Exact stability for Turán's Theorem

Dániel Korándi[*][†] Alexander Roberts[*] Alex Scott[*]

September 23, 2021

Abstract

Turán's Theorem says that an extremal K_{r+1}-free graph is r-partite. The Stability Theorem of Erdős and Simonovits shows that if a K_{r+1}-free graph with n vertices has close to the maximal $t_r(n)$ edges, then it is close to being r-partite. In this paper we determine exactly the K_{r+1}-free graphs with at least m edges that are farthest from being r-partite, for any $m \geq t_r(n) - \delta_r n^2$. This extends work by Erdős, Győri and Simonovits, and proves a conjecture of Balogh, Clemen, Lavrov, Lidický and Pfender.

1 Introduction

Turán's classical theorem [25] from 1941 says that a K_{r+1}-free n-vertex graph maximizing the number of edges (an *extremal graph*) is r-partite; the $r = 2$ case was established earlier by Mantel [17], in 1907. The only extremal n-vertex graph is the *Turán graph* $T_r(n)$, the complete r-partite graph with parts of size $\lfloor n/r \rfloor$ or $\lceil n/r \rceil$, which has $t_r(n) = \left(1 - \frac{1}{r} + o(1)\right)\binom{n}{2}$ edges. Turán's Theorem lay the foundations of extremal graph theory, and has been highly influential in the field ever since.

One of the early discoveries related to Turán's Theorem was that if a K_{r+1}-free graph is "close" to extremal in the number of edges, then it must be "close" to the Turán graph in its structure. Indeed, the famous Stability Theorem of Erdős and Simonovits [9, 22] from the 1960s implies the following: if G is a K_{r+1}-free n-vertex graph with $t_r(n) - o(n^2)$ edges, then it can be made into the Turán graph $T_r(n)$ by changing only $o(n^2)$ edges. It is of little surprise that this powerful structural description of near-extremal graphs has seen many important applications and consequences over the past decades (e.g. [1, 4, 19, 24]).

An alternative form of stability for Turán's Theorem is to look at the distance from being r-partite (rather than the distance to a specific r-partite graph, namely the Turán graph). Thus we are looking for a large r-partite subgraph, which is what is wanted for most applications. The two problems are equivalent if we are only looking for a $o(n^2)$ bound on

---

[*] Mathematical Institute, University of Oxford, Andrew Wiles Building, Radcliffe Observatory Quarter, Woodstock Road, Oxford, United Kingdom.
E-mail: `{korandi, robertsa, scott}@maths.ox.ac.uk`.

[†] Supported by SNSF Postdoc.Mobility Fellowship P400P2_186686.

the distance. However, for graphs that are closer to extremal, we can obtain more structural information by measuring the distance from being r-partite. For example, if we move a constant number of vertices from a smallest vertex class to a largest vertex class of $T_r(n)$ then the resulting graph has $t_r(n) - O(1)$ edges but distance $\Omega(n)$ from the Turán graph. In contrast, a K_{r+1}-free graph on n vertices with at least $t_r(n) - c_r n$ edges must already be r-partite. This phenomenon was first studied by Simonovits [20] and later by many other authors [7, 14, 16, 2, 26]. A tight result was proved by Brouwer [7]:

Theorem 1.1 ([7])**.** *Let $r \geq 2$ and $n \geq 2r+1$ be integers. Every K_{r+1}-free graph with at least $t_r(n) - \lfloor n/r \rfloor + 2$ edges is r-partite.*

Let $f_r(n,t)$ be the smallest number such that any K_{r+1}-free graph G with at least $t_r(n)-t$ edges can be made r-partite by deleting at most $f_r(n,t)$ edges. For fixed r, Theorem 1.1 tells us that $f_r(n,t) = 0$ for $t \leq n/r + O(1)$, while the Stability Theorem tells us that $f_r(n,t) = o(n^2)$ if $t = o(n^2)$. But what happens in between? Better estimates of this function have only been obtained fairly recently. In a short and elegant paper, Füredi [13] proved that $f_r(n,t) \leq t$. Later, Roberts and Scott [18] showed that $f_r(n,t) = O(t^{3/2}/n)$ when $t \leq \delta n^2$, and that this bound is tight up to a constant factor (in fact, they proved much more general results for H-free graphs, where H is edge-critical). Very recently, Balogh, Clemen, Lavrov, Lidický and Pfender [5] determined $f_r(n,t)$ asymptotically, and made a conjecture on its exact value. The main aim of this paper is to prove their conjecture.

When $r = 2$, the exact stability problem was already solved by Erdős, Győri and Simonovits [12]: they proved that for $t \leq n^2/20$ the worst triangle-free graph, defining $f_2(n,t)$, is a blowup of C_5. One can generalize this construction to obtain a family of K_{r+1}-free graphs with many edges as follows. Consider a complete $(r-1)$-partite graph with parts $Z, Z_3, \dots, Z_r$, and insert a blowup of C_5 on Z with independent sets X, Y_1, Y_2, Z_1, Z_2 as in Figure 1 (so $Z = X \cup Y_1 \cup Y_2 \cup Z_1 \cup Z_2$). We will call this a *pentagonal Turán graph* if it further satisfies $|X| \leq |Y_1| = |Y_2| \leq |Z_i|$ for every $i \in [r]$, and each of the sets $X \cup Y_1 \cup Z_1, X \cup Y_2 \cup Z_2, Z_3, \dots, Z_r$ has size $\left\lfloor \frac{n+|X|}{r} \right\rfloor$ or $\left\lceil \frac{n+|X|}{r} \right\rceil$.

Balogh, Clemen, Lavrov, Lidický and Pfender [5] conjectured that $f_r(n,t)$ is witnessed by a pentagonal Turán graph if t is small enough. Our main result is a proof of their conjecture. For a graph G and integer $r \geq 2$, let $D_r(G)$ be the minimum number of edges that must be removed from G to make it r-partite. We prove the following theorem.

Theorem 1.2. *For every $r \geq 2$ there is a $\delta_r > 0$ such that the following holds: If G is a K_{r+1}-free graph on n vertices with $e(G) \geq t_r(n) - \delta_r n^2$ edges, then there is a pentagonal Turán graph G^* on n vertices with $e(G^*) \geq e(G)$ and $D_r(G^*) \geq D_r(G)$.*

The rest of the paper is organized as follows. In Section 2, we present a brief overview of the proof, and collect some necessary tools. We need a special argument when the number of edges in G is very close to $t_r(n)$, and the short proof of this case is presented in Section 3. Section 4 contains the general argument of the proof of Theorem 1.2. We finish the paper with some discussion and open problems in Section 5.

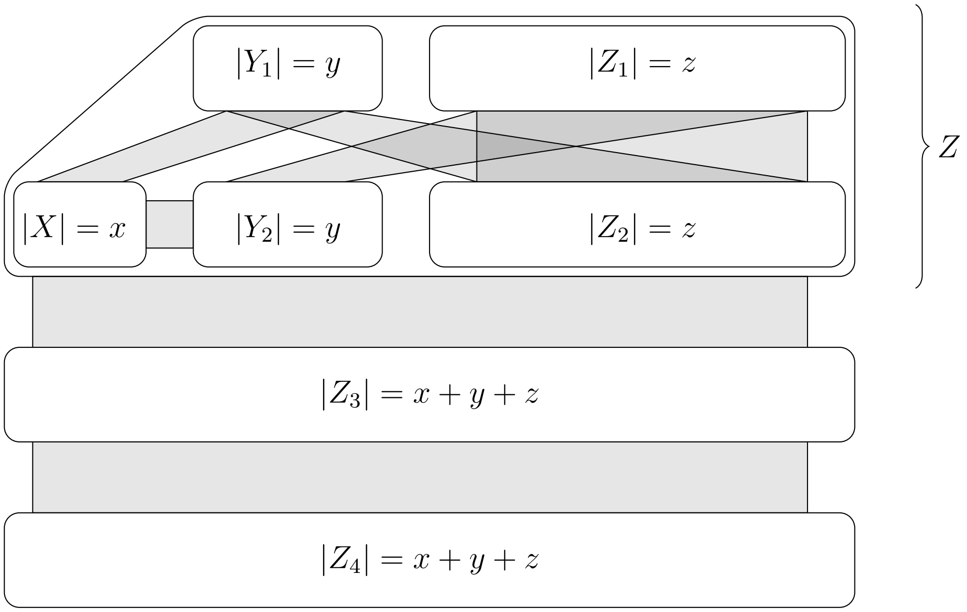

Figure 1: A pentagonal Turán graph with $r = 4$

We follow standard notation throughout. G is always a simple graph with vertex set $V(G)$ and edge set $E(G)$. The number of edges is denoted by $e(G) = |E(G)|$. We write $\Gamma_G(v) \subseteq V(G)$ to denote the neighborhood of a vertex $v \in V(G)$, and $d_G(v) = |\Gamma_G(v)|$ to denote its degree. When the graph in question is clear, we may omit the subscript. For a set of vertices $S \subseteq V(G)$, we write $G - S$ for the subgraph induced on $V(G) \setminus S$. When $S = \{v\}$, we simply write $G - v$.

