\begin{aligned}\sum_{\{u,u'\}\in J}d_H(u,u') &= \sum_{\{u,u'\}\in J}|H(u)\cap H(u')|\\ &=\sum_{v\in V}|J_v| = \sum_{v\in V}(1+O_1(\eta))(1+O_1(\mu))^2p^2|J|\,.\end{aligned}$$

In the last inequality we used that due to preprocessing, cf. (4.1), $|H(v)| = (1 + O_1(\mu))pm_U$ and, hence (a) in Lemma 2.4 gives (note that $(1 + O_1(\mu))p \ge \eta$)

$$|J_v| = (1+O_1(\eta))(1+O_1(\mu))^2p^2|J| = (1+O_1(3\mu))p^2|J|\,.$$

Therefore

$$\sum_{\{u,u'\}\in J} d_H(u,u') = (1+O_1(3\mu))p^2 m_V|J| \ . \tag{4.18}$$

From (4.17) and (4.18) we obtain that

$$\begin{aligned}\sum_{\{u,u'\}\in J} &\big(d_H(u,u') - p^2 m_V\big)^2 \\ &= \sum_{\{u,u'\}\in J} d_H(u,u')^2 - (1+O_1(6\mu))p^4 {m_V}^2|J| \\ &\le \sum_{\{u,u'\}\in J} d_H(u,u')^2 - \left(1-\frac{\delta^2}{2}\right) p^4 {m_V}^2|J| \ .\end{aligned} \tag{4.19}$$

The last inequality holds true due to our choices of δ and μ, cf. (3.7, 3.8). On the other hand, in view of (4.16), we have by the Cauchy–Schwarz inequality that

$$\begin{aligned}\sum_{\{u,u'\}\in J} \big(d_H(u,u') - p^2 m_V\big)^2 &\ge \frac{1}{|J|}\Big(\sum_{\{u,u'\}\in J} |d_H(u,u') - p^2 m_V|\Big)^2 \\ &> \frac{1}{|J|}\big(\delta|J|p^2 m_V\big)^2 = \delta^2 p^4 {m_V}^2|J| \ .\end{aligned} \tag{4.20}$$

Comparing (4.19) and (4.20), we deduce that

$$\sum_{\{u,u'\}\in J} d_H(u,u')^2 \ge \left(1+\frac{\delta^2}{2}\right) p^4 {m_V}^2|J| \ . \tag{4.21}$$

We shall now evaluate the sum on the left-hand side of (4.21) in terms of the matrix $M = (m(e,v))_{e,v}$ defined in §3.1.

Clearly, if $e = \{u,u'\} \in J$, then

$$d_H(u,u') = |H(u)\cap H(u')| = \sum_{v\in V} m(e,v),$$

and hence

$$d_H(u,u')^2 = \sum_{v\in V}\sum_{v'\in V} m(e,v)m(e,v') \ .$$

Therefore

$$\begin{aligned}\sum_{\{u,u'\}\in J} d_H(u,u')^2 &= \sum_{v\in V}\sum_{e\in J} m(e,v) \sum_{v'\in V} m(e,v') \\ &= \sum_{v\in V}\sum_{e\in J_v}\sum_{v'\in V} m(e,v') = \sum_{v\in V}\sum_{v'\in V}\sum_{e\in J_v} m(e,v') \ .\end{aligned} \tag{4.22}$$

Comparing (4.21) and (4.22) we infer that

$$\sum_{v\in V}\sum_{v'\in V}\sum_{e\in J_v} m(e,v') \ge \left(1+\frac{\delta^2}{2}\right) p^4 {m_V}^2|J|$$

and, hence, there is a vertex $v_0 \in V$ for which we have

$$\sum_{v' \in V} \sum_{e \in J_{v_0}} m(e, v') \geq \left(1 + \frac{\delta^2}{2}\right) p^4 m_V |J| \; . \tag{4.23}$$

Following the algorithm, we fix such a vertex v_0. We now *set*

$$V' := \left\{ v' \in V \; : \; \sum_{e \in J_{v_0}} m(e, v') \geq \left(1 + \frac{\delta^2}{4}\right) p^4 |J| \; \right\} . \tag{4.24}$$

As in the algorithm, we put $U' = H(v_0)$. One may prove that both U' and V' are large sets. The proof is postponed for the next section.

