An optimal algorithm for checking regularity (Extended Abstract) *

Y. Kohayakawa † V. Rödl ‡ 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 [18]. 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 [1, 2], 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.

1 Introduction and the Main Result

Szemerédi's regularity lemma [18] is a fundamental result in graph theory (see [15] 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 [7]) is the time required to square a 0–1 matrix over the integers. More recently, Frieze and Kannan [9] (see also [10]) 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 extended abstract, we present a *deterministic* algorithm that solves this problem in $O(m^2)$ time. This algorithm immediately gives an algorithm for Szemerédi's regularity lemma that finds a regular partition of an n-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 $G = (V_1, V_2; E)$ be a bipartite graph. If $\emptyset \neq U_i \subset V_i$ $(i = 1, 2)$, the *density* of (U_1, U_2) in G is $d(U_1, U_2) = e(U_1, U_2)/|U_1||U_2|$, where we write $e(U_1, U_2)$ for the number of edges with one endpoint in U_1 and the other endpoint in U_2. We say that G is *ε-regular* if, for all $U_i \subset V_i$ $(i = 1, 2)$ with $|U_i| \geq \varepsilon|V_i|$, we have

$$|d(U_1, U_2) - d(V_1, V_2)| \leq \varepsilon. \tag{1.1}$$

If G is not ε-regular and a certain pair (U_1, U_2) certifies this fact, then we say that (U_1, U_2) is a *witness* to the *ε-irregularity* of G. If $\Gamma = (V, E)$ is a graph and $V = \bigcup_{i=1}^k V_i$ is a partition of V such that $\big||V_i| - |V_j|\big| \leq 1$ for all i and j, then we say that the partition $(V_i)_{i=1}^k$ is an *ε-regular partition* for Γ if $\geq (1-\varepsilon)\binom{k}{2}$ pairs (V_i, V_j), where $1 \leq i < j \leq k$, are ε-regular. 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 $k_0 \leq k \leq K_0(\varepsilon, k_0)$.*

The algorithmic version of Theorem 1.1, due to

*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 (Proj. 96/04505–2). The second author was partially supported by NSF Grant 0071261. The third author was partially supported by NSF grant DMS–9970622. The collaboration of these authors is supported by a CNPq/NSF cooperative grant.

†Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão 1010, 05508–900 São Paulo, 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

Alon, Duke, Lefmann, Rödl, and Yuster [1, 2], may be stated as follows.

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. *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 [8] suffices.

Theorem 1.3. *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* $G = (V_1, V_2; E)$ *with* $|V_1| = |V_2| = m \geq (2/\varepsilon)^5$, *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 [11] for a somewhat different approach to verifying regularity, based on singular values of matrices.) Our main result is an improvement of Theorem 1.3 above, and may be stated as follows.

Theorem 1.4. (**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* $G = (V_1, V_2; E)$ *with* $|V_1| = |V_2| = m \geq m_0(\varepsilon)$, *it either correctly asserts that* G *is* ε*-regular, or else it returns a witness for the* ε'*-irregularity of* G, *where* $\varepsilon' = \varepsilon'_{\mathcal{A}}(\varepsilon) = \varepsilon^{20}/5^4 2^{32}$. *The running time of* $\mathcal{A}$ *is* $O(m^2)$.

We describe our algorithm $\mathcal{A}$ in §2.2. Deriving an algorithm for the regularity lemma from Theorem 1.4 is standard, and we omit the details.

Corollary 1.1. *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 1.1, one obtains algorithms with optimal time complexity* $O(n^2)$. We also observe that a similar improvement may be obtained from Theorem 1.4 for the subgraph counting algorithm given in Duke, Lefmann, and Rödl [8]. We explicitly state examples of these improvements at the end of this introduction.

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., [12] and [13, 14]). 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 1.3. Again, a $O(n^2)$ time algorithm follows immediately from Theorem 1.4.

Finally, we mention that one may prove a 'non-bipartite version' of Theorem 1.4. 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 [6]. Moreover, if Γ is not quasi-random, then our algorithm will produce a suitable witness proving this (an induced subgraph with $\Omega(n)$ vertices whose density deviates substantially from the density of Γ).

Local conditions for regularity. One may prove Theorem 1.3 by considering a certain 'local condition' on $G = (V_1, V_2; E)$, $|V_1| = |V_2| = m$, 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:

$$(1.2) \qquad \sum_{x,\, y \in V_1} \left| d_G(x, y) - p(G)^2 m \right| \leq \delta p(G)^2 m \ ,$$

where $d_G(x, y) = |N_G(x) \cap N_G(y)|$ is the so-called *codegree* of x and y, and $p(G) = |E|/|V_1||V_2| = |E|/m^2$ 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. Some of the ideas described in this paragraph have appeared in the literature under many guises—basically, we are obtaining a somewhat surprising amount of information from 'pairwise independence'. We do not go into details here.

The key idea in the proof of Theorem 1.4 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 §2.2.3).

Algorithmic applications. As we stated above, our algorithm immediately implies an improvement on deterministic algorithms which are based on the Szemerédi regularity lemma. Here we present three typical examples of algorithms for finding and counting given subgraphs. For more algorithmic applications of the regularity lemma see [2] and [8].

The first two propositions describe problems of finding specific subgraphs in a given graph.

