

PERFECT MATCHINGS IN RANDOM SPARSIFICATIONS OF DENSE HYPERGRAPHS

JIE HAN AND JINGWEN ZHAO

ABSTRACT. The decision problem of perfect matchings in uniform hypergraphs is famously an NP-complete problem. It has been shown by Keevash–Knox–Mycroft [STOC, 2013] that for every $\varepsilon > 0$, such decision problem restricted to k -uniform hypergraphs H satisfying that every $(k-1)$ -set of vertices is in at least $(1/k + \varepsilon)|H|$ edges is tractable, and the quantity $1/k$ is best possible. In this paper we study the existence of perfect matchings in the random p -sparsification of such k -uniform hypergraphs, that is, for $p = p(n) \in [0, 1]$, every edge is kept with probability p independent of others. As a consequence, we give a polynomial-time algorithm that with high probability solves the decision problem; we also derive effective bounds on the number of perfect matchings in such hypergraphs. At last, similar results are obtained for the F -factor problem in graphs.

The key ingredients of the proofs are a strengthened partition lemma for the lattice-based absorption method, and the random redistribution method developed recently by Kelly, Mütters and Pokrovskiy, based on the spread method.

1. INTRODUCTION

Matchings are fundamental objects in Graph Theory and have broad applications in other branches of Science and a variety of practical problems (e.g. the assignment of graduating medical students to their first hospital appointments¹). Applications of matchings in hypergraphs include the ‘Santa Claus’ allocation problem [?]; they also offer a universal framework for many important combinatorial problems, e.g. the Existence Conjecture for designs (see [?, ?]) and Ryser’s conjecture [?] on transversals in Latin squares.

Given $k \geq 2$, a k -uniform hypergraph (or simply a k -graph) is a hypergraph $H = (V(H), E(H))$ in which every edge $e \in E(H)$ has size exactly k . A *matching* M in H is a set of vertex-disjoint edges. A *perfect matching* in H is a matching that covers all the vertices of H .

The question of determining the existence of perfect matchings in k -uniform hypergraphs is a fundamental question in combinatorics. Several well-known conjectures and problems, such as Ryser’s conjecture and the existence of certain combinatorial designs, can be reduced to this type of problems. When $k = 2$, Tutte [?] provided a complete characterization of graphs that contain a perfect matching in 1947, and Edmonds [?] gave a polynomial-time algorithm to determine whether a given graph contains one. However, for $k \geq 3$, the problem becomes NP-complete, as shown by Karp [?]. Naturally, we aim to find sufficient conditions that guarantee the existence of perfect matchings. Among these, the most extensively studied is the minimum degree condition, commonly referred to as the Dirac-type problem.

As an important generalization of perfect matchings, the concept of F -factors has received considerable attention in graph and hypergraph theory. Given two k -graphs F and H , an F -packing of H is a collection of vertex-disjoint copies of F in H . An F -factor, or *perfect F -packing*, is an F -packing that covers all vertices of H . Hell and Kirkpatrick [?] showed that determining whether a graph G contains a perfect F -packing is NP-complete when F has a component with at least three vertices.

¹In 2012, the Nobel Memorial Prize in Economics was awarded to Shapley and Roth “for the theory of stable allocations and the practice of market design.”

Besides studying the containment of an F -factor, it is also interesting to ask how stable this property is. In other words, does the F -factor still exist if the graph is slightly changed? This leads to the study of the robustness of graph properties. There are several measures of robustness that were proposed so far. For example, one can measure the robustness of Dirac graphs (i.e., graphs G with minimum degree at least $|G|/2$) with respect to perfect matchings by computing the number of perfect matchings that a Dirac graph must contain. Indeed, confirming a conjecture of Sárközy, Selkow, and Szemerédi[?], Cuckler and Kahn [?] proved that every Dirac graph contains at least $n!/(2 + o(1))^n$ perfect matchings.

Another measure is the so called resilience. Roughly speaking, for monotone increasing graph properties, the resilience quantifies the robustness in terms of the number of edges one must delete from H , locally or globally, in order to destroy the property \mathcal{P} . For example, Krivelevich, Lee and Sudakov [?] proved that for any Dirac graph H , the random subgraph obtained by keeping each edge of H independently with probability p where $p \geq C \log n/n$ (C is an absolute constant) *asymptotically almost surely* (abbreviated a.a.s.) contains a perfect matching.

In this work, we focus on the robustness of degree conditions that guarantee perfect matchings or F -packings in hypergraphs. While such conditions have traditionally been studied from a structural or algorithmic perspective (particularly in the context of making the decision problem polynomial-time solvable), we show that many of them also exhibit a form of robustness. That is, they continue to guarantee the desired structures even under random perturbations, such as when passing to a random subgraph. In particular, we identify the two settings in which the same minimum degree thresholds that make the problem tractable also lead to robust containment in random subgraphs.

1.1. Perfect matchings in hypergraphs. Given a k -graph H and a subset $S \subseteq V(H)$ of size d , the *degree* of S in H , denoted by $\deg_H(S)$, is the number of edges in H that contain S . The minimum d -degree $\delta_d(H)$ is the minimum of degree over all d -sets of $V(H)$.

One of the central questions in the field is to determine the minimum d -degree threshold that guarantees the existence of a perfect matching. In a landmark result, Rödl, Ruciński, and Szemerédi [?] showed that for all $k \geq 3$, $d = k - 1$, and sufficiently large n , this threshold is $n/2 - k + C$, where $C \in \{3/2, 2, 5/2, 3\}$ depends on the values of n and k . Naturally, one may ask whether the codegree condition can be relaxed while still allowing for a polynomial-time algorithm to decide whether a given k -graph H contains a perfect matching. To formalize this, let $\mathbf{PM}(k, \delta)$ be the decision problem of determining whether a k -graph on $n \in k\mathbb{N}$ vertices with $\delta_{k-1}(H) \geq \delta n$ contains a perfect matching. It follows immediately from the result of Rödl, Ruciński, and Szemerédi [?] that $\mathbf{PM}(k, 1/2)$ is in P. On the other hand, Szymańska [?] showed that for any $\delta < 1/k$, the problem $\mathbf{PM}(k, \delta)$ is NP-complete, via a polynomial-time reduction from $\mathbf{PM}(k, 0)$. Karpiński, Ruciński and Szymańska [?] further proved that $1/2$ is not the threshold for tractability: there exists an $\varepsilon > 0$ such that $\mathbf{PM}(k, 1/2 - \varepsilon)$ is in P. Later, Keevash, Knox and Mycroft [?] showed that $\mathbf{PM}(k, \delta)$ is in P for any $\delta > 1/k$.

Theorem 1.1 ([?]). *Fix $k \geq 3$ and $\gamma > 0$. Let H be an n -vertex k -graph with $\delta_{k-1}(H) \geq n/k + \gamma n$. Then there is an algorithm with running time $O(n^{3k^2-8k})$, which determines whether H contains a perfect matching.*

They indeed proved a stronger result that provides a polynomial-time algorithm that outputs either a perfect matching or a certificate that none exists. Improvements on the minimum degree conditions were made in subsequent works [?, ?]. In this paper, we strengthen this line of work by establishing a robust version of the result. For any hypergraph H and $0 \leq p \leq 1$, let H_p be a spanning random subhypergraph of H obtained by choosing each edge $e \in H$ with probability p independently at random.

Theorem 1.2 (Main result). *Fix $k \geq 3$ and $\gamma > 0$, there exists $C = C_{??}(k, \gamma) > 0$ such that the following holds for all sufficiently large n . Let H be an n -vertex k -graph with $\delta_{k-1}(H) \geq n/k + \gamma n$.*

If $p \geq C \log n / n^{k-1}$, then there exists a deterministic algorithm with running time $O(n^{2^{k-2}k+1})$ such that:

- If the algorithm accepts, then a.a.s. H_p contains a perfect matching;
- If the algorithm rejects, then H_p does not contain a perfect matching.

This result establishes a stronger, probabilistic guarantee concerning rather sparse random subgraphs of the original k -graph. The assumptions are essentially best possible: if $p = o(\log n / n^{k-1})$ then H_p a.a.s. contain no perfect matching even for complete k -graph H ; if $\delta_{k-1}(H) \leq n/k - \gamma n$ then as mentioned above, Szymańska's result implies that the decision problem for perfect matching in H is NP-complete.

A natural and curious question is that given an instance of H that has a perfect matching, what is the probability that the algorithm in Theorem ?? accepts H_p . We shall address this question after we introduce our algorithm Procedure ?? in Section 2.1.

Now we turn to perfect F -packings in graphs and state our second main result, Theorem ??, after some introduction and preparation.

1.2. Perfect packings in graphs. The Hajnal-Szemerédi theorem states that every n -vertex graph with minimum degree at least $(1 - 1/r)n$ and $r \mid n$ contains a K_r -factor, and this bound is tight. More generally, Alon and Yuster [?] proved the following theorem. Given a graph F , let v_F denote its number of vertices, e_F its number of edges, and $\chi(F)$ its chromatic number. Let $M(n)$ be the time needed to multiply two n by n matrices with 0, 1 entries. Determining $M(n)$ is a challenging problem in theoretic computer science, and the best known bound of $M(n) = O(n^{2.3728596})$ was obtained recently by Alman and Williams [?].

Theorem 1.3 ([?]). *For every $\gamma > 0$ and each graph F there exists an integer $n_0 = n_0(\gamma, F)$ such that every graph G whose order $n > n_0$ is divisible by v_F and whose minimum degree is at least $(1 - 1/\chi(F) + \gamma)n$ contains an F -factor. Moreover, there is an algorithm which finds such F -factor in time $O(M(n))$.*

Alon and Yuster also provided examples showing that the error term γn cannot be completely omitted for some graphs, but they conjectured that it could be replaced by a constant depending only on H . This conjecture was later proved by Komlós, Sárközy and Szemerédi [?].

On the other hand, there are graphs for which the bound on the minimum degree can be significantly improved by replacing the chromatic number with a refined parameter known as the *critical chromatic number* [?]. The *critical chromatic number* $\chi_{cr}(F)$ of a graph F is defined as $(\chi(F) - 1)v_F / (v_F - \sigma(F))$, where $\sigma(F)$ denotes the minimum size of the smallest colour class in a colouring of F with $\chi(F)$ colours. Note that $\chi(F) - 1 < \chi_{cr}(F) \leq \chi(F)$ and the equality holds if and only if every $\chi(F)$ -colouring of F has equal colour class sizes. Resolving a conjecture of Komlós, Shokoufandeh and Zhao [?] showed the following result (stated here in a slightly weaker form).

Theorem 1.4 ([?]). *For any F there is an $n_0 = n_0(F)$ so that if G is a graph on $n \geq n_0$ vertices and minimum degree at least $(1 - 1/\chi_{cr}(F))n$, then G contains an F -packing that covers all but at most $5v_F^2$ vertices.*

Given the two types of minimum degree thresholds discussed above, it is natural to consider the corresponding decision problem. Let **Pack**(F, δ) denote the problem of determining whether a graph G with minimum degree $\delta(G) \geq \delta |G|$ contains an F -factor. As mentioned earlier, when F contains a component with at least three vertices, results of Hell and Kirkpatrick [?] show that **Pack**($F, 0$) is NP-complete. More generally, Kühn and Osthus [?] proved that **Pack**(F, δ) is NP-complete for any $\delta \in [0, 1 - 1/\chi_{cr}(F))$, provided that F is either a clique of size at least 3 or a complete k -partite graph where $k \geq 2$ and the second smallest vertex class has size at least 2. In contrast, Theorem ?? implies that **Pack**(F, δ) is in P for all $\delta \in (1 - 1/\chi(F), 1]$. Recently, Han

and Treglown [?] provided a polynomial-time algorithm showing that $\mathbf{Pack}(F, \delta)$ is in \mathbf{P} whenever $\delta \in (1 - 1/\chi_{cr}(F), 1]$, answering a question of Yuster negatively.

Theorem 1.5 ([?]). *Let F be an r -vertex k -chromatic graph. For every n -vertex graph G with minimum degree at least δn , where $\delta \in (1 - 1/\chi_{cr}(F), 1]$, there is an algorithm with running time $O\left(n^{\max\{2^{r^{k-1}-1}r+1, r(2r-1)^r\}}\right)$, which determines whether G contains a perfect F -packing.*

Theorem ?? can efficiently detect the divisibility obstructions when the minimum degree condition guarantees an F -packing that covers all but a constant number of vertices (as established in Theorem ??). Similar to the perfect matching problem, it is natural to seek for a robust version of Theorem ?? similar to Theorem ??. Unfortunately, we are only able to prove partial result which we state below after some preparations.

1.3. Partition, lattices and new result on F -packings. The proofs in [?] and [?] used a lattice-based absorbing method, which combines the absorbing technique with the concept of divisibility barriers, introduced in [?]. It was shown in [?] that a k -graph H either contains a perfect matching or is structurally close to a family of lattice-based constructions known as divisibility barriers. To describe divisibility barriers in general, we begin with the following definition. In this paper, every partition has an implicit order on its parts.

Definition 1.6 (Partition, index vector and lattice). Let $H = (V, E)$ be an n -vertex k -graph, and let $\mathcal{P} = \{V_1, \dots, V_d\}$ be a partition of V . Let F be an r -vertex k -graph. The *index vector* $\mathbf{i}_{\mathcal{P}}(S) \in \mathbb{Z}^d$ of a subset $S \subseteq V$ with respect to \mathcal{P} is the vector whose coordinates are the size of intersection of S with each part of \mathcal{P} , namely, $\mathbf{i}_{\mathcal{P}}(S)|_i = |S \cap V_i|$ for $i \in [d]$, where $\mathbf{v}|_i$ is defined as the i th digit of \mathbf{v} . Then for any $\mu > 0$,

- (L1) $I_{\mathcal{P}, F}^{\mu}(H)$ denotes the set of $\mathbf{i} \in \mathbb{Z}^d$ such that H contains at least μn^r copies of F with index vector \mathbf{i} ;
- (L2) $L_{\mathcal{P}, F}^{\mu}(H)$ denotes the lattice (that is, the additive subgroup) in \mathbb{Z}^d generated $I_{\mathcal{P}, F}^{\mu}(H)$.

We call a copy of F μ -robust if its index vector $\mathbf{i} \in I_{\mathcal{P}}^{\mu}(H)$. For each vertex $v \in V(H)$, let $F_{\mathcal{P}}^{\mu}(v)$ be the collection of r -sets that contain v and span at least one μ -robust copy of F . When F is a single edge, we abbreviate these as $I_{\mathcal{P}}^{\mu}(H)$, $L_{\mathcal{P}}^{\mu}(H)$, and $E_{\mathcal{P}}^{\mu}(v)$, respectively.

Now we give some basic definitions for the absorbing method. We use the reachability arguments introduced by Lo and Markström [?]. We say that two vertices u and v in $V(H)$ are (F, β, i) -reachable in H if there are at least βn^{ir-1} $(ir-1)$ -sets S such that both $H[S \cup \{u\}]$ and $H[S \cup \{v\}]$ have perfect F -packings. We refer to such a set S as a *reachable $(ir-1)$ -set for u and v* . We say a vertex set $U \subseteq V(H)$ is (F, β, i) -closed in H if any two vertices $u, v \in U$ are (F, β, i) -reachable in H . In a similar way, we define a set to be (β, i) -closed when F is the single-edge k -graph.

Now we state a strengthened version of the standard partition lemma (such as Lemmas ?? and ??). While previous versions guarantee two structural properties, namely ?? and ??, our version additionally establishes condition ??. This refinement is not required in earlier applications, but it becomes crucial for proving our robustness result. Throughout the paper, we write $\alpha \ll \beta \ll \gamma$ to mean that there exist increasing functions f and g such that, given γ , any choice of $\beta \leq f(\gamma)$ and $\alpha \leq f(\beta)$ ensures the validity of all calculations in the proof. Moreover, when we have variable of form $1/C$ in the hierarchy, C is always assumed to be a positive integer.

