ON DEFICIENCY PROBLEMS FOR GRAPHS

ANDREA FRESCHI, JOSEPH HYDE AND ANDREW TREGLOWN

Abstract. Motivated by analogous questions in the setting of Steiner triple systems and Latin squares, Nenadov, Sudakov and Wagner [Completion and deficiency problems, Journal of Combinatorial Theory Series B, 2020] recently introduced the notion of *graph deficiency*. Given a global spanning property $\mathcal{P}$ and a graph G, the deficiency $\mathrm{def}(G)$ of the graph G with respect to the property $\mathcal{P}$ is the smallest non-negative integer t such that the join $G * K_t$ has property $\mathcal{P}$. In particular, Nenadov, Sudakov and Wagner raised the question of determining how many edges an n-vertex graph G needs to ensure $G * K_t$ contains a K_r-factor (for any fixed $r \geq 3$). In this paper we resolve their problem fully. We also give an analogous result which forces $G * K_t$ to contain any fixed bipartite $(n+t)$-vertex graph of bounded degree and small bandwidth.

1. Introduction

A natural question dating back to the 1970s asks for the order of the smallest complete Steiner triple system a fixed partial Steiner triple system can be embedded into (see e.g. [3, 12, 13]). Similarly, there has been interest in establishing the order of the smallest Latin square that a fixed partial Latin square can be embedded into (see e.g. [5, 6, 13]).

Motivated by these research directions, Nenadov, Sudakov and Wagner [13] introduced the notion of *graph deficiency*: for a graph G and integer $t \geq 0$, denote by $G * K_t$ the *join* of G and K_t, which is the graph obtained from G by adding t new vertices and adding all edges incident to at least one of the new vertices. Given a global spanning property $\mathcal{P}$ and a graph G, the *deficiency* $\mathrm{def}(G)$ of the graph G with respect to the property $\mathcal{P}$ is the smallest $t \geq 0$ such that the join $G * K_t$ has property $\mathcal{P}$.

Note that the following special type of deficiency problem has been previously studied: given a graph H and $n \in \mathbb{N}$, what is the minimum number of vertices needed to ensure any H-packing on n vertices (i.e., a collection of edge-disjoint copies of H that together form a graph on n vertices) can be extended to an H-design (i.e., an H-packing of a complete graph)? See e.g. [7, 8] for background and results on this problem.

One of the main results in [13] is a bound on $\mathrm{def}(G)$ with respect to the Hamiltonicity property for graphs G of a given density. More precisely, the following result answers the question of how many edges an n-vertex graph G can have such that $G * K_t$ does not contain a Hamilton cycle.

Theorem 1.1 (Nenadov, Sudakov and Wagner [13]). *Let n and t be integers and G an n-vertex graph so that $G * K_t$ does not contain a Hamilton cycle. Then we have the following bounds on $e(G)$.*

- *If $n+t$ is even:*

$$e(G) \leq \binom{n}{2} - \begin{cases} \left(t(n-1) - \binom{t}{2}\right) & \text{if } t \leq (n+4)/5 \\ \left(\binom{\frac{n+t+2}{2}}{2} - 1\right) & \text{if } t \geq (n+4)/5. \end{cases}$$

- *If $n+t$ is odd:*

$$e(G) \leq \binom{n}{2} - \begin{cases} \left(t(n-1) - \binom{t}{2}\right) & \text{if } t \leq (n+1)/5 \\ \binom{\frac{n+t+1}{2}}{2} & \text{if } t \geq (n+1)/5. \end{cases}$$

These bounds on $e(G)$ are sharp.

Another line of inquiry in [13] concerns the deficiency problem for K_r-factors. Given graphs H and G, an *H-factor* in G is a collection of vertex-disjoint copies of H in G that together cover all the vertices of G. Note that H-factors are also often referred to as *perfect H-tilings*, *perfect H-packings* or *perfect H-matchings*. The following seminal result of Hajnal and Szemerédi [9] determines the minimum degree threshold for forcing a K_r-factor in a graph G.

Theorem 1.2 (Hajnal and Szemerédi [9])**.** *Every graph G on n vertices with $r|n$ and whose minimum degree satisfies $\delta(G) \geq (1-1/r)n$ contains a K_r-factor. Moreover, there are n-vertex graphs G with $\delta(G) = (1-1/r)n - 1$ that do not contain a K_r-factor.*

More recently, Kühn and Osthus [11] determined, up to an additive constant, the minimum degree threshold for forcing an H-factor, for any fixed graph H.

The following result of Nenadov, Sudakov and Wagner [13] determines how many edges an n-vertex graph G needs to guarantee that $G * K_t$ contains a K_3-factor (provided that t is not too big compared to n).

Theorem 1.3 (Nenadov, Sudakov and Wagner [13])**.** *There exists $n_0 \in \mathbb{N}$ such that the following holds. Let $n, t \in \mathbb{N}$ so that $n \geq n_0$ and $3|(n+t)$, and let G be an n-vertex graph such that $G * K_t$ does not contain a K_3-factor. If $t \leq n/1000$ then*

$$e(G) \leq \binom{n}{2} - \binom{k}{2} - \begin{cases} k(n-k) & \text{if } t \text{ is odd} \\ k(n-k-1) & \text{if } t \text{ is even,} \end{cases}$$

where $k := \lceil (t+1)/2 \rceil$. This bound on $e(G)$ is sharp.