PROPOSITION. *For every $\varepsilon > 0$ and every integer h, there exists a positive integer $n_0 = n_0(\varepsilon, h)$ such that for every graph H with h vertices and chromatic number $\chi(H)$, there exists a deterministic algorithm of time complexity $O(n^2)$ that, given a graph $G = (V, E)$ with $n > n_0$ vertices and minimum degree $d > \frac{\chi(H)-1}{\chi(H)} \cdot n$, finds a set of $(1-\varepsilon)\frac{n}{h}$ vertex disjoint copies of H in G.*

For the proof of existence see [5].

PROPOSITION. *Let $k \geq 3$ be an integer and $c > 0$, $\varepsilon > 0$ be constants. Then there exists an integer $n_0 = n_0(k, \varepsilon)$ and a function $f(k, \varepsilon)$ such that if $G = (V, E)$ is a graph with $n \geq n_0$ vertices and cn^2 edges, then either*

(i) there exists a subgraph $G' = (V', E')$ of G with $\chi(G') \geq k$ and $|V'| \leq f(k, \varepsilon)$, or

(ii) there exists a set $E'' \subseteq E$ with $|E''| \leq \varepsilon n^2$ such that the subgraph $G'' = (V, E \setminus E'')$ satisfies $\chi(G'') \leq k - 1$.

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

For the algorithms proving both propositions above see [2].

To describe our third example let us introduce some notation. We will follow [8]. Let $G = (V, E)$ be a labeled graph on n vertices whose vertex set $V = \{v_1, \ldots, v_n\}$ is ordered by $v_1 < \cdots < v_n$. Let the set $W = \{w_1, \ldots, w_k\}, k \leq n$, be ordered by $w_1 < \cdots < w_k$. 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, $1 \leq i, j \leq n$, if $w_i < w_j$, then $\phi(w_i) < \phi(w_j)$. Let $H_1, \ldots, H_t$, where $t = 2^{\binom{k}{2}}$, be a list of all labeled graphs on the set W and let $\sigma_k(G) = (h_1, \ldots, h_t)$ be the t-dimensional vector in which for each $i, 1 \leq i \leq t$, 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 [8].

PROPOSITION. *Let $k \geq 4$ be in integer. There is an algorithm whose input is a labeled, ordered graph on n vertices and a list of all labeled graphs on an ordered set of k vertices and whose output in an approximation $\overline{\sigma}_k(G) = (\overline{h}_1, \ldots, \overline{h}_t), t = 2^{\binom{k}{2}}$, to the vector $\sigma_k(G) = (h_1, \ldots, h_t)$ with the property that $\sum_{1 \leq i \leq t} |h_i - \overline{h}_i| \leq \frac{1}{t}\sum_{1 \leq i \leq t} h_i$. The algorithm runs in $O(n^{2+1/\log\log n})$ time.*

2 Preliminaries and the Algorithm

2.1 Auxiliary results on expander graphs. Let $0 < \varrho \leq 1$ and $L > 0$ be fixed. We say that a graph J on m vertices is *(ϱ, L)-uniform* if, for any U, $W \subset V(J)$ such that $U \cap W = \emptyset$ and $1 \leq |U| \leq |W| \leq r|U|$, where $r = \varrho m$, we have

$$\big|e_J(U, W) - \varrho|U||W|\big| \leq L\sqrt{r|U||W|}. \tag{2.3}$$

Here $e_J(U, W)$ stands for the number of edges between the sets U and W in the graph J. The celebrated Ramanujan graphs of Lubotzky, Phillips, and Sarnak [16, 17] are explicitly constructible examples of linear-sized $(\varrho, 2)$-uniform graphs. We shall make crucial use of the fact that linear-sized $(\varrho, O(1))$-uniform graphs may be constructed efficiently.

The following three lemmas are easy.

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

The Ramanujan graphs $X = X^{p,q}$ constructed in [16, 17] depend on primes p and q, which have to satisfy certain simple arithmetical conditions. These graphs $X^{p,q}$ are $(p+1)$-regular and have $n = q(q^2-1)/2$ vertices. 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 integers, which we may suppose to be large. The following lemma asserts that this can be done efficiently.

LEMMA 2.2. *There exists a constant r_1 such that for all $r_0 \geq r_1$ there is an $m_1 = m_1(r_0)$ such that the following holds. Let $m \geq m_1$ be given. Then we may explicitly construct an adjacency list representation of a particular $(\varrho, 5)$-uniform graph J on $V(J) = \{1, \ldots, m\}$ with average degree $r = \varrho m$ satisfying $r_0 \leq r \leq 2r_0$ in time $O(m(\log m)^{O(1)})$.*

To construct the graph J as in Lemma 2.2, one first finds suitable primes p and q so that $n = q(q^2-1)/2$ is only a little larger than m (the existence of such primes follows from Dirichlet's theorem on primes in arithmetic progressions). We have plenty of time for finding these p and q. We now construct $X^{p,q}$ and then take an arbitrary induced subgraph $J \subset X^{p,q}$ on m vertices; the $(\varrho, 5)$-uniformity of J follows from Lemma 2.1. We omit the details.

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

LEMMA 2.3. *Let $\eta_0 > 0$ and $L > 0$ be given. Then there is an $r_0 = r_0(\eta_0, L)$ such that any m-vertex (ϱ, L)-uniform graph J with $\varrho m \geq r_0$ has the following property: if $X \subset V(J)$ is such that $\eta = |X|/n \geq \eta_0$, then $e(J[X]) = (1 + O_1(\eta_0))\eta^2 e(J)$. Here, as usual, $e(J)$ stands for the number of edges of the graph J.*

2.2 The algorithm. Our algorithm $\mathcal{A}$ first preprocesses the input graph G and only then runs a more sophisticated procedure. We first describe the preprocessing stage quite informally and then proceed to describe the heart of Algorithm $\mathcal{A}$ in detail in §2.2.3.