Lemma 1.7 (Partition lemma for F -factors). *Let $1/n \ll \beta \ll \mu \ll \varepsilon \ll \gamma \ll 1/r, 1/k$. Let F be an r -vertex k -chromatic graph and $h := r^{k-1}$. For each n -vertex graph with $\delta(G) \geq (1 - 1/\chi_{cr}(F) + \gamma)n$, we find a partition $\mathcal{P} = \{V_1, \dots, V_d\}$ of $V(G)$ in time $O(n^{2^{h-1}k+1})$ satisfying the following properties:*

- (F1) each part of \mathcal{P} has size at least $(1/r + \gamma/3)n$;
- (F2) for each $i \in [d]$, V_i is $(F, \beta, r2^h + 2)$ -closed in G ;
- (F3) for each vertex $v \in V(G)$, $|F_{\mathcal{P}}^{\mu}(v)| \geq \varepsilon n^{r-1}$.

Now we are ready to state our result for F -factors. For a graph G with at least two vertices, define $d_1(G) := e_G/(v_G - 1)$ and $m_1(G) := \max_{G' \subseteq G: v_{G'} > 1} d_1(G')$. A graph F is *strictly 1-balanced* if $d_1(F') < d_1(F)$ for every $F' \subsetneq F$.

Theorem 1.8 (Robustness of F -factors). *Let $k, r, n \geq 2$ be integers and $\varepsilon, \mu, \beta, \gamma > 0$ where*

$$1/n \ll \beta \ll \mu \ll \varepsilon \ll \gamma \ll 1/r, 1/k.$$

There exists $C = C_{??}(k, r, \gamma) > 0$ such that the following holds. Let F be an r -vertex k -chromatic graph, and G an n -vertex graph such that $\delta(G) \geq (1 - 1/\chi_{cr}(F) + \gamma)n$ with $r \mid n$ and \mathcal{P} found by Lemma ???. If there exists an F -packing M_0 of size at most $\binom{2r-2}{r}$ such that $\mathbf{i}_{\mathcal{P}}(V(G) \setminus V(M_0)) \in L_{\mathcal{P}, F}^{\mu}(G)$, and let $G' := G - V(M_0)$, then a.a.s. G'_p contains an F -factor provided that $p \geq Cn^{-1/m_1(F)} \log n$. In particular, if F is strictly 1-balanced, then a.a.s. G'_p contains an F -factor provided $p \geq Cn^{-1/m_1(F)} \log^{1/e_F} n$. Otherwise, if no such F -packing M_0 exists, then G has no F -factor.

A special case of Theorem ?? is when $M_0 = \emptyset$, and then it says that the p -random sparsification of G a.a.s. contains an F -factor (see Theorem ?? for a stronger result). For the general case, obviously the existence of such M_0 can be tested in time $O(n^{r \binom{2r-2}{r}})$. The difficulty for upgrading Theorem ?? to a result similar to Theorem ?? is that checking the existence of such M_0 may destroy the randomness of G_p and we shall expand on this in the remark section.

For strictly 1-balanced F , Johansson, Kahn, and Vu [?] proved that the threshold for the appearance of an F -factor in the binomial random graph is $\Theta\left(n^{-1/m_1(F)} \log^{1/e_F} n\right)$. Hence, the assumptions on p in Theorem ?? is optimal up to a constant factor for strictly 1-balanced F .

1.4. Spread method and the main technical result. To prove Theorems ?? and ??, we use the so-called *spread* method, which was introduced by Talagrand and used by Frankston, Kahn, Narayanan, and Park [?] to resolve a conjecture of Talagrand. A key contribution of Talagrand's conjecture is the established connection between so-called *spread measure* and threshold in random graph theory. As a consequence, the task of establishing a robustness result reduces to proving the existence of a probability measure with good spread over the desired substructures. Roughly speaking, this means that the probability measure selects edges of \mathcal{H} such that no specific subset of edges is overly likely to appear in an edge of \mathcal{H} . In other words, the distribution avoids concentrating too much weight on any particular edge set. This notion is formalized as follows.

Definition 1.9. Let $q \in [0, 1]$. Let \mathcal{H} be a hypergraph on vertex set V , and let μ be a probability distribution on \mathcal{H} . We say that μ is q -spread if for every set $S \subseteq V$:

$$\mu(S) := \mu(\{A \in \mathcal{H} : S \subseteq A\}) \leq q^{|S|}.$$

In our context, we are primarily concerned with such hypergraph \mathcal{H} where $V := E(H)$ for some host hypergraph H , and \mathcal{H} denotes the collection of F -factors in H .

Now we are ready to state our main technical result, which is a general structural theorem that serves as a robust version of the main result (Theorem ??) from [?], strengthened via the spread-based framework described earlier. Let $k, \ell \in \mathbb{N}$ where $\ell \leq k - 1$. Let F be an r -vertex k -graph and $D \in \mathbb{N}$. Define $\delta(F, \ell, D)$ to be the smallest number δ such that every k -graph H on n vertices with $\delta_{\ell}(H) \geq (\delta + o(1)) \binom{n-\ell}{k-\ell}$ contains an F -packing covering all but at most D vertices. We write $\delta(k, \ell, D)$ for $\delta(F, \ell, D)$ when F is a single edge.

Theorem 1.10 (Main technical result). *Let $k, \ell \in \mathbb{N}$ and let F be an r -vertex k -graph. Define $C'', t, n \in \mathbb{N}$ and $\alpha, \beta, \mu, \gamma, \eta, \varepsilon, c > 0$ where*

$$\alpha, \eta \ll 1/C'' \ll \beta, \mu \ll \gamma, \varepsilon, c, 1/k, 1/\ell, 1/r, 1/t.$$

Then the following holds for sufficiently large $n \in r\mathbb{N}$. Let H be a k -graph on $n \geq n_0$ vertices, and let $\mathcal{P} = \{V_1, \dots, V_d\}$ be a partition of $V(H)$. Suppose that

- (i) for all but at most $\alpha \binom{n}{\ell}$ ℓ -sets S , $\deg_H(S) \geq (\delta(F, \ell, 5r^2) + \gamma) \binom{n-\ell}{k-\ell}$;
- (ii) for every $i \in [d]$, all but $\eta \binom{|V_i|}{2}$ of the pairs $\{u, v\} \subseteq V_i$ are (F, β, t) -reachable in H and $|V_i| \geq cn$;
- (iii) for every $v \in V(H)$, $|F_{\mathcal{P}}^\mu(v)| \geq \varepsilon n^{r-1}$.

If $\mathbf{i}_{\mathcal{P}}(V(H)) \in L_{\mathcal{P}, F}^\mu(H)$, then there exists a $(C''/n^{1/m_1(F)})$ -spread distribution on the set of F -factors in H' . Moreover, if $p \geq K C'' \log n / n^{1/m_1(F)}$, then a.a.s. H_p contains an F -factor.

At last, this randomized construction framework enables us to derive a quantitative enumeration result. For a k -uniform n -vertex hypergraph H with minimum codegree $\delta_{k-1}(H) \geq \delta n$ for some $\delta > 1/2$, several results have established that H contains at least $(\varepsilon n)^{\frac{k-1}{k}n}$ perfect matchings; see, for example, [?, ?, ?]. The current best known result, due to Ferber, Hardiman and Mond [?], proves the number of perfect matchings in such H is at least $\Phi(K_n^k) \cdot (\frac{1}{2} + o(1))^{n/k}$, where $\Phi(K_n^k)$ is the number of perfect matchings in the complete k -graph on n vertices. In this paper, we obtain a similar counting result under the weaker degree condition $\delta > 1/k$: either H contains no perfect matching, or it contains many of them. We note that our result is optimal up to the value of ε .

Corollary 1.11. *For $k \geq 3$, any $\gamma > 0$, there exists $\varepsilon > 0$ such that the following holds for sufficiently large integer n . For any k -graph H with $\delta_{k-1}(H) \geq n/k + \gamma n$, H contains either no perfect matching or at least $(\varepsilon n)^{\frac{k-1}{k}n}$ perfect matchings.*

We also obtain a corresponding result for F -factor.

Corollary 1.12. *For $r \geq 2$, any $\gamma > 0$, there exists $\varepsilon > 0$ such that the following holds for sufficiently large integer n . For any r -vertex graph F and any n -vertex graph G with $\delta(G) \geq (1 - 1/\chi_{cr}(F) + \gamma)n$, G contains either no F -factor or at least $(\varepsilon n)^{\frac{e_F}{rm_1(F)}n}$ F -factors.*

1.5. Related works. As mentioned above, the minimum degree conditions forcing F -factors in graphs are well understood, and the same question for hypergraphs remain a highly challenging problem, even for perfect matchings. On the other hand, the problem for near-perfect packings is slightly easier. Nevertheless, we collect the results we need on $\delta(F, k-1, D)$ here.

Theorem 1.13. [?, ?] *For every graph F , $\delta(F, 1, 5v_F^2) = 1 - 1/\chi_{cr}(F)$. For $k \geq 3$, $\delta(k, k-1, k) = 1/k$.*

For perfect matchings in k -graphs, the minimum degree conditions in Theorem ?? has been improved in subsequent works [?, ?]. Keevash, Knox and Mycroft also proposed a conjecture for perfect matchings under minimum ℓ -degree conditions for $\ell < k-1$, and this has been verified for $\ell \geq 0.4k$ in [?, ?].

Since the work of Krivelevich, Lee and Sudakov (on Hamilton cycles), robust-type results have been proved for cliques factors [?, ?], hypergraph perfect matchings [?], bounded-degree spanning trees [?], F -factors [?], and power of Hamilton cycles [?]. We remark that all of these results focus on minimum degree conditions as sufficient conditions, while in our result the minimum degree conditions are significantly lower.

2. PROOF IDEAS AND AN ALGORITHM

2.1. Algorithm and proof outline for Theorem ??. Theorem ?? is obtained by combining our main structural Theorem ?? with ideas from [?]. The work in [?] provides a necessary and sufficient condition for the existence of a perfect matching in a k -graph under a given minimum codegree condition. For this we also need a similar partition lemma for perfect matchings.

Lemma 2.1 (Partition lemma for perfect matchings). *Let $1/n \ll \beta \ll \mu \ll \varepsilon \ll \gamma \ll 1/k$, where $k \geq 3$ is an integer. Then for each k -graph H on n vertices with $\delta_{k-1}(H) \geq n/k + \gamma n$, we find a partition $\mathcal{P} = \{V_1, \dots, V_d\}$ of $V(H)$ in time $O(n^{2^{k-2}k+1})$ satisfying the following properties:*

- (P1) each part of \mathcal{P} has size at least $(1/k + \gamma/3)n$;
(P2) for each $i \in [d]$, V_i is $(\beta, k2^k + 2)$ -closed in H ;
(P3) for each vertex $v \in V(H)$, $|E_{\mathcal{P}}^{\mu}(v)| \geq \varepsilon n^{k-1}$.

Now we are ready to present our algorithm for Theorem ?? (see Procedure ??).

Procedure PerfectMatching

Data : An n -vertex k -graph H such that $\delta_{k-1}(H) \geq n/k + \gamma n$ and $k \mid n$.

Output: a.a.s. H_p has a perfect matching or none at all.

- 1 Choose constants $1/n_0 \ll \eta \ll 1/C'' \ll \beta \ll \mu \ll \varepsilon \ll \gamma \ll 1/k$. Set $p \geq K_{??} C'' \log n / n^{k-1}$.
 - 2 **if** $n < n_0$ **then**
 - 3 Check whether H_p has a perfect matching by brute force search and halt with appropriate output.
 - 4 Apply Lemma ?? to find a partition \mathcal{P} of $V(H)$ and $L_{\mathcal{P}}^{\mu}(H)$.
 - 5 **for**
every $q' \leq k-1$ and set of r -vectors $\mathbf{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_{q'}\}$ with $\mathbf{i}_{\mathcal{P}}(V(H)) - \sum_{j \in [q']} \mathbf{v}_j \in L_{\mathcal{P}}^{\mu}(H)$
do
 - 6 Let $\mathbf{V}^{\eta} := \mathbf{V} \setminus I_{\mathcal{P}}^{\eta}(H)$.
 - 7 **if** H_p contains a matching M' with $\mathbf{i}_{\mathcal{P}}(V(M')) = \sum_{\mathbf{v} \in \mathbf{V}^{\eta}} \mathbf{v}$ **then**
 - 8 Output “a.a.s. a perfect matching exists.” and halt.
 - 9 Output “no perfect matching” and halt.
-

Suppose we apply Lemma ?? and obtain a partition \mathcal{P} and the lattice $L_{\mathcal{P}}^{\mu}(H)$. The algorithm for decision problem of perfect matchings in H (e.g. in [?]) searches for all possible matchings M of size at most $k-1$ such that $\mathbf{i}_{\mathcal{P}}(V(H) \setminus V(M)) \in L_{\mathcal{P}}^{\mu}(H)$ (and it is shown that this is sufficient). Such searches are deterministic and clearly can be done in polynomial time. In our problem, we need to proceed the search in H_p . If no such M is found in H_p , then we can conclude that H_p has no perfect matching. However, if one such M is found in H_p , we cannot argue as in the dense case by just showing $H - V(M)$ has a perfect matching. The reason is, we may have checked quite many edges of H_p , and after we checked them, they lost their randomness and revealed as non-edges. That is to say, we need to work with the p -random sparsification of $H - V(M)$ with a small set of edges removed. To overcome this issue, we choose a small variable η and do the following

- We do not check η -robust edges, as a.a.s. H_p contains many such edges. In particular, we can find a constant-sized matching of such edges.
- We check the non- η -robust edges, and argue that losing these edges will not quantitatively affect the properties of H too much (that is, the reduced properties still allow us to apply Theorem ??).

For the second bullet point, the new property ?? of the partition lemma is crucial. Then the proof of Theorem ?? reduces to prove the correctness of Procedure ?? and estimate its running time.

Now we can address the question on the acceptance probability of our algorithm, when the input is a k -graph with at least one perfect matching. The relevant quantity is the number of matchings in H with index vector \mathbf{V}^{η} . For instance, if there is a matching M such that $\mathbf{i}_{\mathcal{P}}(V(H) \setminus V(M)) \in L_{\mathcal{P}}^{\mu}(H)$ such that all edges of M are η -robust, then in the algorithm line 6 one can take $\mathbf{V}^{\eta} = \emptyset$ and on line 7 take $M' = \emptyset$, and thus the algorithm has a positive output. Therefore, we have the following corollary by standard concentration results.

Corollary 2.2. *For $\gamma > 0$ and $k \geq 3$, an n -vertex k -graph H with $\delta_{k-1}(H) \geq n/k + \gamma n$ and a partition found by Lemma ??.*

- If there exists \mathbf{V}^η as defined on line 6 of Procedure ?? such that H contains $\omega(p^{-1})$ edges with index vector \mathbf{v} for each $\mathbf{v} \in \mathbf{V}^\eta$, then Procedure ?? a.a.s. accepts H_p ;
- If for every choice of \mathbf{V}^η , there exists $\mathbf{v} \in \mathbf{V}^\eta$ such that H contains $o(p^{-1})$ edges with index vector \mathbf{v} , then Procedure ?? a.a.s. rejects H_p . \square

2.2. Proof ideas for Theorem ??. The main tool for the proof of Theorem ?? is the spread method. For this we first state a celebrated result of Frankston, Kahn, Narayanan and Park [?], conjectured by Talagrand.

Theorem 2.3 ([?]). *There exists a constant $K = K_{??} > 0$ such that for any ℓ -uniform, q -spread hypergraph \mathcal{H} on V , a uniformly random $(Kq \log \ell)$ -element subset of V a.a.s. contains an edge in \mathcal{H} .*

While Theorem ?? reduces the problem to constructing a probability measure with good spread, achieving this on non-complete host graphs remains technically challenging. To address this, for our work we adopt the *randomized embedding algorithm* $G \hookrightarrow H$ proposed by Kelly, Műyesser, and Pokrovskiy [?]. The central concept in their method is the following notion of *vertex spread*.