2 Overview and tools

Given an r-partition of the vertices of a graph G, we say that an edge connecting different parts is *crossing*, and an edge connecting vertices in the same part is *internal*. So $D_r(G)$ is the minimum number of internal edges in an r-partition of the vertices of G.

In their proof of the triangle-free case of Theorem 1.2, Erdős, Győri and Simonovits [12] start with a close to optimal bipartition of G, and construct a pentagonal Turán graph (in this case, a blowup of C_5) with the same number of internal edges, but more crossing edges. An important idea in their proof is to find a large matching of internal edges: as G is triangle-free, this can be used to show that many crossing edges are missing from G.

Our proof for the general case follows a similar spirit, although we need to work harder to find the necessary missing edges when K_{r+1} is forbidden instead of K_3.

We will need several estimates comparing Turán numbers $t_r(n)$ for various r and n. Recall that $t_r(n)$ is the number of edges in the Turán graph $T_r(n)$, which is the complete r-partite graph on an *r-equipartitioned* vertex set, i.e., when each part has size $\lfloor n/r \rfloor$ or $\lceil n/r \rceil$. It is easy to see that $t_r(n) \geq t_r(n-1) + \frac{r-1}{r}(n-1)$, by adding a vertex to a smallest part of $T_r(n-1)$. Similarly, $t_r(n-1) \geq t_r(n) - \frac{r-1}{r}n$ can be obtained by deleting a vertex from a largest part of $T_r(n)$.

The next lemma follows from these inequalities by iterating them, and by noting that $t_r(n)$ is the unique integer between $t_r(n-1) + \frac{r-1}{r}(n-1)$ and $t_r(n-1) + \frac{r-1}{r}n$.

Lemma 2.1. *Let $r \geq 2$ and n be integers. Then:*

1. $t_r(n) = t_r(n-1) + \left\lceil \frac{r-1}{r}(n-1) \right\rceil = t_r(n-1) + \left\lfloor \frac{r-1}{r}n \right\rfloor$,
2. $t_r(n') + \frac{r-1}{r}n(n-n') \geq t_r(n) \geq t_r(n') + \frac{r-1}{r}n'(n-n')$, *for every* $n' \leq n$,
3. $\frac{r-1}{r}\binom{n+1}{2} \geq t_r(n) \geq \frac{r-1}{r}\binom{n}{2}$.

To find a large matching among the internal edges, we will use the following lemma, which follows easily from the Tutte-Berge formula (and is a special case of a theorem of Chvátal and Hanson [8]). We include a sketch of the argument for completeness.

Lemma 2.2. *Let G be a graph on n vertices with maximum degree Δ and let $k \geq 1$ be an integer. If $e(G) > (k-1)\Delta$ and $\Delta \geq 2k-1$, then G contains a matching of size k.*

Proof sketch. If G has no k-matching, then it contains a set S such that $G - S$ has at least $n - 2(k-1) + |S|$ odd components (note that perforce $|S| \leq k-1$). The number of edges in this setup is maximized when $G - S$ is the union of $n - 2(k-1) + |S| - 1$ singletons and a $(2(k-1-|S|)+1)$-clique. Then $G - S$ induces $(k-1-|S|)(2(k-1-|S|)+1) \leq (k-1-|S|)\Delta$ edges, and S touches at most $|S|\Delta$ edges, so G has at most $(k-1)\Delta$ edges, contradicting our assumption. □

For an integer vector $\mathbf{n} = (n_1, \ldots, n_r) \in \mathbb{N}^r$, let $K_{\mathbf{n}}$ be the complete r-partite graph with parts of size $n_1, \ldots, n_r$.

The next lemma will be our main tool for bounding the number of missing crossing edges using the K_{r+1}-freeness of our graph. We will generally apply it to the neighborhood of a vertex. This is a folklore result (see, for example, [6]), but we include a short proof for completeness.

Lemma 2.3. *Let $r \geq 2$ and let $\mathbf{n} = (n_1, \ldots, n_r) \in \mathbb{N}^r$ be such that $n_1 \leq n_2 \leq \cdots \leq n_r$. Then any K_r-free subgraph of $K_{\mathbf{n}}$ contains at most $e(K_{\mathbf{n}}) - n_1 n_2$ edges.*

Proof. There are exactly $\prod_{i=1}^{r} n_i$ copies of K_r in $K_{\mathbf{n}}$. Each edge is contained in at most $\prod_{i=3}^{r} n_i$ of these copies, so a K_r-free subgraph must have at least $n_1 n_2$ missing edges. □

We will also make use of the following classical result saying that every K_{r+1}-free graph with relatively large minimum degree is r-partite.

Theorem 2.4 (Andrásfai-Erdős-Sós [3])**.** *Let $r \geq 2$ and let G be a K_{r+1}-free graph n vertices. If the minimum degree δ of G is strictly greater than $\frac{3r-4}{3r-1}n$, then G is r-partite.*

A *blowup* $H = G[n_1, \dots, n_k]$ of a k-vertex graph G is defined on vertex set $\bigcup_{i\in[k]} W_i$ with $|W_i| = n_i$, where the W_i are disjoint, and $w \in W_i$ and $w' \in W_j$ are adjacent in H if and only if v_i and v_j are adjacent in G. Note that every pentagonal Turán graph is a blowup $L_r[x, y, y, n_1, \dots, n_r]$, where L_r is the graph whose first five vertices induce the pentagon $v_1v_2v_5v_4v_3$, and all other edges are present. Indeed, let us call such a blowup a *complete pentagon-r-partite* (or *CPR*) *graph* if $x \le y \le n_i$ for every $i \in [r]$. A pentagonal Turán graph is then a CPR graph such that the numbers $x + y + n_1, x + y + n_2, n_3, \dots, n_r$ do not differ by more than 1 (i.e., each of them is equal to $\left\lfloor \frac{n+x}{r} \right\rfloor$ or $\left\lceil \frac{n+x}{r} \right\rceil$).

The following statement tells us how to make blowups r-partite. We sketch the proof for completeness.

Theorem 2.5 (Erdős-Győri-Simonovits [12])**.** *Let $H = G[n_1, \dots, n_k]$. Then one can delete $D_r(H)$ edges from H to obtain $G'[n_1, \dots, n_k]$ for some r-partite subgraph G' of G.*

Proof sketch. Take an r-partite subgraph of H obtained by deleting $D_r(H)$ edges from H, and "symmetrize" it, i.e., for $i = 1, \dots, k$, carry out the following: Pick some $v \in W_i$ with $d(v)$ largest. Then for each $w \in W_i \setminus v$, change the edges touching w so that its neighborhood $\Gamma(w)$ becomes the same as $\Gamma(v)$.

Through this process, the graph remains an r-partite subgraph of H, and the number of edges in it does not decrease (thus stays equal to $e(H) - D_r(H)$). At the end, we have $\Gamma(v) = \Gamma(w)$ whenever v and w belong to the same blowup part W_i, so the resulting graph is the blowup of some $G' \subseteq G$ itself. □

Deleting any edge of L_r makes it r-partite, so we get the following.