However, we have to start with setting up some constants.

2.2.1 Definition of the constants. Algorithm $\mathcal{A}$ is given a constant $0 < \varepsilon < 1$. We first let

$$(2.4) \quad \varepsilon_0 = \frac{1}{2}\varepsilon, \qquad \delta = \frac{1}{4}\varepsilon_0^5, \quad \text{and} \quad r_A = 1500\varepsilon_0^{-8}.$$

Now we put

$$(2.5) \quad \theta = \frac{1}{4}\delta^2, \quad \varepsilon'' = \min\left\{\frac{1}{25}\theta, \frac{1}{20}\varepsilon^3, \frac{1}{15}\varepsilon_0^{5/2}\right\},$$
$$\text{and} \quad \varepsilon' = (\varepsilon'')^2 \leq \min\left\{\frac{1}{400}\varepsilon^6, \frac{1}{160}\varepsilon_0^5\right\}.$$

We also let

$$(2.6) \quad \varepsilon_1 = \min\left\{\frac{1}{32}\delta^2\varepsilon^6, \frac{1}{10}\varepsilon^3\theta\right\} \quad \text{and}$$
$$\eta_0 = \min\left\{\frac{1}{3}\varepsilon'', \frac{1}{8}\varepsilon^6, \frac{1}{4}\theta\right\}.$$

We set $L = 5$ and let $r_B = r_0(\eta_0, L)$ be as given in Lemma 2.3. Finally, we let

$$(2.7) \qquad r_0 = \max\{r_A, r_B\}.$$

The reader may find it useful to keep in mind the following relationship among the constants:

$$(2.8) \quad \varepsilon' \ll \varepsilon'' \ll \varepsilon, \qquad \eta_0 \ll \varepsilon'' \ll \theta \ll \delta,$$
$$\text{and} \quad \varepsilon' \ll \varepsilon_1 \ll \delta,\ \varepsilon,\ \theta.$$

2.2.2 The preprocessing stage. We first observe that if the graph $G = (V_1, V_2; E)$ is such that $p(G) = d(V_1, V_2) = |E|/m^2 \leq \varepsilon^3$, then G is ε-regular (we omit the proof of this standard observation). Therefore, we suppose that

$$(2.9) \qquad p(G) > \varepsilon^3.$$

We now observe that we may remove some vertices of our graph G so that we either obtain a subgraph $H \subset G$ 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 in the following lemma (a sketch of the proof is given in the appendix).

LEMMA 2.4. *Suppose $G = (V_1, V_2; E)$ is a balanced bipartite graph with $|V_1| = |V_2| = m$, and suppose that (2.9) holds. There is a procedure that runs in time $O(m^2)$ such that either (i) our procedure produces a bipartite subgraph $H \subset G$, say $H = (U_1, U_2; F)$, such that (a) if $m_i = |U_i|$ $(i = 1, 2)$, then*

$$(2.10) \qquad (1 - 2\varepsilon')m < m_i \leq m,$$

(b) if $p(H) = |F|/m_1 m_2$, then

$$(2.11) \qquad p(H) = p(G) + O_1(5\varepsilon'),$$

and (c) for all $u_i \in U_i$ $(i = 1, 2)$, we have

$$(2.12) \qquad d(u_i) = (p(H) + O_1(10\varepsilon'))m_{3-i},$$

where $d(u_i)$ stands, as usual, for the degree of the vertex u_i in the graph H, or else (ii) our procedure produces a witness to the ε'-irregularity of G.

For convenience, we let $\Xi(m, \varepsilon')$ be the family of subgraphs H of G that satisfy (a)–(c) in (i) of Lemma 2.4 above.

2.2.3 The main procedure. In view of Lemma 2.4, we now suppose that we have a subgraph $H = (U_1, U_2; F) \subset G$ with $H \in \Xi(m, \varepsilon')$. Our algorithm $\mathcal{A}$ now generates a (ϱ, L)-uniform graph J with vertex set U_1 and with average degree $r = \varrho m_1$ satisfying (cf. Lemma 2.2)

$$r_0 \le r \le 2r_0 \ . \tag{2.13}$$

We now introduce some notation. We say that property $\mathcal{P}(J, \delta)$ holds if

$$\sum_{\{x,y\}\in J} \left| d_H(x,y) - p(H)^2 m_2 \right| \le \delta e(J) p(H)^2 m_2, \tag{2.14}$$

where we write $\{x, y\} \in J$ to mean that $\{x, y\}$ is an edge of J, $e(J)$ stands for the number of edges J, and $p(H) = e(H)/|U_1||U_2| = |F|/m^2$. Moreover, let us write $J(u)$ $(u \in U_2)$ for the graph $J[N(u)]$ induced by the neighbourhood $N(u)$ of u in H. We define a 0–1 matrix $M = (M(e, u))_{e,u}$ indexed by $E(J) \times U_2$ as follows:

$$M(e,u) = \begin{cases} 1 & \text{if } e \in J(u) \\ 0 & \text{otherwise.} \end{cases} \tag{2.15}$$

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

Algorithm $\mathcal{A}$ now proceeds as follows:

1. $\mathcal{A}$ verifies whether $\mathcal{P}(J, \delta)$ holds; if it does, then $\mathcal{A}$ claims that G is ε-regular and halts.
2. If $\mathcal{P}(J, \delta)$ does not hold, then $\mathcal{A}$ computes the matrix $M = (M(e, u))_{e,u}$ defined above. Moreover, $\mathcal{A}$ finds and fixes a vertex $u_0 \in U_2$ for which we have

   $$\sum_{u'\in U_2} \sum_{e\in J(u_0)} M(e,u') \ge \left(1 + \frac{\delta^2}{2}\right) p(H)^4 m_2 e(J). \tag{2.16}$$

   [The existence of such a vertex u_0 is proved later.]
3. $\mathcal{A}$ now sets $U = N(u_0)$ and sets

   $$W = W_\theta = \left\{ u' \in U_2 \ : \ u' \text{satisfies (2.18)} \right\} \tag{2.17}$$

   where

   $$\sum_{e\in J(u_0)} M(e,u') \ge (1+\theta) p(H)^4 e(J). \tag{2.18}$$
4. $\mathcal{A}$ outputs (U, W) and claims that this is a witness to the ε'-irregularity of G.