Definition 2.4. Let X and Y be finite sets, and let μ be a probability distribution over injection $\phi : X \rightarrow Y$. For $q \in [0, 1]$, we say that μ is q -vertex-spread if for every two sequences of distinct vertices $x_1, \dots, x_s \in X$ and $y_1, \dots, y_s \in Y$,

$$\mu(\{\phi : \phi(x_i) = y_i \text{ for all } i \in [s]\}) \leq q^s.$$

Based on this notion, they constructed a distribution with provably good spread properties, stated in the result below. A hypergraph embedding $\phi : G \hookrightarrow H$ of a hypergraph G into a hypergraph H is an injective map $\phi : V(G) \hookrightarrow V(H)$ that maps edges of G to edges of H , so there is an embedding of G into H if and only if H contains a subgraph isomorphic to G .

Proposition 2.5 ([?]). *For every $C, k, \Delta > 0$, there exists $C' = C_{??}(C, k, \Delta) > 0$ such that the following holds for all sufficiently large n . Let H and G be n -vertex k -graphs. If there is a (C/n) -vertex-spread distribution on embeddings $G \hookrightarrow H$ and $\Delta(G) \leq \Delta$, then there is a $(C'/n^{1/m_1(G)})$ -spread distribution on subgraphs of H which are isomorphic to G .*

We now briefly outline the proof of Theorem ??; the full proof is presented in Section ?. By Proposition ??, it suffices to construct a (C/n) -vertex-spread distribution on embeddings $G \hookrightarrow H$, where G consists of n/r disjoint copies of F , that is, a perfect F -packing. We begin by randomly partitioning $V(H)$ into constant-sized clusters U_1, \dots, U_m such that each cluster inherits the minimum degree condition and partition properties from H . This is achieved via the *random redistribution* argument introduced by [?]. Although the global structure of H satisfies only approximate conditions, this step ensures that each U_i has exact degree conditions and a locally good partition. We then aim to apply the general structural theorem (Theorem ??) from [?] on each cluster to find a perfect F -packing in $H[U_i]$. However, this theorem requires, in addition to the degree and partition conditions, certain divisibility properties, which a cluster U_i may not initially satisfy. To address this, we slightly adjust each U_i by moving a few vertices from adjacent cluster to ensure the required divisibility condition. Condition ?? plays a crucial role in enabling this adjustment. Let U'_i denote the modified cluster, and applying the structural theorem to it gives a perfect F -packing M_i in $H[U'_i]$. Taking the union of all such packings yields a perfect F -packing in H . Finally, since each cluster is sufficiently small, the probability that a fixed vertex falls in any given cluster is at most C/n . This yields a (C/n) -vertex-spread distribution on embeddings $G \hookrightarrow H$.

The rest of the paper is organized as follows. In the next section, we introduce a series of basic preliminaries that will be used throughout the paper. In Section ??, we prove the partition lemmas ?? and ?. Section ?? is devoted to the proof of our main tool – the random clustering lemma. In Section ??, we prove the main technical result (Theorem ??). Section ?? presents the

proofs of the remaining results, including the main algorithmic result (Theorem ??), the robustness of F -factor (Theorem ??), and two enumeration results (Corollary ?? and ??). We give some concluding remarks and questions at the end.

3. PRELIMINARIES

We need the following concentration inequality due to McDiarmid [?], whose present formulation can be found in [?].

Lemma 3.1 ([?]). *Let $c > 0$ and let f be a function defined on the set of subsets of some set U such that $|f(U_1) - f(U_2)| \leq c$ whenever $|U_1| = |U_2| = m$ and $|U_1 \cap U_2| = m - 1$. Let A be a uniformly random m -subset of U . Then for any $\alpha > 0$ we have*

$$\mathbb{P}[|f(A) - \mathbb{E}[f(A)]| \geq \alpha c \sqrt{m}] \leq 2 \exp(-2\alpha^2).$$

We prove the following easy consequences of the McDiarmid inequality.

Lemma 3.2. *Let $0 < \gamma' < \gamma < 1$ and $1/n, 1/\ell \ll 1/r, \gamma - \gamma'$. Let F be an r -vertex k -graph, and let H be an n -vertex k -graph with vertex set V and a partition \mathcal{P} of V . Suppose that the number of copies F' of F with $\mathbf{i}_{\mathcal{P}}(V(F')) = \mathbf{v}$ is at least γn^r . If $A \subseteq V$ is a vertex set of size ℓ chosen uniformly at random, then with probability at most $2 \exp(-\ell(\gamma - \gamma')^2/2)$, $H[A]$ contains fewer than $\gamma' \ell^r$ copies of F with index vector \mathbf{v} .*

Proof. We verify the assumptions of Lemma ???. Let $f : \binom{V}{\ell} \rightarrow \mathbb{R}$ be defined as the number of copies F' in $H[X]$ of F with index vector \mathbf{v} for each $X \in \binom{V}{\ell}$, and set $\varepsilon = (\gamma - \gamma')/(2\gamma) < 1/2$. Note that $|f(X_1) - f(X_2)| \leq \ell^{r-1}$ for any $X_1, X_2 \in \binom{V}{\ell}$ with $|X_1 \cap X_2| = \ell - 1$. Given a copy F' of F with index vector \mathbf{v} , the probability that F' is contained in $H[A]$ is at least $\frac{\binom{n-r}{\ell-r}}{\binom{n}{\ell}} \geq (1 - \varepsilon) \frac{\ell^r}{n^r}$, where we used $1/n, 1/\ell \ll 1/r, \gamma - \gamma'$. So by linearity of expectation we have $\mathbb{E}[f(A)] \geq \gamma(1 - \varepsilon)\ell^r = \frac{\gamma + \gamma'}{2}\ell^r$. Let E be the event that there are fewer than $\gamma' \ell^r$ copies of F with index vector \mathbf{v} in $H[A]$. We apply Lemma ?? with $c = \ell^{r-1}, m = \ell$ and $\alpha = \sqrt{\ell}(\gamma - \gamma')/2$, and get that $\mathbb{P}[E] \leq 2 \exp(-\ell(\gamma - \gamma')^2/2)$, as desired. \square

For a vertex $v \in V(H)$, a subset $Y \subseteq V(H)$, and a partition \mathcal{P} of $V(H)$, we define $F_{\mathcal{P}}^{\mu}(v, Y) := \{R \in F_{\mathcal{P}}^{\mu}(v) : R \setminus \{v\} \subseteq Y\}$ to be the collection of r -sets that span at least one μ -robust copy of F , contain v , and have all other vertices lying entirely within Y .

Lemma 3.3. *Let $0 < \gamma' < \gamma < 1$ and $1/n, 1/\ell \ll 1/r, \gamma - \gamma'$. Let F be an r -vertex k -graph, and let H be an n -vertex k -graph with vertex set V and a partition \mathcal{P} of V satisfying $|F_{\mathcal{P}}^{\mu}(v)| \geq \gamma n^{r-1}$ for each $v \in V$. If $A \subseteq V$ is a vertex set of size ℓ chosen uniformly at random, then for every $v \in V$ we have*

$$\mathbb{P}[|F_{\mathcal{P}}^{\mu}(v, A)| < \gamma' \ell^{r-1}] \leq 2 \exp(-\ell(\gamma - \gamma')^2/2).$$

Proof. For every vertex $v \in V$, let $f_v : \binom{V}{\ell} \rightarrow \mathbb{R}$ be defined by $f_v(X) = |F_{\mathcal{P}}^{\mu}(v, X)|$ for each $X \in \binom{V}{\ell}$, and set $\varepsilon = (\gamma - \gamma')/(2\gamma) < 1/2$. Note that $|f_v(X_1) - f_v(X_2)| \leq \ell^{r-2}$ for any $X_1, X_2 \in \binom{V}{\ell}$ with $|X_1 \cap X_2| = \ell - 1$. By linearity of expectation we have

$$\begin{aligned} \mathbb{E}[f_v(A)] &= \sum_{F' \in F_{\mathcal{P}}^{\mu}(v)} \mathbb{P}[(V(F') \setminus \{v\}) \subseteq A] \geq \sum_{F' \in F_{\mathcal{P}}^{\mu}(v)} \frac{\binom{n-(r-1)}{\ell-(r-1)}}{\binom{n}{\ell}} \\ &\geq \gamma n^{r-1} (1 - \varepsilon) \frac{\ell^{r-1}}{n^{r-1}} \\ &\geq \gamma(1 - \varepsilon) \ell^{r-1} = \frac{\gamma + \gamma'}{2} \ell^{r-1}. \end{aligned}$$

We can then apply Lemma ?? with $c = \ell^{r-2}$, $m = \ell$ and $\alpha = \sqrt{\ell}(\gamma - \gamma')/2$, and get that $\mathbb{P}[|F_P^\mu(v, A)| < \gamma' \ell^{r-1}] \leq 2 \exp(-\ell(\gamma - \gamma')^2/2)$, as desired. \square

Lemma 3.4. *Let $0 < \beta' < \beta < 1$ and $1/n, 1/\ell \ll 1/r, \beta - \beta'$. Let F be an r -vertex k -graph, and let H be an n -vertex k -graph with vertex set V . Consider the vertex set $U \subseteq V$ where all but $\eta \binom{|U|}{2}$ of the pairs are (F, β, t) -reachable. Let A be a uniformly random subset of ℓ vertices of G . Then with probability at least $1 - \binom{\ell}{2} \left(\eta + 2e^{-(\beta - \beta')^2 \ell/2} \right)$, the random set $U \cap A$ is (F, β', t) -closed in $H[A]$.*

Proof. Let W be the collection of pairs of vertices that are not (F, β, t) -reachable in H . We will prove that for every $u, v \in U \cap A$,

$$\mathbb{P}[u \text{ and } v \text{ are not } (F, \beta/2, t)\text{-reachable in } H[A]] \leq \eta + 2e^{-(\beta - \beta')^2 \ell/2},$$

then the desired result will follow from the union bound.

First, note that for $u, v \in A$, the probability of the event $\{u, v\} \in W$ is at most η . Now, condition on the event that $u, v \in A$ and $\{u, v\} \in W$. Let $A_{u,v}$ be the number of (F, β, t) -reachable sets $S \subseteq A$ for u and v . Set $\varepsilon = (\beta - \beta')/(2\beta) < 1/2$. Note that making any "swap" to our subset A (that is, exchanging any element with an element outside this subset) affects $A_{u,v}$ by at most ℓ^{tr-2} . Given an (F, β, t) -reachable set S , the probability that S is contained in A is at least $\frac{\binom{n-2-(tr-1)}{\ell-2-(tr-1)}}{\binom{n-2}{\ell-2}} \geq (1 - \varepsilon) \frac{\ell^{tr-1}}{n^{tr-1}}$, where we used $1/n, 1/\ell \ll 1/r, \beta - \beta'$. So, by linearity of expectation we have $\mathbb{E}[A_{u,v}] \geq \beta n^{tr-1} \cdot (1 - \varepsilon) \frac{\ell^{tr-1}}{n^{tr-1}} = \frac{\beta + \beta'}{2} \ell^{tr-1}$. We can then apply Lemma ?? with $c = \ell^{tr-2}$, $m = \ell$ and $\alpha = \sqrt{\ell}(\beta - \beta')/2$, and get that the probability that the number of (F, β, t) -reachable sets $S \subseteq A$ for u and v is at most $\beta' \ell^{tr-1}$ is at most $2e^{-\frac{(\beta - \beta')^2}{2} \ell}$. \square

Lemma 3.5 ([?]). *Let $0 < \gamma' < \gamma < 1$. Consider an n -vertex k -graph G where all but $\alpha \binom{n}{d}$ of the d -sets have degree at least $(\delta + \gamma) \binom{n-d}{k-d}$. Let A be a uniformly random subset of $\ell \geq 2d$ vertices of G . Then with probability at least $1 - \binom{\ell}{d} \left(\alpha + e^{-(\gamma - \gamma')^2 \ell/4} \right)$, the random induced subgraph $G[A]$ has minimum d -degree at least $(\delta + \gamma') \binom{\ell-d}{k-d}$.*

We will also use a result due to McDiarmid, as presented in Chapter 16.2 of the textbook by Molloy and Reed [?]. In this context, a *choice* refers to the position that a particular element is mapped to in a permutation.

Lemma 3.6 (McDiarmid's inequality for random permutations). *Let X be a non-negative random variable determined by a random permutation π of $[n]$ such that the following holds for some $c, r > 0$:*

- (1) *Interchanging two elements of π can affect the value of X by at most c .*
- (2) *For any s , if $X \geq s$ then there is a set of at most rs choices whose outcomes certify that $X \geq s$.*

Then, for any $0 \leq t \leq \mathbb{E}[X]$,

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq t + 60c\sqrt{r\mathbb{E}[X]}] \leq 4 \exp(-t^2/(8c^2 r \mathbb{E}[X])).$$

Finally, we need a weaker version of results in [?].

Lemma 3.7 ([?]). *There exists an absolute constant $C_{??}$ with the following property. If G is a balanced bipartite graph on $2n$ vertices with $\delta(G) \geq 3n/4$, then there exists a $(C_{??}/n)$ -spread distribution on perfect matchings of G .*

4. PARTITION LEMMA

In this section, we prove Lemmas ?? and ?. For this we first introduce some useful notation and collect several auxiliary results.

We denote by $K_k^{(k)}(r_1, \dots, r_k)$ the complete k -partite k -graph with parts of sizes r_1, \dots, r_k . If $r_i = r$ for all $1 \leq i \leq k$, we simply write $K_k^{(k)}(r)$. For any $S \subseteq V(H)$, let $N(S) := \{T \subseteq V(H) \setminus S : T \cup S \in E(H)\}$, and for simplicity, we write $N(x)$ for $N(\{x\})$.

We need the following (simple) result of Lo and Markström for verifying the 1-reachability property.

Lemma 4.1 ([?]). *Let $k, r \geq 2$ be integers and let $\alpha > 0$. There exists a constant $\eta = \eta(k, r, \alpha)$ such that the following holds for sufficiently large n . Let H be an n -vertex k -graph. Then any two vertices $u, v \in V(H)$ are $(K_k^{(k)}(r), \eta, 1)$ -reachable to each other if the number of $(k-1)$ -sets $S \in N(u) \cap N(v)$ with $|N(S)| \geq \alpha n$ is at least $\alpha \binom{n}{k-1}$.*

We also require the following simple property of reachability. The fact is clear as we can add copies of F vertex disjoint with the existing reachable sets to enlarge the reachable sets, at the price of getting a smaller β .

Fact 4.2. Let $k, r \geq 2$ be integers and let F be an r -vertex k -graph. Let $i \geq 1$ be an integer and $\beta, \eta > 0$. Then there exists $\beta' = \beta'(r, i, \beta, \eta)$ such that the following holds for sufficiently large n . Let H be an n -vertex k -graph that contains at least ηn^r copies of F . Then any two vertices $u, v \in V(H)$ that are (F, β, i) -reachable are also $(F, \beta', i+1)$ -reachable. \square

The next tool is a supersaturation result of Erdős.

Proposition 4.3 ([?]). *Let $\eta > 0$, $k, d \in \mathbb{N}$ and let $K := K_k^{(k)}(a_1, \dots, a_k)$ be the complete k -partite k -graph with $a_1 \geq \dots \geq a_k$ vertices in each class. There exists $0 < \mu \ll \eta$ such that the following holds for sufficiently large n . Let H be a k -graph on n vertices with a vertex partition $V_1 \cup \dots \cup V_d$ and consider not necessarily distinct $i_1, \dots, i_k \in [d]$. Suppose H contains at least ηn^k edges $e = \{v_1, \dots, v_k\}$ such that $v_1 \in V_{i_1}, \dots, v_k \in V_{i_k}$. Then H contains at least $\mu n^{a_1 + \dots + a_k}$ copies of K whose j th part is contained in V_{i_j} for all $j \in [k]$.*

As mentioned in the introduction, compared to earlier versions, our partition lemmas have one additional structural property. To capture this in a unified framework, we prove the following slightly more general theorem. For convenience, we introduce some definitions and notation that will be used throughout the paper.