1.1. **A deficiency result for K_r-factors.** Nenadov, Sudakov and Wagner [13] state that *'the study of deficiency concept by itself leads to intriguing open problems'*. In particular, the first open problem [13, Section 7] they raise is to extend Theorem 1.3 to the full range of t, and moreover to resolve the analogous question for K_r-factors in general. In this paper we fully resolve this problem via the following theorem.

Theorem 1.4. *Let $n, t, r \in \mathbb{Z}$ with $n \geq 2$, $t \geq 0$ and $r \geq 3$ such that $t < (r-1)n$ and $r|(n+t)$. Further, let $k := \lceil \frac{t+1}{r-1} \rceil$ and q be the integer remainder when t is divided by $r-1$. Let G be a graph on n vertices such that $G * K_t$ does not contain a K_r-factor. Then*

$$e(G) \leq \max\left\{ \binom{n}{2} - \binom{\frac{n+t}{r}+1}{2}, \binom{n}{2} - \binom{k}{2} - k(n-k-(r-2-q)) \right\}.$$

When $(r-1)|(t+1)$, the first term is at most the second term precisely when $t \leq \frac{(r-1)n-r^2}{2r^2-2r+1}$. Note that Theorem 1.4 considers all interesting values of n and t. Indeed, if $t \geq (r-1)n$ and $r|(n+t)$, then $G * K_t$ trivially contains a K_r-factor, even if $e(G) = 0$. Further, in Section 3 we provide extremal examples that demonstrate that the edge condition in Theorem 1.4 cannot be lowered. Perhaps surprisingly, the proof of Theorem 1.4 is short, making use of a couple of vertex-modification tricks (see the proofs of Lemmas 4.1 and 4.2) and Theorem 1.2.

Note that the $t = 0$ case of Theorem 1.4 determines the edge density threshold for forcing a K_r-factor in a graph. In fact, this is an old result due to Akiyama and Frankl [1], so our result can be viewed as a deficiency generalisation of their theorem.

1.2. **A deficiency bandwidth theorem.** One of the central results in extremal graph theory is the so-called *Bandwidth theorem* due to Böttcher, Schacht and Taraz [2]. A graph H on n vertices is said to have *bandwidth at most* b, if there exists a labelling of the vertices of H with the numbers $1, \dots, n$ such that for every edge $ij \in E(H)$ we have $|i-j| \le b$.

Theorem 1.5 (The Bandwidth theorem, Böttcher, Schacht and Taraz [2])**.** *Given any $r, \Delta \in \mathbb{N}$ and any $\gamma > 0$, there exist constants $\beta > 0$ and $n_0 \in \mathbb{N}$ such that the following holds. Suppose that H is an r-chromatic graph on $n \ge n_0$ vertices with $\Delta(H) \le \Delta$ and bandwidth at most βn. If G is a graph on n vertices with*

$$\delta(G) \ge \left(1 - \frac{1}{r} + \gamma\right) n,$$

then G contains a copy of H.

Note that a K_r-factor has bandwidth $r-1$; thus, one can view the bandwidth theorem as a vast asymptotic generalisation of Theorem 1.2.

Following the proof of Theorem 1.1 from [13], and applying a theorem of Knox and the third author [10], one can easily obtain a deficiency result for embedding bipartite graphs of bounded degree and small bandwidth.

Theorem 1.6. *Given any $\Delta \in \mathbb{N}$ and $\varepsilon > 0$, there exist constants $\beta > 0$ and $n_0 \in \mathbb{N}$ such that the following holds. Let $t \in \mathbb{N}$ and $n \ge n_0$. Let H be a bipartite graph on $n+t$ vertices with $\Delta(H) \le \Delta$ and bandwidth at most $\beta(n+t)$. Suppose that G is a graph on n vertices such that $G * K_t$ does not contain a copy of H. Then we have the following bound on $e(G)$.*

$$e(G) \le \binom{n}{2} - \begin{cases} \left(t(n-1) - \binom{t}{2} - \varepsilon n^2\right) & \text{if } t \le \frac{n}{5} \\ \left(\binom{\lceil \frac{n+t}{2} \rceil + 1}{2} - \varepsilon n^2\right) & \text{if } t > \frac{n}{5}. \end{cases}$$

Observe that the bounds on $e(G)$ in Theorem 1.6 are, up to error terms, exactly the same as those in Theorem 1.1. Moreover, the extremal examples that show the condition on $e(G)$ in Theorem 1.1 is sharp also demonstrate that, for many graphs H, the condition on $e(G)$ in Theorem 1.6 is asymptotically best possible (see Section 3). Notice the statement of Theorem 1.6 is only interesting for $t < n-1$. Indeed, if $t \ge n-1$ then even if G has no edges, $G * K_t$ contains all $(n+t)$-vertex bipartite graphs H (we just embed the smallest colour class into K_t).

The paper is organised as follows. We introduce some graph theoretic notation in Section 2. In Section 3 we provide extremal examples for Theorems 1.4 and 1.6. In Section 4 we prove Theorem 1.4 and then in Section 5 we prove Theorem 1.6. Some concluding remarks are given in Section 6.

2. Notation

Let G be a graph. We define $V(G)$ to be the vertex set of G and $E(G)$ to be the edge set of G. Let $X \subseteq V(G)$. Then $G[X]$ is the *graph induced by X on G* and has vertex set X and edge set $E(G[X]) := \{xy \in E(G) : x, y \in X\}$. For each $x \in V(G)$, we define the *neighbourhood of x in G* to be $N_G(x) := \{y \in V(G) : xy \in E(G)\}$ and define $d_G(x) := |N_G(x)|$.

