Linear dependence between hereditary quasirandomness conditions

Xiaoyu He
Department of Mathematics
Stanford University
Stanford, CA, USA
alkjash@stanford.edu

Submitted: Aug 12, 2017; Accepted: Sep 17, 2018; Published: Oct 19, 2018

Abstract

Answering a question of Simonovits and Sós, Conlon, Fox, and Sudakov proved that for any nonempty graph H, and any $\varepsilon > 0$, there exists $\delta > 0$ polynomial in ε, such that if G is an n-vertex graph with the property that every $U \subseteq V(G)$ contains $p^{e(H)}|U|^{v(H)} \pm \delta n^{v(H)}$ labeled copies of H, then G is (p, ε)-quasirandom in the sense that every subset $U \subseteq G$ contains $\frac{1}{2}p|U|^2 \pm \varepsilon n^2$ edges. They conjectured that δ may be taken to be linear in ε and proved this in the case that H is a complete graph. We study a labelled version of this quasirandomness property proposed by Reiher and Schacht. Let H be any nonempty graph on r vertices $v_1, \ldots, v_r$, and $\varepsilon > 0$. We show that there exists $\delta = \delta(\varepsilon) > 0$ linear in ε, such that if G is an n-vertex graph with the property that every sequence of r subsets $U_1, \ldots, U_r \subseteq V(G)$, the number of copies of H with each v_i in U_i is $p^{e(H)} \prod |U_i| \pm \delta n^{v(H)}$, then G is (p, ε)-quasirandom.
Mathematics Subject Classifications: 05C80

1 Introduction

Random-like objects, in particular quasirandom graphs, have become a central object of study in combinatorics and theoretical computer science (see for example the survey article of Krivelevich and Sudakov [6]). In this paper, we will prove a generalization of a result of Conlon, Fox, and Sudakov [3] on quasirandom graphs that ties in with a line of research motivated by two important principles of extremal graph theory. First, that many "natural" properties of random graphs are equivalent; second, that many results provable by Szemerédi's regularity lemma can be more effectively proved directly without it.

Although certain notions of quasirandom graphs were studied earlier, such as in Thomason's work on jumbled graphs [11], the idea that many of these notions are equivalent first appeared in the seminal work of Chung, Graham and Wilson [2].

The Erdős-Rényi random graph $G(n,p)$ is the random graph on n vertices where each of the $\binom{n}{2}$ edges is drawn independently with probability p. A priori, any number of properties of the prototypical random graph $G(n,p)$ could be used to define quasirandomness, but Chung, Graham, and Wilson [2] discovered that many of these properties are qualitatively equivalent, leading to a canonical notion of quasirandomness for graphs.

We will only consider simple, undirected graphs. Write $V(G)$ for the set of vertices of a graph G, $E(G)$ for the set of edges, and define $v(G) = |V(G)|$, $e(G) = |E(G)|$. Also, write $x = y \pm \Delta$ if $|x-y| \leqslant \Delta$. We say that a graph G has edge density p if $e(G) = p\binom{v(G)}{2}$.

Theorem 1. *(Chung, Graham, and Wilson [2]). Let $p \in [0,1]$. The following are equivalent properties of a graph G with edge density p, up to the choice of $\varepsilon > 0$:*

1. *For some $s \geqslant 4$ and every graph H on s vertices, the number of induced subgraphs of G isomorphic to H is*

$$p^{e(H)}(1-p)^{\binom{v(H)}{2}-e(H)}v(G)^{v(H)} \pm \varepsilon v(G)^{v(H)}.$$

2. *There exists a nontrivial graph H such that every induced subgraph $G[U]$ of G contains $p^{e(H)}|U|^{v(H)} \pm \varepsilon v(G)^{v(H)}$ (not necessarily induced) subgraphs isomorphic to H.*

3. *The number of 4-cycles in G is at most $p^4 v(G)^4 + \varepsilon v(G)^4$.*

4. *The two largest (in absolute value) eigenvalues λ_1, λ_2 of the adjacency matrix of G satisfy $\lambda_1 = (p \pm \varepsilon)v(G)$ and $|\lambda_2| \leqslant \varepsilon v(G)$.*

5. *For every vertex subset U of G, the number of edges in U satisfies $e(U) = \frac{1}{2}p|U|^2 \pm \varepsilon v(G)^2$.*

Explicitly, we mean that for any properties $i \neq j \in \{1,2,3,4,5\}$ above, and any $\varepsilon > 0$, there exists $\varepsilon' > 0$ such that any graph G with density p which satisfies property i with error parameter ε' also satisfies property j with error parameter ε.

Although these conditions are equivalent, the relationships between the various ε's are not so well understood. For example, let $\mathcal{P}^*_{H,p}(\varepsilon)$ be property 2 above for some fixed H, p and $\varepsilon > 0$, which is called the "hereditary quasirandomness" condition because it is inherited by all induced subgraphs. Simonovits and Sós [9] were able to prove using Szemerédi's regularity lemma [10] that for any two graphs H, H', $\mathcal{P}^*_{H,p}(\delta) \implies \mathcal{P}^*_{H',p}(\varepsilon)$ where δ^{-1} is growing as a tower function of ε^{-1}. The Simonovits-Sós conjecture is that this dependence can be proved without the regularity lemma.

