Combinatorics, Probability and Computing (2021), **30**, pp. 570–590
doi:10.1017/S0963548320000577

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Near-perfect clique-factors in sparse pseudorandom graphs

Jie Han[1,†], Yoshiharu Kohayakawa[2,‡] and Yury Person[3,*,§,¶]

[1]School of Mathematics and Statistics, Beijing Institute of Technology, Beijing, China, [2]Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brazil and [3]Institut für Mathematik, Technische Universität Ilmenau, 98684 Ilmenau, Germany

*Corresponding author. Email: yury.person@tu-ilmenau.de

(Received 27 May 2019; revised 29 September 2020; accepted 7 July 2020; first published online 11 December 2020)

Abstract

We prove that, for any $t \geqslant 3$, there exists a constant $c = c(t) > 0$ such that any d-regular n-vertex graph with the second largest eigenvalue in absolute value λ satisfying $\lambda \leqslant cd^{t-1}/n^{t-2}$ contains vertex-disjoint copies of K_t covering all but at most $n^{1-1/(8t^4)}$ vertices. This provides further support for the conjecture of Krivelevich, Sudakov and Szábo (*Combinatorica* **24** (2004), pp. 403–426) that (n, d, λ)-graphs with $n \in 3\mathbb{N}$ and $\lambda \leqslant cd^2/n$ for a suitably small absolute constant $c > 0$ contain triangle-factors. Our arguments combine tools from linear programming with probabilistic techniques, and apply them in a certain weighted setting. We expect this method will be applicable to other problems in the field.

2020 MSC Codes: Primary 05C35; Secondary 05C48, 05D40, 90C35

1. Introduction

The study of conditions under which certain spanning or almost spanning structures are forced in random or pseudorandom graphs is one of the central topics in extremal graph theory and in random graphs.

An (n, d, λ)-graph is an n-vertex d-regular graph whose second largest eigenvalue in absolute value is at most λ. Graphs with $\lambda \ll d$ are considered to be *pseudorandom*, that is, they behave in certain respects as random graphs do; for example, the edge count between 'not too small' vertex subsets is close to what one sees in random graphs of the same density. As usual, let $e(A, B) = e_G(A, B)$ denote the number of pairs $(a, b) \in A \times B$ so that ab is an edge of G (note that edges in $A \cap B$ are counted twice). The following result makes the above discussion precise.

†JH was supported by FAPESP (2014/18641-5, 2013/03447-6).
‡YK was partially supported by CNPq (311412/2018-1, 423833/2018-9) and FAPESP (2018/04876-1).
§YP was supported by DFG grant PE 2299/1-1.
¶The cooperation of the authors was supported by a joint CAPES-DAAD PROBRAL project (project 430/15, 57350402, 57391197). This research was partially supported by CAPES (Finance Code 001). FAPESP is the São Paulo Research Foundation. CNPq is the National Council for Scientific and Technological Development of Brazil.

© The Author(s), 2020. Published by Cambridge University Press.

https://doi.org/10.1017/S0963548320000577 Published online by Cambridge University Press

Theorem 1.1 (Expander mixing lemma [4]). *If G is an (n, d, λ)-graph and $A, B \subseteq V(G)$, then*

$$\left| e(A, B) - \frac{d}{n}|A||B| \right| < \lambda\sqrt{|A||B|}. \tag{1.1}$$

As an introduction to the extensive literature on pseudorandom graphs, the reader is referred to [20], [9] or [7, Chapter 9], for example.

It is an interesting problem to understand optimal or asymptotically optimal conditions on the parameter λ in terms of d and n that force an (n, d, λ)-graph to possess a desired property. To demonstrate the optimality of a condition, one needs to show the existence of an (n, d, λ)-graph that certifies that the condition is indeed optimal.

Unfortunately, there are very few examples certifying optimality. A celebrated example is due to Alon, who showed [3] that there are (n, d, λ)-graphs that are K_3-free and yet satisfy $\lambda = cd^2/n$ for some absolute constant $c > 0$. This is in contrast to the fact that, as follows easily from the expander mixing lemma above, for, say, $\lambda \leqslant 0.1d^2/n$, any (n, d, λ)-graph contains a triangle (in fact every vertex lies in a triangle). It turns out that (n, d, λ)-graphs with $\lambda = \Theta(d^2/n)$ must satisfy $d = \Omega(n^{2/3})$. The construction of Alon [3] provides an example of the essentially sparsest possible K_3-free (n, d, λ)-graph with $d = \Theta(n^{2/3})$ and $\lambda = \Theta(n^{1/3})$. The other known example is a generalization of this construction by Alon and Kahale [6] (see also [20, Section 3]) to graphs without odd cycles of length at most $2\ell + 1$.

Our focus here is on spanning or almost spanning structures in (n, d, λ)-graphs. One of the simplest spanning structures is that of a perfect matching. Alon, Krivelevich and Sudakov (see [20]) proved that (n, d, λ)-graphs with $\lambda \leqslant d - 2$ and n even contain perfect matchings. *Factors* generalize perfect matchings: for a graph F, an *F-factor* in a graph G is a collection of vertex-disjoint copies of F in G whose vertex sets cover $V(G)$ (this requires that $v(G) := |V(G)|$ should be divisible by $v(F)$). Motivated by the study of spanning structures in graphs, Krivelevich, Sudakov and Szabó [21] proved that (n, d, λ)-graphs with $\lambda = o(d^3/(n^2 \log n))$ contain a triangle-factor if $3 \mid n$.

A *fractional* triangle-factor in a graph $G = (V, E)$ is a non-negative weight function f on the set $\mathcal{K}_3(G)$ of all triangles T of G, such that, for every $v \in V$, we have $\sum_{T:\, v \in V(T)} f(T) = 1$. Krivelevich, Sudakov and Szabó further proved [21] that (n, d, λ)-graphs with $\lambda \leqslant 0.1d^2/n$ admit a fractional triangle-factor. Moreover, they conjectured the following.

Conjecture 1.2 (Conjecture 7.1 in [21]). *There exists an absolute constant $c > 0$ such that if $\lambda \leqslant cd^2/n$, then every (n, d, λ)-graph G on $n \in 3\mathbb{N}$ vertices has a triangle-factor.*

The tth *power* H^t of a graph H is the graph on the vertex set $V(H)$ where uv $(u \neq v)$ is an edge if there is a u-v-path of length at most t in H. Since the $(t-1)$th power of a Hamilton cycle contains a K_t-factor if $t \mid n$, powers of Hamilton cycles are also of interest when investigating clique-factors.

Allen, Böttcher, Hàn and two of the authors [2] proved that, if $\lambda = o(d^{3t/2}n^{1-3t/2})$ and $t \geqslant 3$, then any (n, d, λ)-graph contains the tth power of the Hamilton cycle (and thus a K_{t+1}-factor if $(t+1) \mid n$). In the case $t = 2$, it was further proved in [2] that the condition $\lambda = o(d^{5/2}/n^{3/2})$ suffices to guarantee squares of Hamilton cycles, and thus K_3-factors, improving on the aforementioned result of Krivelevich, Sudakov and Szabó. For very recent progress, due to Nenadov [23], see Section 7 below.

The construction of Alon of K_3-free (n, d, λ)-graphs shows that the condition on λ in Conjecture 1.2 cannot be weakened. The result from [21] on the existence of fractional triangle-factors supports Conjecture 1.2. As further evidence in support of that conjecture we prove the following result.

https://doi.org/10.1017/S0963548320000577 Published online by Cambridge University Press

Theorem 1.3 (main result[a,b]). *For any $t \geqslant 3$ there exists $n_0 > 0$ for which the following holds. Every (n, d, λ)-graph G with $n \geqslant n_0$ and $\lambda \leqslant (1/(50t4^{t-2}))d^{t-1}/n^{t-2}$ contains vertex-disjoint copies of K_t covering all but at most $n^{1-1/(8t^4)}$ vertices of G.*

We remark that under the condition $\lambda \leqslant cd^{t-1}/n^{t-2}$, for some appropriate $c = c(t) > 0$, Krivelevich, Sudakov and Szabó [21] proved that any (n, d, λ)-graph contains a fractional K_t-factor.

A naïve approach to proving Theorem 1.3 is to pick cliques K_t one after another, each vertex-disjoint from the previous ones, by appealing to the pseudorandomness of G via the expander mixing lemma. However, even for triangles, if $\lambda = cd^2/n$, then all one gets in this way is that G has $(1 - c)n/3$ vertex-disjoint triangles: one can see that a set of cn vertices in G induces a graph of average degree roughly cd, but the condition on λ and the expander mixing lemma do not guarantee that sets of size roughly cd contain an edge, and hence we do not know whether cn vertices necessarily span a triangle. Thus our naïve greedy approach will get stuck, leaving cn vertices uncovered. What our result establishes is that, even for some absolute constant $c > 0$, we can cover all but $o(n)$ vertices of G by vertex-disjoint copies of K_3. Moreover, $o(n)$ can be taken to be of the form $n^{1-\varepsilon}$ for some $\varepsilon > 0$. We have restricted ourselves to triangles in this paragraph, but similar reasoning applies to general cliques K_t as well.

Now let $p = d/n$ and suppose $G = (V, E)$ is an (n, d, λ)-graph with $\lambda \leqslant cd^2/n$. Inequality (1.1) implies that

$$\left| \frac{e(A, B)}{|A||B|} - p \right| < \frac{cp^2 n}{\sqrt{|A||B|}} \leqslant c^{1/2} \tag{1.2}$$

for all $A, B \subseteq V$ with $|A|, |B| \geqslant c^{1/2}n$. Let us now focus on the case in which d is linear in n, that is, $p = d/n$ is a constant independent of n. The powerful blow-up lemma of Komlós, Sárközy and Szemerédi [17] implies that if c is small enough in comparison with p and $1/t$, then *any* graph $G = (V, E)$ on n vertices with minimum degree at least pn that satisfies (1.2) contains a K_t-factor as long as $t \mid n$. Thus Conjecture 1.2 holds for dense graphs.

We remark that the blow-up lemmas for sparse graphs developed recently by Allen, Böttcher, Hàn and two of the authors [1] provide bounds on λ to establish the existence of K_t-factors, but those bounds are worse than those from [2] discussed above.

What is new in this paper is that we further exploit the fractional K_t-factor result of Krivelevich, Sudakov and Szabó [21][c] in a weighted setting, so that we can repeatedly find 'weight-disjoint' fractional K_t-factors in the host graph. Then we show that we can produce an almost K_t-factor from the fractional ones by probabilistic arguments. In fact our method can be seen as a synthesis of some methods in Alon *et al.* [5] and Krivelevich, Sudakov and Szabó [21]. In fact, very recently the first author [12] used this new method to give a regularity-free proof of a remarkable characterization theorem of Keevash and Mycroft [16] on perfect matchings in hypergraphs. We expect that this new method will be applicable to other problems in the field.

Throughout the paper, floor and ceiling signs are omitted for the sake of readability. For graph theory terminology and notation we refer the reader to Bollobás [8].

This paper is organized as follows. In Section 2 we provide an overview of the proof of Theorem 1.3. In Section 3 we collect some of the necessary tools and prove auxiliary results. The proof of Theorem 1.3 splits into two cases (the 'dense' and 'sparse' cases), and we deal with these two cases separately in Sections 4 and 5 and then prove Theorem 1.3 in Section 6. Finally we give some concluding remarks in Section 7.