CLAIM 4.6. $|U'| \geq \varepsilon_1 m_U$ *and* $|V'| \geq \varepsilon_1 m_V$.

Recall that we defined $\varepsilon_1 = \frac{1}{4}\left(\frac{\varepsilon}{2}\right)^{16}$ in (3.11) to satisfy $\varepsilon' \ll \varepsilon_1 \ll \delta^2$. Before we proceed, using the definition of matrix M we observe that for all $v' \in V$ we have

$$\sum_{e \in J_{v_0}} m(e, v') = e(J[H(v_0) \cap H(v')]) \; . \tag{4.25}$$

Combining (4.25) and the fact that the edges of J are extremely well distributed we shall now provide a lower bound on $d_H(v_0, v')$ where v_0 is the vertex fixed above and v' is an arbitrary vertex of V'.

CLAIM 4.7. *For all $v' \in V'$, we have*

$$d_H(v_0, v') \geq \left(1 + \frac{\delta^2}{12}\right) p^2 m_U \; . \tag{4.26}$$

The proof of Claim 4.7 is given in next section. Since $U' = H(v_0)$, it follows immediately from Claim 4.7 that

$$e_H(U', V') \geq \left(1 + \frac{\delta^2}{12}\right) p^2 m_U |V'| \tag{4.27}$$

which implies that

$$\begin{aligned} d_H(U', V') = \frac{e_H(U', V')}{|U'||V'|} &\geq \frac{(1 + \delta^2/12) p^2 m_U |V'|}{(1 + \mu) p m_U |V'|} \\ &\geq \left(1 + \frac{\delta^2}{14}\right) p > p + \varepsilon_1 \; . \end{aligned} \tag{4.28}$$

Since we have already proved that $|U'| \geq \varepsilon_1 m_U$ and $|V'| \geq \varepsilon_1 m_V$ (see Claim 4.6), inequality (4.28) tells us that (U', V') is a witness to the ε_1-irregularity of H.

We shall now prove that (U', V') is in fact a witness to the ε'-irregularity of G. We have

$$|U'| \geq \varepsilon_1 m_U \geq \varepsilon_1 (1 - 2\varepsilon') m \geq \varepsilon' m \tag{4.29}$$

and, similarly,

$$|V'| \geq \varepsilon_1 m_V \geq \varepsilon_1 (1 - 2\varepsilon') m \geq \varepsilon' m \; . \tag{4.30}$$

Because of (3.3) and (4.28), we have

$$d(U', V') > p + \varepsilon_1 \geq p(G) - 5\varepsilon' + \varepsilon_1 \geq p(G) + \varepsilon'. \tag{4.31}$$

In view of (4.29,4.30) inequality (4.31) implies that (U', V') is indeed a witness to the ε'-irregularity of G, as required.

4.2.1. Proofs of Claims 4.6 and 4.7. Here we give proofs of Claims 4.6 and 4.7.

Proof of Claim 4.6. Since $U' = H(v_0)$, estimates (4.1) and our definition of ε_1 imply $|U'| = (1 + O_1(\mu))pm_U \geq \varepsilon_1 m_U$.

Now we will give a lower bound on $|V'|$. By the definition of V', we have

$$\begin{aligned}\sum_{v' \in V} \sum_{e \in J_{v_0}} m(e, v') &= \sum_{v' \notin V'} \sum_{e \in J_{v_0}} m(e, v') + \sum_{v' \in V'} \sum_{e \in J_{v_0}} m(e, v') \\ &< \left(1 + \frac{\delta^2}{4}\right) p^4 |J| (m_V - |V'|) + |V'| e(J_{v'}) \; . \qquad (4.32)\end{aligned}$$

Since $|H(v')| = (1 + O_1(\mu))pm_V$, cf. (4.1), (a) in Lemma 2.4 implies (note $(1 + O_1(\mu))p \geq \eta$)

$$e(J_{v'}) = e(J[H(v')]) = (1 + O_1(\eta))(1 + O_1(\mu))^2 p^2 |J| = (1 + O_1(3\mu))p^2 |J| \; .$$

