Combinatorics, Probability and Computing (2017) 26, 195–200. © Cambridge University Press 2016
doi:10.1017/S0963548316000298

A Sequence of Triangle-Free Pseudorandom Graphs

DAVID CONLON†

Mathematical Institute, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, UK
(e-mail: david.conlon@maths.ox.ac.uk)

Received 1 February 2016; revised 1 August 2016; first published online 13 September 2016

A construction of Alon yields a sequence of highly pseudorandom triangle-free graphs with edge density significantly higher than one might expect from comparison with random graphs. We give an alternative construction for such graphs.

2010 *Mathematics subject classification*: Primary 05C35
Secondary 05B25

1. Introduction

A graph G is said to be (p,β)-jumbled if

$$\left|e(X)-p\binom{|X|}{2}\right|\leqslant\beta|X|$$

for all $X\subseteq V(G)$. For example, the binomial random graph $G_{n,p}$ is (p,β)-jumbled with $\beta=O(\sqrt{pn})$. It is not hard to show [5, 6] that this is essentially best possible, in that a graph with n vertices cannot be (p,β)-jumbled with $\beta=c\sqrt{pn}$ for c sufficiently small. For further information on jumbled graphs and their properties, we refer the reader to the survey [8] or, for more recent developments, the paper [4].

One important class of (p,β)-jumbled graphs is the collection of (n,d,λ)-graphs. These are d-regular graphs on n vertices such that all eigenvalues of the adjacency matrix, save the largest, are bounded in absolute value by λ. By the famous expander mixing lemma, these graphs are (p,β)-jumbled with $p=d/n$ and $\beta=\lambda$.

One of the best known examples of a pseudorandom graph, constructed by Alon [1], is a triangle-free (n,d,λ)-graph with $n=2^{3k}$, $d=2^{k-1}(2^{k-1}-1)$ and $\lambda=O(2^k)$. Taking $p=d/n$, we have $\sqrt{pn}=\sqrt{d}=\Omega(2^k)$, so the graph is close to optimally pseudorandom. Since $p=\Omega(n^{-1/3})$,

† Research supported by a Royal Society University Research Fellowship and by ERC Starting Grant 676632.

https://doi.org/10.1017/S0963548316000298 Published online by Cambridge University Press

the construction also has surprisingly high density. While there are various ways to modify the usual random graph to produce triangle-free graphs with density roughly $n^{-1/2}$ (see, for example, [3]), no such modification can hope to push very far past this density. Nevertheless, Alon's construction does so. The purpose of this note is to give another construction for such graphs.

Theorem 1.1. *There exists a sequence $(n_i)_{i=1}^{\infty}$ of positive integers such that, for each $i \geqslant 1$, there is a triangle-free graph G_i on n_i vertices which is (p,β)-jumbled with $p = \Omega(n_i^{-1/3})$ and $\beta = O(\sqrt{pn_i}\log n_i)$.*

Our construction is weaker than Alon's on several counts: it does not produce regular graphs; it is not completely explicit; and it does not generalize easily. One might also level the accusation that the resulting graphs are not optimally pseudorandom, with the condition $\beta = O(\sqrt{pn}\log n)$ being a logarithmic factor away from the desired bound. However, it seems likely that this extra log factor is simply an artifact of our proof. Countering these disadvantages, we believe that our construction, which we describe in detail below, is more intuitive than Alon's.

For concreteness, we will work with the polarity graph of Lazebnik, Ustimenko and Woldar [9], though the role played by this graph could also be taken by a number of other C_6-free graphs with n vertices and $\Omega(n^{4/3})$ edges. Suppose then that q is an odd power of 2 and let $n = q^3 + q^2 + q + 1$. The polarity graph, which has a small number of loops, is an (n,d,λ)-graph with $d = q+1 \geqslant n^{1/3}$ and, as noted in [8, Section 3.7], $\lambda = \sqrt{2q} = O(n^{1/6})$. Once the loops are removed, the resulting graph, which we label by H, is C_3-, C_4- and C_6-free. For each vertex v in H, we randomly partition its neighbourhood $N_H(v)$ into two sets A_v and B_v and let G_v be the complete bipartite graph between A_v and B_v. We now define G to be the graph with the same vertex set as H and edge set $\cup_{v\in V(H)} G_v$. We will show that G asymptotically almost surely satisfies the requirements of Theorem 1.1.