We write $0 < a \ll b \ll c < 1$ to mean that we can choose the constants a, b, c from right to left. More precisely, there exist non-decreasing functions $f : (0,1] \to (0,1]$ and $g : (0,1] \to (0,1]$ such that for all $a \le f(b)$ and $b \le g(c)$ our calculations and arguments in our proofs are correct. Larger hierarchies are defined similarly.

3. The extremal constructions for Theorem 1.4 and Theorem 1.6

In this section we will give the extremal constructions that match the upper bounds in Theorems 1.4 and 1.6. Firstly, let us consider those for Theorem 1.4.

Definition 3.1. *Let $n, r, t \in \mathbb{Z}$ with $r \geq 3$, $t \geq 0$ and $n \geq 2$ such that $t < (r-1)n$ and $r|(n+t)$. Further, let $k := \lceil \frac{t+1}{r-1} \rceil$ and q be the integer remainder when t is divided by $r-1$. We define graphs $EX_1(n,t,r)$ and $EX_2(n,t,r)$ as follows:*

- *Let $K := K_n$ and $A \subseteq K$ such that $A = K_{\frac{n+t}{r}+1}$. Define $EX_1(n,t,r)$ to be the graph obtained by removing $E(A)$ from K.*
- *Consider a set of isolated vertices B where $|B| = k$ and K_{n-k}, and let $C \subseteq V(K_{n-k})$ where $|C| = r-2-q$. Define $EX_2(n,t,r)$ to be the graph obtained by taking the disjoint union of B and K_{n-k} and adding every edge incident to a vertex in C.*

Observe that

$$e(EX_1(n,t,r)) = \binom{n}{2} - \binom{\frac{n+t}{r}+1}{2}$$

and

$$e(EX_2(n,t,r)) = \binom{n}{2} - \binom{k}{2} - k(n-k-(r-2-q)).$$

Hence

$$\max\left\{\binom{n}{2} - \binom{\frac{n+t}{r}+1}{2}, \binom{n}{2} - \binom{k}{2} - k(n-k-(r-2-q))\right\} = \max\left\{e(EX_1(n,t,r)), e(EX_2(n,t,r))\right\}.$$

Next we show that $EX_1(n,t,r) * K_t$ and $EX_2(n,t,r) * K_t$ do not contain K_r-factors, that is, they are extremal graphs for Theorem 1.4.

Proposition 3.2. *$EX_1(n,t,r) * K_t$ and $EX_2(n,t,r) * K_t$ do not contain K_r-factors.*

Proof. Firstly, let us consider $EX_1(n,t,r) * K_t$. For a contradiction, assume that $EX_1(n,t,r) * K_t$ contains a K_r-factor $\mathcal{T}$. Then each vertex in $V(A)$ belongs to a different copy of K_r in $\mathcal{T}$. This implies $\frac{n+t}{r} = |\mathcal{T}| \geq |V(A)| = \frac{n+t}{r} + 1$, a contradiction. Hence $EX_1(n,t,r) * K_t$ does not contain a K_r-factor.

Now let us consider $EX_2(n,t,r) * K_t$. For a contradiction, assume that $EX_2(n,t,r) * K_t$ contains a K_r-factor $\mathcal{T}$. Then every vertex of B belongs to a different copy of K_r in $\mathcal{T}$. Thus, by construction, the copies of K_r in $\mathcal{T}$ covering B must cover at least

$$\left\lceil \frac{t+1}{r-1} \right\rceil \cdot (r-1) = \left(\frac{t-q}{r-1} + 1\right) \cdot (r-1) = t - q + r - 1 > t + |C|$$

vertices in the copy of K_t and C, a contradiction. Hence $EX_2(n,t,r) * K_t$ does not contain a K_r-factor. □

We now give the extremal constructions which, excluding error terms, match the upper bounds given in Theorem 1.6.

Definition 3.3. *Let $n, t \in \mathbb{N}$ such that $\lceil \frac{n+t}{2} \rceil < n$. We define graphs $EX_1(n,t)$ and $EX_2(n,t)$ as follows:*

- *Let $K := K_n$ and $A \subseteq K$ such that $A = K_{\lceil \frac{n+t}{2} \rceil + 1}$. Define $EX_1(n,t)$ to be the graph obtained by removing $E(A)$ from K.*
- *Define $EX_2(n,t)$ to be the disjoint union of a set of t isolated vertices and a clique of size $n-t$.*

One can see that the extremal examples in Definition 3.3 have the same construction to those in Definition 3.1 for $r = 2$, except that in Definition 3.3 we omit the condition $2|(n + t)$ and add the condition that $\lceil \frac{n+t}{2} \rceil < n$ (in order for $EX_1(n, t)$ to be well-defined).

Observe that $e(EX_1(n, t))$ and $e(EX_2(n, t))$ asymptotically match the upper bounds given in Theorem 1.6. Indeed,

$$e(EX_1(n,t)) = \binom{n}{2} - \binom{\lceil \frac{n+t}{2} \rceil + 1}{2} \quad \text{and} \quad e(EX_2(n,t)) = \binom{n}{2} - \left(t(n-1) - \binom{t}{2}\right).$$

We conclude this section by showing that $EX_1(n, t) * K_t$ and $EX_2(n, t) * K_t$ do not contain certain $(n + t)$-vertex bipartite graphs H.