2.2.4 Correctness and analysis of Algorithm $\mathcal{A}$. The correctness of $\mathcal{A}$ follows from Lemmas 2.5 and 2.6 below. The first lemma says that if Algorithm $\mathcal{A}$ claims that G is ε-regular in Step 1, then this is indeed the case.

LEMMA 2.5. *If property $\mathcal{P}(J, \delta)$ holds, then G is ε-regular.*

LEMMA 2.6. *If property $\mathcal{P}(J, \delta)$ fails, then G is not ε'-regular, and the pair (U, W) produced by Algorithm $\mathcal{A}$ is indeed a witness for the ε'-irregularity of G.*

We shall prove the two lemmas above in §3. The next simple lemma tells us that $\mathcal{A}$ has time complexity $O(m^2)$.

LEMMA 2.7. *Algorithm $\mathcal{A}$ described in §2.2 runs in time $O(m^2)$.*

Proof. This is quite straightforward and we omit the proof. Let us only observe that the fact that J has $O(m)$ edges is crucial, as it is this that allows us to check (2.14) in $O(m^2)$ time.

3 Proofs of the Main Lemmas

3.1 Proof of Lemma 2.5. We say that the graph H has *property $\mathcal{Q}(J, \delta)$* if the following inequality holds true:

$$\sum_{\{x,y\}\in J} \left| d(x,y) + p(H)^2 m_2 - (d(x) + d(y)) p(H) \right| \le \delta p(H)^2 m_2 e(J). \tag{3.19}$$

CLAIM 3.1. *Let $\delta > 0$ be fixed. If a graph $H \in \Xi(m, \varepsilon')$ has property $\mathcal{P}(J, \delta)$, then it has property $\mathcal{Q}(J, 2\delta)$.*

The proof of Claim 3.1 is quite straightforward and can be found in the appendix. In view of the above claim, to prove Lemma 2.5 it is enough to show that $\mathcal{Q}(J, 2\delta)$ implies the ε_0-regularity of H. Let $A = (a_{k,l})_{k,l}$ be a matrix indexed by $U_1 \times U_2$, with entries

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

Moreover, for $i \in U_1$, let $v_i = (a_{i,1}, \dots, a_{i,m_2})$ be the ith row of A. Straightforward calculations show that $\langle v_i, v_j \rangle = d(i, j) + p(H)^2 m_2 - (d(i) + d(j))p(H)$, and hence the following claim follows easily from the definition of property $\mathcal{Q}$ (see (3.19)).

CLAIM 3.2. *For every $H \in \Xi(m, \varepsilon')$ that has property $\mathcal{Q}(J, 2\delta)$, the rows of A are such that*

$$\sum_{\{i,j\}\in J} |\langle v_i, v_j \rangle| \le 2\delta p(H)^2 m_2 e(J). \tag{3.20}$$

Let $U \subset U_1$ and $W \subset U_2$. To shorten our notation $\sum_{ij\in J}^{U}$ will denote summation over $\{i,j\} \in E(J)$ such that i, $j \in U$. Furthermore, for $i \in U$ let w_i be the restriction of the vector v_i to W, i.e., $w_i = (a_{i,j})_{j\in W}$. We clearly have $\sum_{ij\in J}^{U} |\langle v_i, v_j\rangle| \le \sum_{\{i,j\}\in J} |\langle v_i, v_j\rangle|$. We now compare $\sum_{ij\in J}^{U}\langle v_i, v_j\rangle$ with $\sum_{ij\in J}^{U}\langle w_i, w_j\rangle$. We have

$$\sum_{ij\in J}^{U}\langle v_i, v_j\rangle = \sum_{ij\in J}^{U}\langle w_i, w_j\rangle + \sum_{k\notin W}\sum_{ij\in J}^{U} a_{i,k}a_{j,k}. \tag{3.21}$$

For $k \in U_2$ we set $S_k^U = \sum_{ij\in J}^{U} a_{i,k}a_{j,k}$, and proceed to estimate this quantity. For $k \in U_2$ let $U^+ \subset U$ be the set of non-neighbours of the vertex k in U and let $U^- \subset U$ be the set of neighbours of the vertex k in U. Set $\alpha = |U^-|$ and $\beta = |U^+|$. Note $|U| = \alpha + \beta$. Thus, using the definition of the matrix A, we can write

$$\begin{aligned} S_k^U &= e_J(U^-)(1-p(H))^2 + e_J(U^+)p(H)^2 \\ &- e_J(U^-,U^+)p(H)(1-p(H)). \end{aligned} \tag{3.22}$$

The (ϱ, L)-uniformity of J implies the following claim, whose proof is given in the appendix.