Corollary 2.6. *If $G = L_r[x, y, y, n_1, \dots, n_r]$ is a CPR graph with $x \le y \le n_i$ for every $i \in [r]$, then $D_r(G) = xy$.*

This means that an optimal r-partition of a CPR graph (minimizing the number of internal edges) can be obtained by putting $Y_1 \cup Z_1$ in the first part, $X \cup Y_2 \cup Z_2$ in the second, and Z_i in the ith part for every $i \ge 3$. Let us call this the *standard r-partition* of such a graph.

As a benchmark, it will be helpful to understand roughly how many internal edges there are in the conjectured extremal graphs, so that we can cut short some edge cases in our analysis.

Lemma 2.7. *For any integers $r \ge 2$, n and $0 \le s \le \frac{n}{r^4}$, there is a CPR graph G with n vertices and at least $t_r(n) - \frac{sn}{r}(1 + 1/r^3)$ edges such that $D_r(G) \ge \frac{\sqrt{s^3 n}}{r^2}$.*

Proof. If $s = 0$, then $G = T_r(n)$ satisfies the conditions, so we may assume that $s \ge 1$.

Let $t = \left\lceil \frac{\sqrt{sn}}{r^2} \right\rceil$. As $\sqrt{s} \le \frac{\sqrt{n}}{r^2}$, we have $s \le \frac{\sqrt{sn}}{r^2} \le t \le \frac{2\sqrt{sn}}{r^2} \le \frac{2n}{r^4}$. We claim that the graph $G = L_r[s, t, t, n_1, \dots, n_r]$ works if each of the numbers $n_1 + t + \lceil s/2 \rceil, n_2 + t + \lfloor s/2 \rfloor, n_3, n_4, \dots, n_r$ is equal to $\lceil n/r \rceil$ or $\lfloor n/r \rfloor$, in a non-increasing order. This graph is well-defined because, using $s \le t \le \frac{2n}{r^4}$ and $2 \le r$,

$$t + \lceil s/2 \rceil \le t + s \le 2t \le \frac{4n}{r^4} \le \frac{n}{2r}.$$

Moreover, since $\lfloor 2x \rfloor \geq 2\lfloor x \rfloor$ for any $x > 0$, this shows that $s \leq t \leq n_i$ for every $i \in [r]$, so by Corollary 2.6, $D_r(G) = st \geq \frac{\sqrt{s^3 n}}{r^2}$.

To count the edges in G, let us split X into two sets X_1 and X_2 of size $\lceil s/2 \rceil$ and $\lfloor s/2 \rfloor$, respectively, and note that $(X_1 \cup Y_1 \cup Z_1, X_2 \cup Y_2 \cup Z_2, Z_3, Z_4, \ldots, Z_r)$ is an r-equipartition of the vertex set with exactly st internal edges. There are $t_r(n)$ potential crossing edges, but $|X_1|(|X_2| + |Z_2|) + |X_2|(|X_1| + |Z_1|) - |X_1||X_2| + |Y_1||Y_2|$ of them are missing.

Here $|Y_1||Y_2| = t^2 = \left\lceil \frac{\sqrt{sn}}{r^2} \right\rceil^2 \leq \frac{sn}{r^4} + 2t - 1$ because $(\lceil x \rceil - 1)^2 \leq x^2$, and therefore $\lceil x \rceil^2 \leq x^2 + 2\lceil x \rceil - 1$ for every $x \geq 1$. Also, $|X_1||X_2| = \lfloor s/2 \rfloor \lceil s/2 \rceil = \lfloor s^2/4 \rfloor$. Finally, $|X_1| + |Z_1|$ and $|X_2| + |Z_2|$ are both at most $\lceil n/r \rceil - t \leq \frac{n}{r} + 1 - t$, so we get $|X_1|(|X_2| + |Z_2|) + |X_2|(|X_1| + |Z_1|) \leq s(\frac{n}{r} + 1 - t)$.

In total, this gives at least

$$st + t_r(n) - s\left(\frac{n}{r} + 1 - t\right) + \left\lfloor \frac{s^2}{4} \right\rfloor - \left(\frac{sn}{r^4} + 2t - 1\right) = t_r(n) - \frac{sn}{r} - \frac{sn}{r^4} + 2st - 2t + \left\lfloor \frac{s^2}{4} \right\rfloor - s + 1$$

edges in G. We can see that this is at least $t_r(n) - \frac{sn}{r}(1 + 1/r^3)$ using the fact that $2st \geq 2t$ and $\lfloor s^2/4 \rfloor + 1 \geq s$ hold for every integer $s \geq 1$. $\square$

3 Very dense graphs

Theorem 1.1 says that every K_{r+1}-free graph G with very close to $t_r(n)$ edges is r-partite. The next lemma shows that G is at most one vertex away from being r-partite, even if we allow slightly fewer edges.

Lemma 3.1. *Let $r \geq 2$, and suppose G is a K_{r+1}-free graph on $n \geq 9r^4$ vertices with at least $t_r(n) - \frac{n}{r}(1 + 1/r^3)$ edges. Then there is a vertex $v \in V(G)$ such that $G - v$ is r-partite.*

Proof. If the minimum degree of G is greater than $\frac{3r-4}{3r-1}n$, then by Theorem 2.4, G itself is r-partite. Otherwise, there is a vertex v of degree at most $\frac{3r-4}{3r-1}n$, and hence $G - v$ has

$$\begin{aligned} e(G - v) &\geq t_r(n) - \frac{n}{r}(1 + 1/r^3) - \frac{3r-4}{3r-1}n \\ &\geq t_r(n-1) + \frac{r-1}{r}n - \frac{r-1}{r} - \frac{n-1}{r} - \frac{1}{r} - \frac{n}{r^4} - \frac{3r-4}{3r-1}n \\ &= t_r(n-1) - \frac{n-1}{r} + \frac{1}{r(3r-1)}n - \frac{n}{r^4} - 1 \\ &\geq t_r(n-1) - \frac{n-1}{r} + \frac{n}{3r^4} - 1 \\ &\geq t_r(n-1) - \frac{n-1}{r} + 2 \end{aligned}$$

edges, where we used $t_r(n) \geq t_r(n-1) + \frac{r-1}{r}(n-1)$ from Lemma 2.1 in the second line, $3r^2 \leq 3r^4/4$ in the fourth, and $n \geq 9r^4$ in the fifth. But then $G - v$ is r-partite by Theorem 1.1. $\square$

This structural lemma allows us to establish our main result when the number of edges is very close to extremal.

Theorem 3.2. *Let* $r \geq 2$ *and* $n \geq 2^8 r^4$*, and suppose* G *is a* K_{r+1}*-free graph with* n *vertices and at least* $t_r(n) - \frac{n}{r}(1 + 1/r^3)$ *edges. Then there is a CPR graph* G^* *such that* $D_r(G^*) \geq D_r(G)$ *and* $e(G^*) \geq e(G)$*.*

Proof. If G is r-partite, then we can just take $G^* = T_r(n)$, so let us assume that G is not r-partite. By Lemma 3.1, there is a vertex v such that $G - v$ is r-partite, say with parts $U_1, \ldots, U_r$ of size $n_1, \ldots, n_r$. Let a_i be the number of neighbors of v in U_i. We may assume that $a_1 \leq \cdots \leq a_r$. Then clearly, $1 \leq D_r(G) \leq a_1$. We claim that $G^* = L_r[1, a_1, a_1, n_1 - a_1, n_2 - a_1, n_3, n_4 \ldots, n_r]$ works.

To show this, note that G has

$$e(G) \leq \sum_{i<j} n_i n_j - a_1 a_2 + \sum_{i \in [r]} a_i \tag{1}$$

edges. This is because there are $\sum_{i<j} n_i n_j$ potential edges in the r-partite graph induced by $U = U_1 \cup \cdots \cup U_r$, but the neighborhood of v is K_r-free, so by Lemma 2.3, at least $a_1 a_2$ of these edges are missing. The number of edges in G not induced by U is precisely $\sum_{i \in [r]} a_i$.