Roughly speaking, Szemerédi's regularity lemma states that given any $\varepsilon > 0$, every graph G can be decomposed into $K = K(\varepsilon)$ parts $V_1, \ldots, V_K$ (the "regularity partition") such that the edges between most pairs V_i, V_j are within ε of being random. Although the regularity lemma is an extraodinarily powerful tool, the quantative dependence of K on ε is of tower-type growth and this cannot be improved. Thus, tower-type quantitative dependency is indicative of a straightforward application of the regularity lemma.

In practice, the regularity lemma is not always necessary, and we can expect to improve quantitative bounds in various applications by avoiding its use. When this is possible, the tower-type bounds provided by the regularity lemma can usually be replaced by exponential or even polynomial bounds. One of the most important examples of this method is the weak regularity lemma of Frieze-Kannan [5], which proves that a graph G can be decomposed into a regularity partition with only exponentially many parts in ε if we replace the regularity condition by a weaker global version. Problems amenable to the regularity method sometimes only require the Frieze-Kannan weak regularity condition. For example, it is possible to prove the Simonovits-Sós Conjecture with exponential dependence of δ on ε using this method.

The Simonovits-Sós Conjecture was settled by Conlon, Fox, and Sudakov [3] by carefully extracting the useful ingredients of the regularity proof without using its full power, achieving polynomial dependency of δ^{-1} on ε^{-1}. They further conjectured that the true dependence is linear. In this paper we will study a variation of this conjecture introduced by Reiher and Schacht [8], where instead of counting copies of an unlabelled graph H in any induced subgraph $G[U]$ of G, we count labelled homomorphic copies of H with the i-th vertex v_i lying in a prescribed subset $U_i \subset V(G)$. In this situation, we prove the optimal linear dependence by extending a counting argument from [3].

2 Background

We first offer some notation for counting subgraphs. If G is a graph and $U \subseteq V(G)$, write $G[U]$ for the induced subgraph of G on U.

Let H, G be two labelled graphs where H has r vertices $v_1, \dots, v_r$. Let $U_1, U_2, \dots, U_r$ be vertex subsets of G. In this setting, we define $c(H, G; U_1, \dots, U_r)$ to be the number of (labelled graph) homomorphisms $\phi : H \to G$ with $\phi(v_i) \in U_i$. We abbreviate $c(H, G)$ for the total number of homomorphisms $H \to G$. In particular $c(H, G[U]) = c(H, G; U, \dots, U)$ is the number of homomorphisms from H to the induced subgraph of G on U. We think of $c(H, G; U_1, \dots, U_r)$ as the number of (non-induced) labelled copies of H in G with each vertex in a predetermined subset.

Definition 2. If H is a fixed graph, $p \in [0, 1]$ and $\varepsilon > 0$, we say G satisfies the hereditary quasirandomness property $\mathcal{P}^*_{H,p}(\varepsilon)$ if for all $U \subseteq V(G)$,

$$c(H, G[U]) = p^{e(H)}|U|^{v(H)} \pm \varepsilon v(G)^{v(H)}.$$

In other words, the condition $\mathcal{P}^*_{H,p}(\varepsilon)$ is that every induced subgraph contains the right number of copies of H. Note for any fixed $\varepsilon > 0$ and n sufficiently large, the random graph $G(n, p)$ satisfies $\mathcal{P}^*_{H,p}(\varepsilon)$ almost surely. Also, note that $\mathcal{P}^*_{H,p}(\varepsilon)$ is trivially satisfied by every graph if H is empty, so we will only be concerned with nonempty H.

Simonovits and Sós [9] proved using the Szemerédi's regularity lemma that the properties $\mathcal{P}^*_{H,p}(\varepsilon)$ are all equivalent, in the sense that for any nontrivial graphs H, H' and $\varepsilon > 0$, there exists $\delta > 0$ such that $\mathcal{P}^*_{H,p}(\delta) \implies \mathcal{P}^*_{H',p}(\varepsilon)$. Unfortunately, the dependence of δ on ε in their proof is of tower type because of the use of the regularity lemma. Note that it suffices to show that $\mathcal{P}^*_{H,p}(\delta) \implies \mathcal{P}^*_{K_2,p}(\varepsilon)$ where K_2 is the graph with a single edge; the other direction is given by a straightforward counting lemma, which we state as Lemma 10 below.

Conlon, Fox, and Sudakov [3] were able to tailor the regularity method to this problem to prove the same result with polynomial dependence of the form $\delta = \Omega(\varepsilon^{f(p,v(H))})$ where the exponent f depends only on p and $v(H)$. They conjectured that the dependence is in fact linear, and proved it for the case $H = K_n$.

Conjecture 3. For any nonempty graph H, and real numbers $p \in [0,1]$, $\delta > 0$, we have

$$\mathcal{P}^*_{H,p}(\delta) \implies \mathcal{P}^*_{K_2,p}(\varepsilon)$$

for some $\varepsilon = O_{H,p}(\delta)$.

Independently, Reiher and Schacht [8] showed a similar polynomial dependence for a stronger notion of quasirandomness which takes configurations into account.

Definition 4. If H is a fixed graph, we say G satisfies $\mathcal{R}_{H,p}(\varepsilon)$ if for every sequence of $v(H)$ disjoint vertex subsets $U_1, \dots, U_{v(H)} \subseteq G$,

$$c(H, G; U_1, \dots, U_{v(H)}) = p^{e(H)} \prod_{i=1}^{v(H)} |U_i| \pm \varepsilon v(G)^{v(H)}.$$

In this paper we show the linear dependence in Conjecture 3 using the stronger condition $\mathcal{R}_{H,p}(\varepsilon)$. Note that as with $\mathcal{P}^*_{H,p}(\delta)$, the condition $\mathcal{R}_{H,p}(\varepsilon)$ is trivial when H is empty.