For any $\mathbf{v} = \{v_1, \dots, v_d\} \in \mathbb{Z}^d$, let $|\mathbf{v}| := \sum_{i=1}^d v_i$. We say that $\mathbf{v} \in \mathbb{Z}^d$ is an r -vector if it has non-negative coordinates and $|\mathbf{v}| = r$. We write \mathbf{u}_j for the ‘unit’ 1-vector that is 1 in coordinate j and 0 in all other coordinates. Given a partition \mathcal{P} of d parts, we write L_{\max}^d for the lattice generated by all r -vectors, that is, $L_{\max}^d := \{\mathbf{v} \in \mathbb{Z}^d : r \text{ divides } |\mathbf{v}|\}$. Note that $|L_{\max}^d| = \binom{r+d-1}{r}$.

Let $\beta, c > 0$ and $t \in \mathbb{N}$. A partition $\mathcal{P} = \{V_1, \dots, V_d\}$ of $V(H)$ is (F, β, t, c) -good if V_i is (F, β, t) -closed and $|V_i| \geq cn$ for all $i \in [d]$.

Lemma 4.4. *Let $k, r, t \geq 2$ be integers and $\alpha, \beta, \mu, c, c' > 0$ where $\mu \ll \alpha, 1/r, 1/d$ and $c' < c$. There exist β' and ε such that the following holds for sufficiently large integer n . Let F be a k -partite k -graph of order r . For any n -vertex k -graph H with an (F, β, t, c) -good vertex partition $\mathcal{P} = \{U_1, \dots, U_d\}$, suppose for every vertex $v \in V(H)$, the number of $(k-1)$ -sets $S \in N(v)$ with $|N(S)| \geq \alpha n$ is at least $\alpha \binom{n}{k-1}$. Then in time $O(n^{\max\{r, k+1\}})$ we can find an $(F, \beta', 2tr+2, c')$ -good partition \mathcal{P}' of $V(H)$ such that for every vertex $v \in V(H)$, $|F_{\mathcal{P}'}^{\mu/2}(v)| \geq \varepsilon n^{r-1}$.*

Proof. Take $\varepsilon \leq c - c'$ and choose constants satisfying the following hierarchy

$$1/n \ll \mu \ll \varepsilon \ll \eta_1, \eta_2 \ll \alpha \quad \text{and} \quad 1/n \ll \beta' \ll \beta_1 \ll \beta, \mu, \eta_1, \eta_2, c'.$$

Let $K(F) \supseteq F$ be a complete k -partite k -graph on r vertices (take an arbitrary one if there are more than one choices). Then there exist integers $a_1 \geq \dots \geq a_k$ such that $K(F) = K_k^{(k)}(a_1, \dots, a_k)$. We first claim that for every vertex $v \in V(H)$, there exist at least $\eta_1 n^{r-1}$ $(r-1)$ -sets R in H such that

$H[R \cup \{v\}]$ contains a copy of $K(F)$. To see this, duplicate v and denote the resulting graph by H' . By Lemma ??, v and its duplicate v' are $(K(F), 1, \eta_1)$ -reachable in H' , which indeed implies the claim.

Let U_0 be the set of all vertices v such that there are at most $2\varepsilon n^{r-1}$ $(r-1)$ -sets R such that $H[R \cup \{v\}]$ contains a copy of F with $\mathbf{i}_{\mathcal{P}}(R) \in I_{\mathcal{P},F}^\mu(H)$, i.e., $|F_{\mathcal{P}}^\mu(v)| \leq 2\varepsilon n^{r-1}$. Note that for each $v \in U_0$, the vertex v lies in at least $\eta_1 n^{r-1}/2$ copies of F whose index vector is not in $I_{\mathcal{P},F}^\mu(H)$, yielding

$$|U_0| \leq \frac{r \binom{r+d-1}{r} \mu n^r}{\eta_1 n^{r-1}/2} \leq (r+d)^r \mu n / \eta_1 < \varepsilon n,$$

as $\mu \ll \varepsilon, \eta_1, 1/r, 1/d$. To prove the lemma, we shall move every vertex of U_0 greedily to some U_i for $i \in [d]$.

Claim 4.5. For every vertex $v \in U_0$, there exists $\mathbf{v} \in I_{\mathcal{P},F}^\mu(H)$ and $i \in [d]$ such that there are $\eta_2 n^{r-1}/r!$ $(r-1)$ -sets R with index vector $\mathbf{v} - \mathbf{u}_i$ for which $H[R \cup \{v\}]$ contains a copy of F .

Proof. For every vertex $v \in U_0$, by averaging, there exist at least $\alpha \binom{n}{k-1} / \binom{k+d-2}{k-1} \geq \frac{1}{(k+d)^{k-1}} \alpha n^{k-1}$ $(k-1)$ -sets $S \in N(v)$ with same index vector \mathbf{w} and $|N(S)| \geq \alpha n$. Then there exist $i \in [d]$ and at least

$$\frac{1}{kd} \cdot \frac{\alpha}{(k+d)^{k-1}} n^{k-1} \cdot \alpha n \geq \frac{\alpha^2}{(k+d)^{k+1}} n^k$$

edges e such that $e \cup \{v\}$ spans a copy of $K_k^{(k)}(2, 1, \dots, 1)$ and $\mathbf{i}_{\mathcal{P}}(e) = \mathbf{w} + \mathbf{u}_i$, which means that there exist indices $i_1, \dots, i_{k-1} \in [d]$ that are not necessarily distinct and H contains at least $\alpha^2 n^k / (k+d)^{k+1}$ edges with vertex set $\{v_1, \dots, v_k\}$ such that $v_1 \in U_{i_1}, \dots, v_k \in U_{i_k}$, where $i_k = i$, and $\{v_1, \dots, v_{k-1}\} \in N(v)$. Note that by applying Proposition ?? with $l_i := a_i$ for each $i \in [k]$, we get that there are at least $\eta_2 n^r > \mu n^r$ copies of $K(F)$ in H , where the j -th part is contained in U_{i_j} for all $j \in [k]$. This implies that these copies share the index vector $\mathbf{v} = \sum_{j \in [k]} a_{i_j} \mathbf{u}_{i_j}$, and $\mathbf{v} \in I_{\mathcal{P},F}^\mu(H)$. Moreover, by applying Proposition ?? with $l_i := a_i$ for each $i \in [k-1]$ and $l_k := a_k - 1$, we conclude that H contains at least $\eta_2 n^{r-1}$ copies of $K_k^{(k)}(a_1, \dots, a_{k-1}, a_k - 1)$ with index vector $\mathbf{v} - \mathbf{u}_i$ such that each such copy together with v spans a copy of $K(F)$. Since at most $r!$ copies of F can share the same r -set of vertices, this completes the proof of the claim. \square

Now for each $v \in U_0$, we move v to U_i where the index $i \in [d]$ is given by Claim ?? (choose any one if there are more than one such choices of $i \in [d]$). We proceed greedily until all vertices of U_0 are assigned to a new cluster.

Denote the resulting partition of $V(H)$ by $\mathcal{P}' = \{V_1, \dots, V_d\}$. Note that $|V_i| \geq |U_i| - |U_0| \geq c'n$. Now we show that the process above will not decrease the number of robust copies for each vertex significantly. We claim that $I_{\mathcal{P},F}^\mu(H) \subseteq I_{\mathcal{P}',F}^{\mu/2}(H)$. Indeed, since moving any vertex in U_0 decreases the number of “robust copies” of F by at most $r! \cdot 2\varepsilon n^{r-1}$ (that is, turning robust copies to non-robust ones) and by $|U_0| \leq (r+d)^r \mu n / \eta_1$, for every $\mathbf{v} \in I_{\mathcal{P},F}^\mu(H)$, there exists at least $\mu n^r - r! \cdot 2\varepsilon n^{r-1} |U_0| \geq \mu n^r / 2$ copies F' of F with $\mathbf{i}_{\mathcal{P}'}(V(F')) = \mathbf{v}$, which implies $\mathbf{v} \in I_{\mathcal{P}',F}^{\mu/2}(H)$. Then, for every $v \in V(H) \setminus U_0$, we have $|F_{\mathcal{P}'}^{\mu/2}(v)| \geq |F_{\mathcal{P}}^\mu(v)| - |U_0| \binom{n}{r-2} \geq 2\varepsilon n^{r-1} - \varepsilon n \binom{n}{r-2} \geq \varepsilon n^{r-1}$; for $v \in U_0$, $|F_{\mathcal{P}'}^{\mu/2}(v)| \geq (\eta_2 n^{r-1}/r!) - |U_0| \binom{n}{r-2} \geq \varepsilon n^{r-1}$. So we have $|F_{\mathcal{P}'}^{\mu/2}(v)| \geq \varepsilon n^{r-1}$ for every $v \in V(H)$ and it remains to verify the reachability.

Finally, we verify the reachability. Note that we only need to verify for pairs of u, v with at least one of them in U_0 . All index vectors mentioned below are with respect to \mathcal{P} .

- (1) For each $\ell \in [d]$, any $v \in V_\ell \cap U_0$ and $u \in V_\ell \setminus U_0$ are $(F, \beta_1, tr+1)$ -reachable in H . Indeed, by Claim ?? and our process, there exist $\mathbf{v} \in I_{\mathcal{P},F}^\mu(H)$ such that H contains $\eta_2 n^{r-1}/r!$ $(r-1)$ -sets $R = \{v_1, \dots, v_{r-1}\}$, each having index vector $\mathbf{v} - \mathbf{u}_\ell$ with respect to \mathcal{P} , such

that $R \cup \{v\}$ spans a copy of F . Since $\mathbf{v} \in I_{\mathcal{P},F}^\mu(H)$, there are at least μn^r copies of F , each with index vector \mathbf{v} . Fix one such copy of F with vertex set $\{w, w_1, \dots, w_{r-1}\}$, and without loss of generality, we may assume that v_i and w_i are in the same part of \mathcal{P} for all $i \in [k-1]$ and $w \in U_\ell$. Since each part of \mathcal{P} is (F, β, t) -closed in H , there exist βn^{tr-1} reachable $(tr-1)$ -sets S_i for v_i and w_i for all $i \in [r-1]$, and βn^{tr-1} reachable $(tr-1)$ -sets S_0 for u and w . Let $S := \left(\bigcup_{i \in [r-1]} (\{v_i\} \cup S_i \cup \{w_i\}) \right) \cup S_0 \cup \{w\}$. Observe that if $|S| = tr^2 + r - 1$, that is, the sets are all disjoint, then both $H[S \cup \{v\}]$ and $H[S \cup \{u\}]$ have a perfect F -packing. Hence there are at least

$$\frac{(\eta_2 n^{r-1}/r!) \cdot \mu n^r \cdot (\beta n^{tr-1})^r - (tr^2 + r)n^{tr^2+r-2}}{(tr^2 + r - 1)!} \geq \beta_1 n^{tr^2+r-1}$$

reachable $(tr^2 + r - 1)$ -sets for v and u .

- (2) For each $\ell \in [d]$, any $v_1, v_2 \in V_\ell \cap U_0$ are $(F, \beta', 2tr + 2)$ -reachable in H . Indeed, note that there are at least $|V_\ell \setminus U_0| \geq cn - \varepsilon n \geq c'n$ vertices w $(F, \beta_1, tr + 1)$ reachable to both v_1 and v_2 , at least $\beta_1 n^{tr^2+r-1}$ $(tr^2 + r - 1)$ -reachable sets S_i for v_i and w for $i = 1, 2$. Let $S = S_1 \cup \{w\} \cup S_2$ and if $|S| = 2tr^2 + 2r - 1$, then $H[\{v_1\} \cup S]$ and $H[S \cup \{v_2\}]$ have perfect F -packings. Hence there are at least

$$\frac{(c'n) \cdot (\beta_1 n^{tr^2+r-1})^2 - (2tr^2 + 2r)n^{2tr^2+2r-2}}{(2tr^2 + 2r - 1)!} \geq \beta' n^{2tr^2+2r-1}$$

reachable $(2tr^2 + 2r - 1)$ -sets for v_1 and v_2 .

Note that there are at least $\eta_1 n^{r-1} \cdot n/r$ copies of F . Then, by Fact ?? and the choice of β' , each V_ℓ is $(F, \beta', 2tr + 2)$ -closed in H .

We estimate the running time as follows. First, we compute $I_{\mathcal{P},F}^\mu(H)$ by checking each r -set $S \subseteq V(H)$ to determine whether it spans a copy of F , and if so, recording its index vector $\mathbf{i}_{\mathcal{P}}(S)$. This involves listing all possible sets of at most $\binom{r}{k}$ edges on S , and computing part sizes with respect to \mathcal{P} , which takes constant time. Since there are $O(n^r)$ such subsets, the total runtime is $O(n^r)$. Next, we determine the set U_0 by calculating, for every vertex $v \in V(H)$, the number of copies K of $K(F)$ that contain v and satisfy $\mathbf{i}_{\mathcal{P}}(K) \in I_{\mathcal{P},F}^\mu(H)$, which takes time $O(n^r)$. After constructing U_0 , for each vertex $v \in U_0$, we identify the set $N'(v)$ of $(k-1)$ -sets in $N(v)$ that share the same index vector and have at least $\frac{1}{(k+d)^{k-1}} \alpha n^{k-1}$ such tuples, in total time $O(n^k)$. Subsequently, for every $v \in U_0$, we determine the index $i = i(v)$ as in Claim ?? in constant time. This requires time $O(n^{k+1})$. Finally, we move vertices v of U_0 to U_1, \dots, U_d based on Claim ??, which can be done in time $O(n)$. In summary, the overall time complexity for finding such a partition \mathcal{P}' is $O(n^{\max\{r, k+1\}})$. \square

For the rest of this section, we use Lemma ?? to derive Lemmas ?? and ??.

4.1. Perfect matching. We will use the following result in the proof of Lemma ??, which is derived by combining Proposition 3.7 and Lemma 3.8 from [?].

Lemma 4.6. [?] *Let $1/n \ll \beta \ll \gamma \ll 1/k$, where $k \geq 3$ is an integer. For each k -graph H on n vertices with $\delta_{k-1}(H) \geq n/k + \gamma n$, in time $O(n^{2^{k-2}k+1})$, we can find a $(\beta, 2^{k-1}, 1/k + \gamma/2)$ -good partition \mathcal{P} of $V(H)$.*

Now we are ready to prove Lemma ??.

Proof of Lemma ??. Suppose $1/n \ll \beta \ll \mu \ll \varepsilon \ll \gamma \ll 1/k$. Let H be a k -graph on n vertices satisfying $\delta_{k-1}(H) \geq n/k + \gamma n$ with $\mathcal{P} = \{U_1, \dots, U_d\}$ found by Lemma ??. Note that $\delta_1(H) \geq (1/k + \gamma) \binom{n}{k-1}$. Applying Lemma ?? with $\alpha = 1/k + \gamma$, $c = 1/k + \gamma/2$, $c' = 1/k + \gamma/3$, $t = 2^{k-1}$, $r = k$ and F being a single edge, in time $O(n^{k+1})$ we obtain a desired partition $\mathcal{P}' = \{V_1, \dots, V_d\}$. \square

4.2. Perfect packing. In this subsection we prove Lemma ??, and we need slightly more preparation. Let F be an r -vertex k -chromatic graph. By the definition of $\chi_{cr}(F)$, we have

$$\frac{1}{\chi_{cr}(F)} = \frac{r - \sigma(F)}{(k-1)r} \leq \frac{r-1}{(k-1)r}.$$

We can obtain the desired partition by applying the following variant of Lemma ??, which can be easily derived from the original version by defining a k -graph G' where each k -set forms a hyperedge if and only if it spans a copy of K_k in G . For any vertex $u \in V(H)$, let $W(u)$ denote the collection of $(k-1)$ -sets $S \subseteq N(v)$ such that S spans a clique in H . For a set $T \subseteq V(H)$, $N(T) := \bigcap_{v \in T} N(v)$.