[a] This result appears in an extended abstract in the Proceedings of Discrete Mathematics Days 2018 (Seville) [14].
[b] Theorem 1.3 answers a question raised by Nenadov (see [23, Concluding remarks]).
[c] This result is proved for $t = 3$ in [21].

https://doi.org/10.1017/S0963548320000577 Published online by Cambridge University Press

2. Proof outline

In the following we provide a proof overview in the case of triangles, since the general case is similar. Our arguments combine tools from linear programming with probabilistic techniques. In fact they can be seen as a synthesis of some methods in Alon *et al.* [5] and Krivelevich, Sudakov and Szabó [21]. The rôle of linear programming is to obtain some approximate 'fractional' structures, whereas probabilistic techniques allow us to convert them to non-fractional ones.

Let an (n, d, λ)-graph G with $\lambda \leqslant cd^2/n$ be given. From the expander mixing lemma, Theorem 1.1, it follows that every vertex of G lies in $\frac{1}{2}(d^3/n \pm \lambda d) = (d^3/2n)(1 \pm c)$ triangles. The naïve greedy approach outlined in the Introduction (see the discussion soon after Theorem 1.3) does not guarantee a collection of $(1 - o(1))n/3$ vertex-disjoint triangles. Another approach would be to apply some theorem that would tell us that the 3-uniform hypergraph $\mathcal{K}_3(G)$ of the triangles in G contains an almost perfect matching. A theorem of Pippenger (see [11]) would do if we knew that $\mathcal{K}_3(G)$ is pseudorandom enough (roughly speaking, one needs $\mathcal{K}_3(G)$ to be approximately ℓ-regular for some $\ell \to \infty$, and pairs of vertices of $\mathcal{K}_3(G)$ to be contained in $o(\ell)$ triples of $\mathcal{K}_3(G)$, *i.e.* the 'codegrees' should be small). However, for c an absolute constant, this property of $\mathcal{K}_3(G)$ cannot be deduced.

We circumvent the fact that $\mathcal{K}_3(G)$ is not necessarily pseudorandom enough by finding a subhypergraph H of $\mathcal{K}_3(G)$ in which the 'deviation' of the number of triangles at any vertex is 'smoothed out' (thus H will be almost ℓ-regular). This can be done if G has $\ell = n^{\Theta(1)}$ fractional K_3-factors $f_1, \dots, f_\ell$ such that $\sum_{i=1}^{\ell} f_i(T) \leqslant 1$ for each $T \in \mathcal{K}_3(G)$ and, for any edge $e \in E(G)$, the sum of the weights on the triangles containing e across $f_1, \dots, f_\ell$ is at most $\ell^{1-\gamma}$ for some $\gamma \in (0, 1)$. This latter condition helps us force small codegrees.

Indeed, with these fractional K_3-factors, we can select $H \subseteq \mathcal{K}_3(G)$ at random, by including each $T \in \mathcal{K}_3(G)$ in H independently with probability $\sum_{i=1}^{\ell} f_i(T)$. Then Chernoff's inequality guarantees that H satisfies, with high probability, the assumptions of Theorem 3.1, which is a packing result by Kostochka and Rödl from [18] strengthening the aforementioned result of Pippenger. Such a 'randomization' strategy has previously been successfully employed in [5] in the context of perfect matchings in hypergraphs.

Thus it suffices to find such fractional K_3-factors $f_1, \dots, f_\ell$. In fact we find such f_i with the property that, for any $e \in E(G)$, we have $\sum_{e \in E(T)} \sum_{i=1}^{\ell} f_i(T) \leqslant 1$ (hence $\sum_{i=1}^{\ell} f_i(T) \leqslant 1$ for each $T \in \mathcal{K}_3(G)$ is automatically true).

Theorem 1.3 is vacuously true for $d = o(n^{2/3})$ when $t = 3$ (see Fact 3.6 in the next section). We thus suppose $d = \Omega(n^{2/3})$. We consider two cases. We pick any $\beta \in (0, 1/3)$ independent of n. Our first approach (Theorem 4.1) works as long as d is not too small, say $d \geqslant n^{(2/3)+\beta}$. In contrast, the second approach (Theorem 5.1) works as long as d is not too large, say $d \leqslant n^{1-\beta}$.

In the first approach we consider edge-weighted graphs and we repeatedly 'remove' fractional K_3-factors from G (removing from edges e the weights of the triangles T with $e \subseteq V(T)$). This is done by Theorem 4.1, in which we show that we can repeatedly apply Corollary 3.4 in the remaining weighted graph n^β times.

When d is close to $n^{2/3}$, our approach above fails because we cannot execute it sufficiently many times. To circumvent this, we randomly split $E(G)$ into $\ell = n^{\Omega(1)}$ sets $E_1, \dots, E_\ell$, with each subgraph $G_i := (V, E_i)$ distributed as a random subgraph G_p of G, where each edge is included in G_p with probability $p = 1/\ell$, independently of all the other edges. Then we show that with high probability each G_i satisfies the assumptions of Corollary 3.4 and thus contains a fractional K_3-factor f_i. This second approach works only for $d \leqslant n^{1-o(1)}$, which makes both approaches necessary.

More discussion of the limitations of both approaches can be found in the concluding remarks in Section 7.

https://doi.org/10.1017/S0963548320000577 Published online by Cambridge University Press

3. Tools and auxiliary results

3.1 Probabilistic techniques

We shall use the following theorem of Kostochka and Rödl [18] (see also Rödl [24] and Alon and Spencer [7, Theorem 4.7.1]), which asserts the existence of an almost perfect matching in 'pseudorandom' hypergraphs.

Theorem 3.1. *Let integers $t \geqslant 3$ and $k \geqslant 8$ and real numbers δ' and γ with $0 < \delta',\ \gamma < 1$ be fixed. Then there exists D_0 such that the following holds for $D \geqslant D_0$. Let H be a t-uniform hypergraph on a set V of n vertices such that*

(i) *for all vertices $v \in V$, we have $D - k\sqrt{D \log D} \leqslant \deg_H(v) \leqslant D$ and*
(ii) *for any two distinct vertices u and $v \in V$, we have $\deg_H(u, v) \leqslant C < D^{1-\gamma}$.*

Then H contains a matching covering all but $O(n(C/D)^{(1-\delta')/(t-1)})$ vertices.

We shall use the following concentration results.

Theorem 3.2 (Chernoff bounds [15, Corollary 2.4 and Theorems 2.8 and 2.10]). *Suppose X is a sum of a collection of independent Bernoulli random variables. Then, for $\delta \in (0, 3/2)$, we have*

$$\mathbb{P}(X > (1+\delta)\mathbb{E}X) < e^{-\delta^2 \mathbb{E}X/3} \quad \text{and} \quad \mathbb{P}(X < (1-\delta)\mathbb{E}X) < e^{-\delta^2 \mathbb{E}X/2}.$$

Moreover, for any $t \geqslant 6\mathbb{E}X$, we have

$$\mathbb{P}(X \geqslant \mathbb{E}X + t) \leqslant e^{-t}.$$

For a graph $G = (V, E)$ we let G_p denote the spanning random subgraph of G in which each edge from E is included with probability p, independently of all other edges.

Theorem 3.3 (Janson's inequality [15, Theorem 2.14]). *Let $p \in (0, 1)$ be given and consider a family $\{H_i\}_{i \in \mathcal{I}}$ of subgraphs of a graph G. For each $i \in \mathcal{I}$, let X_i denote the indicator random variable for the event that $H_i \subseteq G_p$. Write $H_i \sim H_j$ for each ordered pair $(i, j) \in \mathcal{I} \times \mathcal{I}$ such that $E(H_i) \cap E(H_j) \neq \emptyset$. Let $X = \sum_{i \in \mathcal{I}} X_i$. Then $\mathbb{E}[X] = \sum_{i \in \mathcal{I}} p^{e(H_i)}$. Furthermore, let*

$$\Delta = \sum_{H_i \sim H_j} \mathbb{E}[X_i X_j] = \sum_{H_i \sim H_j} p^{e(H_i)+e(H_j)-e(H_i \cap H_j)}. \tag{3.1}$$

Then, for any $0 < \gamma < 1$, we have

$$\mathbb{P}[X \leqslant (1-\gamma)\mathbb{E}[X]] \leqslant \exp\left(-\frac{\gamma^2 \mathbb{E}[X]^2}{2\Delta}\right). \tag{3.2}$$

3.2 Linear programming techniques

We shall consider *weighted graphs* (G, w), where $G = (V, E)$ is a graph and $w\colon E \to [0, 1]$ is a function on its edges. If $w \equiv 1$, we identify (G, w) with G. For every vertex $v \in V$, we define its *weighted degree* $\deg_w(v)$ to be $\sum_{u \in N(v)} w(uv)$.

Let $\mathcal{K}_t(G)$ denote the set of all copies of K_t in G. A function $f\colon \mathcal{K}_t(G) \to [0, 1]$ is called a *fractional K_t-factor* if $\sum_{T \in \mathcal{K}_t(G)\colon V(T) \ni v} f(T) = 1$ for every $v \in V$, and for every edge $uv \in E(G)$ we have

$$\sum_{T \in \mathcal{K}_t(G)\colon V(T) \supset \{u,v\}} f(T) \leqslant w(uv). \tag{3.3}$$

https://doi.org/10.1017/S0963548320000577 Published online by Cambridge University Press

A fractional K_t-factor in a weighted graph (G, w) thus 'respects' the weight function w. We remark that the definition of a fractional K_t-factor in a graph generalizes the definition of a fractional triangle-factor from the Introduction in a straightforward way, and it is itself generalized by the above one for weighted graphs, since condition (3.3) is satisfied in the case of fractional K_t-factors in unweighted graphs because of our convention that $w \equiv 1$ in that case.

Let A be an $m \times n$ matrix with real entries, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$. The optimization task

$$\max c^T x \text{ subject to } Ax \leqslant b,\ x \geqslant 0, \tag{3.4}$$

is the *primal linear program* defined by (A, b, c), whereas

$$\min b^T y \text{ subject to } A^T y \geqslant c,\ y \geqslant 0 \tag{3.5}$$

is the associated *dual linear program*. We shall use the duality theorem of linear programming in the following form (see *e.g.* [22]).

Theorem 3.4 (duality theorem). *If one of the linear programs* (3.4) *or* (3.5) *has an optimal solution then so does the other, and if x^* is optimal for* (3.4) *and y^* for* (3.5) *then $c^T x^* = b^T y^*$.*

We will also use the following corollary of Theorem 3.4, called the *complementary slackness conditions*, stating that the pair (x^*, y^*) which satisfies both programs (3.4) and (3.5) is optimal if and only if the following holds:

- $(Ax^*)_j = 0$ or $y^*_j = 0$ for all $j \in [m]$, and
- $(A^T y^*)_i = 0$ or $x^*_i = 0$ for all $i \in [n]$.