Theorem 5. *For any nonempty graph H, and real numbers $p \in [0,1]$, $\delta > 0$, we have*

$$\mathcal{R}_{H,p}(\delta) \implies \mathcal{P}^*_{K_2,p}(\varepsilon)$$

for some $\varepsilon = O_H(p^{-3e(H)}\delta)$.

The converse is a standard counting lemma; we show it in Lemma 10.

We will begin by proving that $\mathcal{R}_{H,p}(\varepsilon)$ is equivalent up to linear change of ε to $\mathcal{R}'_{H,p}(\varepsilon)$, which is the same condition with disjointness removed.

Definition 6. If H is a fixed graph, we say G satisfies $\mathcal{R}'_{H,p}(\varepsilon)$ if for every sequence of $v(H)$ (not necessarily disjoint) vertex subsets $U_1, \dots, U_{v(H)} \subseteq G$,

$$c(H, G; U_1, \dots, U_{v(H)}) = p^{e(H)} \prod_{i=1}^{v(H)} |U_i| \pm \varepsilon v(G)^{v(H)}.$$

After this simple argument, we show that the argument of Conlon, Fox, and Sudakov which proves Conjecture 3 for $H = K_n$ extends naturally to all H under the stronger condition $\mathcal{R}'_{H,p}(\varepsilon)$. Note that up to linear change in ε, $\mathcal{R}'_{H,p}(\varepsilon)$ is equivalent to $\mathcal{P}^*_{H,p}(\varepsilon)$ when H is a complete graph.

The original conjecture of Conlon, Fox, Sudakov remains open. By Theorem 5, it would suffice to show that $\mathcal{P}^*_{H,p}(\varepsilon)$ and $\mathcal{R}_{H,p}(\varepsilon)$ are equivalent up to linear change in ε.

Conjecture 7. For any graph H and real numbers $p \in [0,1]$, $\delta > 0$, we have

$$\mathcal{P}^*_{H,p}(\delta) \implies \mathcal{R}_{H,p}(\varepsilon)$$

for some $\varepsilon = O_{H,p}(\delta)$.

The other direction is easy by inclusion-exclusion.

3 Preliminaries

Here we reduce $\mathcal{R}_{H,p}(\varepsilon)$ to $\mathcal{R}'_{H,p}(\varepsilon)$ and then survey some standard lemmas that we will need from graph theory.

Lemma 8. *For any graph H and real numbers $p \in [0,1]$, $\delta > 0$, we have*

$$\mathcal{R}_{H,p}(\delta) \implies \mathcal{R}'_{H,p}(\varepsilon)$$

for some $\varepsilon = O_H(\delta)$.

Proof. If K is a graph on the integers in $[1, v(H)]$, we say a graph G satisfies condition $\mathcal{R}^K_{H,p}(\varepsilon)$ if for every sequence of vertex subsets $U_1, \dots, U_{v(H)} \subseteq G$ such that $U_i \cap U_j = \varnothing$ whenever (i,j) is an edge of K,

$$c(H, G; U_1, \dots, U_{v(H)}) = p^{e(H)} \prod_{i=1}^{v(H)} |U_i| \pm \varepsilon v(G)^{v(H)}.$$

When K is complete, $\mathcal{R}^K_{H,p}(\varepsilon)$ is exactly $\mathcal{R}_{H,p}(\varepsilon)$ and when K is empty it is $\mathcal{R}'_{H,p}(\varepsilon)$. To prove the lemma inductively, it suffices to show that if K' is a graph containing K with has one more edge, $\mathcal{R}^{K'}_{H,p}(\delta) \implies \mathcal{R}^K_{H,p}(\varepsilon)$ for some $\varepsilon = O_H(\delta)$. Let G satisfy $\mathcal{R}^{K'}_{H,p}(\delta)$. Without loss of generality, suppose $K' = K \cup (1,2)$. Let $U_1, \dots, U_{v(H)}$ be a sequence of vertex subsets of G such that $U_i \cap U_j = \varnothing$ whenever (i,j) is an edge of K. We will to show that

$$c(H, G; U_1, \dots, U_{v(H)}) = p^{e(H)} \prod_{i=1}^{v(H)} |U_i| \pm (6 + o(1))\delta v(G)^{v(H)},$$

where the $o(1)$ goes to zero as a function of $v(G)$.

If $U_1 \cap U_2 = \varnothing$ we are immediately done, since $\mathcal{R}^{K'}_{H,p}(\varepsilon)$ already applies. Otherwise, write

$$\begin{aligned} c(H, G; U_1, U_2, \dots) &= c(H, G; U_1 \backslash U_2, U_2, \dots) + c(H, G; U_1 \cap U_2, U_2 \backslash U_1, \dots) \\ &\quad + c(H, G; U_1 \cap U_2, U_1 \cap U_2, \dots). \end{aligned}$$

Directly applying $\mathcal{R}^{K'}_{H,p}(\delta)$ to the first two terms, we obtain

$$\begin{aligned} c(H, G; U_1, U_2, \dots) &= p^{e(H)} \Big(|U_1||U_2| - |U_1 \cap U_2|^2 \Big) \prod_{i=3}^{v(H)} |U_i| \pm 2\delta v(G)^{v(H)} \\ &\quad + c(H, G; U_1 \cap U_2, U_1 \cap U_2, \dots). \end{aligned}$$