Thus continuing with (4.32) we can write

$$\sum_{v' \in V} \sum_{e \in J_{v_0}} m(e, v') < \left(1 + \frac{\delta^2}{4}\right) p^4 |J| m_V + (1 + O_1(3\mu))|V'| p^2 |J| \; . \qquad (4.33)$$

Comparing (4.23) and (4.33), we obtain

$$2|V'| \geq (1 + O_1(3\mu))|V'| \geq \frac{1}{4} \delta^2 p^2 m_V$$

and this, using definition of δ and ε_1, gives

$$|V'| \geq \frac{1}{8} \delta^2 p^2 m_V \geq \frac{1}{128} \left(\frac{\varepsilon}{2}\right)^{10} \varepsilon^6 m_V \geq \varepsilon_1 m_V \; ,$$

as required. □

Proof of Claim 4.7. Suppose to the contrary that (4.26) fails, i.e., $d_H(v_0, v') < (1 + \delta^2/12)p^2 m_U$. We distinguish two cases: If $d_H(v_0, v') \geq \eta m_U$, then using (a) in Lemma 2.4 for $H(v_0) \cap H(v')$ implies

$$e(J[H(v_0) \cap H(v')]) < (1 + O_1(\eta)) \left(1 + \frac{\delta^2}{12}\right)^2 p^4 |J| \; .$$

If, on the other hand, $d(v_0, v') < \eta m_U$ we have (cf. (b) in Lemma 2.4)

$$e(J[H(v_0) \cap H(v')]) < 2\eta^2 |J| \; .$$

In either case, we have

$$e(J[H(v_0) \cap H(v')]) < \left(1 + \frac{\delta^2}{4}\right) p^4 |J| \; . \qquad (4.34)$$

However, in view of the definition of V' (see (4.24, 4.25)), inequality (4.34) cannot hold. This contradiction shows that (4.26) must indeed hold. □

Acknowledgements. The authors are grateful to the referee for his or her detailed comments.

REFERENCES

[1] N. Alon, R. A. Duke, H. Lefmann, V. Rödl, and R. Yuster, *The algorithmic aspects of the regularity lemma* (*extended abstract*), 33rd Annual Symposium on Foundations of Computer Science (Pittsburgh, Pennsylvania), IEEE Comput. Soc. Press, 1992, pp. 473–481.

[2] ———, *The algorithmic aspects of the regularity lemma*, Journal of Algorithms, 16(1) (1994), pp. 80–109.

[3] N. Alon, E. Fischer, M. Krivelevich, and M. Szegedy, *Efficient testing of large graphs* (*extended abstract*), 40th Annual Symposium on Foundations of Computer Science (New York City, NY), IEEE Comput. Soc. Press, 1999, pp. 656–666.

[4] ———, *Efficient testing of large graphs*, Combinatorica, 20(1) (2000), pp. 451–476.

[5] N. Alon and J. Spencer, *The probabilistic method*, Wiley-Interscience Series in Discrete Mathematics, John Wiley & Sons, New York, 1992.

[6] F. R. K. Chung and R. L. Graham, *Quasi-random set systems*, J. Amer. Math. Soc., 4(1) (1991), pp. 151–196.

[7] ———, *Quasi-random tournaments*, J. Graph Theory, 15(2) (1991), pp. 173–198.

[8] ———, *Quasi-random subsets of* Z_n, J. Combin. Theory Ser. A, 61(1) (1992), pp. 64–86.

[9] ———, *Sparse quasi-random graphs*, Combinatorica, 22(2) (2002), pp. 217–244.

[10] F. R. K. Chung, R. L. Graham, and R. M. Wilson, *Quasi-random graphs*, Combinatorica, 9(4) (1989), pp. 345–362.

[11] Don Coppersmith and Shmuel Winograd, *Matrix multiplication via arithmetic progressions*, J. Symbolic Comput., 9(3) (1990), pp. 251–280.