We will apply the theory of linear programming to study fractional K_t-*matchings* in weighted graphs. The maximum weight of a fractional K_t-matching in (G, w) is given by the following linear program:

$$\begin{aligned}
&\max \sum_{T \in \mathcal{K}_t(G)} f(T),\\
&\sum_{T \in \mathcal{K}_t(G):\ V(T) \ni v} f(T) \leqslant 1 \quad \text{for all } v,\\
&\sum_{T \in \mathcal{K}_t(G):\ V(T) \supset \{u,v\}} f(T) \leqslant w(uv) \quad \text{for all } uv \in E(G),\\
&f(T) \geqslant 0 \quad \text{for all } T \in \mathcal{K}_t(G).
\end{aligned} \tag{3.6}$$

The dual of (3.6) is the following linear program:

$$\begin{aligned}
&\min \sum_{v \in V} g(v) + \sum_{uv \in E(G)} h(uv) w(uv),\\
&\sum_{v:\ v \in V(T)} g(v) + \sum_{uv \in E(T)} h(uv) \geqslant 1 \quad \text{for all } T \in \mathcal{K}_t(G),\\
&g(v) \geqslant 0 \quad \text{for all } v,\\
&h(uv) \geqslant 0 \quad \text{for all } uv \in E(G).
\end{aligned} \tag{3.7}$$

Both linear programs above are clearly feasible, and therefore the duality theorem tells us that both admit optimal solutions and that, moreover, these optimal values are equal. Given optimal solutions f and (g, h), the complementary slackness conditions tell us that if $g(v) > 0$ then the corresponding inequality $\sum_{T \in \mathcal{K}_t(G):\ V(T) \ni v} f(T) \leqslant 1$ in the primal linear program (3.6) holds with equality, *i.e.* $\sum_{T \in \mathcal{K}_t(G):\ V(T) \ni v} f(T) = 1$.

https://doi.org/10.1017/S0963548320000577 Published online by Cambridge University Press

Let $t^*(G,w)$ be the optimum value of (3.6) and (3.7). For a collection $\mathcal{F}$ of vertex-disjoint copies of K_t from $\mathcal{K}_t(G)$, let $w(\mathcal{F}) := \sum_{T\in\mathcal{F}} \min_{uv\in E(T)} w(uv)$. Furthermore, let $t(G,w)$ be the maximum possible value of $w(\mathcal{F})$ for such a collection $\mathcal{F}$. Write $|g|$ for $\sum_{v\in V} g(v)$ and $|h|$ for $\sum_{uv\in E(G)} h(uv)w(uv)$ (see the objective function in (3.7)).

The following proposition collects some useful properties of linear programs (3.6) and (3.7). A variant for unweighted uniform hypergraphs and fractional matchings was first stated and proved by Krivelevich in [19, Proposition 2].

Proposition 3.1. *Let $t \geqslant 3$ be given and let (G,w) be a weighted graph. Suppose $G=(V,E)$. Then the following hold.*

(i) $t^*(G,w) \geqslant t(G,w)$.
(ii) $t^*(G,w) \leqslant |V|/t$. *Furthermore, if $t^*(G,w) = |V|/t$, then (G,w) has a fractional K_t-factor.*
(iii) *If $g\colon V \to \mathbb{R}_{\geqslant 0}$ and $h\colon E \to \mathbb{R}_{\geqslant 0}$ form a feasible solution to (3.7), then for every subset $U \subseteq V$ the functions*

$$g' := g \restriction_U \quad \text{and} \quad h' := h \restriction_{E\cap\binom{U}{2}}$$

form a feasible solution to (3.7) with $G[U]$ in place of G; in particular we have $|g'| + |h'| \geqslant t^*(G[U],w)$.
(iv) *If $g\colon V \to \mathbb{R}_{\geqslant 0}$ and $h\colon E \to \mathbb{R}_{\geqslant 0}$ form an optimal solution to (3.7), then $t^*(G,w) \geqslant |V_1|/t$, where $V_1 := \{v \in V\colon g(v) > 0\}$.*

Proof. Let $\mathcal{F}$ be a collection of vertex-disjoint copies of K_t. Let $f(T) := \min_{uv\in E(T)} w(uv)$ for every $T \in \mathcal{F}$ and $f(T) := 0$ for every $T \in \mathcal{K}_t(G) \setminus \mathcal{F}$. Clearly f satisfies (3.6), whence (i) follows.

From

$$\sum_v \sum_{T\in\mathcal{K}_t(G)\colon V(T)\ni v} f(T) \leqslant |V| \quad \text{and} \quad \sum_v \sum_{T\in\mathcal{K}_t(G)\colon V(T)\ni v} f(T) = t \sum_{T\in\mathcal{K}_t(G)} f(T),$$

assertion (ii) follows immediately.

It is also clear that restricting a feasible solution of (3.7) to a subset $U \subseteq V$, we obtain functions for which the relevant inequalities hold. Since the value of a feasible solution of the dual (3.7) is always at least the value of a feasible solution of the primal program (3.6) (if both are feasible, which is the case here), we obtain $|g'| + |h'| \geqslant t^*(G[U],w)$. Thus assertion (iii) follows.

From the complementary slackness conditions we know that

$$\sum_{T\in\mathcal{K}_t(G)\colon V(T)\ni v} f(T) = 1$$

whenever $g(v) > 0$. This implies that

$$t^*(G,w) = \sum_{T\in\mathcal{K}_t(G)} f(T) = \frac{1}{t}\sum_{v\in V} \sum_{T\in\mathcal{K}_t(G)\colon V(T)\ni v} f(T) \geqslant \frac{1}{t}\sum_{v\in V_1} 1 = \frac{|V_1|}{t},$$

which proves (iv). □

3.3 Applications of linear programming to weighted graphs

The main result of this section, Corollary 3.4, generalizes results from Krivelevich, Sudakov and Szabó [21, Section 5] to weighted graphs and fractional K_t-factors (our proof follows their proof strategy). More precisely, Corollary 3.4 will allow us to assert that certain weighted subgraphs of (n,d,λ)-graphs possess fractional K_t-factors by verifying some technical combinatorial conditions in those graphs.

https://doi.org/10.1017/S0963548320000577 Published online by Cambridge University Press

We first state a simple fact. Given $\alpha \geqslant 0$ and a weighted graph (G, w), we call an edge $uv \in E(G)$ *α-rich* if $w(uv) \geqslant 1-\alpha$. Similarly, we call a copy T of K_t in G *α-rich* if $w(T) \geqslant 1-\alpha$, where $w(T) := \min_{uv \in E(T)} w(uv)$.

Fact 3.5. Let (G, w) be a weighted d-regular graph. If $\sum_{u \in N(v)} w(uv) \geqslant d(1-\alpha^2)$ for a vertex v, then v is incident to at least $(1-\alpha)d$ many α-rich edges.

Proof. Let m be the number of non-α-rich edges incident to a vertex v. Then we have $\sum_{u \in N(v)} w(uv) \leqslant d - m\alpha$. If $\sum_{u \in N(v)} w(uv) \geqslant d(1-\alpha^2)$, then $d\alpha^2 \geqslant m\alpha$ and thus $m \leqslant \alpha d$. □

Let a weighted graph (G, w) with $G = (V, E)$ be given. In what follows, we shall consider the spanning subgraph $H = (V, F)$ of G, with $F \subseteq E$ the set of α-rich edges of (G, w), where α will be chosen suitably.

Let $t \geqslant 3$ and $0 < D,\ D' < n$ be given. An n-vertex graph G has property $\mathcal{P}(t, D, D', n)$ if the following holds. For every subset $U \subseteq V(G)$ of cardinality $|U| \geqslant n - D$ and for every subset $U_0 \subseteq U$ with $|U_0| = D/t$, there exists a family $\mathcal{T}_0$ of at least $D'/(t-1)$ copies of K_t in $G[U]$ with the following properties:

(1) $V(T) \subseteq U$ for every $T \in \mathcal{T}_0$,
(2) $|V(T) \cap U_0| = 1$ for every $T \in \mathcal{T}_0$,
(3) $V(T) \cap V(T') \subseteq U_0$ for any distinct T and $T' \in \mathcal{T}_0$.

The copies of K_t specified above thus have the property that they are all edge-disjoint and intersect pairwise in at most one vertex, which must be from U_0.

Lemma 3.2. *Let (G, w) be a weighted graph and let $\alpha \in [0, 1/(20t))$ be given. Suppose H is the spanning subgraph of G formed by the α-rich edges in (G, w). Suppose H is such that*

(i) *every $0.11n/t$ vertices span a copy of K_t in H, and*
(ii) *H has property $\mathcal{P}(t, D, 0.2n, n)$ for some $D \leqslant n/2$.*

Then $t^*(G[U], w) \geqslant t^*(H[U], w) \geqslant (|U| - D/t)/t$ for every $U \subseteq V = V(G)$ with $|U| \geqslant n - D$.

Proof. Let U as in the statement of the lemma be given. Let g and h form an optimal solution of the linear program (3.7) applied to $H[U]$ instead of G. Let $U' = \{u \in U : g(u) = 0\}$. If $|U'| \leqslant D/t$, then we have by Proposition 3.1(iv) that $t^*(H[U], w) \geqslant |U \setminus U'|/t \geqslant (|U| - D/t)/t$, and we are done.

We now assume that $|U'| > D/t$ and derive a contradiction. First we fix a subset $U_0 \subseteq U'$ with $|U_0| = D/t$ and then we consider a family $\mathcal{T}_0$ of cardinality $0.2n/(t-1)$ as given by property $\mathcal{P}$. Let

$$W := \left(\bigcup_{T \in \mathcal{T}_0} V(T) \right) \setminus U_0.$$

We have $|W| = 0.2n$ and, moreover,

$$\sum_{v \in W} g(v) + \sum_{T \in \mathcal{T}_0} \sum_{uv \in E(T)} h(uv) w(uv) \geqslant \sum_{v \in W} g(v) + \sum_{T \in \mathcal{T}_0} \sum_{uv \in E(T)} h(uv)(1-\alpha)$$
$$\geqslant (1-\alpha) \left(\sum_{v \in W} g(v) + \sum_{T \in \mathcal{T}_0} \sum_{uv \in E(T)} h(uv) \right)$$

https://doi.org/10.1017/S0963548320000577 Published online by Cambridge University Press

$$= (1-\alpha)\sum_{T\in\mathcal{T}_0}\Bigg(\sum_{v\in W\cap V(T)} g(v) + \sum_{uv\in E(T)} h(uv)\Bigg)$$
$$\geqslant (1-\alpha)|W|/(t-1)$$

because g and h form an optimal solution of (3.7), $g\restriction_{U_0}\equiv 0$ and

$$\sum_{v:\, v\in V(T)} g(v) + \sum_{V(T)\supset\{u,v\}} h(uv) \geqslant 1 \quad \text{for every } T\in\mathcal{K}_t(H[U]).$$

Since every $0.11n/t$ vertices of H span a copy of K_t, we find at least $(|U\setminus W| - 0.11n/t)/t$ vertex-disjoint copies of K_t in $H[U\setminus W]$. Proposition 3.1(i) then implies that

$$t^*(H[U\setminus W], w) \geqslant t(H[U\setminus W], w) \geqslant (1-\alpha)\frac{|U\setminus W| - 0.11n/t}{t}.$$