Lemma 4.7. *Let $k, r, t \geq 2$ be integers and $\alpha, \beta, \mu, c, c' > 0$ where $\mu \ll \alpha, 1/r, 1/d$ and $c' < c$. There exist β' and ε such that the following holds for sufficiently large integer n . Let F be a k -chromatic graph on r vertices. For any n -vertex graph H , if there exists an (F, β, t, c) -good partition $\mathcal{P} = \{U_1, \dots, U_d\}$ of $V(H)$ and for every vertex $v \in V(H)$, the number of $(k-1)$ -sets $S \in W(v)$ with $|N(S)| \geq \alpha n$ is at least $\alpha \binom{n}{k-1}$, then in time $O(n^{\max\{r, k+1\}})$ we can find an $(F, \beta', 2tr + 2, c')$ -good partition \mathcal{P}' of $V(H)$ such that for every vertex $v \in V(H)$, $|F_{\mathcal{P}'}^\mu(v)| \geq \varepsilon n^{r-1}$. \square*

The following lemma provides a good partition, similar to the perfect matching case.

Lemma 4.8 ([?]). *Let $k, r, n \geq 2$ be integers and $\gamma, \beta > 0$ where $1/n \ll \beta \ll \gamma \ll 1/r, 1/k$. Let F be an r -vertex k -chromatic graph and $h := r^{k-1}$. For each n -vertex graph H with $\delta(H) \geq (1 - 1/\chi_{cr}(F) + \gamma)n$, in time $O(n^{2^{h-1}r+1})$ we can find an $(F, \beta, 2^{h-1}, 1/r + \gamma/2)$ -good partition \mathcal{P} of $V(H)$.*

We now combine these two lemmas and prove Lemma ??.

Proof of Lemma ??. Suppose $1/n \ll \beta \ll \mu \ll \varepsilon \ll \gamma \ll 1/r, 1/k$. Applying Lemma ?? to H , we obtain an $(F, \beta, 2^{h-1}, 1/r + \gamma/2)$ -good partition $\mathcal{P} = \{V_1, \dots, V_d\}$ of $V(H)$ in time $O(n^{2^{h-1}r+1})$, where $h = r^{k-1}$. Since $\delta(H) \geq (1 - 1/\chi_{cr}(F) + \gamma)n$, simple counting shows that for every $(k-1)$ -set S , $|N(S)| \geq (1/r + (k-1)\gamma)n$. This implies that for every vertex $v \in V(H)$, $|W(v)| \geq \binom{n}{k-1}/h$. Therefore, by Lemma ?? applied with $\alpha = 1/h$, $c = 1/r + \gamma/2$, $c' = 1/r + \gamma/3$, and $t = 2^{h-1}$, we obtain an $(F, \beta', r2^h + 2, 1/r + \gamma/3)$ -good partition \mathcal{P}' such that for every vertex $v \in V(H)$, $|F_{\mathcal{P}'}^\mu(v)| \geq \varepsilon n^{r-1}$, in time $O(n^{\max\{r, k+1\}})$. \square

5. RANDOM CLUSTERING LEMMA

Our main tool for the proof of Theorem ?? is Lemma ??, which is proved by the random clustering method of Kelly, Müyesser and Pokrovskiy [?], and we use it to construct a probability distribution over the desired subgraphs with good spread. For $Y \subseteq V(H)$ and a partition \mathcal{P} of $V(H)$, we define the projection of \mathcal{P} on Y as $\mathcal{P}^Y := \{W \cap Y : W \in \mathcal{P}\}$.

Now we are ready to state the main result of this section, which says that we can partition the vertex set of a dense k -graph H randomly, so that all parts inherit the properties of H , and in particular, the probability that a vertex is in a specific part is at most $O(1/n)$. Note that to be applicable in the proof of Theorem ?? (assumption ?? therein), we do not assume that the partition is a good partition – we allow a small fraction of the pairs of vertices in the same part to be not reachable.

Lemma 5.1. *Let $k, \ell \in \mathbb{N}$ and let F be an r -vertex k -graph. Define $d, q, s, t, C, n \in \mathbb{N}$ and $\alpha, \beta, \mu, \gamma, \eta, \varepsilon, c > 0$ where $1/n \ll \alpha, \eta \ll 1/C' \ll 1/C \ll \beta, \mu, \gamma, \varepsilon, c, 1/k, 1/\ell, 1/r, 1/d, 1/q \leq 1$, and $r \mid n$, $r \mid C$. Let H be an n -vertex k -graph such that all but at most $\alpha \binom{n}{\ell}$ ℓ -sets have degree at least $(\delta + \gamma) \binom{n-\ell}{k-\ell}$, and let $\mathcal{P} = \{V_1, \dots, V_d\}$ be a partition of $V(H)$, satisfying the following properties:*

- (P1) for every $i \in [d]$, all but $\eta \binom{|V_i|}{2}$ of the pairs $\{u, v\} \subseteq V_i$ are (F, β, t) -reachable in H and $|V_i| \geq cn$;
(P2) for every $v \in V(H)$, $|F_{\mathcal{P}}^\mu(v)| \geq \varepsilon n^{r-1}$.

Then there exists a random partition $\mathcal{U} = \{U_1, U_2, \dots, U_m\}$ of $V(H)$ with the following properties:

- (R1) for every $2 \leq i \leq m$, we have $|U_1|$ equals $(C-1)C$ plus the remainder when n is divided by $(C-1)C$, and $|U_i| = C$; for every $i \in [m]$ and $j \in [d]$, $|U_i \cap V_j| \geq c|U_i|/2$;
(R2) for every $2 \leq i \leq m$, there exists a set $T_i \subseteq U_i$ with $|T_i \cap V_j| = rq$ for each $j \in [d]$ and a set $L_i \subseteq U_i \setminus T_i$ of size r such that $\mathbf{i}_{\mathcal{P}}(L_i) \in I_{\mathcal{P}, F}^\mu(H)$ and $H[L_i]$ contains a copy of F , denoted by F_i ; let $L_1 = \emptyset$; moreover, for every $2 \leq i \leq m$ and every vertex $v \in T_i$, we have $|F_{\mathcal{P}}^\mu(v, U_{i-1})| \geq \varepsilon |U_{i-1}|^{r-1}/2$;
(R3) for every $i \in [m]$, $\delta_\ell(H[U_i \setminus L_i]) \geq (\delta + \gamma/2) \binom{|U_i \setminus L_i|}{k-\ell}$, and for every $j \in [d]$, $V_j \cap (U_i \setminus L_i)$ is $(F, \beta/2, t)$ -closed in $H[U_i \setminus L_i]$;
(R4) for every $i \in [m]$, $I_{\mathcal{P}, F}^\mu(H) \subseteq I_{\mathcal{P}^{U_i \setminus L_i}, F}^{\mu/2}(H[U_i \setminus L_i])$;
(R5) for every set of distinct vertices $y_1, \dots, y_s \in V(H)$ and every function $f : [s] \rightarrow [m]$,

$$\mathbb{P}[y_i \in U_{f(i)} \text{ for each } i \in [s]] \leq \left(\frac{C'}{n}\right)^s.$$

Now let us explain the properties of the random partition U_i found by Lemma ???. First, ?? reflects the probability of the location of the vertices, which is exactly what we want for getting the vertex-spread. The other properties together are designated so that we could find an F -factor in each cluster U_i : the size of the clusters must be a multiple of r , every part must inherit the minimum degree of H , and the reachability property and the robust-lattice structure should also be inherited. Additional complexities come from the redistribution step, where a small portion of the clusters failing to have these properties will be broken into vertices and distributed randomly to the good ones. In that step it is possible but rather complicated or tedious to show that the new cluster is closed. Here an easier fix is to put the newly added vertex in a robust copy of F , with vertex set denoted by L_i , and then state our reachability and lattice properties only for $U_i \setminus L_i$. Moreover, in ?? we include small sets $T_i \subseteq U_i$, which will be used to correct the divisibility in the proof of Theorem ?? – when $\mathbf{i}_{\mathcal{P}^{U_i \setminus L_i}}(U'_i)$ is not in $L_{\mathcal{P}^{U_i \setminus L_i}}^{\mu/2}(H[U_i \setminus L_i])$ where U'_i denotes the current i -th cluster, we will add a subset of T_{i+1} to U'_i so that the new index vector does belong to the lattice, which allows us to build an F -factor in the i -th cluster. Then this correction can be done in a recursive manner.

Before presenting the full proof, we outline the key ideas behind it. The argument is based on the random redistribution technique introduced in [?], and we modify their framework to suit our setting by adjusting both the technical assumptions and the properties we aim to achieve. We begin with randomly partitioning $V(H)$ into clusters U'_1, \dots, U'_m . This initial partition is, in a sense, already “close” to the one required in Lemma ???: the fifth condition is satisfied, and the remaining conditions are violated by only a small number of clusters. We refer to a cluster U'_i as *bad* if it violates any of the required properties stated in the lemma. Following the idea of [?], we then randomly redistribute the vertices from the bad clusters into the good ones in a way that preserves the key structural properties of the latter.

For the remainder of this section, fix

$$1/n \ll \alpha, \eta \ll 1/C' \ll 1/C \ll \beta, \mu, \gamma, \varepsilon, c, 1/k, 1/\ell, 1/r, 1/d \leq 1$$

where $r \mid n$ and $r \mid C$ as in Lemma ???. Let C_1 be $(C-1)C$ plus the remainder when n is divided by $(C-1)C$. Let $W_1 := [C_1]$, and for $i \geq 2$, let $W_i := [C_1 + (C-1)(i-1)] \setminus \cup_{j=1}^{i-1} W_j = [C_1 + (C-1)(i-2) + 1, C_1 + (C-1)(i-1)]$. Note that $\{W_1, \dots, W_m\}$ is a partition of $[n]$ for $m := (n - C_1)/(C-1) + 1$. Let H be a k -graph with vertex set $V := \{v_1, \dots, v_n\}$ such that all but at most $\alpha \binom{n}{\ell}$ ℓ -sets have degree at least $(\delta + \gamma) \binom{n-\ell}{k-\ell}$, and let $\mathcal{P} = \{V_1, \dots, V_d\}$ be a partition of

V as described in Lemma ?? . Let $\pi : [n] \rightarrow [n]$ be a uniformly random permutation of $[n]$, and let $U'_i := \{v_j : \pi(j) \in W_i\}$.

To prepare for the proof of Lemma ??, we begin by establishing a sequence of auxiliary lemmas, all under the assumptions of Lemma ?? . Each lemma verifies a specific aspect of the desired structure, and together they provide the framework needed for the random redistribution argument to succeed. While these lemmas address the individual components, the complete proof will be presented at the end of this section.

The following lemma ensures that the second part of ?? holds. Let $C^* := |U'_i|$, that is, $C^* = C_1$ if $i = 1$ and $C^* = C - 1$ if $i > 1$. So $C^* \geq C - 1 \geq C/2$. Note that Lemmas ??–?? are under the assumptions of Lemma ??.

Lemma 5.2. *Let E_1 be the event that the following hold:*

- *For at least $(1 - \exp(-c^2C/20))m$ indices $i \in \{2, \dots, m\}$, we have $|V_j \cap U'_i| \geq 2c|U'_i|/3$ for all $j \in [d]$;*
- *The same holds for U'_1 , i.e., $|V_j \cap U'_1| \geq 2c|U'_1|/3$ for all $j \in [d]$.*

Then, $\mathbb{P}[E_1] \geq 99/100$.

Proof. Note that for every $j \in [d]$, $|V_j| \geq cn$ as stated in Lemma ?? . Let $f_j : \binom{V}{C^*} \rightarrow \mathbb{R}$ be defined by $f_j(X) = |V_j \cap X|$ for each $X \in \binom{V}{C^*}$. Note that $|f_j(X_1) - f_j(X_2)| \leq 1$ for any $X_1, X_2 \in \binom{V}{C^*}$ with $|X_1 \cap X_2| = C^* - 1$. By linearity of expectation, $\mathbb{E}[f_j(U'_i)] = \sum_{v \in V_j} \mathbb{P}[v \in U'_i] = |V_j| \cdot \frac{\binom{|U'_i|-1}{n-1}}{\binom{|U'_i|}{n}} \geq c|U'_i|$ for every $i \in [m]$. We can then apply Lemma ?? with $c = 1, m = C^*$ and $\alpha = c\sqrt{C^*}/3$, and get that for $i \in [m]$ and $j \in [d]$

$$\mathbb{P}[f_j(U'_i) = |V_j \cap U'_i| < 2c|U'_i|/3] \leq 2\exp(-2c^2C^*/9) \leq \exp(-c^2C/10).$$

It follows that, with probability at most $\exp(-c^2C/10)d$, there exists some $j \in [d]$ such that $|V_j \cap U'_1| < 2c|U'_1|/3$. On the other hand, let X denote the number of indices $i \in \{2, \dots, m\}$ such that $|V_j \cap U'_i| < 2c|U'_i|/3$ for some $j \in [d]$. Then by linearity of expectation, $\mathbb{E}[X] \leq \exp(-c^2C/10)dm$. By Markov's inequality, $\mathbb{P}[X > \exp(-c^2C/20)m] \leq \exp(-c^2C/20)d$. Combining this with the probability bound for U'_1 , we obtain the desired conclusion. \square

To ensure condition ??, we next establish two lemmas that address its key components. Lemma ?? ensures that, at the random redistribution stage, each vertex has many good clusters into which it can be added.

Lemma 5.3. *Let E_2 be the event that for every vertex $v \in V$, there are at least $(1 - 1/C^2)m$ indices $i \in \{2, \dots, m\}$ such that $|F_{\mathcal{P}}^\mu(v, U'_i)| \geq 2\varepsilon|U'_i|^{r-1}/3$. Then, $\mathbb{P}[E_2] \geq 99/100$.*

Proof. Note that for each vertex $v \in V(H)$, $|F_{\mathcal{P}}^\mu(v)| \geq \varepsilon n^{r-1}$ as stated in Lemma ?? . For every vertex $v \in V$ and $i \in [m]$, by Lemma ?? with U'_i playing the role of A , we have

$$(5.1) \quad \mathbb{P}\left[|F_{\mathcal{P}}^\mu(v, U'_i)| < 2\varepsilon|U'_i|^{r-1}/3\right] \leq 2\exp(-\varepsilon^2C/20).$$

Let X_v denote the number of random sets U'_i such that $|F_{\mathcal{P}}^\mu(v, U'_i)| \geq 2\varepsilon|U'_i|^{r-1}/3$. By linearity of expectation, we have $\mathbb{E}[X_v] \geq (1 - 2\exp(-\varepsilon^2C/20))m$. Note that interchanging two elements of π can affect the value of X_v by at most 2 and if $X_v \geq x$, this can be certified by at most $2C^2x$ choices of the random permutation. Therefore, by Lemma ?? applied with $c = 2, r = 2C^2$, and $t = m/C^3$, we have

$$\mathbb{P}[X_v < (1 - 1/C^2)m] \leq 4\exp\left(-\frac{(m/C^3)^2}{64C^2m}\right) \leq \exp(-n/C^{10}).$$

By the union bound, we have that $X_v \geq (1 - 1/C^2)m$ for every vertex $v \in V$ with probability at least 99/100, as desired. \square

Lemma ?? guarantees that most clusters admit many vertices that can be added while preserving the desired properties.

Lemma 5.4. *Let E_3 be the event that for all but at most $\exp(-\varepsilon^2 C/100)m$ indices $i \in \{2, \dots, m\}$, there are at least $(1 - \exp(-\varepsilon^2 C/50))n$ vertices $v \in V$ such that $|F_{\mathcal{P}}^\mu(v, U'_i)| \geq 2\varepsilon|U'_i|^{r-1}/3$. Then, $\mathbb{P}[E_3] \geq 99/100$.*

Proof. Call v bad for U'_i if $|F_{\mathcal{P}}^\mu(v, U'_i)| < 2\varepsilon|U'_i|^{r-1}/3$. For any U'_i , by Lemma ?? and linearity of expectation, we have $\mathbb{E}[|\{v \in V : v \text{ is bad for } U'_i\}|] \leq 2 \exp(-\varepsilon^2 C/20)n$. Then, for any U'_i , we have $\mathbb{P}[|\{v \in V : v \text{ is bad for } U'_i\}| \geq \exp(-\varepsilon^2 C/50)n] \leq 2 \exp(-\varepsilon^2 C/50)$ by Markov's inequality. Let X be the number of $i \in [2, m]$ such that $|\{v \in V : v \text{ is bad for } U'_i\}| \geq \exp(-\varepsilon^2 C/50)n$. By linearity of expectation, $\mathbb{E}[X] \leq 2 \exp(-\varepsilon^2 C/50)m$. By Markov's inequality, $\mathbb{P}[X \geq \exp(-\varepsilon^2 C/100)m] \leq 2 \exp(-\varepsilon^2 C/100) \leq 1/100$, as C is sufficiently large, implying the desired statement. \square

The following lemma is crucial for establishing conclusion ?? of the theorem.