It is straightforward to verify that G contains no triangles. To begin, note that since H is C_4-free, the only edges of G in $N_H(v)$ are those in G_v. Now suppose that abc is a triangle in G. If a,b and c are all contained in the same neighbourhood $N_H(v)$ they cannot form a triangle, since the only edges of G in $N_H(v)$ are those in G_v and G_v is bipartite. It must then be the case that the edges ab, bc and ca are contained in three different neighbourhoods $N_H(u)$, $N_H(v)$ and $N_H(w)$, respectively. If u,v and w are all distinct from a,b and c, then $aubvcw$ would form a cycle of length 6 in H, contradicting the fact that H is C_6-free. On the other hand, if $u = c$, say, then H must contain the triangle bvc, again contradicting the choice of H.

The remaining claim, that G is asymptotically almost surely (p,β)-jumbled with $p = \Omega(n^{-1/3})$ and $\beta = O(\sqrt{pn}\log n)$, will be verified in the next section.

2. Proving G is jumbled

Let H_0 be the polarity graph with $n = q^3 + q^2 + q + 1$ vertices. This graph is $(q+1)$-regular, has q^2+1 loops and all eigenvalues of the adjacency matrix, save the largest, are bounded in absolute value by $\sqrt{2q}$. We will form G from H_0 by a slightly different procedure to that described in the Introduction, though the two are easily seen to produce the same graph.

https://doi.org/10.1017/S0963548316000298 Published online by Cambridge University Press

In the first step, we form $H_0^{(2)}$, the multigraph with loops on the same vertex set as H_0 where two vertices are joined if there is a walk of length two in H_0 between them, allowing for multiple edges if there is more than one such walk. We note that each vertex has $q+1$ loops, one for each edge in H_0, and, since H, the simple graph formed by removing the loops from H_0, is C_3- and C_4-free, the only parallel edges arise from loops in H_0. In the next step, we turn $H_0^{(2)}$ into a simple graph by removing all loops from $H_0^{(2)}$ and all edges whose corresponding walk in H_0 used a loop. The resulting graph G_1 is easily seen to be the union of n cliques, each clique being $N_H(v)$ for some v in H. We now form the required graph, as before, by randomly partitioning $N_H(v)$ into two sets A_v and B_v and letting G be the union over all v in H of the complete bipartite graphs between A_v and B_v.

Following this plan, we first look at $H_0^{(2)}$. Letting M be the adjacency matrix of H_0, the adjacency matrix of $H_0^{(2)}$ is simply M^2, which implies that the eigenvalues of $H_0^{(2)}$ are the squares of the eigenvalues of H_0. Therefore, $H_0^{(2)}$ is an (n,d,λ)-graph with $d=(q+1)^2$ and $\lambda=2q$. Since each vertex of $H_0^{(2)}$ is contained in exactly $q+1$ loops, the graph G_0 formed by removing these loops has adjacency matrix $M^2-(q+1)I$, implying that G_0 is an (n,d,λ)-graph with $d=q(q+1)$ and $\lambda=q+1$. Therefore, by the expander mixing lemma, for all $X\subseteq V(G_0)$,

$$\left|e_{G_0}(X)-\frac{q}{q^2+1}\binom{|X|}{2}\right|\leqslant (q+1)|X|,$$

where we used the fact that $n=q^3+q^2+q+1=(q^2+1)(q+1)$.

Note now that every loop in H_0 has exactly q other neighbours with which it can form a non-degenerate walk of length two. Therefore, since any $X\subseteq V(H_0)$ contains at most $|X|$ loops, we remove at most $q|X|$ edges from X when forming G_1 from G_0. By the estimate above, this implies that, for all $X\subseteq V(G_1)$,

$$\left|e_{G_1}(X)-\frac{q}{q^2+1}\binom{|X|}{2}\right|\leqslant (2q+1)|X|.$$

Therefore, G_1 is (p,β)-jumbled with $p=q/(q^2+1)$ and $\beta=2q+1$.

Recall that G_1 is the union of cliques, while G is the union of random bipartite graphs, one for each clique in G_1. Our aim now is to show that asymptotically almost surely G is (p,β)-jumbled with $p=q/2(q^2+1)=\Omega(n^{-1/3})$ and $\beta=O(q\log n)=O(\sqrt{pn}\log n)$. We will do this in a rather naive fashion, estimating the probability that

$$\left|e_G(X)-\frac{q}{2(q^2+1)}\binom{|X|}{2}\right|\leqslant Cq|X|\log n \tag{2.1}$$

for any given X and taking a union bound. To do this, we will need the following concentration inequality for quadratic forms in independent random variables due to Hanson and Wright [7] (the exact version we state follows from Theorem 1.1 in [10]).