Claim 3.3. *For all $k \in W$ and $U \subset U_1$, we have*

$$S_k^U = \frac{\varrho}{2}\big[\alpha(1-p(H)) - \beta p(H)\big]^2 + O_1(3L\sqrt{r}\cdot|U|)\,. \tag{3.23}$$

Identity (3.21) and Claims 3.2 and 3.3 above imply an upper and a lower bound on $\sum_{ij\in J}^{U}\langle w_i, w_j\rangle$ as follows (cf. appendix):

$$\begin{aligned} \sum_{ij\in J}^{U}&\langle w_i, w_j\rangle \\ &\le 2\delta p(H)^2 m_2 e(J) + 3L\sqrt{r}|U|(m_2 - |W|) \end{aligned} \tag{3.24}$$

and

$$\begin{aligned} \sum_{ij\in J}^{U}&\langle w_i, w_j\rangle \\ &\ge \frac{\varrho}{2}|U|^2|W|\big|d(U,W) - p(H)\big|^2 - 3L\sqrt{r}|U|\,|W|. \end{aligned} \tag{3.25}$$

Now we are ready to show that H is ε_0-regular. Fix $U \subset U_1$, $|U| \ge \varepsilon_0|U_1| = \varepsilon_0 m_1$, and $W \subset U_2$, $|W| \ge \varepsilon_0|U_2| = \varepsilon_0 m_2$. Set $u = |U|$ and $w = |W|$. Recall that $d(U,W)$ stands for the Szemerédi density of the pair (U,W) and $d(U_1,U_2) = p = p(H)$ in our notation. Combining the lower and upper bounds on $\sum_{ij\in J}^{U}\langle w_i, w_j\rangle$ (cf. (3.24) and (3.25)), we infer

$$\begin{aligned} \big|d(U,W) &- p(H)\big|^2 \\ &\le \frac{2}{\varrho u^2 w}\left(2\delta p(H)^2 m_2 e(J) + 3L\sqrt{r}\cdot u m_2\right) \\ &\le \frac{4\delta p(H)^2 m_2}{\varrho u^2 w}\times\frac{r m_1}{2} + \frac{6L\sqrt{r}\cdot u m_2}{\varrho u^2 w} \\ &\le \frac{2\delta p(H)^2 m_1^2 m_2}{u^2 w} + \frac{6L m_2\sqrt{r}}{\varrho u w} \\ &\le \frac{2\delta p(H)^2}{\varepsilon_0^3} + \frac{6L}{\sqrt{r}\cdot\varepsilon_0^2} \\ &\le \frac{\varepsilon_0^2}{2} + \frac{\varepsilon_0^2}{2} = \varepsilon_0^2. \end{aligned}$$

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

Finally, we observe that the ε_0-regularity of $H = (U_1, U_2; F)$ implies the ε-regularity of $G = (V_1, V_2; F)$. This follows because $\varepsilon' \ll \varepsilon_0 < \varepsilon$. We omit the details.

3.2 Proof of Lemma 2.6.

For this proof, the reader may find it convenient to recall the hierarchy of the constants given in (2.8). Recall we have constants δ and ε with $0 < \delta \ll \varepsilon$. We also have

$$p = p(H) \ge \frac{1}{2}\varepsilon^3. \tag{3.26}$$

Note that (3.26) is guaranteed to hold for the graph H that we obtain after preprocessing H as described in §2.2.

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

$$\sum_{\{x,y\}\in J} |d(x,y) - p^2 m_2| > \delta e(J) p^2 m_2. \tag{3.27}$$

A short calculation (see appendix) shows that we can write

$$\begin{aligned} \sum_{\{x,y\}\in J} &\left(d(x,y) - p^2 m_2\right)^2 \\ &= \sum_{\{x,y\}\in J} d(x,y)^2 - (1 + O_1(6\varepsilon''))p^4 m_2^2 e(J). \end{aligned} \tag{3.28}$$

Applying the Cauchy–Schwarz inequality to the left-hand side of (3.28) and using (3.27) we get

$$\begin{aligned} \sum_{\{x,y\}\in J} &\left(d(x,y) - p^2 m_2\right)^2 \\ &\ge \frac{1}{e(J)}\Big(\sum_{\{x,y\}\in J} |d(x,y) - p^2 m_2|\Big)^2 \\ &> \delta^2 p^4 m_2^2 e(J). \end{aligned} \tag{3.29}$$

Comparing (3.28) and (3.29), and recalling (2.5), we deduce that

$$\sum_{\{x,y\}\in J} d(x,y)^2 \geq \left(1+\frac{\delta^2}{2}\right) p^4 m_2^2 e(J)\ . \tag{3.30}$$

The following claim evaluates the sum on the left-hand side of (3.30) in terms of the matrix $M=(M(e,u))_{e,u}$ defined in §2.2.

CLAIM 3.4.

$$\sum_{\{x,y\}\in J} d(x,y)^2 = \sum_{u\in U_2}\sum_{u'\in U_2}\sum_{e\in J(u)} M(e,u'). \tag{3.31}$$

The proof of (3.31) is included in the appendix. Combining (3.30) and (3.31), we infer that there is a vertex $u_0\in U_2$ for which we have

$$\sum_{u'\in U_2}\sum_{e\in J(u_0)} M(e,u') \geq \left(1+\frac{\delta^2}{2}\right) p^4 m_2 e(J). \tag{3.32}$$

Following the algorithm, we let such a vertex u_0 be fixed. As in the algorithm, we now put $U=N(u_0)$ and consider

$$W=W_\theta=\left\{u'\in U_2\colon \sum_{e\in J(u_0)} M(e,u')\geq (1+\theta)p^4 e(J)\right\}\ . \tag{3.33}$$

One may prove that both U and W are large sets (the proof is given in the appendix).