Lemma 5.5. *Let E_4 be the event that the following hold for all but at most*

$$(\exp(-\gamma^2 C/200) + \exp(-\beta^2 C/200))m$$

indices $i \in \{2, \dots, m\}$ and for $i = 1$: $\delta_\ell(H[U'_i]) \geq (\delta + 2\gamma/3) \binom{|U'_i| - \ell}{k - \ell}$, and $V_j \cap U'_i$ is $(F, 2\beta/3, t)$ -closed in $H[U'_i]$ for all $j \in [d]$. Then, $\mathbb{P}[E_4] \geq 99/100$.

Proof. Fix $i \in [m]$. By Lemma ??, we have

$$\mathbb{P}\left[\delta_\ell(H[U'_i]) < (\delta + 2\gamma/3) \binom{|U'_i| - \ell}{k - \ell}\right] \leq \binom{C^*}{\ell} (\alpha + \exp(-\gamma^2 C^*/36)) \leq \exp(-\gamma^2 C/100)$$

and by Lemma ?? and the union bound, we have

$$\begin{aligned} & \mathbb{P}[\exists j \in [d], V_j \cap U'_i \text{ is not } (F, 2\beta/3, t)\text{-closed in } H[U'_i]] \\ & \leq d \binom{C^*}{2} (\eta + 2 \exp(-\beta^2 C^*/18)) \leq \exp(-\beta^2 C/100), \end{aligned}$$

as $\alpha, \eta \ll 1/C^* \ll \beta, \gamma, 1/\ell, 1/d$. In particular, the above two inequalities hold for U'_1 .

On the other hand, let X_1 be the number of indices $i \in \{2, \dots, m\}$ such that $\delta_\ell(H[U'_i]) < (\delta + 2\gamma/3) \binom{|U'_i| - \ell}{k - \ell}$, and then $\mathbb{E}[X] \leq \exp(-\gamma^2 C/100)m$. By Markov's inequality, we have

$$\mathbb{P}[X_1 \geq \exp(-\gamma^2 C/200)m] \leq \exp(-\gamma^2 C/200).$$

Similarly, let X_2 be the number of indices $i \in \{2, \dots, m\}$ for which there exists $j \in [d]$ such that $V_j \cap U'_i$ is not $(F, 2\beta/3, t)$ -closed in $H[U'_i]$. Thus, we have $\mathbb{P}[X_2 \geq \exp(-\beta^2 C/200)m] \leq \exp(-\beta^2 C/200)$. By the union bound, we get $\mathbb{P}[E_4] \geq 1 - 2 \exp(-\gamma^2 C/200) - 2 \exp(-\beta^2 C/200) \geq 99/100$, as desired. \square

The next lemma will be used in the proof of ??, providing a key probabilistic bound.

Lemma 5.6. *Let E_5 be the event that the following hold for at least $(1 - \exp(-\mu^2 C/50))m$ indices $i \in \{2, \dots, m\}$ and for $i = 1$: every $\mathbf{v} \in I_{\mathcal{P}, F}^\mu(H)$ satisfies $\mathbf{v} \in I_{\mathcal{P}^{U'_i, F}}^{2\mu/3}(H[U'_i])$. Then, $\mathbb{P}[E_5] \geq 99/100$.*

Proof. For every $\mathbf{v} \in I_{\mathcal{P}, F}^\mu(H)$ and $i \in [m]$, we say that $\mathbf{v} \in I_{\mathcal{P}, F}^\mu(H)$ is bad for U'_i if $H[U'_i]$ contains fewer than $2\mu|U'_i|^r/3$ copies of F with index vector \mathbf{v} . Since $\mathcal{P}^{U'_i}$ is the restriction of \mathcal{P} to U'_i , we have $\mathbf{i}_{\mathcal{P}^{U'_i}}(S) = \mathbf{i}_{\mathcal{P}}(S)$ for any subset $S \subseteq U'_i$. It follows that if \mathbf{v} is not bad for U'_i , then $\mathbf{v} \in I_{\mathcal{P}^{U'_i, F}}^{2\mu/3}(H[U'_i])$.

Next, we estimate the probability that a fixed $\mathbf{v} \in I_{\mathcal{P}, F}^\mu(H)$ is bad for some U'_i ; by Lemma ?? (with U'_i playing the role of A), this probability is at most $2 \exp(-\mu^2 C/20)$. Note that $|I_{\mathcal{P}, F}^\mu(H)| \leq$

$|L_{\max}^d| \leq \binom{r+d-1}{d}$. Then, by the union bound, the probability that there exists some $\mathbf{v} \in I_{\mathcal{P},F}^\mu(H)$ bad for U_1' is at most

$$2 \exp(-\mu^2 C/20) \binom{r+d-1}{d} \leq \exp(-\mu^2 C/50).$$

On the other hand, let X denote the number of indices $i \in \{2, \dots, m\}$ for which there exists some $\mathbf{v} \in I_{\mathcal{P},F}^\mu(H)$ that is bad for U_i' . By linearity of expectation, $\mathbb{E}[X] \leq 2 \exp(-\mu^2 C/20) \binom{r+d-1}{d} m$. By Markov's inequality,

$$\mathbb{P}[X \geq \exp(-\mu^2 C/50)m] \leq 2 \exp(-\mu^2 C/50) \binom{r+d-1}{d} \leq \exp(-\mu^2 C/100).$$

Combining this with the bound for U_1' , and applying the union bound, we obtain the desired conclusion. \square

We shall show that U_1' satisfies the properties ??-?? with $U_1 = U_1'$. The next lemma is important for establishing that U_1' satisfies ??. For every $i \in [m]$, choose some $T_i \subseteq U_i'$ of size rqd as follows. Let \mathcal{T}_i denote the collection of all rqd -subsets $T_i \subseteq U_i'$ such that $|T_i \cap V_j| = rq$ for all $j \in [d]$.

- If $\mathcal{T}_i \neq \emptyset$, then we fix $T_i \in \mathcal{T}_i$ arbitrarily.
- If $\mathcal{T}_i = \emptyset$, then we choose any $T_i \in \binom{U_i'}{rqd}$ arbitrarily.

Given such a choice of T_1, \dots, T_m , we consider an auxiliary digraph $D = D(T_1, \dots, T_m)$ with vertex set $[m]$, where $(i, j) \in E(D)$ if for every vertex $v \in T_i$, we have $|F_{\mathcal{P}}^\mu(v, U_j')| \geq 2\varepsilon|U_i'|^{r-1}/3$. Note that condition ?? in Lemma ?? requires that the sets T_i lie in \mathcal{T}_i . Although it is possible that $\mathcal{T}_i = \emptyset$ for some i , Lemma ?? implies that the number of such “bad” indices is small with high probability. In the proof of Lemma ??, we will reassign these bad clusters and retain only those U_i for which $\mathcal{T}_i \neq \emptyset$, thereby ensuring that the final selection of T_i satisfies the desired condition.

Lemma 5.7. *Let E_6 be the event that U_1' satisfies $d_D^-(1) \geq (1 - \exp(-\varepsilon^2 C/50))(m-1)$. Then, $\mathbb{P}[E_6] \geq 99/100$.*

Proof. Recall that for every $v \in V$, (??) says $\mathbb{P}[|F_{\mathcal{P}}^\mu(v, U_1')| < 2\varepsilon|U_1'|^{r-1}/3] \leq 2 \exp(-\varepsilon^2 C/20)$. This implies that for every $i \in \{2, \dots, m\}$, $\mathbb{P}[(i, 1) \in E(D)] \geq 1 - 2 \exp(-\varepsilon^2 C/20)rqd$. By linearity of expectation, $\mathbb{E}[d_D^-(1)] \geq (1 - 2 \exp(-\varepsilon^2 C/20)rqd)(m-1)$. Then by Markov's inequality,

$$\mathbb{P}[m-1 - d_D^-(1) \geq \exp(-\varepsilon^2 C/50)(m-1)] \leq 2 \exp(-\varepsilon^2 C/50)rqd,$$

implying the desired result. \square

We now proceed to prove Lemma ?? by considering the distribution of U_1', \dots, U_m' conditioning on the events established in the previous lemmas, and then applying the random redistribution argument.

Proof of Lemma ??. We condition on $E_1 \cap \dots \cap E_6$, which holds with probability at least $94/100$ by Lemmas ??-??. Note that $d_D^+(i) \geq (1 - rqd/C^2)m$ for every $i \in [m]$ by Lemma ?? assuming that E_2 holds.

For each permutation $\pi : [n] \rightarrow [n]$, define a set of “bad” clusters $\mathcal{F}_\pi \subseteq \{U_2', \dots, U_m'\}$ which includes U_i' if any of the following holds:

- (A1) $|V_j \cap U_i'| < 2c|U_i'|/3$ for some $j \in [d]$,
- (A2) $d_D^-(i) < (1 - rqd/\sqrt{C})m$,
- (A3) $|F_{\mathcal{P}}^\mu(v, U_i')| < 2\varepsilon|U_i'|^{r-1}/3$ for at least $\exp(-\varepsilon^2 C/50)n$ vertices $v \in V$,
- (A4) $\delta_\ell(H[U_i']) < (\delta + 2\gamma/3) \binom{|U_i'| - \ell}{k - \ell}$ or there exists $j \in [d]$ such that $V_j \cap U_i'$ is not $(F, 2\beta/3, t)$ -closed in $H[U_i']$, and

$$(A5) \quad I_{\mathcal{P},F}^\mu(H) \not\subseteq I_{\mathcal{P}^{U'_i,F}}^{2\mu/3}(H[U'_i]).$$

Let $m' := |U'_2 \cup \dots \cup U'_m|/C$, and note that m' is a positive integer by the choice of r . We claim that if $\pi \in E_1 \cap \dots \cap E_6$, then $|\mathcal{F}_\pi| \leq m - 1 - m' = m - 1 - (m - 1)(C - 1)/C = (m - 1)/C$. Indeed, there are at most $\exp(-c^2C/20)m$ $U'_i \in \mathcal{F}_\pi$ of type ?? by Lemma ?? assuming $\pi \in E_1$, at most $m/C^{3/2}$ $U'_i \in \mathcal{F}_\pi$ of type ?? since $\sum_{i=1}^m d_D^-(i) = \sum_{i=1}^m d_D^+(i) \geq (1 - rqd/C^2)m^2$ by Lemma ?? assuming $\pi \in E_2$, at most $\exp(-\varepsilon^2C/100)m$ $U'_i \in \mathcal{F}_\pi$ of type ?? by Lemma ?? assuming $\pi \in E_3$, at most $(\exp(-\gamma^2C/200) + \exp(-\beta^2C/200))m$ $U'_i \in \mathcal{F}_\pi$ of type ?? by Lemma ?? assuming $\pi \in E_4$, and there are at most $\exp(-\mu^2C/50)m$ $U'_i \in \mathcal{F}_\pi$ of type ?? by Lemma ?? assuming $\pi \in E_5$. Now we add arbitrary clusters to \mathcal{F}_π to ensure $|\mathcal{F}_\pi| = (m - 1)/C$.

By possibly relabeling the sets U'_1, \dots, U'_m , we may assume that $\mathcal{F}_\pi = \{U'_i : i \in [m] \setminus [m+1-|\mathcal{F}_\pi|]\}$. Now consider random sets U'_i given by a random permutation π conditioned on $E_1 \cap \dots \cap E_6$.

Next, we define a bipartite graph G_π between $A := \bigcup_{U'_i \in \mathcal{F}_\pi} U'_i$ and $B := \{U'_2, \dots, U'_m\} \setminus \mathcal{F}_\pi$. Note that by the choice of m' , we have $(m - 1 - m')(C - 1) = m'$, which implies that both two parts have size m' . We put an edge between $u \in A$ and $U'_i \in B$ if there exists a copy of F in $H[(U'_i \setminus T_i) \cup \{u\}]$ containing u and with index vector in $I_{\mathcal{P},F}^\mu(H)$. Conditioning on E_2 , we have $d_{G_\pi}(v) \geq (1 - 1/C^2)m - |\mathcal{F}_\pi| \geq 3m'/4$ for all $v \in A$ as $s \ll C$. Since all sets satisfying ?? are in \mathcal{F}_π , for every $U'_i \in B$, we have $d_{G_\pi}(U'_i) \geq m' - \exp(\gamma^2C/50)n \geq 3m'/4$. Thus, $\delta(G_\pi) \geq 3m'/4$, and by Lemma ??, there is a $(C??/n)$ -edge-spread distribution on perfect matchings of G_π .

Now we define the random sets U_i as follows. First, sample π from the uniform distribution on permutations of $[n]$ conditioning on $E_1 \cap \dots \cap E_6$. Then, sample M_π from the $(C??/m)$ -edge-spread distribution on perfect matchings of G_π . For each $U'_i \in B$, let $u_i \in A$ be the vertex that it is assigned to in M_π , and define $U_i := U'_i \cup \{u_i\}$. Let $U_1 := U'_1$, where u_1 denotes an arbitrary vertex selected from U_1 . This concludes the algorithm that defines $\mathcal{U} := \{U_1, \dots, U_{m'+1}\}$ with m in the lemma statement being $m' + 1$. Finally, we verify the properties ??-?? required by the theorem (Note that U_1 already satisfies ??, ?? and ??).

- (1) For $i \geq 2$, we have $|U_i| = |U'_i| + 1 = C + 1$, as required, and the second condition in ?? holds by the choice of \mathcal{F}_π that contain all sets satisfying ??.
- (2) Now we explain ??. Note that for $i \in [m' + 1]$, we have $|U'_i \cap V_j| \geq c|U_i|/2$ for all $j \in [d]$, which implies that $T_i \neq \emptyset$. Thus, all T_i satisfy the requirement in ??. For the first property in ??, we take $L_i \subseteq U_i \setminus T_i$ as follows: the case $i = 1$ is trivial since $L_1 = \emptyset$; for $i \geq 2$, we take $L_i = V(F_i)$ where F_i is a copy of F containing u_i with $\mathbf{i}_{\mathcal{P}}(V(F_i)) \in I_{\mathcal{P},F}^\mu(H)$ guaranteed by the definition of the edges of G_π . Next, we verify the second half of ??. For $i = 1$, $d_D^-(1) \geq (1 - \exp(\varepsilon^2C/50))(m - 1)$ because we condition on E_6 , and for $i \in [2, m' + 1]$, we have $d_D^-(i) \geq (1 - rqd/\sqrt{C})m$ since these U'_i are good sets and so satisfy the negation of ??. Let D' be the subgraph of D induced on the first $m' + 1$ vertices, and then we have $d_{D'}^-(i) \geq m' - rqdm/\sqrt{C} \geq (m' + 1)/2$. Note that for every $i \in [m' + 1]$, $d_{D'}^+(i) \geq m' - rqdm/C^2 \geq (m' + 1)/2$. Thus, $\delta^0(D') \geq |D'|/2$. By a theorem of Ghouila-Houri[?], D' contains a directed Hamilton cycle, which we may assume, without loss of generality, have edges $(i, i - 1)$ for all $2 \leq i \leq m' + 1$. Then by the definition of D and the assumption, we have for every $2 \leq i \leq m' + 1$ and every vertex $v \in T_i$, $|F_{\mathcal{P}}^\mu(v, U_{i-1})| \geq |F_{\mathcal{P}}^\mu(v, U'_{i-1})| \geq 2\varepsilon|U'_{i-1}|^{r-1}/3 \geq \varepsilon|U_{i-1}|^{r-1}/2$.
- (3) For $i \geq 2$, ?? holds by the choice of \mathcal{F}_π that contain all sets satisfying ?? and C is sufficiently large.
- (4) For $i \geq 2$, ?? holds by the choice of \mathcal{F}_π that contain all sets satisfying ?? and C is sufficiently large.