Definition 3.4. *Let $\mathcal{H}_1$ be the class of bipartite graphs H on $n+t$ vertices with largest independent set of size $\lceil \frac{n+t}{2} \rceil$. Let $\mathcal{H}_2$ be the class of bipartite graphs H on $n + t$ vertices which do not have a tripartition (A, B, C) of $V(H)$ such that $|A| = n - t$, $|B| = |C| = t$ and every vertex in C is only adjacent to vertices in B.*

For example, the Hamilton cycle (when $n + t$ is even), the disjoint union of an isolated vertex and a cycle on $n + t - 1$ vertices (when $n + t$ is odd) and $K_{s,s}$-factors (for any fixed $s \in \mathbb{N}$) belong to $\mathcal{H}_1$; the Hamilton cycle (when $n+t$ is even), $K_{s,s}$-factors (for any fixed $s \in \mathbb{N}$ so that s does not divide t) and any bipartite graph with minimum degree at least $t + 1$ belong to $\mathcal{H}_2$.

Proposition 3.5. *$EX_1(n, t) * K_t$ does not contain any graph in $\mathcal{H}_1$ and $EX_2(n, t) * K_t$ does not contain any graph in $\mathcal{H}_2$.*

Proof. Firstly, let us consider $EX_1(n, t) * K_t$. Since $EX_1(n, t)$ contains an independent set of size $\lceil \frac{n+t}{2} \rceil + 1$, any bipartite graph from $\mathcal{H}_1$ cannot be in $EX_1(n, t)$.

Now let us consider $EX_2(n, t) * K_t$. Since $EX_2(n, t)$ has a set of t isolated vertices and $|K_t| = t$, we require that any bipartite graph H spanning $EX_2(n, t) * K_t$ must have a tripartition (A, B, C) of $V(H)$ such that $|A| = n - t$, $|B| = |C| = t$ and every vertex in C is only adjacent to vertices in B, where $B = V(K_t)$ and C is the set of t isolated vertices in $EX_2(n, t)$. Hence $EX_2(n, t) * K_t$ does not contain any graph in $\mathcal{H}_2$. □

4. Proof of Theorem 1.4

The proof of Theorem 1.4 follows an inductive argument on the number of vertices n of G. Given a graph G such that $G * K_t$ does not contain a K_r-factor, we apply an appropriate vertex-modification procedure which, roughly speaking, allows us to assume G is locally isomorphic to one of the two extremal examples. This allows us to remove such local structure from G and apply induction.

The vertex-modification procedures are described by the following two structural lemmas regarding graphs G with the property that $G * K_t$ does not contain a K_r-factor. Lemma 4.1 allows us to assume that the degree of a vertex is either $n - 1$ (which is the degree of all vertices in $V(EX_1(n, t, r)) \setminus A$ and $C \subset V(EX_2(n, t, r))$) or at most $n - 1 - \lceil \frac{t+1}{r-1} \rceil$ (which is the degree of all vertices in $V(EX_2(n, t, r)) \setminus (B \cup C)$). Lemma 4.2 allows us to assume that either each edge of G belongs to some r-clique or there is a vertex with degree $n - 1$.

Lemma 4.1. *Let $t \geq 0$, $r \geq 3$ and G be a graph on n vertices such that $e(G)$ is maximal with respect to the property that $G * K_t$ does not contain a K_r-factor. Then for every vertex $v \in V(G)$ either $d_G(v) = n - 1$ or $d_G(v) \leq n - 1 - \lceil \frac{t+1}{r-1} \rceil$.*

Proof. Suppose there exists a vertex $v \in V(G)$ such that

$$n - 1 - \left\lceil \frac{t+1}{r-1} \right\rceil < d_G(v) < n - 1.$$

Let G' be the graph obtained from G by adding every possible edge incident to v, that is, $d_{G'}(v) = n-1$. Since $e(G)$ is maximal with respect to $G * K_t$ not containing a K_r-factor, we must have that $G' * K_t$ contains a K_r-factor $\mathcal{T}'$. Using $\mathcal{T}'$, we will now construct a K_r-factor $\mathcal{T}$ in $G * K_t$, giving us a contradiction.

Let K^v be the copy of K_r in $\mathcal{T}'$ covering v. If $K^v \subseteq G * K_t$ then we can take $\mathcal{T} := \mathcal{T}'$. Hence assume $K^v \nsubseteq G * K_t$. If there exists a copy K' of K_r in $\mathcal{T}'$ that lies entirely in K_t, then, for any $u \in V(K')$, we can take

$$\mathcal{T} := \big(\mathcal{T}' \setminus \{K^v, K'\}\big) \cup \{(G * K_t)[\{u\} \cup (V(K^v) \setminus \{v\})], (G * K_t)[\{v\} \cup \big(V(K') \setminus \{u\}\big)]\}.$$

Hence assume no such copy of K_r in $\mathcal{T}'$ exists. Since $K^v \nsubseteq G * K_t$ and no copy of K_r in $\mathcal{T}'$ lies entirely in K_t, the number of copies of K_r in $\mathcal{T}'$ which cover some vertex of $\{v\} \cup V(K_t)$ in $G' * K_t$ is at least $\lceil \frac{t+1}{r-1} \rceil$. But $d_G(v) > n - 1 - \lceil \frac{t+1}{r-1} \rceil$, hence there exists a copy $\hat{K}$ of K_r in $\mathcal{T}'$ which intersects $\{v\} \cup V(K_t)$ and whose vertices are all neighbours of v in $G * K_t$ or v itself. Now, if $v \in V(\hat{K})$, then $K^v = \hat{K} \subseteq G * K_t$, a contradiction to our previous assumption. Hence $V(\hat{K}) \cap V(K_t) \neq \emptyset$ and, for any $u \in V(\hat{K}) \cap V(K_t)$, we can take

$$\mathcal{T} := (\mathcal{T}' \setminus \{K^v, \hat{K}\}) \cup \{(G * K_t)[\{u\} \cup (V(K^v) \setminus \{v\})], (G * K_t)[\{v\} \cup (V(\hat{K}) \setminus \{u\})]\}.$$