[12] A. Czumaj and C. Sohler, *Testing hypergraph coloring*, Proc. of ICALP, 2001, pp. 493–505.

[13] A. Czygrinow, S. Poljak, and V. Rödl, *Constructive quasi-Ramsey numbers and tournament ranking*, SIAM Journal on Discrete Mathematics, 12(1) (1999), pp. 48–63.

[14] W. Fernandez de la Vega, *MAX-CUT has a randomized approximation scheme in dense graphs*, Random Structures Algorithms, 8(3) (1996), pp. 187–198.

[15] R. A. Duke, H. Lefmann, and V. Rödl, *A fast approximation algorithm for computing the frequencies of subgraphs in a given graph*, SIAM Journal on Computing, 24(3) (1995), pp. 598–620.

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

[17] P. Erdős and J. Spencer, *Probabilistic methods in combinatorics*, Akademiai Kiado, Budapest, 1974, 106pp.

[18] Alan Frieze and Ravi Kannan, *The regularity lemma and approximation schemes for dense problems*, 37th Annual Symposium on Foundations of Computer Science (Burlington, VT, 1996), IEEE Comput. Soc. Press, Los Alamitos, CA, 1996, pp. 12–20.

[19] ———, *Quick approximation to matrices and applications*, Combinatorica, 19(2) (1999), pp. 175–220.

[20] ———, *A simple algorithm for constructing Szemerédi's regularity partition*, Electronic Journal of Combinatorics, 6(1) (1999), Research Paper 17, 7 pp. (electronic).

[21] Oded Goldreich, *Combinatorial property testing* (*a survey*), Randomization methods in algorithm design (Princeton, NJ, 1997), Amer. Math. Soc., Providence, RI, 1999, pp. 45–59.

[22] Oded Goldreich, Shafi Goldwasser, and Dana Ron, *Property testing and its connection to learning and approximation*, 37th Annual Symposium on Foundations of Computer Science (Burlington, VT, 1996), IEEE Comput. Soc. Press, Los Alamitos, CA, 1996, pp. 339–348.

[23] ———, *Property testing and its connection to learning and approximation*, Journal of the Association for Computing Machinery, 45(4) (1998), pp. 653–750.

[24] Y. Kohayakawa and V. Rödl, *Szemerédi's regularity lemma and quasi-randomness*, Recent advances in algorithms and combinatorics, CMS Books Math./Ouvrages Math. SMC, vol. 11, Springer, New York, 2003, pp. 289–351.

[25] J. Komlós and M. Simonovits, *Szemerédi's regularity lemma and its applications in graph theory*, Combinatorics—Paul Erdős is eighty, vol. 2 (Keszthely, 1993) (D. Miklós, V. T. Sós, and T. Szőnyi, eds.), Bolyai Society Mathematical Studies, vol. 2, János Bolyai Mathematical Society, Budapest, 1996, pp. 295–352.

[26] A. Lubotzky, R. Phillips, and P. Sarnak, *Explicit expanders and the Ramanujan conjectures*, Proceedings of the Eighteenth Annual ACM Symposium on Theory of Computing (STOC '86) (Berkeley, California), ACM, 1986, pp. 240–246.

[27] ______, *Ramanujan graphs*, Combinatorica, 8(3) (1988), pp. 261–277.
[28] M. Luby and A. Wigderson, *Pairwise independence and derandomization*, Tech. Report 95-035, International Computer Science Institute, July 1995.
[29] J. Spencer, *Optimal ranking of tournaments*, Networks, 1 (1971), pp. 135–138.
[30] ______, *Nonconstructive methods in discrete mathematics*, Studies in Combinatorics (G.C. Rota, ed.), Mathematical Association of America, Washington, D.C., 1978, pp. 142–178.
[31] E. Szemerédi, *Regular partitions of graphs*, Problèmes Combinatoires et Théorie des Graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976) (Paris), Colloques Internationaux CNRS n. 260, 1978, pp. 399–401.
[32] A. G. Thomason, *Pseudorandom graphs*, Random graphs '85 (Poznań, 1985), North-Holland Math. Stud., vol. 144, North-Holland, Amsterdam–New York, 1987, pp. 307–331.