By Proposition 3.1(iii), $g\restriction_{U\setminus W}$ and $h\restriction_{E(G)\cap\binom{U\setminus W}{2}}$ form a feasible solution to the linear program (3.7) applied to $H[U\setminus W]$ instead of H and, moreover,

$$|g\restriction_{U\setminus W}| + |h\restriction_{E(H)\cap\binom{U\setminus W}{2}}| \geqslant t^*(H[U\setminus W], w) \geqslant (1-\alpha)\frac{|U\setminus W| - 0.11n/t}{t}.$$

Thus we have

$$\begin{aligned}
t^*(H[U], w) &\geqslant \Bigg(\sum_{v\in W} g(v) + \sum_{T\in\mathcal{T}_0}\sum_{uv\in E(T)} h(uv)w(uv)\Bigg) + |g\restriction_{U\setminus W}| + |h\restriction_{E(H)\cap\binom{U\setminus W}{2}}| \\
&\geqslant (1-\alpha)\left(\frac{|W|}{t-1} + \frac{|U| - |W| - 0.11n/t}{t}\right) \\
&\geqslant (1-\alpha)\frac{|U| + |W|/t - 0.11n/t}{t} \\
&= (1-\alpha)\frac{|U| + 0.09n/t}{t} \\
&= \frac{|U|}{t} + \frac{0.09n}{t^2} - \alpha\frac{1.1n}{t} \\
&> \frac{|U|}{t},
\end{aligned}$$

which contradicts Proposition 3.1(ii). □

We let $\mathcal{T}_v$ denote a family of copies of K_t in G with $v\in T$ for every $T\in\mathcal{T}_v$ and with $V(T)\cap V(T') = \{v\}$ for all distinct T and T' in $\mathcal{T}_v$. The next lemma establishes a sufficient condition for the existence of a fractional K_t-factor in a weighted graph.

Lemma 3.3. *Let $\alpha\geqslant 0$ be a real number and let t, n and $D\geqslant 3$ be integers with $\alpha < 1/t^2$ and $D\leqslant n/2$. Suppose (G, w) is an n-vertex weighted graph such that*

(i) *for every $v\in V(G)$, there exists a family $\mathcal{T}_v$ of at least $D/(t-1)$ α-rich copies of K_t, and*
(ii) *for every $U\subseteq V$ of size $|U|\geqslant n-D$, we have $t^*(G[U], w)\geqslant (|U| - D/t)/t$.*

Then (G, w) contains a fractional K_t-factor.

Proof. Let $g\colon V = V(G)\to\mathbb{R}_{\geqslant 0}$ and $h\colon E = E(G)\to\mathbb{R}_{\geqslant 0}$ form an optimal solution to (3.7). If $g(v) > 0$ for each $v\in V$, we obtain by Proposition 3.1(iv) that $t^*(G, w)\geqslant |V|/t$. By Proposition 3.1(ii), we deduce that $t^*(G, w) = |V|/t$ and conclude that (G, w) has a fractional K_t-factor.

https://doi.org/10.1017/S0963548320000577 Published online by Cambridge University Press

Assume now for a contradiction that there is a vertex x with $g(x) = 0$. Let $\mathcal{T}_x$ be a family of $D/(t-1)$ α-rich copies of K_t. Let

$$W := \left(\bigcup_{T\in\mathcal{T}_x} V(T)\right) \setminus \{x\}$$

and observe that $|W| = D$. Since the copies of K_t from $\mathcal{T}_x$ pairwise intersect only at x and they are α-rich, it follows that

$$\begin{aligned}\sum_{v\in W} g(v) + \sum_{T\in\mathcal{T}_x}\sum_{uv\in E(T)} h(uv)w(uv) &\geqslant \sum_{v\in W} g(v) + \sum_{T\in\mathcal{T}_x}\sum_{uv\in E(T)} h(uv)(1-\alpha)\\ &\geqslant (1-\alpha)\left(\sum_{v\in W} g(v) + \sum_{T\in\mathcal{T}_x}\sum_{uv\in E(T)} h(uv)\right)\\ &= (1-\alpha)\sum_{T\in\mathcal{T}_x}\left(\sum_{v\in W\cap V(T)} g(v) + \sum_{uv\in E(T)} h(uv)\right)\\ &\geqslant (1-\alpha)D/(t-1).\end{aligned}$$

Proposition 3.1(iii) yields that

$$g_1 := g\restriction_{V\setminus W} \quad \text{and} \quad h_1 := h\restriction_{E\cap\binom{V\setminus W}{2}}$$

is a feasible solution to (3.7) with G replaced by $G[V\setminus W]$. Therefore we have $|g_1| + |h_1| \geqslant t^*(G[V\setminus W], w)$. By assumption (ii) we get

$$|g_1| + |h_1| \geqslant t^*(G[V\setminus W], w) \geqslant \frac{|V\setminus W| - D/t}{t} = \frac{n - D - D/t}{t}.$$

Together we obtain

$$\begin{aligned}t^*(G,w) &= |g| + |h|\\ &\geqslant \sum_{v\in W} g(v) + \sum_{T\in\mathcal{T}_x}\sum_{uv\in E(T)} h(uv)w(uv) + |g_1| + |h_1|\\ &\geqslant \frac{(1-\alpha)D}{t-1} + \frac{n-D-D/t}{t}\\ &> \frac{n}{t}.\end{aligned}$$

This contradicts $t^*(G,w) \leqslant n/t$ (see Proposition 3.1(ii)). Thus this case never happens, *i.e.* $g(v) > 0$ for all $v \in V$. We conclude that (G,w) contains a fractional K_t-factor. □

The following corollary of Lemmas 3.2 and 3.3 is our main tool for finding fractional K_t-factors.

Corollary 3.4. *Let $\alpha \geqslant 0$ be a real number and let t, n and $D \geqslant 3$ be integers with $\alpha < 1/(7t^2)$ and $D \leqslant n/2$. Suppose (G,w) is an n-vertex weighted graph and suppose H is the spanning subgraph of G formed by the α-rich edges of (G,w). Suppose*

(i) *for every $v \in V(G)$ there exists a family $\mathcal{T}_v$ of at least $D/(t-1)$ α-rich copies of K_t,*
(ii) *every $0.11n/t$ vertices span a copy of K_t in H,*
(iii) *H has property $\mathcal{P}(t, D, 0.2n, n)$ for some $D \leqslant n/2$.*

Then (G,w) contains a fractional K_t-factor.

https://doi.org/10.1017/S0963548320000577 Published online by Cambridge University Press

3.4 Further useful properties of (n, d, λ)-graphs

We will use the following auxiliary results.

Proposition 3.5 (Proposition 2.3 in [21]). *Let G be an (n, d, λ)-graph with $d \leqslant n/2$. Then $\lambda \geqslant \sqrt{d/2}$.*

Fact 3.6. Let G be an (n, d, λ)-graph with $d \leqslant n/2$. Suppose $\lambda \leqslant d^{t-1}/n^{t-2}$ for some $t \geqslant 3$. Then $d \geqslant n^{1-1/(2t-3)}/2$.

Proof. Proposition 3.5 tells us that $\lambda \geqslant \sqrt{d/2}$. Thus $\lambda \leqslant d^{t-1}/n^{t-2}$ implies that $d^{2t-3} \geqslant n^{2t-4}/2$, whence $d \geqslant n^{1-1/(2t-3)}/2^{1/(2t-3)}$ follows. □

Given two graphs G and G' on the same vertex set V, let $G \setminus G' = (V, E(G) \setminus E(G'))$. The following proposition gives a rough estimate for the number of copies of K_{t-1} in induced subgraphs of (n, d, λ)-graphs, even after removing a small number of edges incident to each vertex.

Proposition 3.6. *For any integer $t \geqslant 3$, there exists n_0 such that every (n, d, λ)-graph G with $n \geqslant n_0$ satisfies the following. Suppose $\lambda(4n/d)^{t-2} \leqslant m \leqslant d$. Let G' be a graph on $V(G)$ with maximum degree $(d/(4n))^{t-2}m$. Then, for any $2 \leqslant i \leqslant t-1$ and any set U of at least $(d/(4n))^{t-i-1}m$ vertices of G, the number of copies of K_i in $(G \setminus G')[U]$ is at least $2^{-i^2} i!^{-1} |U|^i (d/n)^{\binom{i}{2}}$ and at most $2^{i^2} i!^{-1} |U|^i (d/n)^{\binom{i}{2}}$.*

Proof. Let $m_i = (d/(4n))^{t-i-1}m$ for $2 \leqslant i \leqslant t-1$. Hence $\lambda \leqslant m(d/(4n))^{t-2} = m_i(d/(4n))^{i-1}$. Suppose $|U| \geqslant m_i$. We prove by induction on i that the stated estimates hold. Note that since $i \geqslant 2$, we have

$$\Delta(G') \leqslant (d/(4n))^{t-2}m \leqslant dm_i/(4n) \leqslant d|U|/(4n).$$

First let $i = 2$. Theorem 1.1 implies that $|2e_G(U) - (d/n)|U|^2| \leqslant \lambda|U|$. Since $\lambda \leqslant |U|(d/(4n))$, we have

$$(3/8)(d/n)|U|^2 \leqslant e_G(U) \leqslant (5/8)(d/n)|U|^2.$$

Hence the number of edges in $(G \setminus G')[U]$ is at most $(5/8)(d/n)|U|^2$ and at least

$$(3/8)(d/n)|U|^2 - |U| \cdot d|U|/(4n) = (1/8)(d/n)|U|^2,$$

which verifies our claim for $i = 2$. Now suppose $3 \leqslant i \leqslant t-1$ and that the estimates hold for smaller values of i. Note that, in particular, we have $t \geqslant 4$. Let X_1 be the set of vertices $v \in U$ such that $\deg(v, U) \geqslant 2d|U|/n$. By the definition of X_1 and Theorem 1.1, we have

$$\frac{2d}{n}|U||X_1| \leqslant e(U, X_1) \leqslant \frac{d}{n}|U||X_1| + \lambda\sqrt{|U||X_1|}.$$

Together with $\lambda \leqslant m_i(d/(4n))^{i-1} \leqslant |U|(d/(4n))^{i-1}$ and $i \geqslant 3$, this implies that

$$|X_1| \leqslant \left(\frac{\lambda n}{d}\right)^2 \frac{1}{|U|} \leqslant \frac{|U|}{16}\left(\frac{d}{4n}\right)^{2i-4} \leqslant \frac{|U|}{16}\left(\frac{d}{4n}\right)^{i-1}.$$

Note that $2d|U|/n \geqslant 2dm_i/n \geqslant m_{i-1}$. By the inductive hypothesis, the number of copies of K_i in $(G \setminus G')[U]$ is at most

$$\frac{1}{i}\left(|X_1| \cdot \frac{2^{(i-1)^2}}{(i-1)!}|U|^{i-1}\left(\frac{d}{n}\right)^{\binom{i-1}{2}} + |U| \cdot \frac{2^{(i-1)^2}}{(i-1)!}\left(\frac{2d|U|}{n}\right)^{i-1}\left(\frac{d}{n}\right)^{\binom{i-1}{2}}\right) \leqslant \frac{2^{i^2}}{i!}|U|^i\left(\frac{d}{n}\right)^{\binom{i}{2}},$$

https://doi.org/10.1017/S0963548320000577 Published online by Cambridge University Press

where the $1/i$ factor avoids counting i times each copy of K_i.