□

Lemma 4.2. *Let $t \geq 0$, $r \geq 3$. Let G be a graph on n vertices such that $G * K_t$ does not contain a K_r-factor and suppose G contains an edge which is not contained in any copy of K_r in G. Then there exists a graph G' on n vertices such that $G' * K_t$ does not contain a K_r-factor, $e(G) \leq e(G')$ and G' has a vertex of degree $n - 1$.*

Proof. Let xy be an edge in G which is not contained in any copy of K_r. Let Q be a clique of maximal size containing xy and set $\ell := |V(Q)|$. Observe that every vertex in G has at most $\ell - 1$ neighbours in Q as otherwise xy would lie in an $(\ell + 1)$-clique. Thus

$$\sum_{v \in V(Q)} d_G(v) \leq n(\ell - 1).$$

Let G' be the graph obtained by deleting all edges between x and vertices in $V(G) \setminus V(Q)$ and, subsequently, adding any missing edge incident to any vertex in $V(Q) \setminus \{x\}$. Note that Q is still an ℓ-clique in G' and

$$\sum_{v \in V(Q)} d_{G'}(v) = n(\ell - 1).$$

Hence $e(G) \leq e(G')$. Also, G' is a graph on n vertices and $d_{G'}(y) = n - 1$. It remains to show that $G' * K_t$ does not contain a K_r-factor. Suppose, for a contradiction, that $G' * K_t$ does contain a K_r-factor $\mathcal{T}'$. We will now use $\mathcal{T}'$ to construct a new K_r-factor $\mathcal{T}$ in $G' * K_t$ which does not contain any edge vw where $v \in V(Q)$ and $w \in V(G') \setminus V(Q)$. Such a K_r-factor $\mathcal{T}$ is also a K_r-factor in $G * K_t$, giving us a contradiction.

Suppose that the copy K^x of K_r in $\mathcal{T}'$ covering x has exactly s vertices in K_t. Then K^x has exactly $r - s$ vertices in Q. Thus there are $\ell - r + s$ remaining vertices in $V(Q) \setminus V(K^x)$. Note that $\ell - r + s \leq s$, hence there is an injection $f : V(Q) \setminus V(K^x) \to V(K^x) \cap V(K_t)$. We construct our K_r-factor $\mathcal{T}$ as follows: for every copy of K_r in $\mathcal{T}'$ intersecting Q other than K^x, we substitute all vertices lying in its intersection with Q with their images under f. Finally, we take the copy of K_r formed by the ℓ-clique Q and the $s-(\ell-r+s)=r-\ell$ vertices in $V(K^x)\cap V(K_t)$ which do not appear in the image of f. Observe that $\mathcal{T}$ does not use any edge vw where $v\in V(Q)$ and $w\in V(G)\setminus V(Q)$, and we are done. $\square$

Before proceeding to the proof of Theorem 1.4, we prove the following technical lemma.

Lemma 4.3. *Let $n,t,r\in\mathbb{N}$ such that the following holds: $n,r\geq 3$; $r-1$ divides $t+1$; r divides $n+t$; $t+1<(r-1)(n-1)$. If*

$$e(EX_1(n-1,t+1,r))<e(EX_2(n-1,t+1,r))^1 \tag{1}$$

then

$$e(EX_2(n-1,t+1,r))+(n-1)\leq e(EX_2(n,t,r)). \tag{2}$$

Proof. Let $k:=\frac{t+1}{r-1}$. Since $r-1$ divides $t+1$, we can compute $e(EX_2(n,t,r))$ and $e(EX_2(n-1,t+1,r))$ explicitly:

$$e(EX_2(n,t,r))=\binom{n}{2}-\binom{k}{2}-k(n-k),$$

$$e(EX_2(n-1,t+1,r))=\binom{n-1}{2}-\binom{k+1}{2}-(k+1)(n-k-r).$$

It follows that

$$e(EX_2(n,t,r))-e(EX_2(n-1,t+1,r))=(n-1)+k+(n-k-r)-kr.$$

Rearranging this, one obtains that (2) holds precisely when

$$t\leq n-r-\frac{n}{r}=:g(n,r).$$

Further, one can calculate that $e(EX_2(n-1,t+1,r))\leq e(EX_1(n-1,t+1,r))$ precisely when

$$f_1(n,r):=\frac{n(r-1)}{(2r^2-2r+1)}-r\leq t\leq n(r-1)-r^2=:f_2(n,r). \tag{3}$$

Since (1) holds, we have that (3) implies that $t<f_1(n,r)$ or $t>f_2(n,r)$. Observe that $f_1(n,r)\leq g(n,r)$. Thus, if $t<f_1(n,r)$ then $t<g(n,r)$ and the claim holds.