It remains to show that for any $U \subseteq V(G)$,

$$c(H, G; U, U, U_3, \ldots, U_{v(H)}) = p^{e(H)}|U|^2 \prod_{i=3}^{v(H)} |U_i| \pm (4 + o(1))\delta v(G)^{v(H)}.$$

For this, we may assume (by adding a vertex if necessary) that $|U|$ is even, since a single vertex lies in $O(v(G)^{v(H)-1})$ copies of H. Pick U_1', U_2' to be a uniform random equitable bipartition of U, i.e. dividing U into two subsets of equal order. The number of homomorphisms $\phi : H \to G$ with $\phi(v_1) = \phi(v_2)$ is $O(v(G)^{v(H)-1})$. Apart from these, each homomorphism in $c(H, G, U, U, U_3, \ldots)$ is counted in $c(H, G; U_1', U_2', U_3, \ldots)$ with probability $1/4$, and so by linearity of expectation,

$$\mathbb{E}[c(H, G; U_1', U_2', U_3, \ldots)] = \frac{1}{4} c(H, G; U, U, U_3, \ldots) + O(v(G)^{v(H)-1}).$$

On the other hand, because U_1' and U_2' are disjoint, the expression inside the expectation is bounded by the property $\mathcal{R}_{H,p}^{K'}(\delta)$. Thus, reversing the last equation gives

$$\begin{aligned} c(H, G; U, U, U_3, \ldots) \; &= 4\mathbb{E}[c(H, G; U_1', U_2', U_3, \ldots)] + O(v(G)^{v(H)-1}) \\ &= p^{e(H)} \textstyle\prod_{i=1}^{v(H)} |U_i| \pm (4 + o(1))\delta v(G)^{v(H)}, \end{aligned}$$

as desired. In particular, since the induction takes $\binom{v(H)}{2}$ steps and at each step the constant factor is at most $6 + o(1)$, we have proved that $\mathcal{R}_{H,p}(\delta) \implies \mathcal{R}'_{H,p}(\varepsilon)$ for some $\varepsilon > 0$ satisfying

$$\varepsilon \leqslant \left(6^{\binom{v(H)}{2}} + o(1)\right)\delta. \qquad \square$$

The next lemma is needed to give a preliminary lower bound on the edge density of a graph satisfying $\mathcal{R}_{H,p}(\delta)$. It is a corollary of a stronger result of Alon [1], who determined the asymptotic order of the maximum number of copies of H in any graph G with a prescribed number of edges.

Lemma 9. *If H is a graph with no isolated vertices, then for any graph G,*

$$c(H, G) = O(e(G)^{v(H)}).$$

We also recall a standard counting lemma, see for example Section 10.5 of Lovász' problem book [7]. It tells us how to count copies of H given quasirandomness. If $A, B \subseteq V(G)$, let $e(A, B) = c(K_2, G; A, B)$ be the number of edges between A and B, defined so that if A and B intersect we count each edge within $G[A \cap B]$ twice. In particular, $e(A, A) = 2e(A)$, since the former counts labelled edges.

Lemma 10. *If G is a graph which satisfies $\mathcal{P}^*_{K_2,p}(\delta)$ then G satisfies $\mathcal{R}'_{H,p}(4e(H)\delta)$ for all graphs H.*

Proof. Suppose G satisfies $\mathcal{P}^*_{K_2,p}(\delta)$. We wish to show that for any graph H (on vertices $v_1, \dots, v_{v(H)}$) and any sets $U_1, \dots, U_{v(H)} \subseteq V(G)$,

$$c(H, G; U_1, \dots, U_{v(H)}) = p^{e(H)} \prod_{i=1}^{v(H)} |U_i| \pm 4e(H)\delta v(G)^{v(H)}. \tag{1}$$

Let $c_H = c(H, G; U_1, \dots, U_{v(H)})$ be the desired homomorphism count. To prove (1), we will expand c_H as a sum involving the indicator functions of edges, and prove that it is possible to approximate these indicator functions with the constant function p. Since

$$e(A, B) = e(A \cup B) + e(A \cap B) - e(A \backslash B) - e(B \backslash A),$$

the fact that G satisfies $\mathcal{P}^*_{K_2,p}(\delta)$ implies that for every pair of vertex subsets $A, B \subseteq V(G)$,

$$\Big| e(A, B) - p|A||B| \Big| \leqslant 4\delta v(G)^2.$$

Let $1_G(u, v)$ be the indicator function of edges of G. Another way of writing the above inequality is that for any two functions $f, g : V(G) \to \{0, 1\}$ (which will be the indicator functions of some two sets A and B),

$$\Big| \sum_{u,v \in V(G)} f(u)g(v)(1_G(u, v) - p) \Big| \leqslant 4\delta v(G)^2. \tag{2}$$