Similarly, let X_2 be the set of vertices $v \in U$ such that $\deg(v, U) \leqslant d|U|/(2n)$. By the definition of X_2 and Theorem 1.1, we have

$$\frac{d}{n}|U||X_2| - \lambda\sqrt{|U||X_2|} \leqslant e(U, X_2) \leqslant \frac{d}{2n}|U||X_2|.$$

This implies that $|X_2| \leqslant (2\lambda n/d)^2/|U| \leqslant |U|/4$. Note that $d|U|/(4n) \geqslant dm_i/(4n) = m_{i-1}$. By the inductive hypothesis, the number of copies of K_i in $(G \setminus G')[U]$ is at least

$$\frac{1}{i} \cdot (|U| - |X_2|)\frac{1}{2^{(i-1)^2}(i-1)!}\left(\frac{d|U|}{2n} - \frac{d|U|}{4n}\right)^{i-1}\left(\frac{d}{n}\right)^{\binom{i-1}{2}} \geqslant \frac{1}{2^{i^2}i!}|U|^i\left(\frac{d}{n}\right)^{\binom{i}{2}}$$

(the $1/i$ factor takes care of the fact that each copy of K_i is counted at most i times). □

We remark that a better estimate of the number of copies of K_i in m-vertex subgraphs of (n, d, λ)-graphs G, of the form $(1 + o(1))\binom{m}{i}(d/n)^{\binom{i}{2}}$, was obtained by Alon (see [20, Theorem 4.10]). Furthermore, a Turán-type result was proved by Sudakov, Szabó and Vu [25] for the containment of a copy of K_i in sufficiently dense subgraphs of (n, d, λ)-graphs. However, both results require a stronger condition on λ than the one we will have available in our applications of Proposition 3.6 (see Sections 4 and 5).

4. Fractional K_t-factors: the dense case

In this section we prove the following theorem.

Theorem 4.1. *For any integer $t \geqslant 3$ and $\beta \in (0, 1/(2t-3)]$, there exists n_0 such that every (n, d, λ)-graph G with $n \geqslant n_0$, $d \geqslant n^{1-1/(2t-3)+\beta}$ and $\lambda \leqslant (1/(20t \cdot 4^{t-2}))d^{t-1}/n^{t-2}$ contains $\ell = n^\beta$ fractional K_t-factors $f_1, \ldots, f_\ell$ such that*

$$\sum_{i=1}^{\ell} \sum_{T \in \mathcal{K}_t(G):\ V(T) \supset \{u,v\}} f_i(T) \leqslant 1 \quad \text{for every edge } uv \in E(G). \tag{4.1}$$

4.1 Proof idea

Our main idea for proving Theorem 4.1 is to view the (n, d, λ)-graph $G = (V, E)$ as a graph equipped with the weight function $w\colon E \to [0, 1]$. Once we manage to find a fractional K_t-factor f in (G, w) (by Corollary 3.4), we update the weight function w as follows:

$$w(uv) := w(uv) - \sum_{T \in \mathcal{K}_t(G):\ V(T) \supset \{u,v\}} f(T),$$

which remains non-negative by condition (3.3) from Section 3.2. Moreover, the weighted degree of every vertex decreases by *exactly* $t - 1$, since

$$\sum_{u \in N(v)} \sum_{T \in \mathcal{K}_t(G):\ V(T) \supset \{u,v\}} f(T) = (t-1) \sum_{T \in \mathcal{K}_t(G):\ V(T) \ni v} f(T) = t - 1.$$

Therefore, if in our graph (G, w) all weighted degrees were the same, then, after updating w, the weighted degrees would stay the same. To prove Theorem 4.1, it suffices to show that we can iterate this procedure $\ell = n^\beta$ times.

https://doi.org/10.1017/S0963548320000577 Published online by Cambridge University Press

4.2 Cliques in weighted subgraphs of (n, d, λ)-graphs

In the next two propositions we use Proposition 3.6 to derive the assumptions of Corollary 3.4. Recall that, given a graph G and $v \in V(G)$, we let $\mathcal{T}_v$ denote a family of copies of K_t in G with $v \in T$ for every $T \in \mathcal{T}_v$ and with $V(T) \cap V(T') = \{v\}$ for all distinct T and T' in $\mathcal{T}_v$.

Proposition 4.1. *For any integer $t \geqslant 3$, there exists n_0 such that every (n, d, λ)-graph G with $n \geqslant n_0$ and $\lambda \leqslant (1/(20t \cdot 4^{t-2}))d^{t-1}/n^{t-2}$ satisfies the following. Let G' be a graph on $V(G)$ with maximum degree at most $(d/(4n))^{t-2}d/(20t)$. Then, for every $v \in V(G)$, there exists a family $\mathcal{T}_v$ of at least $d/(2t-2)$ copies of K_t in $G \setminus G'$.*

Proof. Fix a vertex $v \in V(G)$ and let $U = N_{G\setminus G'}(v)$. Then $|U| \geqslant (1 - d/(80tn))d \geqslant 0.9d$. Let $\mathcal{T}_v$ be a collection of copies of K_t in $G \setminus G'$ such that $|\mathcal{T}_v| < d/(2t-2)$. Then

$$\left|U \setminus \bigcup_{T\in\mathcal{T}_v} V(T)\right| \geqslant 0.9d - d/2 \geqslant d/(20t).$$

By Proposition 3.6, $G[U \setminus \bigcup_{T\in\mathcal{T}_v} V(T)]$ contains a copy of K_{t-1} in $G \setminus G'$, which, together with v, gives a copy of K_t in $G \setminus G'$. The proposition follows. □

The following proposition establishes property $\mathcal{P}(t, D, D', n)$ from Section 3.3 in subgraphs of (n, d, λ)-graphs for certain values of the parameters.

Proposition 4.2. *For any integer $t \geqslant 3$, there exists n_0 such that every (n, d, λ)-graph G with $n \geqslant n_0$ and $\lambda \leqslant (1/(20t \cdot 4^{t-2}))d^{t-1}/n^{t-2}$ satisfies the following. Let G' be a graph on $V(G)$ with maximum degree at most $(d/(4n))^{t-2}d/(20t)$. Then every $0.11n/t$ vertices of $G \setminus G'$ span a copy of K_t in $G \setminus G'$. Moreover, $G \setminus G'$ has property $\mathcal{P}(t, d/2, 0.2n, n)$.*

Proof. We first prove that $G \setminus G'$ has property $\mathcal{P}(t, d/2, 0.2n, n)$. Set $c := 1/(20t \cdot 4^{t-2})$ so that $\lambda \leqslant cd^{t-1}/n^{t-2}$. Let $\mathcal{T}_0$ be a family of copies of K_t in $G \setminus G'$ of maximum cardinality satisfying properties (1)–(3) in the definition of $\mathcal{P}(t, d/2, 0.2n, n)$. If $|\mathcal{T}_0| < 0.2n/(t-1)$, then let

$$W := \left(\bigcup_{T\in\mathcal{T}_0} V(T)\right) \setminus U_0.$$

It follows that $|W| = (t-1)|\mathcal{T}_0| < 0.2n$. Because of the maximality of $\mathcal{T}_0$, there are no copies of K_t in $G \setminus G'$ with one vertex from U_0 and the other $t-1$ from $U \setminus (U_0 \cup W)$. Note that

$$|U \setminus (U_0 \cup W)| \geqslant n - d/2 - d/(2t) - 0.2n \geqslant n/(10t).$$

Take any subset $U' \subseteq U \setminus (U_0 \cup W)$ of cardinality $n/(10t)$ and note that, by Theorem 1.1, we have

$$e_G(U_0, U') \geqslant \frac{d}{n} \cdot \frac{n}{10t} \cdot \frac{d}{2t} - \lambda\sqrt{\frac{n}{10t} \cdot \frac{d}{2t}} \geqslant \frac{d^2}{20t^2} - c\frac{d^{t-1}}{n^{t-3}}\sqrt{\frac{d}{20t^2 n}} \geqslant \frac{d^2}{30t^2}.$$

Thus there is a vertex $v \in U_0$ of degree at least $d/(15t)$ into U' in G and hence v is connected to at least $d/(15t) - \Delta(G') \geqslant d/(15t) - d/(80t) \geqslant d/(20t)$ vertices in U' via edges in $G \setminus G'$. Let $R := N_{G\setminus G'}(v) \cap U'$. We have $|R| \geqslant d/(20t)$. By Proposition 3.6, $(G \setminus G')[R]$ contains a copy of K_{t-1}, which, together with v, gives a copy of K_t in $G \setminus G'$. This contradicts the maximality of $\mathcal{T}_0$. This shows that $|\mathcal{T}_0| \geqslant 0.2n/(t-1)$, as required.

It remains to prove that every set U of $0.11n/(t-1)$ vertices of $G \setminus G'$ spans a copy of K_t in $G \setminus G'$. This can be done in a similar way by showing that Theorem 1.1 implies that the average degree in $(G \setminus G')[U]$ is at least $7d/(80t)$ and then finding a copy of K_{t-1} in the neighbourhood of a vertex of maximum degree in $(G \setminus G')[U]$. We omit the details. □

We are now ready to prove Theorem 4.1.

https://doi.org/10.1017/S0963548320000577 Published online by Cambridge University Press

4.3 Proof of Theorem 4.1

We set $\alpha := (d/(4n))^{t-2}/(20t)$ and choose n_0 sufficiently large. We start with the graph G and at the beginning we set all edge weights to one, *i.e.* $w(e) := 1$ for all $e \in E(G)$. We shall iteratively apply Corollary 3.4 to find $\ell = n^{\beta}$ fractional K_t-factors $f_1, \dots, f_\ell$. In doing so we will iteratively update the weights of the edges in G. By Fact 3.5 and Propositions 4.1 and 4.2, the assumptions of Corollary 3.4 will be satisfied after each iteration, and we shall be able to find a fractional K_t-factor in the weighted graph at hand. Recall that the 'weighted degree' $\deg_w(v)$ of a vertex $v \in V = V(G)$ is defined to be $\sum_{u \in N(v)} w(uv)$. We observe that this degree is exactly d at the beginning (when $w \equiv 1$). Then we update the edge weights for all $uv \in E(G)$ in the ith iteration as follows:

$$w(uv) := w(uv) - \sum_{T \in \mathcal{K}_t(G):\ V(T) \supset \{u,v\}} f_i(T). \tag{4.2}$$

Observe that the weighted degree of *every* vertex decreased by $t-1$, since

$$\sum_{u \in N(v)} \sum_{T \in \mathcal{K}_t(G):\ V(T) \supset \{u,v\}} f_i(T) = (t-1) \sum_{T \in \mathcal{K}_t(G):\ V(T) \ni v} f_i(T) = t-1,$$

because f_i is a fractional K_t-factor. Now suppose $\sum_{u \in N(v)} w(uv) \geqslant (1-\alpha^2)d$ for any $v \in V$ throughout the process. Let G' be the graph on V consisting of edges of G that are not α-rich. By Fact 3.5, $\Delta(G') \leqslant \alpha d$. Thus, by Propositions 4.1 and 4.2, we can apply Corollary 3.4 with $D = d/2$ to (G, w) iteratively, updating the weights

$$\frac{d\alpha^2}{t-1} = \frac{d}{400t^2(t-1)}\left(\frac{d}{4n}\right)^{2t-4} = \frac{d^{2t-3}}{400t^2(t-1)(4n)^{2t-4}} \geqslant \frac{n^{(2t-3)\beta}}{400t^2(t-1)4^{2t-4}} \geqslant n^{\beta} = \ell$$

times. Because of our update rule (4.2), condition (4.1) holds for the f_i $(1 \leqslant i \leqslant \ell)$ that we have obtained. □

5. Fractional K_t-factors: the sparse case

In this section we prove the following theorem.