Suppose $t>f_2(n,r)$. We will show that in this case the hypothesis of the lemma cannot actually hold. In particular, under the assumptions that $r-1$ divides $t+1$, r divides $n+t$ and $t>f_2(n,r)$, $EX_2(n-1,t+1,r)$ is undefined. Indeed, for a contradiction let us assume that $EX_2(n-1,t+1,r)$ is well-defined in this case, thus the inequality $t+1<(r-1)(n-1)$ must hold. By assumption, t satisfies the following modular equations:

$$t\equiv -1 \pmod{r-1} \quad\text{and}\quad t\equiv -n \pmod{r}. \tag{4}$$

For n and r fixed, the solution of (4) is unique modulo $r(r-1)$ by the Chinese Remainder Theorem, since r and $r-1$ are coprime. Note that $t'=(r-1)(n-1)-1$ is a solution of (4), hence $t=(r-1)(n-1)-1-kr(r-1)$ for some integer k. The constraint $t+1<(r-1)(n-1)$ forces $k\geq 1$, hence

$$t\leq (r-1)(n-1)-1-r(r-1)=n(r-1)-r^2=f_2(n,r).$$

This contradicts the assumption that $t>f_2(n,r)$. $\square$

[1] Note that $EX_1(n-1,t+1,r)$, $EX_2(n-1,t+1,r)$ and $EX_2(n,t,r)$ are well-defined since the assumptions in Definition 3.1 are satisfied.

Proof of Theorem 1.4. We prove Theorem 1.4 by induction on n. If $n = 2$ then $e(G) \in \{0, 1\}$. Since $t < 2(r-1)$ and $r|(n+t)$, we must have $t = r-2$. If $e(G) = 1$ then $G * K_t$ is a copy of K_r, contradicting our choice of G. Thus $e(G) = 0$. Observe that $k = \lceil \frac{t+1}{r-1} \rceil = 1$ and $q = r-2$. Thus

$$\max\left\{\binom{n}{2} - \binom{\frac{n+t}{r}+1}{2}, \binom{n}{2} - \binom{k}{2} - k(n-k-(r-2-q))\right\} = \max\{0, 0\} = 0.$$

Thus Theorem 1.4 holds for $n = 2$.

For the inductive step, we may assume without loss of generality that G is a graph on n vertices such that $e(G)$ is maximal with respect to the property that $G * K_t$ does not contain a K_r-factor.

Case (i): G contains an isolated vertex v. If $t < r-2$, then one could add a single edge to v and G would still not contain a K_r-factor, contradicting our choice of G. Hence $t \geq r-2$. If $t = r-2$, then $k = \lceil \frac{t+1}{r-1} \rceil = 1$ and $q = r-2$. Hence

$$\begin{aligned} &\max\left\{\binom{n}{2} - \binom{\frac{n+t}{r}+1}{2}, \binom{n}{2} - \binom{k}{2} - k(n-k-(r-2-q))\right\} \\ =& \max\left\{\binom{n}{2} - \binom{\frac{n-2}{r}+2}{2}, \binom{n-1}{2}\right\}. \end{aligned}$$

Since G contains at least one isolated vertex,

$$e(G) \leq \binom{n-1}{2} \leq \max\left\{\binom{n}{2} - \binom{\frac{n-2}{r}+2}{2}, \binom{n-1}{2}\right\},$$

and we are done. If $t \geq r-1$, consider the graph G' obtained by deleting v from G. Since $G * K_t$ does not contain a K_r-factor, $G' * K_{t-r+1}$ does not contain a K_r-factor. Thus, by our inductive hypothesis

$$e(G') \leq \max\{e(EX_1(n-1, t-r+1, r)), e(EX_2(n-1, t-r+1, r))\}.$$

It follows from $e(G) = e(G')$ and $e(EX_i(n-1, t-r+1, r)) \leq e(EX_i(n, t, r))$[2] for $i = 1, 2$ that $e(G) \leq \max\{e(EX_1(n, t, r)), e(EX_2(n, t, r))\}$.

Case (ii): G contains a vertex of degree $n-1$. Consider the graph G' obtained by deleting such a vertex from G. Note that $G * K_t = G' * K_{t+1}$. If $t+1 \geq (r-1)(n-1)$, then trivially $G' * K_{t+1}$ contains a K_r-factor, a contradiction. So we must have that $t+1 < (r-1)(n-1)$ and hence by induction

$$e(G') \leq \max\{e(EX_1(n-1, t+1, r)), e(EX_2(n-1, t+1, r))\}.$$

Observe that $e(G) = e(G') + n - 1$. We aim to show that

$$e(G) \leq \max\{e(EX_1(n, t, r)), e(EX_2(n, t, r))\}. \tag{5}$$

If $e(G') \leq e(EX_1(n-1, t+1, r)$ then

$$e(G * K_t) = e(G' * K_{t+1}) \leq e(EX_1(n-1, t+1, r) * K_{t+1}) = e(EX_1(n, t, r) * K_t)$$

and thus

$$e(G) \leq e(EX_1(n, t, r)).$$

Similarly, if $e(G') \leq e(EX_2(n-1, t+1, r)$ and $r-1$ does not divide $t+1$ then

$$e(G * K_t) = e(G' * K_{t+1}) \leq e(EX_2(n-1, t+1, r) * K_{t+1}) = e(EX_2(n, t, r) * K_t)$$

and thus

$$e(G) \leq e(EX_2(n, t, r)).$$

[2] This holds since $EX_i(n-1, t-r+1, r)$ can be obtained by removing an appropriate vertex from $EX_i(n, t, r)$, for $i = 1, 2$.

It remains to check that (5) holds in the case when $(r-1)|(t+1)$ and

$$e(EX_1(n-1,t+1,r)) < e(EX_2(n-1,t+1,r)).$$

In this case Lemma 4.3 implies that

$$e(G) = e(G') + n - 1 \leq e(EX_2(n-1,t+1,r)) + (n-1) \leq e(EX_2(n,t,r)),$$

as desired.

Case (iii): G contains no vertex of degree 0 or $n-1$. If G contains an edge which is not contained in any copy of K_r, then by Lemma 4.2 there exists a graph G' on n vertices such that $G' * K_t$ does not contain a K_r-factor, $e(G) \leq e(G')$ and G' has a vertex of degree $n-1$. The argument from Case (ii) then implies that $e(G) \leq e(G') \leq \max\{e(EX_1(n,t,r)), e(EX_2(n,t,r))\}$.