- (5) Let $y_1, \dots, y_s \in V(H)$ and $f : [s] \rightarrow [m]$. For $i \in [s]$, let D_i be the event that $y_i \in U_{f(i)}$, let D_i^1 be the event that $y_i \in U'_{f(i)}$, and let D_i^2 be the event that $y_i = u_{f_i}$. Note that

$$(5.2) \quad \bigcap_{i \in [s]} D_i = \bigcap_{i \in [s]} (D_i^1 \cup D_i^2) = \bigcup_{S \subseteq [n]} \left(\bigcap_{i \in S} D_i^1 \cap \bigcap_{i \in [s] \setminus S} D_i^2 \right).$$

Next, we will estimate the probability that the event $\bigcap_{i \in S} D_i^1 \cap \bigcap_{i \in [n] \setminus S} D_i^2$ holds. Note that D_i^1 is completely determined by the permutation π , and D_i^2 holds only if y_i is matched to U'_i by M_π . Then, for every $S \subseteq [n]$, we have

$$\mathbb{P} \left[\bigcap_{i \in S} D_i^1 \right] \leq \left(\frac{C^{2|S|} (n - |S|)!}{n!} \right) / \mathbb{P}[E_1 \cap \dots \cap E_6] \leq \frac{100}{94} \left(\frac{eC^2}{n} \right)^{|S|},$$

and for every $\pi' \in E_1 \cap \dots \cap E_6$,

$$\mathbb{P} \left[\bigcap_{i \in [s] \setminus S} D_i^2 \mid \pi = \pi' \right] \leq \left(\frac{C_{??}}{m'} \right)^{s - |S|}.$$

Therefore, for every $S \subseteq [n]$, as C' is sufficiently large, we obtain that

$$\mathbb{P} \left[\bigcap_{i \in S} D_i^1 \cap \bigcap_{i \in [s] \setminus S} D_i^2 \right] = \mathbb{P} \left[\bigcap_{i \in S} D_i^1 \right] \mathbb{P} \left[\bigcap_{i \in [s] \setminus S} D_i^2 \mid \bigcap_{i \in S} D_i^1 \right] \leq \frac{100}{94} \left(\frac{eC^2}{n} \right)^{|S|} \left(\frac{C_{??}}{m'} \right)^{s - |S|} \leq \left(\frac{C'}{2n} \right)^s.$$

Finally, the result follows by (??) and the union bound over the 2^s choices of $S \subseteq [s]$.

The proof is completed. \square

6. PROOF OF THEOREM ??

In this section, we prove Theorem ?. First we introduce some definitions.

Suppose $L \subseteq L_{\max}^{|\mathcal{P}|}$ is a lattice in $\mathbb{Z}^{|\mathcal{P}|}$, where \mathcal{P} is a partition of a set V . The *coset group* of (\mathcal{P}, L) is $Q = Q(\mathcal{P}, L) := L_{\max}^{|\mathcal{P}|} / L$. For any $\mathbf{i} \in L_{\max}^{|\mathcal{P}|}$, the *residue* of \mathbf{i} in Q is $R_Q(\mathbf{i}) := \mathbf{i} + L$. For any $A \subseteq V$ of size divisible by r , the *residue* of A in Q is $R_Q(A) := R_Q(\mathbf{i}_{\mathcal{P}}(A))$.

Let $q \in \mathbb{N}$. A (possibly empty) F -packing M in H of size at most q is a q -*solution* for $(\mathcal{P}, L_{\mathcal{P}, F}^\mu(H))$ (in H) if $\mathbf{i}_{\mathcal{P}}(V(H) \setminus V(M)) \in L_{\mathcal{P}, F}^\mu(H)$; we say that $(\mathcal{P}, L_{\mathcal{P}, F}^\mu(H))$ is q -*soluble* if it has a q -solution.

The following theorem from [?] provides a necessary and sufficient condition for H to contain a perfect F -packing.

Theorem 6.1 ([?]). *Let $k, \ell \in \mathbb{N}$ where $\ell \leq k - 1$ and let F be an r -vertex k -graph. Define $D, q, t, n_0 \in \mathbb{N}$ and $\beta, \mu, \gamma, c > 0$ where*

$$1/n_0 \ll \beta, \mu \ll \gamma, c, 1/r, 1/D, 1/q, 1/t.$$

Let H be a k -graph on $n \geq n_0$ vertices where r divides n . Suppose that

- (i) $\delta_\ell(H) \geq (\delta(F, \ell, D) + \gamma) \binom{n - \ell}{k - \ell}$;
- (ii) $\mathcal{P} = \{V_1, \dots, V_d\}$ is an (F, β, t, c) -good partition of $V(H)$;
- (iii) $|Q(\mathcal{P}, L_{\mathcal{P}, F}^\mu(H))| \leq q$.

Then H contains an F -factor if and only if $(\mathcal{P}, L_{\mathcal{P}, F}^\mu(H))$ is q -soluble.

For the value of q , it is easy to see a trivial upper bound $|L_{\max}^{|\mathcal{P}|}| = \binom{d+r-1}{r}$, which suffices for the proof of our main technical result.

Proof of Theorem ??. Suppose we have the constants satisfying the following hierarchy

$$1/n_0 \ll \alpha, \eta \ll 1/C'' \ll 1/C' \ll 1/C \ll \beta, \mu \ll \gamma, \varepsilon, c, 1/k, 1/\ell, 1/r, 1/d \leq 1.$$

Let $n \geq n_0$ be a multiple of r , and let the k -graph G be an F -packing on n vertices. Let H be an n -vertex k -graph satisfying the assumptions of the theorem. The main task of this proof is to define $\psi : G \hookrightarrow H$, a random embedding of an F -factor in H , and we achieve it in a few steps.

Step 1: Sample random clusters. Applying Lemma ?? with $q = \binom{d+r-1}{r}$, we obtain a random partition $\mathcal{U} = \{U_1, \dots, U_m\}$ of $V(H)$ that satisfies properties ??-??.

Step 2: Adjust the random partition. Our goal is to apply Theorem ?? to each random cluster and obtain an F -factor in each of them. By ?? and ??, it suffices to establish that each cluster is q -soluble. For this we adjust the partition by moving at most rq vertices from the vertex set T_{i+1} to U_i so that the solubility condition is satisfied. Below we explain how this is achieved.

We do this in an inductive manner. Let $i \in [m-1]$ and suppose we have corrected the first $i-1$ clusters, that is, the clusters $\{U'_j : j < i\}$ satisfy $\mathbf{i}_{\mathcal{P}}(U'_j) \in L_{\mathcal{P},F}^\mu(H)$ and $U'_j \subseteq U_j \cup T_{j+1}$. Let J_i denote the set of vertices removed from U_i in the previous step (they were added to U'_{i-1}). Now we consider the residues of the remaining set of vertices in the i -th cluster with respect to \mathcal{P} . Note that $\mathbf{i}_{\mathcal{P}}(V(H)) \in L_{\mathcal{P},F}^\mu(H)$ and by the assumption, we have $R_G(U_i \setminus J_i) + \sum_{j>i} R_G(U_j) = \mathbf{0} + L_{\mathcal{P},F}^\mu(H)$. Suppose $R_G(U_i \setminus J_i) = \mathbf{v}_0 + L_{\mathcal{P},F}^\mu(H)$ for some $\mathbf{v}_0 \in L_{\max}^d$ and we get $\sum_{j>i} R_G(U_j) = -\mathbf{v}_0 + L_{\mathcal{P},F}^\mu(H)$.

Claim 6.2. There exists r -sets $S_1, \dots, S_{d'} \subseteq T_{i+1}$ for some $d' \leq q-1$ such that

$$\sum_{j \in [d']} R_G(S_j) = -\mathbf{v}_0 + L_{\mathcal{P},F}^\mu(H).$$

Proof. Fix any family of r -sets $S_1, \dots, S_h \subseteq \bigcup_{j>i} U_j$ for $h \geq q$ that satisfies $\sum_{j \in [h]} R_G(S_j) = -\mathbf{v}_0 + L_{\mathcal{P},F}^\mu(H)$, consider $h+1$ partial sums $\sum_{j \in [h']} R_G(S_j)$ for $0 \leq h' \leq h$. Since $|G(\mathcal{P}, L_{\mathcal{P},F}^\mu(H))| \leq q$, by the pigeonhole principle, there exists $j_1 < j_2$ such that $\sum_{j_1 \leq j \leq j_2} R_G(S_j) = \mathbf{0} + L_{\mathcal{P},F}^\mu(H)$. Then we delete them from the left hand side of the equation, and we repeat this process until there are at most $q-1$ r -sets left. By possibly relabeling the sets S_i 's, we may assume that $\sum_{j \in [d']} R_G(S_j) = -\mathbf{v}_0 + L_{\mathcal{P},F}^\mu(H)$. Recall that the size of intersection of T_{i+1} with each part of \mathcal{P} is rq , then we can take disjoint r -sets $S'_1, \dots, S'_{d'} \subseteq T_{i+1}$ such that $\mathbf{i}_{\mathcal{P}}(S'_j) = \mathbf{i}_{\mathcal{P}}(S_j)$ for all $j \in [d']$, which implies $\sum_{j \in [d']} R_G(S'_j) = -\mathbf{v}_0 + L_{\mathcal{P},F}^\mu(H)$. \square

So we have $\left| \bigcup_{i \in [d']} S_i \right| \leq rq$ and $\sum_{i \in [d']} \mathbf{i}_{\mathcal{P}}(S_i) + \mathbf{i}_{\mathcal{P}}(U_i \setminus J_i) \in L_{\mathcal{P},F}^\mu(H)$. Let $U'_i := \bigcup_{i \in [d']} S_i \cup (U_i \setminus J_i)$, and we have $\mathbf{i}_{\mathcal{P}}(U'_i) \in L_{\mathcal{P},F}^\mu(H)$. Finally let $U'_m := U_m \setminus J_m$, then we get $\mathbf{i}_{\mathcal{P}}(U'_m) \in L_{\mathcal{P},F}^\mu(H)$ as $\mathbf{i}_{\mathcal{P}}(V(H)) \in L_{\mathcal{P},F}^\mu(H)$.

Step 3: Find F -factor in each new cluster.

Let $J_{m+1} := \emptyset$. Then for every $i \in [m]$, we have $U'_i = (U_i \setminus J_i) \cup J_{i+1}$, where $J_i \subseteq T_i$, $J_{i+1} \subseteq T_{i+1}$, and $|J_i|, |J_{i+1}| \leq rq$. Note that by ??, for $2 \leq i \leq m$, F_i is a copy of F in $H[U'_i]$ such that $\mathbf{i}_{\mathcal{P}}(V(F_i)) \in I_{\mathcal{P},F}^\mu(H)$, and for every vertex $v \in J_i$, we have $|F_{\mathcal{P}}^\mu(v, U_{i-1})| \geq \varepsilon |U_{i-1}|^{r-1}/2$. Thus, we find an F -packing M_1 in U'_i : take F_i (if $i \geq 2$) and for each vertex $v \in J_{i+1}$ (if $i \leq m-1$), pick a copy F_v of F that contains v and satisfies $\mathbf{i}_{\mathcal{P}}(V(F_v)) \in I_{\mathcal{P},F}^\mu(H)$. The copies of F can be chosen greedily, as $H[U'_i]$ contains at least $\varepsilon |U'_i|^{r-1}/3$ such copies of F for every vertex $v \in J_{i+1}$, and we have $|V(M_1)| \leq r + r^2q$.

Next, consider the induced subgraph $H_i := H[U'_i \setminus V(M_1)]$ and its partition $\mathcal{P}_i := \mathcal{P}^{H_i}$. Note that $|V(H_i)| \geq |U'_i| - rq - (r + r^2q) \geq |U'_i| - 2r^2q$. We claim that H_i and \mathcal{P}^{H_i} satisfy the properties described in Theorem ??. By applying Theorem ?? to H_i and \mathcal{P}^{H_i} , we obtain an F -factor M_2 in H_i . Then $M_1 \cup M_2$ is an F -factor of $H[U'_i]$.

To establish the claim, we first show that $(\mathcal{P}_i, L_{\mathcal{P}_i, F}^{\mu/3}(H_i))$ is 0-soluble, i.e. that $\mathbf{i}_{\mathcal{P}_i}(V(H_i)) \in L_{\mathcal{P}_i, F}^{\mu/3}(H_i)$. Indeed, observe that $\mathbf{i}_{\mathcal{P}}(V(H_i)) = \mathbf{i}_{\mathcal{P}}(U'_i \setminus V(M_1)) \in L_{\mathcal{P}, F}^{\mu}(H)$. By ?? and the definition of index vector, it follows that $\mathbf{i}_{\mathcal{P}_i}(V(H_i)) = \mathbf{i}_{\mathcal{P}}(V(H_i)) \in L_{\mathcal{P}, F}^{\mu}(H) \subseteq L_{\mathcal{P}^{U_i \setminus L_i}, F}^{\mu/2}(H[U_i \setminus L_i])$. Moreover, for every $\mathbf{v} \in I_{\mathcal{P}^{U_i \setminus L_i}, F}^{\mu/2}(H[U_i \setminus L_i])$, there exist at least $\frac{\mu}{2}|U_i \setminus L_i|^r - 2r^2q|V(H_i)|^{r-1} \geq \frac{\mu}{3}|V(H_i)|^r$ copies of F in H_i with index vector \mathbf{v} . Therefore, $\mathbf{v} \in I_{\mathcal{P}_i, F}^{\mu/3}(H_i)$, which implies $L_{\mathcal{P}^{U_i \setminus L_i}, F}^{\mu/2}(H[U_i \setminus L_i]) \subseteq L_{\mathcal{P}_i, F}^{\mu/3}(H_i)$. Therefore the claim is proved.

Furthermore, $|Q(\mathcal{P}_i, L_{\mathcal{P}_i, F}^{\mu/3}(H_i))| \leq q$ as \mathcal{P}_i has d parts. In addition, by ?? and that C is sufficiently large, we have $\delta_{\ell}(H_i) \geq \delta_{\ell}(H[U_i \setminus L_i]) - 2r^2q \binom{C-\ell}{k-\ell} \geq (\delta(F, \ell, D) + \gamma/3) \binom{|V(H_i)|-\ell}{k-\ell}$. Similarly, also by ??, for every $j \in [d]$, $V_j \cap V(H_i)$ is $(F, \beta/3, t)$ -closed in H_i and $|V_j \cap V(H_i)| \geq c|V(H_i)|/3$. Therefore, for $i \in [m]$, each pair (H_i, \mathcal{P}_i) satisfies the assumptions of Theorem ??, and thus has an F -factor. The union of these F -factors form a (random) F -factor of H .

Defining ψ . Now we are ready to define the random embedding function and conclude the proof. Assume without loss of generality that $V(G) = [n]$ and $G[\{(j-1)r+1, \dots, jr\}] \cong F$ for every $j \in [n/r]$. For every $i \in [n]$, let $\psi(i) := u_{x,y}$ where x is the largest integer such that $i > \sum_{j=1}^{x-1} |U'_j|$ and $y = i - \sum_{j=1}^{x-1} |U'_j|$.

To complete the proof, by Proposition ??, it suffices to show that the random embedding ψ is $(2C'/n)$ -vertex-spread. Let $s \in [n]$. For every two sequences of distinct vertices $x_1, \dots, x_s \in [n]$ and $y_1, \dots, y_s \in V(H)$, we need to show that $\mathbb{P}[\psi(x_i) = y_i \text{ for } i \in [s]] \leq (2C'/n)^s$.

To that end, let $C_1 = |U_1|$, $W_1 := [C_1]$, and for $i \geq 2$, let $W_i := [C_1 + C(i-2) + 1, C_1 + C(i-1)] = [C_1 + C(i-1)] \setminus \bigcup_{j=1}^{i-1} W_j$. Let $w : [s] \rightarrow [m]$ where $w(i)$ be the unique index such that $x_i \in W_{w(i)}$.