Lemma 2.1. *Let $Z=(Z_1,\dots,Z_t)\in\{-1,+1\}^t$ be a random vector with independent components, each of which is equal to 1 or -1 with probability $1/2$. Let M be a $t\times t$ real matrix. Then*

$$\mathbb{P}[|Z^TMZ-\mathbb{E}(Z^TMZ)|>\varepsilon]\leqslant 2\exp\left\{-c\min\left(\frac{\varepsilon^2}{\|M\|_F^2},\frac{\varepsilon}{\|M\|}\right)\right\},$$

https://doi.org/10.1017/S0963548316000298 Published online by Cambridge University Press

where $\|M\|_F = (\sum_{i,j} m_{ij}^2)^{1/2}$ *is the Frobenius norm and* $\|M\| = \sup_{x \neq 0} \|Mx\|_2 / \|x\|_2$ *is the spectral norm.*

Suppose now that $G[X]$ is the union of s cliques $T_1, \dots, T_s$, of orders $t_1, \dots, t_s$, and let $t = t_1 + \cdots + t_s$. We define t random variables $X_1, \dots, X_t$, each equal to 1 or -1 with probability $1/2$, and assign one of these random variables to every vertex of every clique, noting that any given vertex may receive multiple random variables, but only one relative to any given clique.

Suppose that the random variables assigned to the clique T_i are $Z_{i1}, \dots, Z_{it_i}$. If v_i is the vertex whose neighbourhood in H contains T_i, the value of Z_{ij} determines whether its corresponding vertex $T_i(j)$ is placed in A_{v_i} or B_{v_i}, with $T_i(j)$ placed in A_{v_i} if $Z_{ij} = 1$ and B_{v_i} if $Z_{ij} = -1$. The number of edges in $G[T_i]$ is then

$$e_G(T_i) = |A_{v_i}||B_{v_i}| = \left(\frac{t_i}{2} + \frac{1}{2}\sum_{j=1}^{t_i} Z_{ij}\right)\left(\frac{t_i}{2} - \frac{1}{2}\sum_{j=1}^{t_i} Z_{ij}\right) = \frac{t_i^2}{4} - \frac{1}{4}\sum_{j=1}^{t_i}\sum_{k=1}^{t_i} Z_{ij}Z_{ik}.$$

Summing over i, we have

$$\begin{aligned} e_G(X) = \sum_{i=1}^{s} e_G(T_i) &= \sum_{i=1}^{s} \frac{t_i^2}{4} - \frac{1}{4}\sum_{i=1}^{s}\sum_{j=1}^{t_i}\sum_{k=1}^{t_i} Z_{ij}Z_{ik} \\ &= \frac{1}{2}\sum_{i=1}^{s}\binom{t_i}{2} + \frac{1}{4}\sum_{i=1}^{s} t_i - \frac{1}{4}\sum_{i=1}^{s}\sum_{j=1}^{t_i}\sum_{k=1}^{t_i} Z_{ij}Z_{ik} \\ &= \frac{e_{G_1}(X)}{2} - \frac{1}{4}\sum_{i=1}^{s}\sum_{1 \leqslant j \neq k \leqslant t_i} Z_{ij}Z_{ik}. \end{aligned}$$

Therefore, by our estimate on $e_{G_1}(X)$, it only remains to show that

$$Q = \sum_{i=1}^{s}\sum_{1 \leqslant j \neq k \leqslant t_i} Z_{ij}Z_{ik}$$

is smaller than $Cq|X|\log n$ with sufficiently high probability.

Let M be the $t \times t$ matrix whose entry m_{jk} is equal to 1 if $j \neq k$ and j and k (each of which represents a vertex associated to a particular clique) are from the same clique T_i. Otherwise, we take m_{jk} to be 0. Then $Q = Z^T M Z$ and it follows from the Hanson–Wright bound that

$$\mathbb{P}[|Q| > Cq|X|\log n] \leqslant 2\exp\left\{-c\min\left(\frac{C^2q^2|X|^2\log^2 n}{\|M\|_F^2}, \frac{Cq|X|\log n}{\|M\|}\right)\right\}.$$

But it is straightforward to verify that

$$\|M\|_F^2 = \sum_{i=1}^{s} t_i(t_i - 1) = 2e_{G_1}(X) \leqslant \frac{|X|^2}{q} + 6q|X| \leqslant 2\max\left\{\frac{|X|^2}{q}, 6q|X|\right\}$$

and, writing $\rho(A)$ for the spectral radius of a matrix A,

$$\|M\| = \sqrt{\rho(M^*M)} = \rho(M) \leqslant \sup_{x \neq 0} \frac{\|Mx\|_\infty}{\|x\|_\infty} = \max_{1 \leqslant j \leqslant t} \sum_{k=1}^{t} |m_{jk}| \leqslant q.$$

https://doi.org/10.1017/S0963548316000298 Published online by Cambridge University Press