We may therefore assume every edge of G is contained in some copy of K_r. Moreover, as no vertex in G has degree 0, every vertex in G is contained in a copy of K_r. Let w be a vertex of smallest degree in G. If $d_G(w) \geq n - \frac{n+t}{r}$ then

$$\delta(G * K_t) \geq n - \frac{n+t}{r} + t = \frac{r-1}{r}(n+t).$$

Hence by Theorem 1.2, we have that $G * K_t$ contains a K_r-factor, a contradiction.

Thus $d_G(w) < n - \frac{n+t}{r}$. Let K be a copy of K_r in G containing w. Consider the graph G' obtained by removing K and all its vertices from G. Then $G' * K_t$ does not contain a K_r-factor; this implies that $t < (r-1)(n-r)$. Hence, by our inductive hypothesis, we have

$$e(G') \leq \max\{e(EX_1(n-r,t,r)), e(EX_2(n-r,t,r))\}.$$

If $e(G') \leq e(EX_1(n-r,t,r))$ then $d_G(w) < n - \frac{n+t}{r}$ implies $e(G) \leq e(EX_1(n,t,r))$, as desired: this follows from the fact that $EX_1(n-r,t,r)$ can be obtained by removing a clique Q of size r from $EX_1(n,t,r)$ where Q has one vertex of degree $n - \frac{n+t}{r} - 1$. By Lemma 4.1, we have that $\Delta(G) \leq n - 1 - \lceil \frac{t+1}{r-1} \rceil$ since we assumed G contains no vertices of degree $n-1$ and $e(G)$ is maximal with respect to the property that $G * K_t$ does not contain a K_r-factor. Thus, if $e(G') \leq e(EX_2(n-r,t,r))$ then applying this observation yields that $e(G) \leq e(EX_2(n,t,r))$, as desired: similarly as before, this follows from the fact that $EX_2(n-r,t,r)$ can be obtained by removing a clique Q of size r from $EX_2(n,t,r)$ where all vertices in Q have degree $n-1-\lceil \frac{t+1}{r-1} \rceil$. □

5. Proof of Theorem 1.6

In the proof of Theorem 1.6 we will make use of the following theorem of Knox and Treglown [10].

Theorem 5.1 (Knox and Treglown [10]). *Given any $\Delta \in \mathbb{N}$ and any $\gamma > 0$, there exists constants $\beta > 0$ and $n_0 \in \mathbb{N}$ such that the following holds. Suppose that H is a bipartite graph on $n \geq n_0$ vertices with $\Delta(H) \leq \Delta$ and bandwidth at most βn. Let G be a graph on n vertices with degree sequence $d_1 \leq \cdots \leq d_n$. If*

$$d_i \geq i + \gamma n \quad \text{for all} \quad i < n/2$$

then G contains a copy of H.

In fact, Knox and Treglown proved a more general result for robust expanders (see [10, Theorem 1.8]). We use Theorem 5.1 in a similar way to how Chvátal's theorem [4] is used in the proof of Theorem 1.1 in [13].

Proof. Let $\Delta \in \mathbb{N}$ and $\varepsilon > 0$. Define $\gamma > 0$ such that $\gamma \ll \varepsilon, 1/\Delta$. Apply Theorem 5.1 with Δ and γ to produce constants $\beta > 0$ and $n_0 \in \mathbb{N}$ such that

$$0 < \frac{1}{n_0} \ll \beta \ll \gamma \ll \varepsilon, \frac{1}{\Delta}.$$

Let $t \in \mathbb{N}$, $n \geq n_0$ and H be an $(n+t)$-vertex graph as in the statement of the theorem. Suppose that G is a graph on n vertices such that $G * K_t$ does not contain H. Let $m(G)$ denote the number of missing edges in G, that is, $m(G) := \binom{n}{2} - e(G)$. Then proving Theorem 1.6 is equivalent to proving the following bound on $m(G)$:

$$m(G) \geq \begin{cases} t(n-1) - \binom{t}{2} - \varepsilon n^2 & \text{if } \ t \leq \frac{n}{5} \\ \binom{\lceil \frac{n+t}{2} \rceil + 1}{2} - \varepsilon n^2 & \text{if } \ t > \frac{n}{5}. \end{cases} \tag{6}$$

Let $G' := G * K_t$ and label the vertices of G' as $v_1, \ldots, v_{n+t}$ such that $d_{G'}(v_i) := d_i$ is the degree of vertex v_i and $d_1 \leq d_2 \leq \ldots \leq d_{n+t}$. Since G' does not contain H as a subgraph, it does not satisfy the degree sequence condition of Theorem 5.1. Moreover, $\delta(G') \geq t$, hence there must exist $t - \gamma(n+t) < i \leq \lceil (n+t)/2 \rceil - 1$ such that $d_i < i + \gamma(n+t)$. From $d_1 \leq \ldots \leq d_i < i + \gamma(n+t)$ we deduce that the number of edges missing from G' is at least

$$\begin{aligned} m(G') \geq \sum_{j=1}^{i} (n+t-1-d_j) - \binom{i}{2} &> i(n+t-1-i-\gamma(n+t)) - \binom{i}{2} \\ &\geq i((1-2\gamma)(n+t)-i) - \binom{i}{2} =: f(i). \end{aligned} \tag{7}$$