On the other hand,

$$e(G^*) = \sum_{i<j} n_i n_j - a_1^2 + 2a_1 + \sum_{i=3}^{r} n_i \geq \sum_{i<j} n_i n_j - a_1^2 + a_1 - a_2 + \sum_{i \in [r]} a_i.$$

As $a_1 a_2 \geq a_1^2 - a_1 + a_2$ for any positive integers $a_2 \geq a_1$, we get $e(G^*) \geq e(G)$.

To conclude the argument, it is enough to prove that $n_i \geq 2a_1$ for every $i \in [r]$. Indeed, this will establish that G^* is a CPR graph, and, using Corollary 2.6, imply that $D_r(G^*) = a_1$. We can show this through a fairly straightforward calculation.

As the number of edges in an r-partite graph is maximized by the Turán graph, we have $\sum_{i<j} n_i n_j \leq t_r(n)$. Combining this with (1), we get $e(G) \leq t_r(n) - a_1 a_2 + n$. But we assumed that $e(G) > t_r(n) - n$, so $a_1 \leq \sqrt{2n}$.

On the other hand, suppose that $n_{i'} \leq 3\sqrt{n}$ for some $i' \in [r]$. Let $\mathbf{n} = (n_1, \ldots, n_r)$ and $\mathbf{n}' = (n_1, \ldots, n_{i'-1}, n_{i'+1}, \ldots, n_r)$. Once again, the maximality of Turán graphs gives

$$\sum_{i<j} n_i n_j = e\left(K_{\mathbf{n}}\right) \leq e\left(K_{\mathbf{n}'}\right) + n_{i'} n \leq t_{r-1}(n - n_{i'}) + n_{i'} n \leq t_{r-1}(n) + n_{i'} n.$$

We can therefore further bound (1) as

$$e(G) \leq t_{r-1}(n) + 3n^{3/2} + n \leq \frac{r-2}{r-1} \cdot \frac{n^2}{2} + 4n^{3/2} \leq \frac{r-1}{r} \cdot \frac{n^2}{2} - \frac{n^2}{2r^2} + \frac{n^2}{4r^2} \leq t_r(n) - n,$$

using $n \geq 2^8 r^4$ and $\frac{r-1}{r} \cdot \frac{n^2}{2} + n \geq t_r(n) \geq \frac{r-1}{r} \cdot \frac{n^2}{2} - n$ from Lemma 2.1. But this contradicts our assumption on $e(G)$, so indeed, $n_i \geq 3\sqrt{n} \geq 2a_1$ for every $i \in [r]$. $\square$

4 Proof of Theorem 1.2

It will be more convenient for us to prove the following, slightly weaker analog of Theorem 1.2.

Theorem 4.1. *For every $r \geq 2$ there is a $\delta_r > 0$ such that the following holds: If G is a K_{r+1}-free graph on n vertices with $e(G) \geq t_r(n) - \delta_r n^2$ edges, then there is a CPR graph G^* on n vertices with $e(G^*) \geq e(G)$ and $D_r(G^*) \geq D_r(G)$.*

This statement easily implies the full theorem:

Proof of Theorem 1.2. Theorem 4.1 shows the existence of a CPR graph G^*, such that $e(G^*) \geq e(G)$ and $D_r(G^*) \geq D_r(G)$. Let us choose such a G^* so that $e(G^*)$ is maximum. We claim that this G^* is in fact a pentagonal Turán graph.

We know that $G^* = L_r[x, y, y, n_1, \ldots, n_r]$ such that $x \leq y \leq n_i$ for every $i \in [r]$. Note that $e(G^*) \leq t_r(n) - y^2 - xn_1 + xy \leq t_r(n) - y^2$, so if $\delta_r < r^{-10}$, then $y \leq \frac{n}{4r}$. To show that G^* is a pentagonal Turán graph, we just need to check that the numbers $x+y+n_1, x+y+n_2, n_3, \ldots, n_r$ do not differ by more than 1. Suppose that the i-th of these quantities is the largest among them, and the j-th is the smallest. If their difference was at least 2, then the graph $\widetilde{G} = L_r[x, y, y, n_1, \ldots, n_i - 1, \ldots, n_j + 1, \ldots, n_r]$ would have more edges than G^*. Also, $x + y + n_i \geq \frac{n}{r}$ and $x \leq y \leq \frac{n}{4r}$, so $\widetilde{G}$ is a CPR graph with $D_r(\widetilde{G}) = xy = D_r(G^*)$. This contradicts the maximality of G^* and establishes the theorem. $\square$

Our proof of Theorem 4.1 divides into two main parts: defining a CPR graph G^* based on our G, and comparing the number of edges in G and G^*. In the first part of the proof, we find an appropriate r-partition of G, with a large enough matching of internal edges, and use structural considerations to construct a G^* that has at least as many *internal* edges in its standard r-partition as G. Then in the second part, we use the K_{r+1}-freeness of G to prove that it misses many of its *crossing* edges, and ultimately show that G^* has more crossing edges in its standard r-partition.

4.1 The candidate CPR graph G^*

Proof of Theorem 4.1. We will start with defining an r-partition on G.

Let $\delta_r = r^{-60}$, and suppose our K_{r+1}-free graph $G = (V, E)$ has $t_r(n) - \delta n^2$ edges for some $\delta \in (0, r^{-60})$. We may assume that $\delta n^2 \geq 1$, and hence $n \geq \delta^{-1/2} \geq r^{20}$. Now if $\delta n^2 \leq \frac{n}{r}(1 + 1/r^3)$, then we can apply Theorem 3.2, noting that $n \geq r^{20} \geq 2^8 r^4$, to obtain the desired G^*. So we may also assume that $\delta n^2 > \frac{n}{r}(1 + 1/r^3)$, and in particular, $n \geq \frac{1}{\delta r}$.

We first show that G contains a large induced subgraph with high minimum degree.

Proposition 4.2. *There is a vertex subset $S \subseteq V$ with $|S| \leq 2\delta r^{10} n$ such that for all $v \in V \setminus S$,*

$$d_{G-S}(v) \geq n\left(\tfrac{r-1}{r} - r^{-10}\right).$$

Proof. Let us iteratively remove vertices of degree less than $n(\frac{r-1}{r} - r^{-10})$. If this procedure stops with at most $2\delta r^{10} n$ removals, then we are done by choosing S to be the set of removed vertices. So suppose otherwise, and let B be the set of the first $\lceil 2\delta r^{10} n \rceil$ vertices deleted. Then the number of edges in the graph $J = G - B$ can be bounded by

$$e(J) \geq e(G) - n\left(\tfrac{r-1}{r} - r^{-10}\right)|B| = t_r(n) - \delta n^2 - n\left(\tfrac{r-1}{r} - r^{-10}\right)|B|.$$

By Lemma 2.1, we have $t_r(n) \geq t_r(n - |B|) + \frac{r-1}{r}(n - |B|)|B|$, and hence

$$e(J) \geq t_r(|J|) - \delta n^2 + r^{-10} n |B| - \tfrac{r-1}{r}|B|^2.$$

Note that $|B| \geq 2$ (as $2\delta r^{10} n \geq 2r^9 > 1$), so $2\delta r^{10} n \leq |B| \leq 4\delta r^{10} n$. Using $1 > \delta r^{60} > 16\delta r^{20}$, this yields

$$r^{-10} n |B| \geq 2\delta n^2 > \delta n^2 + 16\delta^2 r^{20} n^2 \geq \delta n^2 + |B|^2.$$

But then $e(J) > t_r(|J|)$, contradicting the fact that J is K_{r+1}-free. $\square$

Theorem 2.4 implies that $G - S$ is r-partite. Let $U_1 \cup \cdots \cup U_r$ be an r-partition of $G - S$. By the minimum degree condition of $G - S$, every vertex $x \in U_i$ has at least $n(\frac{r-1}{r} - r^{-10})$ neighbors in $G - S - U_i$, so $|U_i| \leq n(\frac{1}{r} + r^{-10}) - |S|$ for each i. On the other hand, $|U_i| \geq n - |S| - \sum_{j \neq i} |U_j|$, so we get that for every i,

$$|U_i| \geq n\left(\tfrac{1}{r} - (r-1)r^{-10}\right). \tag{2}$$

This also means that the neighborhood of each vertex in U_i misses at most $r^{-9} n$ vertices in $\bigcup_{j \neq i} U_j$ and so the number of crossing edges missing between the U_i is at most $r^{-9} n^2$.