Therefore,

$$\mathbb{P}[|Q| > Cq|X|\log n] \leqslant 2\exp\left\{-c\min\left(\frac{1}{2}C^2q^3\log^2 n, \frac{1}{12}C^2q|X|\log^2 n, C|X|\log n\right)\right\}$$
$$\leqslant 2\exp\{-2|X|\log n\}$$

for C sufficiently large in terms of c. Applying the union bound, we see that the probability that there exists a set X such that (2.1) fails is at most

$$\sum_{|X|=1}^{n}\binom{n}{|X|}2e^{-2|X|\log n} \leqslant 2\sum_{|X|=1}^{n} n^{|X|}e^{-2|X|\log n} = 2\sum_{|X|=1}^{n} e^{-|X|\log n} \leqslant \frac{2}{n-1}.$$

The result follows.

3. Concluding remarks

Our construction also extends to give a sequence $(n_i)_{i=1}^{\infty}$ of positive integers such that, for each $i \geqslant 1$, there is a C_5-free graph G_i on n_i vertices which is (p,β)-jumbled with $p = \Omega(n_i^{-3/5})$ and $\beta = O(\sqrt{pn_i}\log n_i)$. The construction starts with the C_{10}-free polarity graph of Lazebnik, Ustimenko and Woldar [9], but follows the proof of Theorem 1.1 in all other respects. Because we lack optimal constructions of $C_{2\ell}$-free graphs for $\ell \geqslant 6$, our method does not extend further to give constructions of pseudorandom C_{2k+1}-free graphs for $k \geqslant 3$. As mentioned in the Introduction, this is a distinct weakness of our method when compared to Alon's, which does extend to longer odd cycles [2, 8].

An alternative method for proving the jumbledness of our construction G might be to estimate the eigenvalues of its adjacency matrix M by using the fact that $\mathrm{Tr}(M^{2k})$ is both the sum of the $2k$th powers of its eigenvalues and the number of walks of length $2k$ in G. However, G is constructed by starting from a graph G_1 which is a union of cliques and then taking a random bipartite graph within each clique. This process creates an imbalance between odd and even cycles within each clique, deleting all odd cycles, while doubling the proportion of even cycles relative to the density. This makes it difficult to count the number of degenerate walks of length $2k$ without having close control over the counts of degenerate walks of different types in the base graph G_1. Nevertheless, it is plausible that this could be done, and may even allow one to save the lost logarithmic factor in β.

Another definite weakness of our method is that it is not explicit and so, unlike Alon's example, cannot be used to give a constructive lower bound for the off-diagonal Ramsey number $r(3,t)$. It remains to decide whether there is some more explicit method for choosing large bipartite subgraphs of the cliques in G_1 which also produces a highly pseudorandom subgraph.

Acknowledgements

I would like to thank the anonymous referee for a number of helpful remarks.

References

[1] Alon, N. (1994) Explicit Ramsey graphs and orthonormal labelings. *Electron. J. Combin.* **1** R12.

https://doi.org/10.1017/S0963548316000298 Published online by Cambridge University Press

[2] Alon, N. and Kahale, N. (1998) Approximating the independence number via the θ-function. *Math. Program.* **80** 253–264.
[3] Bohman, T. (2009) The triangle-free process. *Adv. Math.* **221** 1653–1677.
[4] Conlon, D., Fox, J. and Zhao, Y. (2014) Extremal results in sparse pseudorandom graphs. *Adv. Math.* **256** 206–290.
[5] Erdős, P., Goldberg, M., Pach, J. and Spencer, J. (1988) Cutting a graph into two dissimilar halves. *J. Graph Theory* **12** 121–131.
[6] Erdős, P. and Spencer, J. (1971/72) Imbalances in k-colorations. *Networks* **1** 379–385.
[7] Hanson, D. L. and Wright, F. T. (1971) A bound on tail probabilities for quadratic forms in independent random variables. *Ann. Math. Statist.* **42** 1079–1083.
[8] Krivelevich, M. and Sudakov, B. (2006) Pseudo-random graphs. In *More Sets, Graphs and Numbers*, Vol. 15 of *Bolyai Society Mathematical Studies*, Springer, pp. 199–262.
[9] Lazebnik, F., Ustimenko, V. A. and Woldar, A. J. (1999) Polarities and 2k-cycle-free graphs. *Discrete Math.* **197/198** 503–513.
[10] Rudelson, M. and Vershynin, R. (2013) Hanson–Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.* **18** 1–9.

https://doi.org/10.1017/S0963548316000298 Published online by Cambridge University Press