We can expand c_H in terms of the indicator function 1_G, giving

$$c_H = \sum_{(u_1, \dots, u_{v(H)})} \prod_{(v_i, v_j) \in E(H)} 1_G(u_i, u_j),$$

where the sum is over all $v(H)$-tuples of vertices $(u_i)_{i=1}^{v(H)}$ with $u_i \in U_i$. For a spanning subgraph $H' \subseteq H$, define

$$c_{H,H'} = \sum_{(u_1, \dots, u_{v(H)})} p^{e(H) - e(H')} \prod_{(v_i, v_j) \in E(H')} 1_G(u_i, u_j),$$

the sum obtained by replacing $1_G(u_i, u_j)$ by p for all the edges (v_i, v_j) of H not in H'. Let $H_0 \subset H_1 \subset \cdots \subset H_{e(H)} = H$ be a maximal filtration of H by spanning subgraphs, so that for each $1 \leqslant k \leqslant e(H)$, H_k has exactly one more edge than H_{k-1}. Let $e_k = (v_{i_k}, v_{j_k})$ be the edge introduced in H_k. We will show that for all $1 \leqslant k \leqslant e(H)$,

$$|c_{H,H_k} - c_{H,H_{k-1}}| \leqslant 4\delta v(G)^{v(H)}. \tag{3}$$

We also know that

$$c_{H,H_0} = \sum_{(u_1, \dots, u_{v(H)})} p^{e(H)} = p^{e(H)} \prod_{i=1}^{v(H)} |U_i|,$$

so since $c_{H,H_{e(H)}} = c_H$, summing inequality (3) over k and applying the triangle inequality would complete the proof of (1).

Notice that

$$c_{H,H_k} - c_{H,H_{k-1}} = \sum_{(u_1,\dots,u_{v(H)})} p^{e(H)-k}(1_G(u_{i_k}, u_{j_k}) - p) \prod_{(v_i,v_j)\in E(H_{k-1})} 1_G(u_i, u_j). \quad (4)$$

In the product on the right, all factors depend on at most one of u_{i_k} and u_{j_k}. Let $T_k = \prod_{i\notin\{i_k,j_k\}} U_i$ be the set of all $(v(H)-2)$-tuples of choices of all u_i except for u_{i_k} and u_{j_k}. Write $1_U(\cdot)$ for the indicator function of a vertex set U. For each tuple $t \in T_k$, we may define $\{0,1\}$-valued functions

$$\begin{aligned} f_t(u) &= 1_{U_{i_k}}(u) \prod_{(v_{i_k},v_j)\in E(H_{k-1})} 1_G(u, u_j) \\ g_t(u) &= 1_{U_{j_k}}(u) \prod_{(v_i,v_{j_k})\in E(H_{k-1})} 1_G(u_i, u) \end{aligned}$$

which depend on the choice of $t \in T_k$ via the choices of u_i. Define H_k^* to be the graph obtained by removing the two vertices u_{i_k} and u_{j_k} from H_k. Then, we can separate out the factors in the sum in (4) that depend on u_{i_k}, u_{j_k} to find

$$c_{H,H_k} - c_{H,H_{k-1}} = p^{e(H)-k} \sum_{t\in T_k} \prod_{(v_i,v_j)\in H_k^*} 1_G(u_i, u_j) \sum_{u_{i_k},v_{i_k}\in V(G)} (1_G(u_{i_k}, u_{j_k}) - p) f_t(u_{i_k}) g_t(u_{j_k}).$$

The absolute value of the inner double sum is bounded by $4\delta v(G)^2$ because of inequality (2), and the product term takes values in $\{0,1\}$. Thus,

$$|c_{H,H_k} - c_{H,H_{k-1}}| \leqslant p^{e(H)-k} v(G)^{v(H)-2} \cdot 4\delta v(G)^2 \leqslant 4\delta v(G)^{v(H)},$$

which proves (3). It follows that

$$c_H = c_{H,H_{e(H)}} = c_{H,H_0} + \sum_{k=1}^{e(H)} (c_{H,H_k} - c_{H,H_{k-1}}) = p^{e(H)} \prod_{i=1}^{v(H)} |U_i| \pm 4e(H)\delta v(G)^{v(H)},$$

as desired. □

Finally, we require a lemma of Erdős, Goldberg, Pach, and Spencer [4] on discrepancy.

Lemma 11. *Let G be a graph with edge density q. If there is a subset $S \subseteq V(G)$ for which $|e(S) - q\binom{|S|}{2}| \geqslant D$, then there exists a set $S' \subseteq V(G)$ of order $\frac{1}{2}v(G)$ such that*

$$\left| e(S') - q\binom{|S'|}{2} \right| \geqslant (\frac{1}{4} - o(1))D,$$

where $o(1)$ goes to zero as $D \to \infty$.

In fact, we will only need the following corollary.

Corollary 12. *Let G be a graph with edge density q. If there is a subset $S \subseteq V(G)$ for which $|e(S) - q\binom{|S|}{2}| \geqslant D$, then there exist two disjoint subsets $X, Y \subseteq V(G)$ of size $\frac{1}{4}v(G)$ such that*

$$|e(X) - e(Y)| \geqslant \frac{1}{16}D - o(v(G)^2),$$

where the error term is a function of $v(G) \to \infty$.

Proof. Apply Lemma 11 to find a set S' with the stated properties. Now, pick a uniformly random subset $A \subseteq V(G)$, obtained by independently picking each vertex of G with probability 1/2. Let $X = A \cap S'$, and pick a uniformly random subset $Y \subseteq V(G) \backslash A$. We now prove that the sizes of $|X|, |Y|, e(X)$, and $e(Y)$ are tightly concentrated about their expected values.

The marginal distribution of X is the uniform distribution on subsets of S'. Thus, we expect $|X|$ to be tightly concentrated about $\frac{1}{2}|S'| = \frac{1}{4}v(G)$ and $e(X)$ to be tightly concentrated about $\frac{1}{4}e(S')$. Meanwhile, the marginal distribution of Y is the distribution obtained by independently adding each $v \in V(G)$ to Y with probability $\frac{1}{4}$, so we expect $|Y|$ to be tightly concentrated about $\frac{1}{4}v(G)$ and $e(Y)$ to be tightly concentrated about $\frac{1}{16}e(G)$.