Now let us extend this partition into an r-partition $V = V_1 \cup \cdots \cup V_r$ of the entire vertex set of G that maximizes the number of crossing edges, assuming $U_i \subseteq V_i$. In particular, each vertex of S has at most as many neighbors in its own part as in any other part, i.e., for $s \in S \cap V_i$,

$$|\Gamma(s) \cap V_i| = \min_{j \in [r]} |\Gamma(s) \cap V_j|. \tag{3}$$

Let us define Δ to be the maximum internal degree of G in this partition, i.e.,

$$\Delta = \max_{i \in [r]} \max_{v \in V_i} |\Gamma(v) \cap V_i|$$

Claim 4.3. We may assume that Δ is the internal degree of some vertex $u \in S$, and that

$$6|S| \leq \Delta \leq 2r^{-4.5} n.$$

Proof. Note that all internal edges are incident with S and so $D_r(G) \leq |S|\Delta$. If Δ is smaller than $6|S|$, then $D_r(G) \leq 6|S|^2 \leq 24\delta^2 r^{20} n^2$. We claim that there is a CPR graph G^* with at least $t_r(n) - \delta n^2$ edges such that $D_r(G^*)$ is larger than this. Indeed, apply Lemma 2.7 with $s = \left\lfloor \frac{\delta r n}{1 + 1/r^3} \right\rfloor$ to obtain the graph G^* with at least $t_r(n) - \delta n^2$ edges and $D_r(G^*) \geq \frac{\sqrt{s^3 n}}{r^2}$.

Our previous assumption that $\delta n^2 > \frac{n}{r}(1+1/r^3)$ implies that $s \geq 1$, and therefore $s \geq \frac{\delta r n}{4}$. This means that

$$D_r(G^*) \geq \frac{\delta^{3/2} r^{3/2} n^2}{8r^2} > \frac{\delta^2 r^{29} n^2}{8} > 24\delta^2 r^{20} n^2 \geq D_r(G),$$

as required (we used $1 > \sqrt{\delta} r^{30}$ and $r \geq 2$).

So we may assume that $\Delta \geq 6|S|$. In particular, as the internal degree of each vertex in $V \setminus S$ is at most $|S|$, a vertex of maximum internal degree Δ must lie in S. Let u be any such vertex.

Now we see from (3) that $|\Gamma(u) \cap U_i| \geq \Delta - |S| \geq \frac{5\Delta}{6}$ for every $i \in [r]$. Since $\Gamma(u)$ is K_r-free, Lemma 2.3 tells us that there are at least $\left(\frac{5\Delta}{6}\right)^2 \geq \Delta^2/2$ crossing edges missing between the U_i. On the other hand, we have seen that there are at most $r^{-9}n^2$ such edges missing, so $\Delta \leq 2r^{-4.5}n$. □

Let $u \in S$ be the vertex from Claim 4.3. By (3), it has at least Δ neighbors in each V_i. For each $i \in [r]$, fix a set $P_i \subseteq \Gamma(u) \cap V_i$ with $|P_i| = \Delta$.

We now come to finding a suitable matching consisting of internal edges. Let $H = \bigcup_{i\in[r]} G[V_i]$ be the subgraph of G containing only the internal edges. Then H has at most $\Delta|S|$ edges and maximum degree Δ. Let $k = \left\lceil \frac{e(H)}{\Delta} \right\rceil$ and note that $k \leq |S|$, so $\Delta \geq 6|S| \geq 2k$. Therefore, by Lemma 2.2, we can find a matching M of size k in H.

For each $i \in [r]$, let $M_i = M[V_i]$ be the set of matching edges in V_i. Further split each M_i into three sets $M_i = A_i \cup B_i \cup C_i$ according to the matching pairs' interaction with P_i:

$$\begin{aligned} A_i &= \{uv \in M_i : u, v \notin P_i\}, \\ B_i &= \{uv \in M_i : u \in P_i, v \notin P_i\}, \\ C_i &= \{uv \in M_i : u, v \in P_i\}. \end{aligned}$$

Then define $a_i = |A_i|$, $b_i = |B_i|$ and $c_i = |C_i|$, and set $a = \sum_{i\in[r]} a_i$, $b = \sum_{i\in[r]} b_i$, and $c = \sum_{i\in[r]} c_i$ (so we have $k = a + b + c$). Note that if V_i^A, V_i^B, V_i^C and V_i^M denote the vertex sets of the matchings A_i, B_i, C_i and M_i respectively, then $|V_i^A| = 2a_i$, $|V_i^B| = 2b_i$ and $|V_i^C| = 2c_i$. We denote the unions over $i \in [r]$ by V^A, V^B, V^C and V^M, so $|V^M| = 2k$ (see Figure 2).

Finally, we set $R_i = V_i \setminus (P_i \cup V_i^M)$ and $\kappa_i = |R_i|$. With this notation at hand, we note that $|V_i| = \kappa_i + \Delta + 2a_i + b_i$ for each $i \in [r]$. To bound κ_i from below, recall that $U_i \subseteq V_i$ is an independent set, so at most $k \leq |S|$ of its vertices are covered by M. So by (2), $\delta < r^{-60}$, Proposition 4.2 and Claim 4.3, we have

$$\begin{aligned} \kappa_i \geq |U_i| - |S| - \Delta &\geq n\left(\frac{1}{r} - (r-1)r^{-10}\right) - 2\delta r^{10} n - 2r^{-4.5}n \\ &\geq n\left(r^{-1} - r^{-9} - 2r^{-50} - 2r^{-4.5}\right) \\ &\geq r^{-4.5}n\left(r^{3.5} - 3\right) \geq 8r^{-4.5}n \geq 4\Delta. \end{aligned}$$

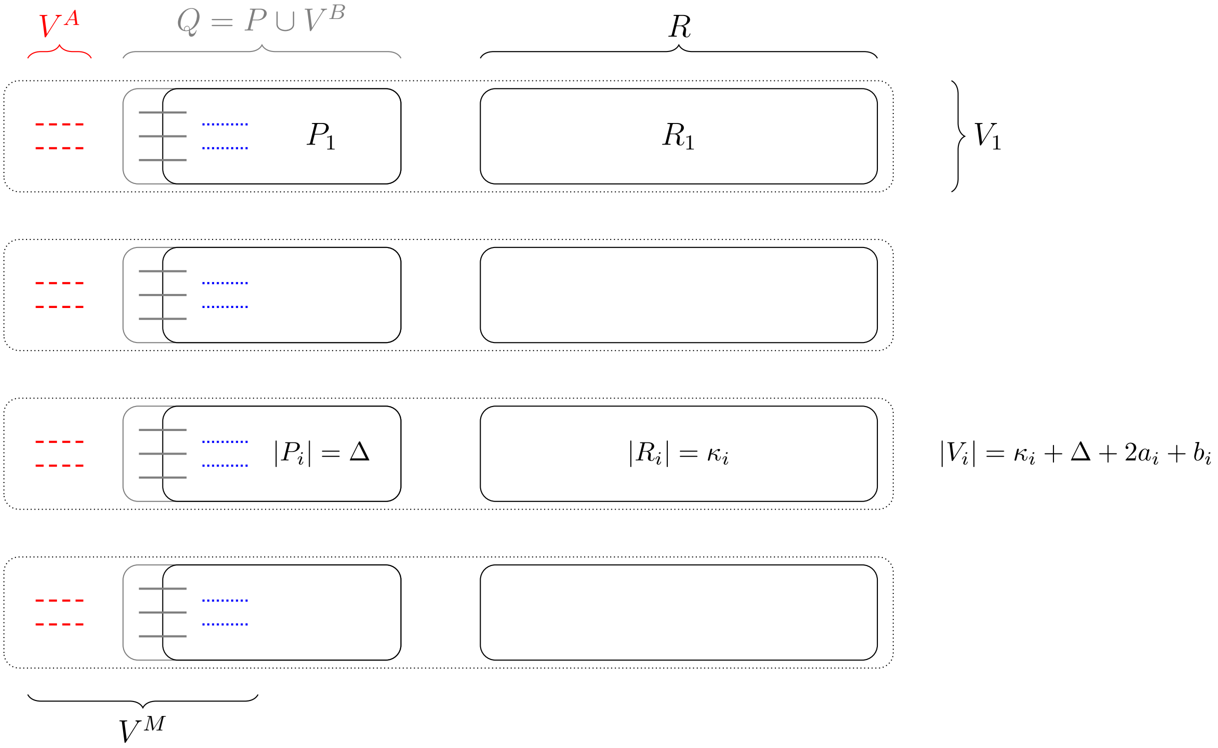

Figure 2: The structure of G

We may assume without loss of generality that $\kappa_1 \le \kappa_2 \le \cdots \le \kappa_r$. Together with Claim 4.3, we get the following relationship between our quantities, which we will use throughout the rest of the proof:

