Proceedings of the 12th European Conference on Combinatorics, Graph Theory and Applications
EUROCOMB'23
Prague, August 28 - September 1, 2023

Forcing Generalized Quasirandom Graphs Efficiently

(Extended abstract)

Andrzej Grzesik* Daniel Král'† Oleg Pikhurko‡

Abstract

We study generalized quasirandom graphs whose vertex set consists of q parts (of not necessarily the same sizes) with edges within each part and between each pair of parts distributed quasirandomly; such graphs correspond to the stochastic block model studied in statistics and network science. Lovász and Sós showed that the structure of such graphs is forced by homomorphism densities of graphs with at most $(10q)^q + q$ vertices; subsequently, Lovász refined the argument to show that graphs with $4(2q+3)^8$ vertices suffice. Our results imply that the structure of generalized quasirandom graphs with $q \geq 2$ parts is forced by homomorphism densities of graphs with at most $4q^2 - q$ vertices, and, if vertices in distinct parts have distinct degrees, then $2q+1$ vertices suffice. The latter improves the bound of $8q-4$ due to Spencer.

DOI: `https://doi.org/10.5817/CZ.MUNI.EUROCOMB23-070`

1 Introduction

Quasirandom graphs play an important role in structural and extremal graph theory. The notion of quasirandom graphs can be traced to the works of Rödl [38], Thomason [42, 43]

*Faculty of Mathematics and Computer Science, Jagiellonian University, Łojasiewicza 6, 30-348 Kraków, Poland. E-mail: Andrzej.Grzesik@uj.edu.pl. Supported by the National Science Centre grant number 2021/42/E/ST1/00193.

†Faculty of Informatics, Masaryk University, Botanická 68A, 602 00 Brno, Czech Republic. E-mail: kral@fi.muni.cz. Supported by the MUNI Award in Science and Humanities (MUNI/I/1677/2018) of the Grant Agency of Masaryk University.

‡Mathematics Institute and DIMAP, University of Warwick, Coventry CV4 7AL, United Kingdom. E-mail: o.pikhurko@warwick.ac.uk. Supported by ERC Advanced Grant 101020255 and Leverhulme Research Project Grant RPG-2018-424.

and Chung, Graham and Wilson [9] in the 1980s, and is also deeply related to Szemerédi's Regularity Lemma [40]. Indeed, the Regularity Lemma asserts that each graph can be approximated by partitioning into a bounded number of quasirandom bipartite graphs. There is also a large body of literature concerning quasirandomness of various kinds of combinatorial structures such as groups [24], hypergraphs [5, 6, 22, 23, 28, 31, 37, 39], permutations [4, 10, 32, 33], Latin squares [11, 17, 20, 25], subsets of integers [8], tournaments [3, 7, 13, 14, 27, 26], etc. Many of these notions have been treated in a unified way in the recent paper by Coregliano and Razborov [15].

The starting point of our work is the following classical result [9] on quasirandom graphs: a sequence of graphs $(G_n)_{n\in\mathbb{N}}$ is quasirandom with density p if and only if the homomorphism densities of the single edge K_2 and the 4-cycle C_4 in $(G_n)_{n\in\mathbb{N}}$ converge to p and p^4, i.e., to their expected densities in the Erdős-Rényi random graph with density p. In particular, quasirandomness is forced by homomorphism densities of graphs with at most 4 vertices. We consider a generalization of quasirandom graphs, which corresponds to the stochastic block model in statistics. In this model, the edge density of a (large) graph is not homogeneous as in the Erdős-Rényi random graph model, however, the graph can be partitioned into q parts such that the edge density is homogeneous inside each part and between each pair of the parts. Lovász and Sós [35] established that the structure of such graphs is forced by homomorphism densities of graphs with at most $(10q)^q + q$ vertices. Lovász [34, Theorem 5.33] refined this result by showing that homomorphism densities of graphs with at most $4(2q + 3)^8$ vertices suffice. Our main result (Theorem 1) improves this bound: the structure of generalized quasirandom graphs with $q \geq 2$ parts is forced by homomorphism densities of graphs with at most $4q^2 - q$ vertices. Our line of arguments substantially differs from that in [35, 34], in particular, it is more explicit and so of a more constructive nature, which is of importance in relation to applications [2, 19, 29, 30].

Spencer [41] considered generalized quasirandom graphs with q parts with an additional assumption that vertices in distinct parts have distinct degrees, and established that the structure of such graphs is forced by homomorphism densities of graphs with at most $8q-4$ vertices. Addressing a question posed in [41], we show (Theorem 2) that graphs with at most $\max\{2q + 1, 4\}$ vertices suffice in this restricted setting.