We first check the tight concentrations of orders. Note that $|X| \sim B(\frac{1}{2}v(G), \frac{1}{2})$ and $|Y| \sim B(v(G), \frac{1}{4})$ are both binomial distributions with the same mean. By a standard application of Chernoff bounds, we have that for any $\delta \in [0, 1]$,

$$\Pr\left[\left||X| - \frac{1}{4}v(G)\right| \leqslant \frac{1}{4}\delta v(G)\right] \leqslant e^{-\delta^2 v(G)/12},$$

and the same bound holds for $|Y|$. In particular, w.h.p. $|X| = |Y| = (\frac{1}{4} + o(1))v(G)$.

Next, let A_i be the event that the i-th vertex of S' lies in A, and define

$$E_i = \mathbb{E}[e(X)|A_1, \ldots, A_i]$$

to be the expected number of edges in S', conditioned on the first i events of the A_i. Notice that $E_{|S'|} = e(X)$ and $E_0, \ldots, E_{|S'|}$ is a martingale satisfying the $v(G)$-Lipschitz condition, since adding or removing a single vertex from X changes the number of edges by at most $v(G)$. The expected value of $e(X)$ is $\frac{1}{4}e(S')$ by linearity of expectation, so the Azuma-Hoeffding inequality gives

$$\Pr[|e(X) - \frac{1}{4}e(S')| \geqslant t] \leqslant 2\exp\Big(-\frac{t^2}{v(G)^3}\Big).$$

Thus, w.h.p. $e(X) = \frac{1}{4}e(S') + o(v(G)^2)$. By the same argument, w.h.p. $e(Y) = \frac{1}{16}e(G) + o(v(G)^2)$.

Since all four concentration events happen with high probability, there exists some two disjoint sets $X \subseteq S'$ and $Y \subseteq V(G)$ for which $|X| = |Y| = (\frac{1}{4} + o(1))v(G)$, $e(X) = \frac{1}{4}e(S') + o(v(G)^2)$, and $e(Y) = \frac{1}{16}e(G) + o(v(G)^2)$. Add or remove $o(v(G))$ vertices (and therefore $o(v(G)^2)$ edges) to X and Y, we obtain two sets of the desired order $\frac{1}{4}v(G)$. Their edge counts still satisfy

$$|e(X) - e(Y)| = |\frac{1}{4}e(S') - \frac{1}{16}e(G)| + o(v(G)^2) = \frac{1}{16}D + o(D) + o(v(G)^2)$$

as desired. $\square$

4 The main lemma

With the stronger quasirandomness condition $\mathcal{R}'_{H,p}(\delta)$, we are ready to prove the main lemma, which is motivated by an argument of Conlon, Fox, Sudakov [3]. Fix a vertex $v_0 \in H$ with degree r. We will want to control the average difference of $|d^r(u) - d^r(v)|$ over all pairs $u, v \in G$, where $d^r(u)$ is the r-th power of the degree of u.

Definition 13. Let $H' = H \backslash \{v_1\}$. Define $c(u, v)$ to be the number of pairs of homomorphisms $\phi, \psi : H \to G$ such that $\phi(v_1) = u$, $\psi(v_1) = v$, and $\phi|_{H'} = \psi|_{H'}$, i.e. the number of copies of H' in G that extend to a copy of H when we add either u or v for the first vertex.

In the language of Reiher and Schacht [8], $c(u, v)$ counts copies of the graph K obtained from H by doubling the first vertex and fixing the images of these two doubles to be u and v.

Lemma 14. *Let H be a graph with a vertex v_0 of degree r, and let G be a graph that satisfies $\mathcal{R}_{H,p}(\delta)$. Then,*

$$\sum_{u,v \in V(G)} |d^r(u) - d^r(v)| = O(p^{e(H)}\delta v(G)^{r+2}).$$

Proof. First we give a heuristic for the average value of $c(u, v)$. Counting pairs of homomorphisms ϕ, ψ with the given property is the same as counting the number of homomorphisms $\psi : H \to G$ for which $\psi(v_i) \in N(u)$ whenever $(v_1, v_i) \in E(H)$. If the r such vertices are $v_2, \dots, v_{r+1}$ (without loss of generality), then

$$c(u, v) = c(H, G; \{v_1\}, \overbrace{N(u), \dots, N(u)}^{r}, \overbrace{V(G), \dots, V(G)}^{v(H)-r-1}).$$

By this formula, we expect that on average,

$$c(u, v) \approx p^{e(H)} d^r(u) v(G)^{v(H)-r-1}.$$

Turning this approximation on its head, we expect $d^r(u) \approx p^{-e(H)} v(G)^{-v(H)+r+1} c(u, v)$, and so it is natural to bound

$$|d^r(u) - d^r(v)| \leqslant |d^r(u) - p^{-e(H)} v(G)^{-v(H)+r+1} c(u, v)| + |d^r(v) - p^{-e(H)} v(G)^{-v(H)+r+1} c(u, v)|$$

via the triangle inequality. Summing over all pairs (u, v), we obtain

$$\begin{aligned}\sum_{u,v\in V(G)} |d^r(u) - d^r(v)| &\leqslant 2 \sum_{u,v\in V(G)} |d^r(u) - p^{-e(H)} v(G)^{-v(H)+r+1} c(u,v)| \\ &= 2p^{-e(H)} v(G)^{-v(H)+r+1} \sum_{u,v\in V(G)} |p^{e(H)} d^r(u) v(G)^{v(H)-r-1} - c(u,v)|. \end{aligned} \tag{5}$$