Theorem 5.1. *For any integer $t \geqslant 3$ and $\delta \in (0, 1/(2t-3))$, there exists n_0 such that every (n, d, λ)-graph G with $n \geqslant n_0$, $d \leqslant n^{1-\delta}$ and $\lambda \leqslant (1/(50t \cdot 4^{t-2}))d^{t-1}/n^{t-2}$ contains $\ell = n^{\delta/(4t^2)}$ fractional K_t-factors $f_1, \dots, f_\ell$ such that*

$$\sum_{i=1}^{\ell} \sum_{T \in \mathcal{K}_t(G):\ V(T) \supset \{u,v\}} f_i(T) \leqslant 1 \quad \textit{for every edge } uv \in E(G). \tag{5.1}$$

5.1 Proof overview

Our proof strategy this time will be to *partition* G randomly into $\ell = n^{\delta/(4t^2)}$ edge-disjoint subgraphs G_i, and then to show that, with high probability, each such random subgraph G_i satisfies the assumptions of Corollary 3.4 and thus contains a fractional K_t-factor f_i. Because the host graphs G_i of such f_i are edge-disjoint, condition (5.1) will hold because, for any $T \in \mathcal{K}_t(G)$, there is at most one $i \in [\ell]$ such that $f_i(T) > 0$. Note that, in this section, when we use results from Sections 3.2 and 3.3, we always use them on standard graphs, *i.e.* with $w \equiv 1$. In particular, here, every edge will be α-rich for any $\alpha \geqslant 0$. For the sake of definiteness, we always take $\alpha = 0$ in this section.

https://doi.org/10.1017/S0963548320000577 Published online by Cambridge University Press

5.2 Probabilistic lemmas

Recall that for a graph $G=(V,E)$ the random subgraph G_p of G is a spanning subgraph of G in which each edge from E is included with probability p, independently of all other edges. In this subsection we show that, for suitable (n,d,λ)-graphs G, with high probability G_p satisfies the assumptions of Corollary 3.4 and thus contains a fractional K_t-factor (Theorem 5.2 below). We first show that, with high probability, G_p satisfies assumption (i) of Corollary 3.4, with $D=p^{\binom{t}{2}}d/4$.

Proposition 5.1. *For any $t\geqslant 3$ and $\delta\in(0,1/(2t-3))$ there exists n_0 such that the following holds. Suppose G is an (n,d,λ)-graph with $n\geqslant n_0$, $d\leqslant n^{1-\delta}$ and $\lambda\leqslant(1/(20t\cdot 4^{t-2}))d^{t-1}/n^{t-2}$. Let $p=d^{-\eta}$ for some $\eta\in(0,1/t^2)$. Then, with probability at least $1-n\exp(-\sqrt{d})$, for any vertex $v\in V(G_p)$ there exists a family $\mathcal{T}_v$ in G_p with $|\mathcal{T}_v|\geqslant p^{\binom{t}{2}}d/(4t-4)$.*

Proof. For any $v\in V(G_p)=V(G)$, by Proposition 4.1, there is a family $\mathcal{T}'_v$ in G with $|\mathcal{T}'_v|=d/(2t-2)$. Let X be the number of cliques K_t from $\mathcal{T}'_v$ in G_p. Since the cliques in $\mathcal{T}_v$ are edge-disjoint, $X\sim\mathrm{Bin}(d/(2t-2),p^{\binom{t}{2}})$. By Chernoff's inequality (Theorem 3.2),

$$\mathbb{P}[X<\mathbb{E}X/2]<e^{-\mathbb{E}X/12}=\exp(-p^{\binom{t}{2}}d/(24t))\leqslant\exp(-d^{1-\binom{t}{2}\eta}/(24t))\leqslant\exp(-\sqrt{d}).$$

The union bound over all vertices yields that, with probability at least $1-n\exp(-\sqrt{d})$, every vertex v lies in some family $\mathcal{T}_v$ in G_p of at least $p^{\binom{t}{2}}d/(4t-4)$ copies of K_t. □

We now state and prove a technical lemma that will be required to show that G_p is very likely to have property $\mathcal{P}(t,D,D',n)$ for certain values of D and D'. The proof of this lemma is based on Janson's inequality (Theorem 3.3). However, to have 'weak enough dependence' when applying inequality (3.2), we shall have to employ an additional trick.

Lemma 5.2. *For any $t\geqslant 3$ and $\delta\in(0,1/(2t-3))$ there exists n_0 such that the following holds. Suppose G is an (n,d,λ)-graph with $n\geqslant n_0$, $d\leqslant n^{1-\delta}$ and $\lambda\leqslant(1/(50t\cdot 4^{t-2}))d^{t-1}/n^{t-2}$. Let $p=d^{-\eta}$ for some $\eta\in(0,\delta/(2t^2)]$. Then, with probability $1-2^{-n}$, for every pair of disjoint subsets U_0 and U' of $V(G)$ with $|U_0|=D=p^{\binom{t}{2}}d/(4t)$ and $|U'|=n/(10t)$, there is a clique K on t vertices in G_p with $|V(K)\cap U_0|=1$ and $|V(K)\cap U'|=t-1$.*

Proof. Let $c:=1/(50t\cdot 4^{t-2})$. We shall use Proposition 3.6 to show that there are sufficiently many cliques K_t in G with one vertex from U_0 and $t-1$ from U', so that after keeping each edge at random, the probability that none of the cliques survives is very small. Taking the union bound over all possible choices of U_0 and U' will then finish the proof. To estimate the survival probability of some clique, we shall apply Janson's inequality, Theorem 3.3. However, some special care needs to be taken, as otherwise the parameter Δ in (3.1) may turn out to be too large. Therefore we shall restrict our attention to certain 'nice' cliques between U_0 and U'.

Step 1 (good and bad vertices). For a given set U_0 we say that a vertex u from U' is *bad with respect to* U_0 if $\deg_{U_0}(u)\geqslant 2Dd/n$, and otherwise we say it is *good*. Let B be the set of bad vertices from U' with respect to U_0. Then $e(B,U_0)\geqslant 2|B|Dd/n$, whereas the expander mixing lemma (Theorem 1.1) asserts that

$$e(B,U_0)\leqslant|B|Dd/n+\lambda\sqrt{|B|D}\leqslant|B|Dd/n+cd^{t-1}\sqrt{|B|D}/n^{t-2}.$$

From this we infer that $|B|Dd/n\leqslant cd^{t-1}\sqrt{|B|D}/n^{t-2}$, and therefore

$$|B|\leqslant c^2d^{2t-4}/(Dn^{2t-6})\leqslant(4t)c^2d^{2t-5+\binom{t}{2}\eta}/n^{2t-6}<n/(30t),$$

https://doi.org/10.1017/S0963548320000577 Published online by Cambridge University Press

by the choice of $\eta < \delta/\binom{t}{2}$ and c. Thus the number of good vertices in U' is $|U' \setminus B| \geqslant n/(15t)$.

Next we estimate the number of edges $e_G(U_0, U' \setminus B)$ between U_0 and $U' \setminus B$ in G (we omit the subscript G whenever it is clear from the context). Note that $Dn/d^2 \geqslant n^\delta p^{\binom{t}{2}}/(4t) \geqslant n^\delta n^{-\eta t^2}/(4t) \geqslant 1$, because $\eta \leqslant \delta/(2t^2)$. By $|U' \setminus B| \geqslant n/(15t)$ and the expander mixing lemma, we have

$$e(U_0, U' \setminus B) \geqslant \frac{d}{n}\frac{Dn}{15t} - \lambda\sqrt{\frac{Dn}{15t}} \geqslant \frac{dD}{15t} - \frac{Dd}{60t}\sqrt{\frac{d^2}{15tDn}} \geqslant \frac{Dd}{20t},$$

where we used that $\lambda \leqslant cd^{t-1}/n^{t-2} \leqslant cd^2/n < d^2/(60tn)$ and $d^2/(Dn) \leqslant 1$. Thus, since $\Delta(G) \leqslant d$, we have at least $D/(40t)$ vertices in U_0 of degree at least $d/(40t)$ into $U' \setminus B$ in G.

Step 2 (expensive and inexpensive vertices). Given a vertex $u \in U_0$ with $\deg_{U' \setminus B}(u) \geqslant d/(40t)$, we call a vertex $v \in N(u) \cap (U' \setminus B)$ *expensive with respect to* u if $\deg_{N(u)\cap(U'\setminus B)}(v) \geqslant 4d^2/n$, and otherwise we call it *inexpensive*. Let R be the set of expensive vertices with respect to u. Then $e(R, N(u) \cap (U' \setminus B)) \geqslant 4d^2|R|/n$, whereas Theorem 1.1 asserts that

$$e(R, N(u) \cap (U' \setminus B)) \leqslant \deg_{U' \setminus B}(u)|R|\frac{d}{n} + c\frac{d^{t-1}}{n^{t-2}}\sqrt{\deg_{U' \setminus B}(u)|R|} \leqslant \frac{d^2|R|}{n} + \frac{cd^2}{n}\sqrt{d|R|},$$

where we used that $\deg_{U' \setminus B}(u) \leqslant d$. Thus we have $d^2|R|/n \leqslant cd^2\sqrt{d|R|}/n$ and therefore $|R| \leqslant c^2 d$.

Step 3 (nice cliques). We now introduce the notion of 'nice cliques' in G. Let K be a copy of K_t in G with $K \cap U_0 = \{u\}$ and $|K \cap (U' \setminus B)| = t - 1$. We call K *nice* if $\deg_{U' \setminus B}(u) \geqslant d/(40t)$, and any other vertex $v \in V(K) \setminus \{u\}$ is inexpensive with respect to u. Since there are at least $d/(50t)$ such inexpensive vertices in $N(u) \cap (U' \setminus B)$, we can deduce the following lower bound on the number of nice cliques in G. For $u \in U_0$ with $\deg_{U' \setminus B}(u) \geqslant d/(40t)$, let R_u be the set of inexpensive vertices with respect to u and notice that $|R_u| \geqslant d/(40t) - c^2 d \geqslant d/(50t)$. Let $\mathcal{C}$ be the family of nice cliques in G. Then we have

$$|\mathcal{C}| = \sum |\mathcal{K}_{t-1}(R_u)| \overset{\text{Proposition 3.6}}{\geqslant} \frac{D}{40t}\frac{(d/(50t))^{t-1}}{2^{(t-1)^2}(t-1)!}(d/n)^{\binom{t-1}{2}} \geqslant \frac{Dd^{t-1}}{2^{(t-1)^2}(50t)^t(t-1)!}(d/n)^{\binom{t-1}{2}},$$

where the sum is over all $u \in U_0$ such that $\deg_{U' \setminus B}(u) \geqslant d/(40t)$.