We present our arguments using the language of the theory of graph limits, which is introduced in Section 2. We remark that similarly to arguments presented in [35, 34], although not explicitly stated there, our arguments also apply in a more general setting of kernels in addition to graphons (see Section 2 for the definitions of the two notions). In Section 3, we state our main results and sketch the main ideas of their proofs.

2 Notation

We now introduce the notions and tools from the theory of graph limits that we need to present our results; we refer the reader to the monograph by Lovász [34] for a more comprehensive introduction. We also rephrase results concerning quasirandom graphs and generalized quasirandom graphs with q parts presented in Section 1 in the language of the theory of graph limits.

If H and G are two graphs, the *homomorphism density* of H in G, denoted by $t(H,G)$, is the probability that a random mapping of the vertex set of H to the vertex set of G is a homomorphism of H to G. A sequence $(G_n)_{n\in\mathbb{N}}$ of graphs is *convergent* if the number of vertices of G_n tends to infinity and the values of $t(H,G_n)$ converge for every graph H as $n\to\infty$. A sequence $(G_n)_{n\in\mathbb{N}}$ of graphs is *quasirandom with density* p if it is convergent and the limit of $t(H,G_n)$ is equal to $p^{|E(H)|}$ for every graph H, where $E(H)$ denotes the edge set of H. If the particular value of p is irrelevant, we just say that a sequence of graphs is quasirandom instead of quasirandom with density p.

The theory of graph limits provides analytic ways of representing sequences of convergent graphs. A *kernel* is a bounded measurable function $U:[0,1]^2\to\mathbb{R}$ that is *symmetric*, i.e., $U(x,y)=U(y,x)$ for all $(x,y)\in[0,1]^2$. The points in the domain $[0,1]$ of a kernel are often referred to as *vertices*. A *graphon* is kernel whose values are restricted to $[0,1]$. The *homomorphism density* of a graph H in a kernel U is defined as follows:

$$t(H,U)=\int_{[0,1]^{V(H)}}\prod_{uv\in E(H)}U(x_u,x_v)\mathrm{d}x_{V(H)};$$

we often just briefly say the *density* of a graph H in a kernel U rather than the homomorphism density of H in U. A graphon W is a *limit* of a convergent sequence $(G_n)_{n\in\mathbb{N}}$ of graphs if $t(H,W)$ is the limit of $t(H,G_n)$ for every graph H. Every convergent sequence of graphs has a limit graphon and every graphon is a limit of a convergent sequence of graphs [36]; also see [16] for relation to exchangeable arrays. Two kernels (or graphons) U_1 and U_2 are *weakly isomorphic* if $t(H,U_1)=t(H,U_2)$ for every graph H. Note that any two limits of a convergent sequence of graphs are weakly isomorphic, and we refer particularly to [1] for results on the structure of weakly isomorphic graphons and more generally kernels.

We now revisit the notion of quasirandom graphs using the language of the theory of graph limits. Observe that a sequence of graphs is quasirandom with density p if and only if the sequence is convergent and its limit is the graphon equal to p everywhere. Hence, the following holds for every graphon W: W is weakly isomorphic to the graphon equal to p everywhere if and only if $t(K_2,W)=p$ and $t(C_4,W)=p^4$. More strongly, we say that a kernel (or graphon) U is *forced* by graphs contained in a set $\mathcal{H}$ if every kernel U' such that $t(H,U')=t(H,U)$ for every graph $H\in\mathcal{H}$ is weakly isomorphic to U. In particular, the constant graphon is forced by the graphs K_2 and C_4.

A *q-step kernel* U is a kernel such that $[0,1]$ can be partitioned to q non-null measurable sets $A_1,\dots,A_q$ such that U is constant on $A_i\times A_j$ for all $1\le i,j\le q$ but no such partition into $q-1$ parts exists; a *q-step graphon* is a q-step kernel that is also a graphon. If the value of q is not important, we just briefly say a *step kernel* or a *step graphon*. Observe that step graphons correspond to stochastic block models and so to generalized quasirandom graphs discussed in Section 1. In particular, the result of Lovász and Sós [35] mentioned in Section 1 asserts that every q-step graphon is forced by graphs with at most $(10q)^q+q$ vertices, and the result of Lovász [34, Theorem 5.33] that every q-step graphon is forced by graphs with at most $4(2q+3)^8$ vertices.