For a fixed vertex u, the relevant sum over v is

$$\Sigma_u = \sum_{v\in V(G)} |c(u,v) - p^{e(H)} d^r(u) v(G)^{v(H)-r-1}|.$$

Break up $V(G) = V^+ \cup V^-$ so that $c(u,v) - p^{-e(H)} d^r(u) v(G)^{v(H)-r-1}$ is nonnegative on $v \in V^+$ and negative on $v \in V^-$, and let Σ_u^+, Σ_u^- be the pieces of the sum Σ_u supported on V^+, V^- respectively. For the positive side, we get

$$\begin{aligned}\Sigma_u^+ &= \sum_{v\in V^+} \left(c(u,v) - p^{e(H)} d^r(u) v(G)^{v(H)-r-1}\right) \\ &= \sum_{v\in V^+} c(u,v) - p^{e(H)} |V^+| d^r(u) v(G)^{v(H)-r-1}.\end{aligned}$$

On the other hand, for a fixed u, the sum of $c(u,v)$ over $v \in V^+$ is exactly

$$\sum_{v\in V^+} c(u,v) = c(H, G; V^+, \overbrace{N(u), \ldots, N(u)}^{r}, \overbrace{V(G), \ldots, V(G)}^{v(H)-r-1}),$$

a quantity which we can estimate using the given quasirandomness condition. By $\mathcal{R}_{H,p}(\delta)$ it follows that

$$|\Sigma_u^+| \leqslant \delta v(G)^{v(H)},$$

and the same argument shows that

$$|\Sigma_u^-| \leqslant \delta v(G)^{v(H)}.$$

Returning to (5), we get the desired bound by the triangle inequality:

$$\begin{aligned}\sum_{u,v} |d^r(u) - d^r(v)| &\leqslant 2p^{-e(H)} v(G)^{-v(H)+r+1} \sum_u (|\Sigma_u^+| + |\Sigma_u^-|) \\ &= O(p^{-e(H)} \delta v(G)^{r+2}).\end{aligned}$$

□

5 Proof of the main theorem

Conlon, Fox, and Sudakov [3] prove the following elementary inequality to fully exploit the degree bound obtained in Lemma 14.

Lemma 15. *(Corollary 2.2 from [3].) Let $a_1, \dots, a_n$ and $b_1, \dots, b_n$ be two sets of n nonnegative integers. Then, for any $r \in \mathbb{N}$,*

$$\sum_{i,j=1}^{n} |b_j^r - a_i^r| \geqslant \sum_{j=1}^{n} b_j^{r-1} \cdot \Big(\sum_{j=1}^{n} b_j - \sum_{i=1}^{n} a_i \Big).$$

Now we finish the proof of Theorem 5 in the following form.

Lemma 16. *If H is a nonempty graph and G is a graph satisfying $\mathcal{R}'_{H,p}(\delta)$, then G satisfies $\mathcal{P}^*_{K_2,p}(\varepsilon)$ for some $\varepsilon = O_H(p^{-2e(H)}\delta)$.*

Proof. It is easy to check that removing isolated vertices from H has no effect on the quasirandomness condition $\mathcal{R}'_{H,p}(4e(H)\delta)$. We thus assume that H has no isolated vertices.

We may assume that $\delta = o(p^{2e(H)})$ is small compared to $p^{e(H)}$, or else the desired result would be immediately true with $\varepsilon = 1$. Also, because G satisfies $\mathcal{R}'_{H,p}(\delta)$,

$$c(H, G) \geqslant p^{e(H)} v(G)^{v(H)} - \delta v(G)^{v(H)}.$$

Thus, by Lemma 9,

$$p^{e(H)} v(G)^{v(H)} = O(c(H, G)) = O(e(G)^{v(H)}),$$

which gives a lower bound

$$q = \Omega(p^{e(H)/v(H)}) \tag{6}$$

on the edge density q of G. Here and henceforth all implicit constants are allowed to depend only on H.

Let r be the minimum degree of H. We show that G satisfies $\mathcal{P}^*_{K_2,q}(\gamma)$ where $\gamma = O(q^{-r+1}p^{-e(H)}\delta)$, and then that $|p - q|$ is small. Suppose G does not satisfy $\mathcal{P}^*_{K_2,q}(\gamma)$ for some $\gamma = Cq^{-r+1}p^{-e(H)}\delta$. It must therefore contain a subset S such that

$$|c(K_2, G[S]) - q|S|^2| > \gamma v(G)^2.$$

But $c(K_2, G[S]) = 2e(S)$, so it follows that G satisfies the conditions of Corollary 12 with $\delta = \frac{1}{2}\gamma - o(1)$. Applying Corollary 12 with this δ, we may pick $X, Y \subseteq V(G)$ of size $v(G)/4$ for which

$$|e(X) - e(Y)| \geqslant \Big(\frac{1}{32}\gamma - o(1)\Big) v(G)^2. \tag{7}$$

Without loss of generality, assume $e(X) > e(Y)$. The induced subgraph $G' = G[X \cup Y]$ satisfies $\mathcal{R}_{H,p}(\delta)$ for the same δ as G since this property is hereditary, so applying Lemma 14 to G' and writing $d_{G'}(u)$ for the degree in G' of u, we find that

$$\sum_{u,v \in V(G')} |d^r_{G'}(u) - d^r_{G'}(v)| = O(p^{-e(H)}\delta v(G')^{r+2}). \tag{8}$$