CLAIM 3.5. *We have $|U|\geq \varepsilon_1 m_1$ and $|W|\geq \varepsilon_1 m_2$.*

Before we proceed, we observe, using the definition of matrix M, that for all $u'\in U_2$ we have

$$\sum_{e\in J(u_0)} M(e,u') = e(J[N(u_0)\cap N(u')]). \tag{3.34}$$

We now infer a lower bound for the codegree $d(u_0,u')$ for the case in which $u'\in W$ using (3.34) and the fact that the edges of J are extremely well distributed.

CLAIM 3.6. *For all $w\in W$, we have $d(u_0,w)\geq (1+\theta/3)\,p^2 m_1$.*

See the appendix for the proof of Claim 3.6. It follows immediately from Claim 3.6 that $e(U,W)\geq \left(1+\frac{\theta}{3}\right)p^2 m_1|W|$, which implies

$$\begin{aligned} d(U,W) &= \frac{e(U,W)}{|U||W|} \geq \frac{(1+\theta/3)p^2 m_1|W|}{(1+\varepsilon'')pm_1|W|} \\ &\geq \left(1+\frac{\theta}{4}\right)p > p+\varepsilon_1. \end{aligned} \tag{3.35}$$

Since we have already proved that $|U|\geq\varepsilon_1 m_1$ and $|W|\geq\varepsilon_1 m_2$ (see Claim 3.5), inequality (3.35) tells us that (U,W) is a witness to the ε_1-irregularity of H.

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

$$|U|\geq\varepsilon_1 m_1\geq\varepsilon_1(1-2\varepsilon')m\geq\varepsilon' m \tag{3.36}$$

and

$$|W|\geq\varepsilon_1 m_2\geq\varepsilon_1(1-2\varepsilon')m\geq\varepsilon' m\ . \tag{3.37}$$

Because of (2.11) and (3.35), we have

$$d(U,W)>p+\varepsilon_1\geq p(G)-5\varepsilon'+\varepsilon_1\geq p(G)+\varepsilon'. \tag{3.38}$$

In view of (3.36), inequality (3.38) tells us that (U,W) is indeed a witness to the ε'-irregularity of G, as required.

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] N. Alon, R. A. Duke, H. Lefmann, V. Rödl, and R. Yuster, *The algorithmic aspects of the regularity lemma*, Journal of Algorithms **16** (1994), no. 1, 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] N. Alon, E. Fischer, M. Krivelevich, and M. Szegedy, *Efficient testing of large graphs*, Combinatorica **20** (2000), no. 4, pp. 451–476.

[5] N. Alon and R. Yuster, *Almost H-factors in dense graphs*, Graphs and Combinatorics **8** (1992), no. 2, pp. 95–102.

[6] F. R. K. Chung, R. L. Graham, and R. M. Wilson, *Quasi-random graphs*, Combinatorica **9** (1989), no. 4, pp. 345–362.

[7] D. Coppersmith and S. Winograd, *Matrix multiplication via arithmetic progressions*, J. Symbolic Comput. **9** (1990), no. 3, pp. 251–280.

[8] 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** (1995), no. 3, pp. 598–620.

[9] A. Frieze and R. 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.

[10] A. Frieze and R. Kannan, *Quick approximation to matrices and applications*, Combinatorica **19** (1999), no. 2, pp. 175–220.

[11] A. Frieze and R. Kannan, *A simple algorithm for constructing Szemerédi's regularity partition*, Electronic Journal of Combinatorics **6** (1999), no. 1, Research Paper 17, 7 pp. (electronic).

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

[13] O. Goldreich, S. Goldwasser, and D. 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.

[14] O. Goldreich, S. Goldwasser, and D. Ron, *Property testing and its connection to learning and approximation*, Journal of the Association for Computing Machinery **45** (1998), no. 4, pp. 653–750.

[15] 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.

[16] 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.

[17] A. Lubotzky, R. Phillips, and P. Sarnak, *Ramanujan graphs*, Combinatorica **8** (1988), pp. 261–277.

[18] 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.

Appendix: Some proofs

We summarize all the postponed proofs and calculations in this appendix.

Proof of Lemma 2.4. (Sketch) We first omit the vertices u in V_2 for which the condition

$$d(u) = (p(G) + O_1(\varepsilon'))m \tag{3.39}$$

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 (ii) in the statement of our lemma. Suppose therefore that the number of such vertices is $< 2\varepsilon' m$. Let $U_2 \subset V_2$ be the resulting subset of V_2, and let $m_2 = |V_2| > (1-2\varepsilon')m$. We now omit the vertices $u \in V_1$ for which the condition

$$d(u) = (p(G) + O_1(\varepsilon'))m_2 \tag{3.40}$$

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 (i) in the statement of our lemma.

Deferred proofs for Lemma 2.5

Here we present the proofs of the claims and identities used in the proof of Lemma 2.5.

Proof of Claim 3.1. Let H satisfy $\mathcal{P}(J,\delta)$ and $H \in \Xi(m,\varepsilon')$. We have

(3.41)
$$\begin{aligned}\sum_{\{x,y\}\in J} &\left|d(x,y) + p(H)^2 m_2 - (d(x)+d(y))p(H)\right| \\ &\leq \sum_{\{x,y\}\in J} \left|d(x,y) - p(H)^2 m_2\right| \\ &\quad + 2p(H)\sum_{x\in U_1}\left|p(H)m_2 - d(x)\right| d_J(x) \\ &\leq 2\delta p(H)^2 m_2 e(J)\end{aligned}$$

since $\varepsilon' \leq \delta/40$ (see (2.5)).