Next we aim to employ Janson's inequality to show that at least one of these nice cliques survives in G_p with 'sufficiently high' probability. Let X be the number of nice cliques that are contained in G_p. By the above bound on $|\mathcal{C}|$,

$$\mathbb{E}X = p^{\binom{t}{2}}|\mathcal{C}| \geqslant \frac{p^{\binom{t}{2}}Dd^{t-1}}{2^{(t-1)^2}(50t)^t(t-1)!}(d/n)^{\binom{t-1}{2}} \geqslant \frac{p^{t(t-1)}d^t}{2^{t^2}(50t)^t t!}(d/n)^{\binom{t-1}{2}}. \tag{5.2}$$

Step 4 (estimating parameter Δ in Janson's inequality). Towards the estimate of Δ we consider how two nice cliques can intersect in at least one edge (outside U_0 or inside both U_0 and U').

Let a clique K from $\mathcal{C}$ be given. First we estimate the number of nice cliques K' with $K \cap K' \cap U_0 = \emptyset$ and $|K \cap K' \cap U'| \geqslant 2$. Note that $|K' \cap U_0| = 1$ and denote this single vertex in $K' \cap U_0$ by v. We need to choose at least two vertices to lie in the intersection $K \cap K' \cap U'$, and these have to be connected to v. Since every vertex from $K \cap U'$ is good, we have at most $2Dd/n$ choices for v. Moreover, we need to specify a further $t - 3$ vertices from U' to belong to the copy of K' and these have to form a $(t-3)$-clique in $N(v)$. For $t \geqslant 5$, since $|N(v)| = d \geqslant (d/4n)^2 d$, by Proposition 3.6 with $m = d$ and $i = t - 3$, there are at most

$$\binom{t-1}{2} \cdot \frac{2Dd}{n} \cdot 2^{(t-3)^2} d^{t-3}(d/n)^{\binom{t-3}{2}} \leqslant (d^{t-1}/n)(d/n)^{\binom{t-3}{2}} \tag{5.3}$$

potential nice cliques K'. Note that the estimates in (5.3) also hold for $t = 3$ and 4.

https://doi.org/10.1017/S0963548320000577 Published online by Cambridge University Press

Next we estimate the number of nice cliques K' with $K \cap K' \cap U_0 \neq \emptyset$ and $|K \cap K' \cap U'| \geqslant 1$ for our fixed clique K from $\mathcal{C}$. We have $t-1$ choices for a common vertex x from $K \cap K' \cap U'$. Since x is inexpensive, its common neighbourhood with the vertex y from $K \cap K' \cap U_0$ is at most $4d^2/n$. It remains to estimate the number of possible extensions of x and y to a clique of size t. For this we would like to count the number of copies of K_{t-2} on a set of size $4d^2/n$. For $t \geqslant 4$, since $4d^2/n \geqslant d^2/n$, by Proposition 3.6 with $m = d$ and $i = t-2$, there are at most

$$(t-1) \cdot 2^{(t-2)^2} (4d^2/n)^{t-2} (d/n)^{\binom{t-2}{2}} \leqslant (d^{t-1}/n)(d/n)^{\binom{t-3}{2}} \tag{5.4}$$

potential nice cliques K'. Note that the estimates in (5.4) also hold for $t = 3$.

Step 5 (application of Janson's inequality). By (5.3) and (5.4), we get

$$\Delta \leqslant 2p^{\binom{t}{2}} |\mathcal{C}| (d^{t-1}/n)(d/n)^{\binom{t-3}{2}} = 2\mathbb{E}X (d^{t-1}/n)(d/n)^{\binom{t-3}{2}}.$$

Then Janson's inequality yields

$$\begin{aligned}
\mathbb{P}[X=0] &\leqslant \exp\left(-\frac{\mathbb{E}[X]^2}{2\Delta}\right) \\
&\leqslant \exp\left(-\frac{\mathbb{E}[X]}{4(d^{t-1}/n)(d/n)^{\binom{t-3}{2}}}\right) \\
&\overset{(5.2)}{\leqslant} \exp\left(-\frac{p^{t(t-1)} d^t (d/n)^{\binom{t-1}{2}}}{2^{t^2+2}(50t)^t t! \cdot (d^{t-1}/n) \cdot (d/n)^{\binom{t-3}{2}}}\right) \\
&= \exp\left(-\frac{p^{t(t-1)}}{2^{t^2+2}(50t)^t t!} n^2 (d/n)^{2t-4}\right).
\end{aligned}$$

Note that $p = d^{-\eta} \geqslant n^{-\eta} \geqslant n^{-\delta/(2t^2)}$. Together with $d \geqslant n^{1-1/(2t-3)}/2$ (Fact 3.6) and $\delta < 1/(2t-3)$, we get

$$\mathbb{P}[X=0] \leqslant \exp\left(-\frac{p^{t(t-1)}}{2^{t^2+2t-2}(50t)^t t!} n^{2-(2t-4)/(2t-3)}\right) \leqslant \exp\left(-n^{1+1/(4t-6)}\right).$$

Taking the union bound over all choices for U_0 and U' (of which there are at most 4^n), we obtain the desired claim with probability $1 - 4^n \exp\left(-n^{1+1/(4t-6)}\right) \geqslant 1 - 2^{-n}$ for n large enough. □

Now we show that it is very likely that G_p satisfies assumptions (ii) and (iii) of Corollary 3.4 with $D = p^{\binom{t}{2}} d/4$ and $\alpha = 0$.

Proposition 5.3. *For any $t \geqslant 3$ and $\delta \in (0, 1/(2t-3))$ there exists n_0 such that the following holds. Suppose G is an (n, d, λ)-graph with $n \geqslant n_0$, $d \leqslant n^{1-\delta}$ and $\lambda \leqslant (1/(50t \cdot 4^{t-2}))d^{t-1}/n^{t-2}$. Let $p = d^{-\eta}$ for some $\eta \in (0, \delta/(2t^2))$. Then, with probability $1 - 2^{-n}$, every $0.11n/t$ vertices of G_p span a copy of K_t and G_p has property $\mathcal{P}(t, p^{\binom{t}{2}} d/4, 0.2n, n)$.*

Proof. Let t and δ as in the statement be fixed. Let $n_0 = n_0(t, \delta)$ be given by Lemma 5.2. Then, for any $\eta \in (0, \delta/(2t^2))$, the conclusion in Lemma 5.2 holds with probability at least $1 - 2^{-n}$. Write $D = p^{\binom{t}{2}} d/4$.

Let $\mathcal{T}_0$ be a family of t-cliques in G_p of maximum cardinality satisfying properties (1)–(3) in the definition of $\mathcal{P}(t, p^{\binom{t}{2}} d/4, 0.2n, n)$ (see Section 3.3). If $|\mathcal{T}_0| < 0.2n/(t-1)$, then let

$$W := \left(\bigcup_{K \in \mathcal{T}_0} V(K)\right) \setminus U_0.$$

https://doi.org/10.1017/S0963548320000577 Published online by Cambridge University Press

It follows that $|W| = (t-1)|\mathcal{T}_0| < 0.2n$. Because of the maximality of $\mathcal{T}_0$, there are no cliques in G_p with one vertex from U_0 and the other $t-1$ vertices from $U' := U \setminus (U_0 \cup W)$. Observe that

$$|U'| \geqslant n - D - |U_0| - |W| \geqslant 0.5n.$$

But then, by the assertion of Lemma 5.2, there must be yet another copy of K_t in G_p between U_0 and U', contradicting the maximality of $\mathcal{T}_0$. Hence $|\mathcal{T}_0| \geqslant 0.2n/(t-1)$.

It remains to prove that every set U of $0.11n/t$ vertices of G_p spans a clique K_t. But this is immediate since $D < 0.01n/t$ and thus any two disjoint subsets of U of cardinality D and $0.1n/t$ span a copy of K_t. □

Next we show that G_p is very likely to contain a fractional K_t-factor.

Theorem 5.2. *For any $t \geqslant 3$ and $\delta \in (0, 1/(2t-3))$ there exists n_0 such that the following holds. Suppose G is an (n, d, λ)-graph with $n \geqslant n_0$, $d \leqslant n^{1-\delta}$ and $\lambda \leqslant (1/(50t \cdot 4^{t-2}))d^{t-1}/n^{t-2}$. Let $p = d^{-\eta}$ for some $\eta \in (0, \delta/(2t^2))$. Then G_p contains a fractional K_t-factor with probability at least $1 - n\exp(-\sqrt{d}) - 2^{-n}$.*

Proof. Let t and δ as in the statement be fixed. Let $n_0 = n_0(t, \delta)$ be given by Propositions 5.1 and 5.3 for the given parameters t and δ. Then, for any fixed $\eta \in (0, \delta/(2t^2)]$, the conclusions in Propositions 5.1 and 5.3 hold with probability at least $1 - n\exp(-\sqrt{d}) - 2^{-n}$. Let $D := p^{\binom{t}{2}}d/4$. The conclusion in our theorem follows from Corollary 3.4 (with $\alpha = 0$). □

We are now ready to prove Theorem 5.1.

5.3 Proof of Theorem 5.1

Let $p = d^{-\eta}$, where $\eta = \delta/(2t^2)$. Let $\ell := d^{\eta} = p^{-1}$. By Fact 3.6, we have $d \geqslant n^{1-1/(2t-3)}/2 \geqslant n^{1/2}$. Thus $\ell = d^{\eta} \geqslant n^{\delta/(4t^2)}$. Consider the random variable I that takes values uniformly at random in $[\ell] = \{1, \ldots, \ell\}$. For each $e \in E(G)$, let $I_e \sim I$ be an independent copy of I. We randomly partition the edge set of G into spanning subgraphs $G_1, \ldots, G_\ell$ of G, where each $e \in E(G)$ is put into G_{I_e}. Observe that each G_i is distributed as G_p. By Theorem 5.2, the probability that a given G_i should *not* contain a fractional K_t-factor is at most $n\exp(-\sqrt{d}) + 2^{-n}$. Since $d \geqslant n^{1-1/(2t-3)}/2$, it follows that, with probability $1 - \ell(n\exp(-\sqrt{d}) + 2^{-n}) = 1 - o(1) > 0$, there is a partition of G into deterministic edge-disjoint spanning subgraphs $G_1, \ldots, G_\ell$ such that each G_i ($i \in [\ell]$) possesses a fractional K_t-factor f_i. Since $G_1, \ldots, G_\ell$ are edge-disjoint, condition (5.1) holds because for any $T \in \mathcal{K}_t(G)$, there is at most one $i \in [\ell]$ such that $f_i(T) > 0$. □

6. Proof of Theorem 1.3

Let $t \geqslant 3$ be fixed and let G be an (n, d, λ)-graph as in the statement of Theorem 1.3. Let $\delta = 4t^2/((4t^2+1)(2t-3))$ and $\beta = 1/((4t^2+1)(2t-3))$, so that $n^{\beta} = n^{\delta/(4t^2)}$. Let $\ell = n^{\beta}$. By Theorems 4.1 and 5.1 (depending on whether $d \geqslant n^{1-1/(2t-3)+\beta}$ or $d \leqslant n^{1-\delta}$), our graph G contains fractional K_t-factors $f_1, \ldots, f_\ell$ such that, for every edge $uv \in E(G)$, we have

$$\sum_{i=1}^{\ell} \sum_{T \in \mathcal{K}_t(G):\ V(T) \supset \{u,v\}} f_i(T) \leqslant 1.$$

Clearly, $f(T) := \sum_i^{\ell} f_i(T) \leqslant 1$ for any $T \in \mathcal{K}_t(G)$. Let $H := \mathcal{K}_t(G)$ be the t-uniform hypergraph with copies of K_t in G as hyperedges and $V(H) = V(G)$. Thus, for every $v \in V = V(H)$ and $i \in [\ell]$, we

https://doi.org/10.1017/S0963548320000577 Published online by Cambridge University Press

have $\sum_{T\in E(H):\, v\in T} f_i(T) = 1$ and thus $\sum_{T\in E(H):\, v\in T} f(T) = \ell$. Moreover, for any distinct vertices u and v, we have

$$\sum_{T\in E(H):\, \{u,v\}\subseteq T} f(T) = \sum_{i\in[\ell]} \sum_{T\in E(H):\, \{u,v\}\subseteq T} f_i(T) \leqslant 1.$$

Let H_f be the random spanning subhypergraph of H such that each edge T of H is included in H_f independently with probability $f(T)$.