$$\kappa_r \ge \cdots \ge \kappa_2 \ge \kappa_1 \ge 4\Delta \ge 24k. \tag{4}$$

We are now ready to introduce our candidate CPR graph that will satisfy Theorem 4.1. Let $G^* = L_r[k, \Delta, \Delta, n_1, \ldots, n_r]$ be the graph on vertex set $X \cup Y_1 \cup Y_2 \cup Z_1 \cup \cdots \cup Z_r$ as defined in the introduction, where $|X| = k$, $|Y_1| = |Y_2| = \Delta$, $|Z_j| = n_j = \kappa_j + \Delta + 2a_j + b_j$ for $j \ge 3$, and

$$\begin{aligned} |Z_1| &= n_1 = \kappa_1 + a_1 - c_1 \\ |Z_2| &= n_2 = \kappa_2 + a_1 + b_1 + c_1 + 2a_2 + b_2 - k. \end{aligned}$$

Note that $|V_j| = |Z_j|$ for $j \ge 3$, and $|V_1| + |V_2| = |X| + |Y_1| + |Y_2| + |Z_1| + |Z_2|$, so G and G^* have an equal number of vertices.

The proof of Theorem 4.1 therefore reduces to establishing Proposition 4.4 below. □

Proposition 4.4. *G^* satisfies both $e(G^*) \ge e(G)$ and $D_r(G^*) \ge D_r(G)$.*

4.2 Comparing G and G^*

Proof of Proposition 4.4. Corollary 2.6 and (4) give $D_r(G^*) = k\Delta$. Here the definition of k implies $k\Delta \geq e(H)$ and we clearly have $e(H) \geq D_r(G)$, thus $D_r(G^*) \geq D_r(G)$, and G^* has at least as many internal edges in its standard r-partition as G. It is therefore enough to show that G^* also has at least as many crossing edges as G. We start with a lower bound for G^*.

Proposition 4.5. *The number of crossing edges in G^* is at least*

$$\sum_{i<j} |V_i||V_j| - \left(\Delta^2 + b_1 b_2 + (a_1 + b_1 + c_1)\kappa_2 + (k - a_1 - b_1 - c_1)\kappa_1 + (a_1 + a_2)\frac{\Delta}{2}\right).$$

Proof. First of all, as $|V_1 \cup V_2| = |Z_1 \cup Z_2 \cup Y_1 \cup Y_2 \cup X|$, and $|V_i| = |Z_i|$ for every $i \geq 3$, there are exactly $\sum_{i<j} |V_i||V_j| - |V_1||V_2|$ crossing edges in G^* incident to $\bigcup_{i=3}^{r} Z_i$.

As for the edges induced by $Z = Z_1 \cup Z_2 \cup Y_1 \cup Y_2 \cup X$, there are

$$\begin{aligned}(|Y_1| + |Z_1|)(|X| + |Y_2| + |Z_2|) &= \big(|V_1| - (a_1 + b_1 + c_1)\big)\big(|V_2| + (a_1 + b_1 + c_1)\big) \\ &= |V_1||V_2| - (|V_2| - |V_1| + a_1 + b_1 + c_1)(a_1 + b_1 + c_1)\end{aligned}$$

potential crossing edges in the standard r-partition of G^* (see Figure 1), out of which

$$|Y_1||Y_2| + |X||Z_1| = \Delta^2 + k(\kappa_1 + a_1 - c_1)$$

are missing. Here $|V_2| - |V_1| + a_1 + b_1 + c_1 = \kappa_2 - \kappa_1 + 2a_2 + b_2 - a_1 + c_1$, so by rearranging, we get that the number of crossing edges in G^* is

$$\sum_{i<j} |V_i||V_j| - \big(\Delta^2 + b_1 b_2 + (a_1 + b_1 + c_1)\kappa_2 + (k - a_1 - b_1 - c_1)\kappa_1 + \Lambda\big),$$

where

$$\Lambda = a_1(k + 2a_2 + b_2 - a_1 - b_1) + a_2(2b_1 + 2c_1) - c_1(k - b_1 - b_2 - c_1) \leq (a_1 + a_2) \cdot 3k,$$

where we used that $k = a + b + c \geq a_2 + b_1 + b_2 + c_1$. The result then follows from $\Delta \geq 6k$. $\square$

Recall that there are exactly $\sum_{i<j} |V_i||V_j|$ potential crossing edges in G. It therefore suffices to show that at least

$$\Delta^2 + b_1 b_2 + (a_1 + b_1 + c_1)\kappa_2 + (k - a_1 - b_1 - c_1)\kappa_1 + (a_1 + a_2)\frac{\Delta}{2} \tag{5}$$

of them are missing from G.

It will be easier to split the graph into two, and bound the number of missing edges separately. Let $Q_i = P_i \cup V_i^B$ be the set obtained by extending P_i with the vertices of the matching B_i for every $i \in [r]$ (see Figure 2), so that V_i^A, Q_i and R_i partition V_i, and let $Q = \bigcup_{i \in [r]} Q_i$. We first count the number of crossing edges with both endpoints in Q, and then the ones with at most one end in Q.

Lemma 4.6. *G misses at least $\Delta^2 + b_1 b_2$ of the crossing edges induced by Q.*

Proof. We use a similar argument to the proof of Lemma 2.3. Let $\mathcal{F}$ be the family of all r-sets $\{v_1, \dots, v_r\}$ such that $v_i \in P_i$ for every $i = 1, \dots, r$, but $v_1 \notin V_1^B$ or $v_2 \notin V_2^B$. Then $|\mathcal{F}| = \Delta^r - b_1 b_2 \Delta^{r-2}$. Similarly, let $\mathcal{G}$ be the family of all $(r+2)$ sets $\{v_1, \dots, v_r, v_1', v_2'\}$ such that $v_1 v_1' \in B_1$, $v_2 v_2' \in B_2$, and $v_i \in P_i$ for every $i = 3, \dots, r$. Then $|\mathcal{G}| = b_1 b_2 \Delta^{r-2}$.

Recall that $P_1, \dots, P_r$ were all in the neighborhood of some vertex u. This means that there must be a (crossing) edge missing in $G[X]$ for every $X \in \mathcal{F}$. Also, for $Y \in \mathcal{G}$, $G[Y]$ is a K_{r+1}-free graph on $r+2$ vertices and so must be missing at least two edges. As $v_1 v_1'$ and $v_2 v_2'$ are both present in G, the missing edges in $G[Y]$ are also crossing.

Summing over the sets in $\mathcal{F} \cup \mathcal{G}$ gives at least $\Delta^r + b_1 b_2 \Delta^{r-2}$ missing crossing edges in total. It is easy to check that each missing edge $v_i v_j$ (or $v_i' v_j$ or $v_i' v_j'$) in G is contained in exactly Δ^{r-2} sets from $\mathcal{F} \cup \mathcal{G}$, so $G[Q]$ misses at least $\Delta^2 + b_1 b_2$ crossing edges. □

Lemma 4.7. *G misses at least*

$$(a_1 + b_1 + c_1)\kappa_2 + (k - a_1 - b_1 - c_1)\kappa_1 + (a_1 + a_2)\Delta/2 \tag{6}$$

crossing edges with at most one endvertex in Q.