3 Results

We now state our two main results and sketch the ideas behind their proofs.

Theorem 1. *The following holds for every $q \geq 2$ and every q-step kernel U: if the density of each graph with at most $4q^2 - q$ vertices in a kernel U' is the same as in U, then the kernels U and U' are weakly isomorphic.*

To sketch the proof of Theorem 1, we need to recall the notion of a quantum graph: a *quantum graph* is a finite linear combination of graphs (called *constituents*) and the *density* of a quantum graph G in a kernel U is the linear combination of densities of graphs forming G in U with the coefficients as in G. Fix now a q-step kernel U, and let U' be another kernel such that the density of each graph with at most $4q^2 - q$ vertices in U' is the same as in U. Lovász [34, Proposition 14.44] established the existence of a quantum graph Q_k with constituents having $k(k+1)$ vertices such that $t(Q_k, U'') = 0$ if and only if U'' is weakly isomorphic to a step kernel with at most $k-1$ parts. It follows that $t(Q_q, U) \neq 0$ and $t(Q_{q+1}, U) = 0$ and so $t(Q_q, U') \neq 0$ and $t(Q_{q+1}, U') = 0$, which yields that U' is a q-step kernel.

The main step of our argument is a construction of a quantum graph $P_{s_1,\ldots,s_q}$ with $s_1 + \cdots + s_q$ roots, which are split into q groups of $s_1, \ldots, s_q$ roots, with the following property: when each root of $P_{s_1,\ldots,s_q}$ is assigned a vertex of a q-step kernel, i.e., a point of $[0,1]$, the rooted quantum graph $P_{s_1,\ldots,s_q}$ evaluates to zero unless the roots in each of the q groups are chosen from the same part of the step kernel. We show that there exists a quantum rooted graph $P_{s_1,\ldots,s_q}$ for each choice of parameters $s_1, \ldots, s_q$ between $q+2$ and $2q+2$ such that

- each constituent of $P_{s_1,\ldots,s_q}$ has at most $s_1 + \cdots + s_q + 2q(q-1)$ vertices, and
- if the roots in the same group are chosen from the same part but roots from different groups are from different parts, then the value of $P_{s_1,\ldots,s_q}$ is non-zero and does not depend on the parameters $s_1, \ldots, s_q$.

By introducing edges between some of the roots of $P_{s_1,\ldots,s_q}$, it is possible to extract the values of the densities of U' within the q parts and between the pairs of the parts, and so these values need to be the same as the corresponding values in U. If we consider different choices of the parameters $s_1, \ldots, s_q$ in addition to introducing edges between the roots, it is also possible to extract a system of q equations that determines the sizes of the parts of U' uniquely, which yields that the kernels U and U' are weakly isomorphic. Finally, the analysis of the range of parameters $s_1, \ldots, s_q$ needed in the argument yields the bound given in Theorem 1 on the number of vertices of graphs that need to be considered.

To state our second result, recall that if U is a kernel and $x \in [0,1]$ is a vertex of U, then the *degree* of x is

$$\int_{[0,1]} U(x,y)\mathrm{d}y.$$

Theorem 2. *The following holds for every $q \geq 2$ and every q-step kernel U such that the degrees of vertices in different parts are different: if the density of each graph with at most $2q+1$ vertices in a kernel U' is the same as in U, then the kernels U and U' are weakly isomorphic.*

We now sketch the proof of Theorem 2. Fix $q \geq 2$ and a q-step kernel U with properties given in the statement of Theorem 2 and let U' be another kernel such that the density of each graph with at most $2q+1$ vertices in U' is the same as in U. To prove Theorem 2, we construct for every choice of reals $d_1, \dots, d_q \in \mathbb{R}$ a quantum graph $G_{d_1,\dots,d_q}$ with $2q+1$ vertices such that the density of $G_{d_1,\dots,d_q}$ in a kernel is zero if and only if the degree of almost every vertex of the kernel is equal to one of the values $d_1, \dots, d_q$. The assumption of Theorem 2 now yields that the sets of the degrees of the vertices of the kernels U and U' are the same. We next construct a quantum graph with q vertices, one of them being a root, which forces the root to be from a part of a step kernel with a specific degree. These rooted quantum graphs are then used to force the sizes of the parts, the densities within the parts and between all pairs of the parts. Finally, we use the fact that a step kernel (see [12, Lemma 11], also see [34, Proposition 14.14]) is the minimizer of the density of C_4 among all partitioned kernels with same sizes of the parts, densities within the parts and between the pairs of the parts, to conclude that the kernels U and U' are weakly isomorphic.