Proof of Claim 3.3. Recall that for two numbers a, b we write $a = O_1(b)$ if and only if $|a| \leq b$.

Since J is a (ϱ, L)-uniform graph, we get

$$\begin{aligned} e_J(U^-) &= \varrho\binom{|U^-|}{2} + O_1\left(2L\sqrt{r}\cdot|U^-|\right) \\ &= \varrho\binom{\alpha}{2} + O_1\left(2L\sqrt{r}\cdot\alpha\right) \\ &= \frac{\varrho\alpha^2}{2} - \frac{\varrho\alpha}{2} + O_1\left(2L\sqrt{r}\cdot\alpha\right) \\ e_J(U^+) &= \frac{\varrho\beta^2}{2} - \frac{\varrho\beta}{2} + O_1\left(2L\sqrt{r}\cdot\beta\right) \\ e_J(U^-,U^+) &= \varrho\alpha\beta + O_1\left(L\sqrt{r\alpha\beta}\right)\,. \end{aligned}$$

To bound S_k^U we are going to use the inequalities $\alpha(1-p(H))^2 + \beta p(H)^2 \leq \alpha+\beta$ and $2(\alpha+\beta)+\sqrt{\alpha\beta} \leq \frac{5}{2}(\alpha+\beta)$. We get for the upper bound on S_k^U

$$\begin{aligned} &S_k^U - \frac{1}{2}\varrho\left[\alpha(1-p(H)) - \beta p(H)\right]^2 \\ &\leq -\varrho\frac{\alpha}{2}(1-p(H))^2 - \varrho\frac{\beta}{2}p(H)^2 \\ &\quad +2L\sqrt{r}\cdot\alpha(1-p(H))^2 + 2L\sqrt{r}\cdot\beta p(H)^2 \\ &\quad +L\sqrt{r}\cdot\sqrt{\alpha\beta}\cdot p(H)(1-p(H)) \\ &\leq 2L\sqrt{r}(\alpha+\beta) + L\sqrt{r}\cdot\sqrt{\alpha\beta} \\ &\leq 3L\sqrt{r}(\alpha+\beta) \\ &= 3L\sqrt{r}\cdot|U|\,. \end{aligned}$$

We note that we got a bound linear in $|U|$ since L and r are constants.

Similarly, we get for the lower bound

$$\begin{aligned}
& S_k^U - \frac{1}{2}\varrho\,[\alpha(1-p(H)) - \beta p(H)]^2 \\
& \geq -\varrho\frac{\alpha}{2}(1-p(H))^2 - \varrho\frac{\beta}{2}p^2 \\
& \quad -2L\sqrt{r}\cdot\alpha(1-p(H))^2 - 2L\sqrt{r}\cdot\beta p(H)^2 \\
& \quad -L\sqrt{r}\cdot\sqrt{\alpha\beta}\cdot p(H)(1-p(H)) \\
& \geq -\varrho\frac{\alpha}{2}(\alpha+\beta) - 2L\sqrt{r}\cdot(\alpha+\beta) - L\sqrt{r}\cdot\sqrt{\alpha\beta} \\
& \geq -\frac{\varrho}{2}(\alpha+\beta) - L\sqrt{r}\cdot\frac{5}{2}(\alpha+\beta) \\
& \geq -\frac{1}{2}|U|\left(\frac{r}{m} + 5L\sqrt{r}\right) \\
& \geq -3L\sqrt{r}\cdot|U|\ .
\end{aligned}$$

Derivation of the upper and lower bounds (3.24) *and* (3.25) *on* $\sum_{ij\in J}^U \langle w_i, w_j\rangle$. To derive an upper bound on $\sum_{ij\in J}^U \langle w_i, w_j\rangle$ we use (3.21). We get

$$\sum\nolimits_{ij\in J}^U \langle w_i, w_j\rangle = \sum\nolimits_{ij\in J}^U \langle v_i, v_j\rangle - \sum_{k\notin W} S_k^U.$$

Note that (3.23) implies $S_k^U \geq -3L\sqrt{r}|U|$. Thus Claim 3.2 and this lower bound on S_k^U imply

$$\begin{aligned}
\sum\nolimits_{ij\in J}^U \langle w_i, w_j\rangle \quad \leq \quad & 2\delta p(H)^2 m_2 e(J) \\
& +3L\sqrt{r}|U|(m_2 - |W|).
\end{aligned}$$

To estimate $\sum_{ij\in J}^U \langle w_i, w_j\rangle$ from below, we first write

$$\sum\nolimits_{ij\in J}^U \langle w_i, w_j\rangle = \sum_{k\in W}\sum\nolimits_{ij\in J}^U a_{i,k}a_{j,k} = \sum_{k\in W} S_k^U.$$

Set $u = |U|$, $w = |W|$, and write $d_U(k) = |U^-| = \alpha$ for the size of the intersection of the neighbourhood of the vertex k with the set U and $\beta = |U\setminus U^-|$ (as introduced in the main text). We have

$$\begin{aligned}
& \sum\nolimits_{ij\in J}^U \langle w_i, w_j\rangle = \sum_{k\in W} S_k^U \\
& \geq \sum_{k\in W}\left(\frac{\varrho}{2}\big[\alpha(1-p(H)) - \beta p(H)\big]^2 - 3L\sqrt{r}\cdot u\right),
\end{aligned}$$

which, by Cauchy–Schwarz, is at least

$$\begin{aligned}
& \frac{\varrho}{2w}\left[\sum_{k\in W}(\alpha(1-p(H)) - \beta p(H))\right]^2 - 3L\sqrt{r}uw \\
& \geq \frac{\varrho}{2w}\left[\sum_{k\in W}(d_U(k) - up(H))\right]^2 - 3L\sqrt{r}\cdot uw \\
& = \frac{\varrho}{2w}\big|e(U,W) - p(H)uw\big|^2 - 3L\sqrt{r}uw \\
& = \frac{\varrho}{2}u^2 w\big|d(U,W) - p(H)\big|^2 - 3L\sqrt{r}\cdot uw.
\end{aligned}$$

Deferred proofs for Lemma 2.6

This section contains the proofs of the claims and identities used in the proof of Lemma 2.6.