On the other hand, the differences $d^r_{G'}(u) - d^r_{G'}(v)$ appear in this sum for every pair $(u,v) \in X \times Y$, and these are large on average by Lemma 15:

$$\begin{aligned} \sum_{u,v \in V(G')} |d^r_{G'}(u) - d^r_{G'}(v)| &\geqslant \sum_{u \in X}\sum_{v \in Y} |d^r_{G'}(u) - d^r_{G'}(v)| \\ &\geqslant \sum_{u \in X} d^{r-1}_{G'}(u)\Big(\sum_{u \in X} d_{G'}(u) - \sum_{v \in Y} d_{G'}(v)\Big) \\ &\geqslant |X|\Big(\frac{1}{|X|}\sum_{u \in X} d_{G'}(u)\Big)^{r-1}(e(X) - e(Y)). \end{aligned}$$

Now, it is easy to see from the proof of Corollary 12 that we choose X to have more edges than average, so that

$$\frac{1}{|X|}\sum_{u \in X} d_{G'}(u) \geqslant \frac{2e(X)}{|X|} \geqslant (q - o(1))|X|.$$

Together with (7), it follows that

$$\sum_{u,v \in V(G')} |d^r_{G'}(u) - d^r_{G'}(v)| \geqslant \Omega(q^{r-1}|X|^r \cdot \gamma v(G)^2) \geqslant \Omega(q^{r-1}\gamma v(G)^{r+2}),$$

since $|X| = \frac{1}{4}v(G)$. Together with the upper bound (8), we have shown

$$q^{r-1}\gamma = O(p^{-e(H)}\delta),$$

which contradicts our choice of $\gamma = Cq^{-r+1}p^{-e(H)}\delta$ if the constant C is sufficiently large. Therefore, G must satisfy $\mathcal{P}^*_{K_{2,q}}(\gamma)$ for some $\gamma = O(q^{-r+1}p^{-e(H)}\delta)$. But then Lemma 10 tells us that G also satisfies $\mathcal{R}'_{H,q}(4e(H)\gamma)$.

Since G is to satisfy both $\mathcal{R}'_{H,p}(\delta)$ and $\mathcal{R}'_{H,q}(4e(H)\gamma)$, we have simultaneously that $c(H,G) = p^{e(H)}v(G)^{e(H)} \pm \delta v(G)^{e(H)}$ and $c(H,G) = q^{e(H)}v(G)^{e(H)} \pm 4e(H)\gamma v(G)^{e(H)}$. By the triangle inequality,

$$\begin{aligned} |p^{e(H)} - q^{e(H)}| &\leqslant 4e(H)\gamma + \delta, \\ &= O(q^{-r+1}p^{-e(H)}\delta). \end{aligned}$$

To compare q to p, we have the lower bound (6). Recalling that r is the degree of a single vertex of H, we find that $r \leqslant v(H)$, so $q^{-r+1} = O(p^{-e(H)})$. As a result,

$$|p^{e(H)} - q^{e(H)}| \leqslant O(p^{-2e(H)}\delta).$$

Since G satisfies $\mathcal{P}^*_{K_{2,q}}(\gamma)$, it must also satisfy $\mathcal{P}^*_{K_{2,p}}(\varepsilon)$ for

$$\varepsilon = \gamma + |p^{e(H)} - q^{e(H)}| = O(p^{-2e(H)}\delta),$$

as desired. □

Combining Lemma 8 with Lemma 16, Theorem 5 is proved.

Acknowledgements

The author would like to thank Jacob Fox for calling attention to this problem and for many helpful discussions, and David Conlon and Mathias Schacht for stimulating conversations. The referee's comments were also very useful.

References

[1] N. Alon, On the number of subgraphs of prescribed type of graphs with a given number of edges, *Israel J. Math.* **38** (1981), 116–130.

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

[3] D. Conlon, J. Fox, and B. Sudakov, Hereditary quasirandomness without regularity, *Mathematical Proceedings of the Cambridge Philosophical Society* **164**, no. 3 (2018), pp. 385–399, https://doi.org/10.1017/S0305004116001055

[4] P. Erdős, M. Goldberg, J. Pach and J. Spencer, Cutting a graph into two dissimilar halves, *J. Graph Theory* **12** (1988), 121–131.

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

[6] M. Krivelevich and B. Sudakov, Pseudo-random graphs, in More sets, graphs and numbers, 199–262, Bolyai Soc. Math. Stud. 15, Springer, Berlin, 2006.

[7] L. Lovász, Combinatorial problems and exercises, 2nd edition. AMS Chelsea Publishing, Providence, RI, 2007.

[8] C. Reiher and M. Schacht, in preparation.

[9] M. Simonovits and V. T. Sós, Hereditarily extended properties, quasi-random graphs and not necessarily induced subgraphs, *Combinatorica* **17** (1997), 577–596.

[10] E. Szemerédi, Regular partitions of graphs, Problèmes combinatoires et théorie des graphes, *Colloq. Internat. CNRS* **260** (1978), 399–401.

[11] A. Thomason, Pseudo-random graphs, in: Proceedings of Random Graphs, Poznań 1985, M. Karoński, ed., *Annals of Discrete Math.* **33** (1987), 307–331.