We conclude by stating as an open problem whether it suffices in Theorem 1 to consider homomorphism densities of graphs with $o(q^2)$ vertices. To supplement the open problem, we show that the order of graphs needs to be at least linear in q. Our argument is similar to that used in analogous scenarios, e.g., in [18, 21]. For reals $a_1, \dots, a_q > 0$ such that $a_1 + \dots + a_q < 1$, let $U_{a_1,\dots,a_q}$ be the $(q+1)$-step graphon with parts whose sizes are $a_1, \dots, a_q$ and $1 - a_1 - \dots - a_q$, and that is equal to one within each of the first q parts and to zero elsewhere. Observe that if H is a graph that, after removing isolated vertices, consists of k components with respectively $n_1, \dots, n_k$ vertices then

$$t\left(H, U_{a_1,\dots,a_q}\right) = \prod_{i=1}^{k} \sum_{j=1}^{q} a_j^{n_i}.$$

It follows that if

$$t\left(K_{\ell+1}, U_{a_1,\dots,a_q}\right) = t\left(K_{\ell+1}, U_{a'_1,\dots,a'_q}\right) \text{ for every } \ell = 1, \dots, q-1, \tag{1}$$

then the homomorphism density of every graph with at most q vertices is the same in $U_{a_1,\dots,a_q}$ and in $U_{a'_1,\dots,a'_q}$. View $(t(K_{\ell+1}, U_{a_1,\dots,a_q}))_{\ell=1}^{q-1} \in \mathbb{R}^{q-1}$ as a function of $a_1, \dots, a_{q-1}$. If its arguments $a_1, \dots, a_{q-1}$ are distinct, then the Jacobian matrix can be shown to be invertible and the Implicit Function Theorem gives, for every a'_q sufficiently close to a_q, a vector $(a'_1, \dots, a'_{q-1})$ close to $(a_1, \dots, a_{q-1})$ such that (1) holds. It follows that there are two non-weakly-isomorphic $(q+1)$-step graphons that have the same homomorphism density of every graph with at most q vertices.

References

[1] C. Borgs, J. Chayes and L. Lovász: *Moments of two-variable functions and the uniqueness of graph limits*, Geom. Funct. Anal. **19** (2010), 1597–1619.

[2] C. Borgs, J. T. Chayes, H. Cohn and S. Ganguly: *Consistent nonparametric estimation for heavy-tailed sparse graphs*, Ann. Statist. **49** (2021), 1904–1930.

[3] M. Bucić, E. Long, A. Shapira and B. Sudakov: *Tournament quasirandomness from local counting*, Combinatorica **41** (2021), 175–208.

[4] T. Chan, D. Král', J. A. Noel, Y. Pehova, M. Sharifzadeh and J. Volec: *Characterization of quasirandom permutations by a pattern sum*, Random Structures & Algorithms **57** (2020), 920–939.

[5] F. R. K. Chung and R. L. Graham: *Quasi-random hypergraphs*, Random Structures & Algorithms **1** (1990), 105–124.

[6] F. R. K. Chung and R. L. Graham: *Quasi-random set systems*, Journal of the American Mathematical Society **4** (1991), 151–196.

[7] F. R. K. Chung and R. L. Graham: *Quasi-random tournaments*, Journal of Graph Theory **15** (1991), 173–198.

[8] F. R. K. Chung and R. L. Graham: *Quasi-random subsets of* Z_n, Journal of Combinatorial Theory, Series A **61** (1992), 64–86.

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

[10] J. N. Cooper: *Quasirandom permutations*, Journal of Combinatorial Theory, Series A **106** (2004), 123–143.

[11] J. W. Cooper, D. Král', A. Lamaison and S. Mohr: *Quasirandom Latin squares*, Random Structures & Algorithms **61** (2022), 298–308.

[12] J. W. Cooper, D. Král' and T. Martins: *Finitely forcible graph limits are universal*, Adv. Math. **340** (2018), 819–854

[13] L. N. Coregliano, R. F. Parente and C. M. Sato: *On the maximum density of fixed strongly connected subtournaments*, The Electronic Journal of Combinatorics **26** (2019), P1.44.

[14] L. N. Coregliano and A. A. Razborov: *On the density of transitive tournaments*, Journal of Graph Theory **85** (2017), 12–21.

[15] L. N. Coregliano and A. A. Razborov: *Natural quasirandomness properties*, preprint arXiv:2012.11773 (2020).