Derivation of identity (3.28). Let us first observe that

$$\begin{aligned}
(3.42)\qquad & \sum_{\{x,y\}\in J}\big(d(x,y) - p^2 m_2\big)^2 \\
& = \sum_{\{x,y\}\in J}\big(d(x,y)^2 - 2d(x,y)p^2 m_2 + p^4 m_2^2\big) \\
& = \sum_{\{x,y\}\in J} d(x,y)^2 - 2p^2 m_2 \sum_{\{x,y\}\in J} d(x,y) + p^4 m_2^2 e(J).
\end{aligned}$$

However,

$$\begin{aligned}
\sum_{\{x,y\}\in J} d(x,y) \quad & = \quad \sum_{\{x,y\}\in J} |N(x)\cap N(y)| \\
& = \quad \sum_{u_2\in U_2} e(J[N(u_2)]) \\
& = \quad \sum_{u_2\in U_2} (1+O_1(\eta_0))(1+O_1(\varepsilon''))^2 p^2 e(J),
\end{aligned}$$

where we used that $|N(u_2)| = d(u_2) = (p + O_1(10\varepsilon'))m_1 = (1+O_1(\varepsilon''))pm_1 \geq \eta_0 m_1$. Therefore

$$(3.43)\qquad \sum_{\{x,y\}\in J} d(x,y) = (1+O_1(3\varepsilon''))p^2 m_2 e(J)\ .$$

From (3.42) and (3.43) we obtain that

$$\begin{aligned}
& \sum_{\{x,y\}\in J}\big(d(x,y) - p^2 m_2\big)^2 \\
& \qquad = \sum_{\{x,y\}\in J} d(x,y)^2 - (1+O_1(6\varepsilon''))p^4 m_2^2 e(J)\ ,
\end{aligned}$$

as claimed in (3.28).

Proof of Claim 3.4. Clearly, if $e = \{x,y\}\in J$, then

$$(3.44)\qquad d(x,y) = |N(x)\cap N(y)| = \sum_{u\in U_2} M(e,u)\ ,$$

and hence

$$d(x,y)^2 = \sum_{u\in U_2}\sum_{u'\in U_2} M(e,u)M(e,u') \,. \tag{3.45}$$

Therefore

$$\begin{aligned}
\sum_{\{x,y\}\in J} d(x,y)^2 &= \sum_{u\in U_2}\sum_{e\in J} M(e,u) \sum_{u'\in U_2} M(e,u') \\
&= \sum_{u\in U_2}\sum_{e\in J(u)}\sum_{u'\in U_2} M(e,u') \\
&= \sum_{u\in U_2}\sum_{u'\in U_2}\sum_{e\in J(u)} M(e,u').
\end{aligned}$$

Proof of Claim 3.5. Using (2.12) we, clearly, have

$$|U| = (p+O_1(10\varepsilon'))m_1 = (1+O_1(\varepsilon''))pm_1 \geq \varepsilon_1 m_1. \tag{3.46}$$

By the definition of W, we have

$$\begin{aligned}
\sum_{u'\in U_2}\sum_{e\in J(u_0)} M(e,u') \\
&< (1+\theta)p^4 m_2 e(J) \\
&\quad +|W|(1+O_1(\eta_0))(p+O_1(10\varepsilon'))^2 p^2 e(J) \\
&= (1+\theta)p^4 m_2 e(J) + (1+O_1(3\varepsilon''))|W|p^4 e(J).
\end{aligned} \tag{3.47}$$

Comparing (3.32) and (3.47), we obtain

$$2|W|p^2 e(J) \geq (1+O_1(3\varepsilon''))|W|p^2 e(J) \geq \frac{1}{4}\delta^2 p^4 m_2 e(J), \tag{3.48}$$

and hence

$$|W| \geq \frac{1}{8}\delta^2 p^2 m_2 \geq \frac{1}{32}\delta^2\varepsilon^6 m_2 \geq \varepsilon_1 m_2, \tag{3.49}$$

as required.

Proof of Claim 3.6. Suppose to the contrary that our claim fails. Then we have

$$e(J[N(u_0)\cap N(w)]) < (1+O_1(\eta_0))\left(1+\frac{\theta}{3}\right)^2 p^6 e(J) \tag{3.50}$$

if $d(u_0,w) \geq \eta_0 m_1$, and we have

$$e(J[N(u_0)\cap N(w)]) < 2\eta_0^2 e(J) \tag{3.51}$$

if $d(u_0,w) < \eta_0 m_1$. In either case, we have

$$e(J[N(u_0)\cap N(w)]) < (1+\theta)p^6 e(J) \,. \tag{3.52}$$

However, in view of the definition of W (see (3.33)) and (3.34), inequality (3.52) cannot hold. This contradiction shows that our claim must indeed hold.