Proof. As a first attempt, we try to find a set of missing crossing edges for each matching edge in M so that they are all disjoint and not induced by Q. More specifically, we want to show that for every edge $e \in M_1$, there are κ_2 missing edges between e and $R = \bigcup_{i \in [r]} R_i$, and for every remaining edge $e \in M \setminus M_1$, there are κ_1 missing edges between e and R. Moreover, for every $e \in A_1 \cup A_2$, we want $\Delta/2$ additional missing edges between e and Q. As $|M| = k$ and $|M_1| = a_1 + b_1 + c_1$, this would be exactly the amount we need.[1]

Of course, it may well be that some edge in M is incident to fewer missing edges. Let $M_1' = M_1$ and $M_2' = M \setminus M_1$. To first bound the number of crossing edges between M and R, we define τ to be the largest "deficit" in the above counting, i.e., the smallest *nonnegative* integer such that for each $i = 1, 2$ and every edge $vv' \in M_i'$, there are at least $\kappa_{3-i} - \tau$ missing edges between $\{v, v'\}$ and $R \setminus R_i$.

To count the missing edges between $A_1 \cup A_2$ and Q, we split A_i into $A_i^g \cup A_i^b$ for each $i = 1, 2$ as follow. A_i^g is the set of "good" edges vv', such that there are at least $\kappa_{3-i} - \tau + \Delta/2$ edges missing between $\{v, v'\}$ and $(Q \cup R) \setminus (Q_i \cup R_i)$, and A_i^b is the set of "bad" edges, where this is not the case.

So far this gives at least

$$|A_1^g|(\kappa_2 - \tau) + |A_2^g|(\kappa_1 - \tau) + (|A_1^g| + |A_2^g|)\Delta/2$$

missing crossing edges between the good edges of A and $Q \cup R$, and another

$$(|M_1'| - |A_1^g|)(\kappa_2 - \tau) + (|M_2'| - A_2^g)(\kappa_1 - \tau)$$

---

[1] The reader might find it helpful to check what the bound means when G is a CPR graph: the r-partition $V_1 \cup \dots \cup V_r$ is much like the standard r-partition, except the set X might be split between V_1 and V_2. In any case, we always have $M = B_1 \cup B_2$ (in particular, $a_1 = a_2 = 0$), and every edge in B_i contributes exactly κ_{3-i} missing edges: one to each vertex of R_{3-i}.

between all other edges of M and R. This is a total of

$$|M_1'|(\kappa_2 - \tau) + |M_2'|(\kappa_1 - \tau) + (|A_1^g| + |A_2^g|)\Delta/2 \tag{7}$$

missing edges between V^M and R. To get (6), we need to analyze the structure a bit.

Let $vv' \in A_i^b$ be some fixed bad edge for some i. Then there are at most $\kappa_{3-i} - \tau + \Delta/2$ missing edges from $\{v, v'\}$ to $(R \cup Q) \setminus (R_i \cup Q_i)$, and by the definition of τ, at least $\kappa_{3-i} - \tau$ of these are incident with $R \setminus R_i$. So vv' must have at least $\Delta/2$ common neighbors in each P_j with $j \neq i$. In particular, as $k \leq \Delta/6$ and hence $|V^M| = 2k \leq \Delta/3$, we get that for every $j \neq i$ there is a set $N_j \subseteq P_j \setminus (V_j^B \cup V_j^C)$ of at least $\Delta/6$ common neighbors in P_j that is disjoint from V^M.

Choose $i' \neq i$ so that $\Gamma(v) \cap \Gamma(v') \cap R_{i'}$ is smallest. Then for every $j \neq i, i'$,

$$\begin{aligned}|\Gamma(v) \cap \Gamma(v') \cap R_j| &\geq \frac{|\Gamma(v) \cap \Gamma(v') \cap (R_{i'} \cup R_j)|}{2} \\ &\geq \frac{(\kappa_{i'} + \kappa_j) - (\kappa_{3-i} + \Delta/2)}{2} \geq \frac{\kappa_2}{2} - \frac{\Delta}{4} \geq \frac{7\kappa_2}{16}\end{aligned} \tag{8}$$

because there are at most $\kappa_{3-i} + \Delta/2$ missing edges from $\{v, v'\}$ to $R_{i'} \cup R_j$, and we also used $\kappa_{i'} + \kappa_j \geq \kappa_2 + \kappa_{3-i}$ and $\Delta \leq \kappa_2/4$, which follow from (4) for any distinct $i \in \{1, 2\}$, i' and j.

Observation 4.8. We may assume that every triangle induced by $V_i \cup V_i'$ has at most $\kappa_2/4$ common neighbors in some R_j with $j \neq i, i'$.

Indeed, the common neighborhood of this triangle is κ_{r-2}-free. The case $r \leq 3$ is then vacuously true, so suppose $r \geq 4$. Then if the triangle has at least $\kappa_2/4$ common neighbors in every R_j with $j \neq i, i'$, then by Lemma 2.3, $G[R]$ misses at least $\kappa_2^2/16$ crossing edges. But $\kappa_2^2/16 \geq k(\kappa_2 + \Delta) \geq (6)$, so we are done.

This means that for the above bad edge vv', we can assume that every triangle $vv'w$ with $w \in N_{i'}$ has at most $\kappa_2/4$ common neighbors in some R_j with $j \neq i, i'$. Using (8), we see that there are at least $\frac{7\kappa_2}{16} - \frac{\kappa_2}{4} = \frac{3\kappa_2}{16} > 4k$ missing edges between w and $R \setminus (R_i \cup R_j)$. Summing over all $w \in N_{i'}$, we find at least

$$4k\Delta/6 \geq k\Delta/2 \geq (|A_1^b| + |A_2^b|)\Delta/2 \tag{9}$$

missing edges between $Q \setminus V^M$ and R.

If $\tau = 0$, then we are already done: (7) and (9) together give enough edges for (6). So let us assume that $\tau > 0$, i.e., there is an edge $vv' \in M_i'$ for some i such that there are exactly $\kappa_{3-i} - \tau$ missing edges between $\{v, v'\}$ and $R \setminus R_i$.

Once again, choose $i' \neq i$ so that $\Gamma(v) \cap \Gamma(v') \cap R_{i'}$ is smallest. Then, similarly to (8),

$$|\Gamma(v) \cap \Gamma(v') \cap R_{i'}| \geq \kappa_{i'} - (\kappa_{3-i} - \tau) \geq \tau$$

and for every $j \neq i, i'$,

$$|\Gamma(v) \cap \Gamma(v') \cap R_j| \geq \frac{\kappa_{i'} + \kappa_j - (\kappa_{3-i} - \tau)}{2} \geq \frac{\kappa_2}{2}.$$

By Lemma 2.3, there must be at least

$$\frac{\kappa_2}{2} \cdot \tau \geq k\tau \tag{10}$$

missing edges induced by R. Adding (7), (9) and (10) together, we get (6). $\square$

Putting Proposition 4.5 and Lemmas 4.6 and 4.7 together yields Proposition 4.4, and finishes the proof of our main result. $\square$

5 Concluding remarks

With Theorem 1.2 in hand, finding the exact pentagonal Turán graph G that maximizes $D_r(G)$ assuming $e(G) \geq t_r(n) - \delta n^2$ is a matter of calculation. The result of Balogh, Clemen, Lavrov, Lidický and Pfender [5] shows that among pentagonal Turán graphs with $t_r(n) - \delta n^2$ edges, $D_r(G)$ is maximized when $x \approx \frac{2r}{3}\delta n$, $y \approx \sqrt{\frac{\delta}{3}}n$, $n_j \approx (\frac{1}{r} + \frac{2}{3}\delta)n$ for $j \geq 3$, and $n_i \approx (\frac{1}{r} - \frac{2(r-1)}{3}\delta - \sqrt{\frac{\delta}{3}})n$ for $i = 1, 2$, and the maximum is $D_r(G) \approx \frac{2r}{3\sqrt{3}}\delta^{3/2}n^2$.