[16] P. Diaconis and S. Janson: *Graph limits and exchangeable random graphs*, Rend. Mat. Appl. **28** (2008), 33–61.

[17] S. Eberhard, F. Manners and R. Mrazović: *Transversals in quasirandom Latin squares*, preprint arXiv:2209.02180 (2022).

[18] P. Erdős, L. Lovász and J. Spencer: *Strong independence of graphcopy functions*, Graph theory and related topics (1979), 165–172.

[19] C. Gao, Y. Lu and H. H. Zhou: *Rate-optimal graphon estimation*, Ann. Statist. **43** (2015), 2624–2652.

[20] F. Garbe, R. Hancock, J. Hladký and M. Sharifzadeh: *Limits of Latin squares*, preprint arXiv:2010.07854 (2020).

[21] R. Glebov, A. Grzesik, T. Klimošová and D. Kráľ: *Finitely forcible graphons and permutons*, J. Combin. Theory Ser. B **110** (2015), 112–135.

[22] W. T. Gowers: *Quasirandomness, counting and regularity for 3-uniform hypergraphs*, Combinatorics, Probability and Computing **15** (2006), 143–184.

[23] W. T. Gowers: *Hypergraph regularity and the multidimensional Szemerédi theorem*, Annals of Mathematics, Second Series **166** (2007), 897–946.

[24] W. T. Gowers: *Quasirandom groups*, Combinatorics, Probability and Computing **17** (2008), 363–387.

[25] W. T. Gowers and J. Long: *Partial associativity and rough approximate groups*, Geometric and Functional Analysis **30** (2020), 1–65.

[26] A. Grzesik, D. Il'kovič, B. Kielak and D. Kráľ: *Quasirandom forcing orientations of cycles*, preprint arXiv:2212.09343 (2022).

[27] R. Hancock, A. Kabela, D. Kráľ, T. Martins, R. Parente, F. Skerman and J. Volec: *No additional tournaments are quasirandom-forcing*, European Journal of Combinatorics **108** (2023), 103632.

[28] J. Haviland and A. Thomason: *Pseudo-random hypergraphs*, Discrete Math. **75** (1989), 255–278.

[29] O. Klopp, A. B. Tsybakov and N. Verzelen: *Oracle inequalities for network models and sparse graphon estimation*, Ann. Statist. **45** (2017), 316–354.

[30] O. Klopp and N. Verzelen: *Optimal graphon estimation in cut distance*, Probab. Theory Related Fields **174** (2019), 1033–1090.

[31] Y. Kohayakawa, V. Rödl and J. Skokan: *Hypergraphs, quasi-randomness, and conditions for regularity*, Journal of Combinatorial Theory, Series A **97** (2002), 307–352.

[32] D. Král' and O. Pikhurko: *Quasirandom permutations are characterized by 4-point densities*, Geometric and Functional Analysis **23** (2013), 570–579.

[33] M. Kurečka: *Lower bound on the size of a quasirandom forcing set of permutations*, Combinatorics, Probability and Computing **31** (2022), 304–319.

[34] L. Lovász: Large Networks and Graph Limits, *Colloquium Publications*, volume 60, 2012.

[35] L. Lovász and V. Sós: *Generalized quasirandom graphs*, J. Combin. Theory Ser. B **98** (2008), 146–163.

[36] L. Lovász and B. Szegedy: *Limits of dense graph sequences*, J. Combin. Theory Ser. B **96** (2006), 933–957.

[37] B. Nagle, V. Rödl and M. Schaht: *The counting lemma for regular k-uniform hypergraphs*, Random Structures & Algorithms **28** (2006), 113–179.

[38] V. Rödl: *On universality of graphs with uniformly distributed edges*, Discrete Mathematics **59** (1986), 125–134.

[39] V. Rödl and J. Skokan: *Regularity lemma fork-uniform hypergraphs*, Random Structures & Algorithms **25** (2004), 1–42.

[40] M. Simonovits and V. T. Sós: *Szemerédi's partition and quasirandomness*, Random Structures & Algorithms **2** (1991), 1–10.

[41] J. Spencer: *Quasirandom multitype graphs*, in: An irregular mind, *Bolyai Soc. Math. Stud.*, volume 21 (2010), 607–617.

[42] A. Thomason: *Pseudo-random graphs*, Annals of Discrete Mathematics **144** (1987), 307–331.

[43] A. Thomason: *Random graphs, strongly regular graphs and pseudo-random graphs*, in: Surveys in Combinatorics, *London Mathematical Society Lecture Note Series*, volume 123 (1987), 173–196.