Next we let $k := 8\beta^{-1/2}$. For $v \in V$, let X_v be the degree of v in H_f. We have

$$\mathbb{E}X_v = \sum_{T\in E(H):\, v\in T} f(T) = \ell.$$

Then Chernoff's inequality (Theorem 3.2) implies that

$$\mathbb{P}[|X_v - \ell| > (k/2)\sqrt{\ell \ln \ell}] \leqslant 2e^{-k^2(\ln \ell)/12} < 2e^{-5\ln n} = 2n^{-5}.$$

On the other hand, for any distinct vertices u and $v \in V$, let Y_{uv} be the collective degree $\deg_{H_f}(u, v)$ of u and v in H_f. Then

$$\mathbb{E}Y_{uv} = \sum_{T\in E(H):\, \{u,v\}\subseteq T} f(T) \leqslant 1$$

(in fact $Y_{uv} = 0$ if $uv \notin E(G)$). By Chernoff's inequality again, we have

$$\mathbb{P}[Y_{uv} \geqslant 1 + 3\ln n] \leqslant e^{-3\ln n} = n^{-3}.$$

The union bound over all $u \in V$ and all pairs $\{u, v\} \in \binom{V}{2}$ yields that, with probability at least $1 - 1/n$, we have

$$\deg_{H_f}(u) = \ell \pm (k/2)\sqrt{\ell \ln \ell} \quad \text{and} \quad \deg_{H_f}(u, v) \leqslant 1 + 3\ln n$$

for all u and $v \in V$.

It is easy to check now that the assumptions of Theorem 3.1 are satisfied with $\delta' = 1/t$, $\gamma = 0.9$, $D = (1 + (k/2)\sqrt{(\ln \ell)/\ell})\ell \leqslant 2n^{\beta}$ and $C = 1 + 3\ln n$. Thus an application of that theorem gives us a matching in H_f covering all but at most

$$O(n(C/D)^{(1-\delta')/(t-1)}) = O\left(n\left(\frac{1+3\ln n}{2n^{\beta}}\right)^{1/t}\right) \leqslant n^{1-1/(8t^4)}$$

vertices. These edges correspond to cliques K_t in the original (n, d, λ)-graph G and thus they correspond to a collection of vertex-disjoint t-cliques covering all but at most $n^{1-1/(8t^4)}$ vertices of the graph. This completes the proof of Theorem 1.3. □

7. Concluding remarks

In this paper we have studied near-perfect K_t-factors in sparse (n, d, λ)-graphs. We have presented two different approaches for finding many 'weight-disjoint' fractional K_t-factors: one for 'large' d (Theorem 4.1) and one for 'small' d (Theorem 5.1).

We believe that the first approach is more powerful, since it can be extended to the whole range of d for all $t \geqslant 4$ as follows. In the proof of Theorem 4.1 we used Proposition 3.6 to embed a copy of K_{t-1} in a set of $\Omega(d)$ vertices, even after removing a small number of edges from each vertex, say $o((d/n)^{t-2}d)$. We can even allow removal from each vertex of $o(d^2/n)$ edges if we just ask for the existence of a single copy of K_{t-1}, which would suffice for our first approach. To establish this, we could have used a Turán-type result in (n, d, λ)-graphs due to Sudakov, Szabó and Vu [25, Theorem 3.1]. However, a drawback is that the pseudorandomness condition in

https://doi.org/10.1017/S0963548320000577 Published online by Cambridge University Press

[25, Theorem 3.1] is not numerically explicit (it was stated as $\lambda \ll d^{t-1}/n^{t-2}$). Thus we chose to present a self-contained and numerically explicit proof. Moreover, the first approach seems not to provide Theorem 1.3 for triangle-factors in the whole range, while triangle-factors can be considered as the most important testbed at the moment.

As for the second approach, we observe that it can be pushed to work in a range of d from $\Omega(n^{1-1/(2t-3)})$ to $n/\operatorname{polylog} n$, although near the upper bound we would only be able to assert that one can cover all but $n/\operatorname{polylog} n$ vertices with vertex-disjoint copies of K_t. Also, for the second approach one could probably replace an application of our Corollary 3.4 in Section 5 with a recent result of Devlin and Kahn [10, Proposition 4].[d] However, the merit of Corollary 3.4 is that it applies to a more general weighted scenario and is also indispensable for the more powerful first approach as discussed above.

Recent developments

When finalizing this paper, we learned that very recently Nenadov [23] proved the existence of K_t-factors in (n, d, λ)-graphs with $\lambda \leqslant \varepsilon d^{t-1}/(n^{t-2}\log n)$ for a suitably small $\varepsilon = \varepsilon_t > 0$, which is at most a $\log n$ factor away from Conjecture 1.2. Independently, Morris and the authors of this paper [13] proved a similar result with $\lambda \leqslant \varepsilon_t d^t/n^{t-1}$, which is a stronger requirement, except when $d \geqslant c_t n/\log n$ for a suitable constant $c_t > 0$.

An n-vertex (λ, p)-bijumbled graph is a graph that satisfies inequality (1.1) of the expander mixing lemma (Theorem 1.1). This notion is slightly more general: any (n, d, λ)-graph is $(\lambda, d/n)$-bijumbled, but in a bijumbled graph not every vertex has necessarily degree d. The result of [23] applies to $(\varepsilon_t p^2 n/\log n, p)$-bijumbled graphs with minimum degree $\Omega(pn)$. The result of the present paper as well as the one in [13] can be easily shown to hold in appropriately bijumbled graphs as well.

References

[1] Allen, P., Böttcher, J., Hàn, H., Kohayakawa, Y. and Person, Y. (2016) Blow-up lemmas for sparse graphs. arXiv:1612.00622

[2] Allen, P., Böttcher, J., Hàn, H., Kohayakawa, Y. and Person, Y. (2017) Powers of Hamilton cycles in pseudorandom graphs. *Combinatorica* **37** 573–616.

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

[4] Alon, N. and Chung, F. R. K. (1988) Explicit construction of linear sized tolerant networks. *Discrete Math.* **72** 15–19.

[5] Alon, N., Frankl, P., Huang, H., Rödl, V., Ruciński, A. and Sudakov, B. (2012) Large matchings in uniform hypergraphs and the conjectures of Erdős and Samuels. *J. Combin. Theory Ser. A* **119** 1200–1215.

[6] Alon, N. and Kahale, N. (1998) Approximating the independence number via the ϑ-function. *Math. Program.* **80** 253–264.

[7] Alon, N. and Spencer, J. H. (2016) *The Probabilistic Method*, fourth edition. Wiley.

[8] Bollobás, B. (1998) *Modern Graph Theory*, Vol. 184 of Graduate Texts in Mathematics. Springer.

[9] Conlon, D., Fox, J. and Zhao, Y. (2014) Extremal results in sparse pseudorandom graphs. *Adv. Math.* **256** 206–290.

[10] Devlin, P. and Kahn, J. (2017) Perfect fractional matchings in k-out hypergraphs. *Electron. J. Combin.* **24** P3.60.

[11] Füredi, Z. (1988) Matchings and covers in hypergraphs. *Graphs Combin.* **4** 115–206.

[12] Han, J. (2020) On perfect matchings in k-complexes. *Int. Math. Res. Not.* rnz343.

[13] Han, J., Kohayakawa, Y., Morris, P. and Person, Y. (2019) Clique-factors in sparse pseudorandom graphs. *European J. Combin.* **82** 102999.

[14] Han, J., Kohayakawa, Y. and Person, Y. (2018) Near-perfect clique-factors in sparse pseudorandom graphs. In *Discrete Mathematics Days 2018: Extended abstracts of the 11th 'Jornadas de matemática discreta y algorítmica' (JMDA), Seville 2018*, pp. 221–226. Elsevier.

[15] Janson, S., Łuczak, T. and Ruciński, A. (2000) *Random Graphs*. Wiley-Interscience.

[16] Keevash, P. and Mycroft, R. (2015) *A Geometric Theory for Hypergraph Matching*, Vol. 233, no. 1098 of Memoirs of the American Mathematical Society. AMS.

[d] This was pointed out to us by an anonymous referee.

https://doi.org/10.1017/S0963548320000577 Published online by Cambridge University Press

[17] Komlós, J., Sárközy, G. N. and Szemerédi, E. (1997) Blow-up lemma. *Combinatorica* **17** 109–123.
[18] Kostochka, A. V. and Rödl, V. (1998) Partial Steiner systems and matchings in hypergraphs. *Random Struct. Algorithms* **13** 335–347.
[19] Krivelevich, M. (1996) Perfect fractional matchings in random hypergraphs. *Random Struct. Algorithms* **9** 317–334.
[20] Krivelevich, M. and Sudakov, B. (2006) Pseudo-random graphs. In *More Sets, Graphs and Numbers*, Vol. 15 of Bolyai Society Mathematical Studies, pp. 199–262. Springer.
[21] Krivelevich, M., Sudakov, B. and Szabó, T. (2004) Triangle factors in sparse pseudo-random graphs. *Combinatorica* **24** 403–426.
[22] Matoušek, J. and Gärtner, B. (2007) *Understanding and Using Linear Programming*. Springer.
[23] Nenadov, R. (2019) Triangle-factors in pseudorandom graphs. *Bull. Lond. Math. Soc.* **51** 421–430.
[24] Rödl, V. (1985) On a packing and covering problem. *European J. Combin.* **6** 69–78.
[25] Sudakov, B., Szabó, T. and Vu, V. (2005) A generalization of Turán's theorem. *J. Graph Theory* **49** 187–195.

Cite this article: Han J, Kohayakawa Y and Person Y (2021). Near-perfect clique-factors in sparse pseudorandom graphs. *Combinatorics, Probability and Computing* **30**, 570–590. https://doi.org/10.1017/S0963548320000577

https://doi.org/10.1017/S0963548320000577 Published online by Cambridge University Press