It would be very interesting to find exact stability results for other classes of graphs. Of course, this is generally a harder problem than determining the exact extremal graphs, which is often already a difficult task on its own. A natural next step is to consider H-free graphs where H is a graph with a critical edge, that is, there is an edge $e \in E(H)$ such that the deletion of e from H reduces the chromatic number. Examples of such graphs include cliques and odd cycles.

An old theorem of Simonovits [21] says that when H is an $(r+1)$-chromatic graph with a critical edge, the Turán graph $T_r(n)$ is the unique H-free graph maximizing the number of edges, provided n is large enough. But even in this case, it seems unclear what the right conjecture should be for the set of H-free graphs G that maximize $D_r(G)$ when $e(G) \geq t_r(n) - t$. We think that the theorem of Erdős, Győri and Simonovits should at least generalize to odd cycles in the following sense: Among C_{2k-1}-free graphs of close to extremal size, some C_{2k+1}-blowup is farthest from being bipartite.

Unfortunately, this might fail when the number of edges is very close to the extremal number. For example, let G be the graph obtained from $C_6[1, 1, 1, 1, n/2 - 2, n/2 - 3]$ by adding a vertex adjacent to the first three (singly blown up) vertices. Then G is a C_5-free graph satisfying $D_2(G) = 1$, but with strictly more edges than any blowup of C_7 (itself being a supergraph of the densest C_7-blowup). Nevertheless, we believe that the existence of such examples is an artifact of the small blowup factors, and C_{2k+1}-blowups are still optimal when the density of G is bounded away from $1/4$.

Conjecture 5.1. *Fix $k \geq 2$ and let δ be small enough. Then for any $\delta > \delta_0 > 0$ and large enough n, the following holds. For every C_{2k-1}-free graph G on n vertices with $(\frac{1}{4} - \delta_0)n^2 \geq e(G) \geq (\frac{1}{4} - \delta)n^2$ edges, there is a C_{2k+1}-blowup G^* satisfying $e(G^*) \geq e(G)$ and $D_2(G^*) \geq D_2(G)$.*

Blowups of C_{2k+1} might also be optimal for every 3-chromatic graph H with a critical edge, whose shortest odd cycle has length $2k-1$. Such graphs are certainly H-free, and results of Roberts and Scott [18] imply that the bound they give on $D_2(G)$ (with $e(G)$ fixed) is tight up to a constant factor.

It is also tempting to guess that when H is a general $(r+1)$-chromatic graph with a critical edge, then the optimum $D_r(G)$ is attained by complete C_{2k+1}-Turán graphs (defined analogously to pentagonal Turán graphs by inserting a blowup of C_{2k+1} into a part of a complete $(r-1)$-partite graph), where k is some parameter depending only on H.

A closely related problem, which served as the main motivation for the paper of Erdős, Győri and Simonovits [12], is the old conjecture of Erdős [10] claiming $D_2(G) \leq \frac{n^2}{25}$ for every K_3-free graph G on n vertices. This trivially holds when $e(G) \leq \frac{2n^2}{25}$, and was proved for $e(G) \geq \frac{n^2}{5}$ by Erdős, Faudree, Pach and Spencer [11]. If true, the conjecture is tight for a balanced blowup of C_5.

This problem led to further research into how far K_{r+1}-free graphs can be from being bipartite. Sudakov [23] proved a variant of the conjecture for 4-cliques, showing that $D_2(G)$ is maximized by $G = T_3(n)$ among K_4-free graphs. Sudakov conjectured that this generalizes to larger cliques (i.e., among K_{r+1}-free graphs, $D_2(G)$ is maximum when $G = T_r(n)$). A proof of this for K_6 has been announced by Hu, Lidický, Martins, Norin and Volec [15]. The remaining cases remain wide open.

References

[1] N. Alon, J. Balogh, P. Keevash and B. Sudakov, The number of edge colorings with no monochromatic cliques, *J. Lond. Math. Soc.* 70 (2004), 273–288.

[2] K. Amin, J. Faudree, R. J. Gould and E. Sidorowicz, On the non-$(p-1)$-partite K_p-free graphs, *Discuss. Math. Graph Theory* 33 (2013), 9–23.

[3] B. Andrásfai, P. Erdős and V.T. Sós, On the connection between chromatic number, maximal clique and minimal degree of a graph, *Discrete Math.* 8 (1974), 205–218.

[4] J. Balogh, B. Bollobás and M. Simonovits, The number of graphs without forbidden subgraphs, *J. Combin. Theory Ser. B* 91 (2004), 1–24.

[5] J. Balogh, F.C. Clemen, M. Lavrov, B. Lidický and F. Pfender, Making K_{r+1}-free graphs r-partite, *Combin. Probab. Comput.* 30 (2021), 609–618.

[6] J. Balogh, R. Morris, W. Samotij and L. Warnke, The typical structure of sparse K_{r+1}-free graphs, *Trans. Amer. Math. Soc.* 368 (2016), 6439–6485.

[7] A.E. Brouwer, Some lotto numbers from an extension of Turán's theorem, *Math. Centr. report* ZW152, Amsterdam (1981), 6pp.

[8] V. Chvátal and D. Hanson, Degrees and matchings, *J. Combin. Theory Ser. B* 20 (1976), 128–138.

[9] P. Erdős, Some recent results on extremal problems in graph theory (Results), *Theory of Graphs (Internl. Symp. Rome)* (1966), 118–123.

[10] P. Erdős, Problems and results in graph theory and combinatorial analysis, in: *Proc. Fifth British Combinatorial Conference (Univ. Aberdeen, Aberdeen, 1975),* Congress. Numer. XV Utilitas Math., Winnipeg, Man. (1976) 169–192.

[11] P. Erdős, R. Faudree, J. Pach and J. Spencer, How to make a graph bipartite, *J. Combin. Theory Ser. B* 45 (1988), 86–98.

[12] P. Erdős, E. Győri and M. Simonovits, How many edges should be deleted to make a triangle-free graph bipartite?, in: *Sets, graphs and numbers (Budapest, 1991),* Colloq. Math. Soc. János Bolyai, North-Holland, Amsterdam, 60 (1992) 239–263.

[13] Z. Füredi, A proof of the stability of extremal graphs, Simonovits' stability from Szemerédi's regularity, *J. Combin. Theory Ser. B* 115 (2015), 66–71.

[14] D. Hanson and B. Toft, k-saturated graphs of chromatic number at least k, *Ars Combin.* 31 (1991), 159–164.

[15] P. Hu, B. Lidický, T. Martins, S. Norin and J. Volec, Large multipartite subgraphs in H-free graphs, manuscript.

[16] M. Kang and O. Pikhurko, Maximum K_{r+1}-free graphs which are not r-partite, *Mat. Stud.* 24 (2005), 12–20.

[17] W. Mantel, Problem 28, *Wiskundige Opgaven* 10 (1907), 60–61.

[18] A. Roberts and A. Scott, Stability results for graphs with a critical edge, *European J. Combin.* 94 (2018), 27–38.

[19] W. Samotij, Stability results for random discrete structures, *Random Structures Algorithms* 44 (2014), 269–289.

[20] M. Simonovits, Extrém gráfok struktúrájáról (On the structure of extremal graphs, in Hungarian), *CSc Thesis,* Eötvös Loránd University, Budapest (1969), 112pp.

[21] M. Simonovits, Extremal graph problems with symmetrical extremal graphs. Additional chromatic conditions, *Discrete Math.* 7 (1974), 349–376.

[22] M. Simonovits, A method for solving extremal problems in graph theory, stability problems, in: *Theory of Graphs (Proc. Colloq. Tihany, 1966),* Academic Press, New York (1968), 279–319.

[23] B. Sudakov, Making a K_4-free graph bipartite, *Combinatorica* 27 (2007), 509–518.

[24] K. J. Swanepoel, Unit distances and diameters in Euclidean spaces, *Discrete Comput. Geom.* 41 (2009), 1–27.

[25] P. Turán, Eine Extremalaufgabe aus der Graphentheorie, *Mat. Fiz. Lapok* 48 (1941), 436–452.

[26] M. Tyomkyn and A.J. Uzzel, Strong Turán stability, *Electron. J. Combin.* 22 (2015), P3.9, 24pp.