Set $u := \lceil (n+t)/2 \rceil - 1$. Now $f(i)$ is a quadractic in i and $\frac{d^2(f(i))}{di^2} < 0$. Also, note that $m(G) = m(G')$. Hence, as $t - \gamma(n+t) < i \leq u$, we have from (7) that

$$m(G) \geq \min\{f(t-\gamma(n+t)), f(u)\}. \tag{8}$$

One can calculate that

$$f(t-\gamma(n+t)) \leq f(u) \ \text{ if and only if } \ t \leq \begin{cases} \frac{n-2\gamma n+8}{5+2\gamma} & \text{if } n+t \text{ is even} \\ \frac{n-2\gamma n+5}{5+2\gamma} & \text{if } n+t \text{ is odd.} \end{cases} \tag{9}$$

As $\frac{1}{n} \ll \gamma \ll \varepsilon$ we have

$$f(t-\gamma(n+t)) \geq t(n-1) - \binom{t}{2} - \varepsilon n^2 \tag{10}$$

and

$$f(u) \geq \binom{\lceil \frac{n+t}{2} \rceil + 1}{2} - \frac{\varepsilon n^2}{2}. \tag{11}$$

Moreover, for $\frac{n-2\gamma n+5}{5+2\gamma} \leq t \leq \frac{n}{5}$ we have

$$f(u) \geq \binom{\lceil \frac{n+t}{2} \rceil + 1}{2} - \frac{\varepsilon n^2}{2} \geq t(n-1) - \binom{t}{2} - \varepsilon n^2. \tag{12}$$

Regardless of the parity of $n+t$, using (8)–(12) we conclude that (6) holds. $\square$

6. Concluding remarks

In this paper we resolved the deficiency problem for K_r-factors. For a general fixed graph H, it would be interesting to prove deficiency results regarding H-factors. As a starting point for this problem we pose the following question. Let $\alpha(H)$ denote the size of the largest independent set in H.

Question 6.1. *Let $K := K_n$ and $A \subseteq K$ such that $A = K_{\frac{\alpha(H)(n+t)}{|H|}+1}$. Define $EX_H(n,t)$ to be the graph obtained by removing $E(A)$ from K. Does there exist a constant $c := c(H) > 0$ such that if $t \geq cn$ and G is an n-vertex graph so that $G * K_t$ does not contain an H-factor then $e(G) \leq e(EX_H(n,t)) + o(n^2)$?*

Note that Theorem 1.6 answers this question in the affirmative e.g. for $H = K_{s,s}$ (for fixed $s \in \mathbb{N}$). On the other hand, at least for some H one cannot remove the $o(n^2)$ term in Question 6.1 completely. Indeed, let $H = K_{1,s}$ where $s \geq 2$ and consider the n-vertex graph $EX'_H(n,t)$ obtained from $EX_H(n,t)$ by adding a maximal matching in $V(A)$. It is easy to see that $EX'_H(n,t) * K_t$ does not contain a $K_{1,s}$-factor. This example suggests it might be rather challenging to resolve the H-factor deficiency problem completely for all graphs H.

It would also be interesting to prove bandwidth deficiency results in the vein of Theorem 1.6 for non-bipartite graphs H.

Acknowledgment

The authors are grateful to the referee for a helpful and careful review.

References

[1] A. Akiyama and P. Frankl, On the Size of Graphs with Complete-Factors, *J. Graph Theory* **9** (1985), 197–201.
[2] J. Böttcher, M. Schacht and A. Taraz, Proof of the bandwidth conjecture of Bollobás and Komlós, *Math. Ann.* **343** (2009), 175–205.
[3] D. Bryant and D. Horsley, A proof of Lindner's conjecture on embeddings of partial Steiner triple systems, *J. Combin. Des.* **17** (2009), 63–89.
[4] V. Chvátal, On Hamilton's ideals, *J. Combin. Theory Ser. B* **12** (1972), 163–168.
[5] D. Daykin and R. Häggkvist, Completion of sparse partial Latin square, in: Proc. Cambridge Conference in Honour of Paul Erdős, 1983.
[6] T. Evans, Embedding in complete Latin squares, *Am. Math. Mon.* **67** (1960), 958–961.
[7] Z. Füredi and J. Lehel, Tight embeddings of partial quadrilateral packings, *J. Combin. Theory Ser. A* **117** (2010), 466–474.
[8] Z. Füredi, A. Riet and M. Tyomkyn, Completing partial packings of bipartite graphs, *J. Combin. Theory Ser. A* **118** (2011), 2463–2473.
[9] A. Hajnal and E. Szemerédi, Proof of a conjecture of Erdős, *Combinatorial Theory and its Applications vol. II* **4** (1970), 601–623.
[10] F. Knox and A. Treglown, Embedding spanning bipartite graphs of small bandwidth, *Combin. Probab. Comput.* **22** (2013), 71–96.
[11] D. Kühn and D. Osthus, The minimum degree threshold for perfect graph packings, *Combinatorica* **29** (2009), 65–107.
[12] C.C. Lindner, A partial Steiner triple system of order n can be embedded in a Steiner triple system of order $6n+3$, *J. Combin. Theory Ser. A* **18** (1975), 349–351.
[13] R. Nenadov, B. Sudakov and A.Z. Wagner, Completion and deficiency problems, *J. Combin. Theory Ser. B* **145** (2020), 214–240.

Andrea Freschi, Joseph Hyde & Andrew Treglown
School of Mathematics
University of Birmingham
Birmingham
B15 2TT
UK

E-mail addresses: $\{$`axf079, jfh337, a.c.treglown`$\}$`@bham.ac.uk`