Recall that $\mathcal{U} = \{U_1, \dots, U_m\}$ is the collection of random sets. Since $U'_i \subseteq U_i \cup T_{i+1}$, the embedding ψ induces injections between W_i and $U_i \cup U_{i+1}$ for each $i \in [m]$. Hence, for every $s \in [n]$, $\psi(x_i) = y_i$ implies $y_i \in U_{w(i)} \cup U_{w(i)+1}$. Thus, by ??,

$$\mathbb{P}[\forall i \in [s], \psi(x_i) = y_i] \leq \mathbb{P}[\forall i \in [s], y_i \in U_{w(i)} \cup U_{w(i)+1}] \leq \left(\frac{2C'}{n}\right)^s,$$

that is, the random embedding ψ is $(2C'/n)$ -vertex-spread.

By Proposition ??, we conclude that there is a $(C''/n^{1/m_1(F)})$ -spread distribution on subgraphs of H which are isomorphic to G (note that $m_1(F) = m_1(G)$). The *moreover* part of the theorem follows from Theorem ??. \square

7. PROOF OF THE REMAINING THEOREMS

In this section, we deduce Theorems ??, and ?? as consequences of Theorem ??.

7.1. Algorithm: proof of Theorem ??. To reduce the time complexity, we present the following result whose proof is similar to that of Lemma 6.4 in [?] and is deferred to the appendix. We say that a set I of k -vectors is *full* if for every $(k-1)$ -vector \mathbf{v} there is some $i \in [d]$ such that $\mathbf{v} + \mathbf{u}_i \in I$.

Lemma 7.1. *Suppose $k \geq 3$ and L is a lattice in $\mathbb{Z}^{\mathcal{P}}$, where \mathcal{P} is a partition of a set V . If L contains a full set of k -vectors, then $|Q(\mathcal{P}, L)| \leq |\mathcal{P}|$.*

The above lemma yields the following fact, which provides a tighter bound on the relevant coset group.

Fact 7.2. Fix an integer $k \geq 3$. Suppose $\mu \ll c, 1/k$. Let H be a k -graph on n vertices such that $\delta_{k-1}(H) \geq n/k$, and let \mathcal{P} be a partition of $V(H)$ in which each part has size at least cn . Then $|Q(\mathcal{P}, L_{\mathcal{P}, F}^{\mu}(H))| \leq |\mathcal{P}|$.

Proof. By Lemma ??, it suffices to show that $L_{\mathcal{P},F}^\mu(H)$ contains a full set of k -vectors. Fix any $(k-1)$ -vector \mathbf{v} , and let $\mathcal{P} = \{V_1, \dots, V_d\}$. Since there are at least $\binom{\min_{i \in [d]} |V_i|}{k-1}$ $(k-1)$ -sets with index vector \mathbf{v} and $\delta_{k-1}(H) \geq n/k$, the number of edges containing such a set is at least $\frac{1}{k} \cdot \frac{n}{k} \cdot \binom{\min_{i \in [d]} |V_i|}{k-1} \geq \frac{1}{k} \cdot \frac{n}{k} \cdot \binom{cn}{k-1} \geq d\mu n^k$. By averaging, there exists $i \in [d]$ such that at least μn^k edges with index vector $\mathbf{v} + \mathbf{u}_i$. \square

We are now ready to prove Theorem ??.

Proof of Theorem ??. We show correctness of Algorithm ?? and estimate its running time. Set $t = k2^k + 2$ and $q = \left\lceil Q(\mathcal{P}, L_{\mathcal{P},F}^\mu(H)) \right\rceil$.

First, we apply Lemma ?? which finds \mathcal{P} satisfying ??–?? in time $O(n^{2^{k-2}k+1})$. Note that by Theorem ??, if H_p does not contain any of the q -solutions of H , then H_p contains no perfect matching. By line 5, for every set of r -vectors $\mathbf{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_{q'}\}$ with $\mathbf{i}_{\mathcal{P}}(V(H)) - \sum_{j \in [q']} \mathbf{v}_j \in L_{\mathcal{P}}^\mu(H)$, we check if H_p contains a matching with index vector $\sum_{j \in [q']} \mathbf{v}_j$.

Claim 7.3. For $\mathbf{v} \in I_{\mathcal{P}}^\eta(H)$, a.a.s. H_p contains a matching of size qk with index vector of all edges equal to \mathbf{v} .

Proof. Let $h = qk$, and randomly partition $V(H)$ into h parts of nearly equal size U_1, \dots, U_h . For every $i \in [h]$, let X_i be the number of edges in $H[U_i]$ with index vector \mathbf{v} , and let Y_i be the number of such edges that appear in H_p . It suffices to show that $\mathbb{P}[\forall i \in [h], Y_i \geq 1] = 1 - o(1)$.

By Lemma ?? with U_i playing the role of A , we have

$$\mathbb{P}[X_i < \frac{\eta}{2} |U_i|^k] \leq \exp\left(-\frac{\eta^2 n}{10h}\right) = o(1).$$

Conditioning on $X_i \geq \frac{\eta}{2} |U_i|^k$, we have $\mathbb{E}[Y_i] \geq \frac{\eta}{2} |U_i|^k p$, so by Chernoff's bound, $\mathbb{P}[Y_i = 0 \mid X_i \geq \frac{\eta}{2} |U_i|^k] \leq \exp\left(-\frac{\eta |U_i|^k p}{4}\right) = o(1)$. Thus, we get $\mathbb{P}[Y_i = 0] = o(1)$ and the claim follows by the union bound over all $i \in [h]$. \square

By the claim, it suffices to check if H contains a matching M' with $\mathbf{i}_{\mathcal{P}}(V(M')) = \sum_{\mathbf{v} \in \mathbf{V}^\eta} \mathbf{v}$ – we can a.a.s. extend M' to a desired matching because $|V(M')| \leq k|M'|$. If no such matching is found, on line 9 the algorithm terminates with negative output, and indeed H_p has no perfect matching as it does not contain any q -solution. Otherwise, suppose the algorithm finds a matching M' . Note that M' a.a.s. can be extended to a matching M with $\mathbf{i}_{\mathcal{P}}(V(H) \setminus V(M)) \in L_{\mathcal{P}}^\mu(H)$ by the claim. Indeed, for every $\mathbf{v} \in \mathbf{V} \cap I_{\mathcal{P}}^\eta(H)$, we expose exactly ηn^k edges in H with index vector \mathbf{v} and by the claim, a.a.s. we find a matching of size qk . Note that these edges have the property that given any set of $(q-1)k$ vertices in H , we can still find an edge disjoint from these vertices. Therefore we can construct M from M' greedily, and the probability of failure is $o(1)$ as there are constantly many such vectors. Now suppose we find such M successfully and it remains to show that a.a.s. the remaining k -graph has a perfect matching. To this end, we shall verify that the assumptions ??–?? of Theorem ?? hold for the remaining k -graph.

Recall that $\delta(k, k-1, k) = 1/k$ by Theorem ???. Note that the algorithm never reveals any μ -robust edges of H , and may have revealed at most ηn^k edges for every $\mathbf{v} \notin I_{\mathcal{P}}^\mu(H)$. Let E' be the set of these revealed edges. Indeed we need to work on the k -graph $H - V(M)$ with the edges of E' removed, and let us denote this k -graph by H' . Note that E' contains at most $\eta n^k |L_{\max}^{\mathcal{P}}| \leq \eta(2kn)^k$ edges of $H - V(M)$ (possibly revealed as non-edges). As $\eta \ll 1/C \ll \gamma$, we get that at most $\frac{\eta(2kn)^k}{\gamma n/2} \leq \sqrt{\eta} \binom{n}{k-1}$ $(k-1)$ sets S satisfy that $\deg_{H'}(S) < n/k + \gamma n/2$, verifying ??. Moreover, recall that all parts of \mathcal{P} are (β, t) -closed, that is, for vertices u, v from the same part of \mathcal{P} , there are at least βn^{tk-1} reachable $(tk-1)$ -sets for u, v . Note that at most $|E'| n^{tk-k} \leq \eta(2k)^k n^{tk}$ (tk) -sets of vertices contain edges of E' . Note that for vertices u, v and one of their reachable $(tk-1)$ -sets S ,

S may no longer be a reachable set in H' only if $\{u\} \cup S$ or $\{v\} \cup S$ coincide with such a (tk) -set. Partition $\bigcup_{i \in [d]} \binom{V_i}{2}$ into n stars, and we say a star loses a reachable set if any of its edges uv loses a reachable set by passing from H to H' . Then we see that there are at most $\frac{\eta(2k)^k n^{tk}}{\beta n^{tk-1}/2} \leq \sqrt{\eta}n$ of the stars each losing at least $\beta n^{tk-1}/2$ reachable sets, where we used $\eta \ll 1/C \ll \beta$. Note that if a star loses less than $\beta n^{tk-1}/2$ reachable sets, all of its edges remain $(\beta/2, t)$ -reachable in H' ; otherwise, all edges of the star may become not $(\beta/2, t)$ -reachable in H' . Therefore, all but at most $\sqrt{\eta}n^2 \leq \eta^{1/4}(n/k)^2$ of the pairs of vertices of each V_i are $(\beta/2, t)$ -reachable in H' , verifying ???. Finally, since the exposed edges are not μ -robust, this does not affect the condition ??, verifying ??.

Therefore, we can apply Theorem ?? with $\eta^{1/4}$ in place of η and $\beta/2$ in place of β to H' and conclude that a.a.s. H'_p contains a perfect matching.

For the running time, note that Lemma ?? finds \mathcal{P} in time $O(n^{2^{k-2}k+1})$ and there are constantly many choices for the vector set \mathbf{V} and for each of them the search takes time $O(n^{kq}) = O(n^{k(k-1)})$ which follows from $q \leq |\mathcal{P}| \leq k-1$ by Fact ??. Thus, the overall running time is $O(n^{2^{k-2}k+1})$. \square

7.2. Proof of Theorem ??. Now we prove Theorem ??. To have a more precise threshold for strictly 1-balanced F , we consider the spread properties of the F -complex of G . For hypergraphs F and G , the F -complex of G , denoted by G_F , is the $|V(F)|$ -uniform multi-hypergraph with vertex set $V(G)$ in which every copy of F in G corresponds to a distinct hyperedge of G_F on the same set of vertices. Note that G has an F -factor if and only if G_F contains a perfect matching. Let $\mathbb{G}_F(n, p)$ denote the binomial random multi-hypergraph on n vertices where every edge of the F -complex of the complete hypergraph is included independently with probability p .

To prove Theorem ??, we need the following result of Riordan [?, Theorem 18].

Theorem 7.4 ([?]). *For every $r \in \mathbb{N}$, there exists $a = a_{??}(r)$ such that the following holds. Let F be a fixed strictly 1-balanced graph. If $p = p(n) \leq \log^2(n)/n^{1/m_1(F)}$, then, for some $\pi = \pi(n) \sim ap^{e_F}$, there exists a coupling for $G = \mathbb{G}(n, p)$ and $G_F = \mathbb{G}_F(n, \pi)$ such that, a.a.s. for every F -edge present in G_F the corresponding copy of F is present in G .*

Now we are ready to prove Theorem ??.

Proof of Theorem ??. To prove the theorem we consider two cases.

We first prove the second part of the theorem – if G contains an F -factor then H contains an F -packing M_0 of size at most $\binom{2r-2}{r}$ such that $\mathbf{i}_{\mathcal{P}}(V(G) \setminus V(M_0)) \in L_{\mathcal{P}, F}^\mu(G)$. Note that $|\mathcal{P}| \leq r-1$, so $|\mathcal{Q}(\mathcal{P}, L_{\mathcal{P}, F}^\mu(G))| \leq |L_{\max}^\mu| \leq \binom{2r-2}{r}$. It then follows from Theorem ?? that $(\mathcal{P}, L_{\mathcal{P}, F}^\mu(G))$ is $\binom{2r-2}{r}$ -soluble, which implies the existence of a desired F -packing.

Conversely, assume that there exists an F -packing M_0 of size at most $\binom{2r-2}{r}$ satisfying $\mathbf{i}_{\mathcal{P}}(V(G) \setminus V(M_0)) \in L_{\mathcal{P}, F}^\mu(G)$. Let $G' := G - V(M_0)$. Recall that $\delta(F, 1, 5r^2) = 1 - 1/\chi_{cr}(F)$ by Theorem ??. Since by removing M_0 we lose at most $\binom{2r-2}{r}r$ vertices, the minimum degree, the number of reachable sets between any two vertices and the number of robust copies containing any vertex only reduced by a small amount. So we can apply Theorem ?? to H' with $\gamma/2, \beta/2, \varepsilon/2$ in place of $\gamma, \beta, \varepsilon$, and obtain a $(C_1/n^{1/m_1(F)})$ -spread distribution on the set of F -factors in G' , which implies that a.a.s. G'_p contains an F -factor if $p \geq Cn^{-1/m_1(F)} \log n$.

We now turn to the case when F is strictly 1-balanced. It suffices to prove the result for $p = Cn^{-1/m_1(F)} \log^{1/e_F} n$. Denote the F -complex of G' by \mathcal{G}' . As there is an $(C_1/n^{1/m_1(F)})$ -spread distribution λ on the family of F -factors in G' , it gives rise to a $(C_1^{e_F}/n^{r-1})$ -spread distribution λ' on the perfect matchings of \mathcal{G}' . Indeed, suppose S is a matching of size ℓ in \mathcal{G}' (if S does not form a matching then $\lambda'(S) = 0$), then we have

$$\lambda'(S) = \lambda(T) \leq C_1^{e_F \cdot \ell} n^{-\frac{1}{m_1(F)} e_F \cdot \ell} = C_1^{e_F \cdot \ell} n^{-(r-1)\ell},$$

where T is the F -packing of size ℓ in G corresponding to S and we used $|T| = e_F \cdot \ell$ and $m_1(F) = e_F/(r-1)$. Therefore, Theorem ?? implies that a.a.s. \mathcal{G}'_π contains a perfect matching if $\pi \geq K??C_1^{e_F} \log n/n^{(r-1)} = ap^{e_F}$. Thus, by Theorem ??, a.a.s. G'_p contains an F -factor. \square

7.3. The enumeration aspect. Corollary ?? and Corollary ?? follow directly from the fact below and Theorem ??. Define $\mathcal{M}(H)$ to be the set of perfect F -packings of H . If H satisfies the divisibility conditions described in Theorem ?? (i.e., there exists a matching M_0 satisfying the required properties), then we can obtain a good spread in the subgraph $H' := H - V(M_0)$ by Theorem ??. By the fact below, it follows that $|\mathcal{M}(H')| \geq (\varepsilon(n-c))^{\frac{e_F}{v_F m_1(F)}(n-c)}$, where $c \leq v_F \binom{2v_F-2}{v_F}$ is a constant. Observe that every perfect F -packing M in H' can be extended to a perfect F -packing $M \cup M_0$ in H . Hence, $|\mathcal{M}(H)| \geq |\mathcal{M}(H')| \geq (\varepsilon n/2)^{\frac{e_F}{v_F m_1(F)}n}$.

Fact 7.5. Let F be an r -vertex k -graph, and H an n -vertex k -graph. If there is a $(C/n^{1/m_1(F)})$ -spread distribution on the set of F -factors in H , then H contains at least $(\varepsilon n)^{\frac{e_F}{r m_1(F)}n}$ F -factors for $\varepsilon = C^{-1/m_1(F)}$.

Proof. By assumption, we have a $(C/n^{1/m_1(F)})$ -spread distribution \mathbf{M} on the set of F -factors in H . Fix an F -factor $M \in \mathcal{M}(H)$ consisting of $e_F n/r$ edges. Since \mathbf{M} is $(C/n^{1/m_1(F)})$ -spread, we have $\mathbb{P}[\mathbf{M} = M] \leq (C/n^{1/m_1(F)})^{e_F n/r}$. Thus,

$$1 = \sum_{M \in \mathcal{M}(H)} \mathbb{P}[\mathbf{M} = M] \leq |\mathcal{M}(H)| (C/n^{1/m_1(F)})^{e_F n/r},$$

which implies $|\mathcal{M}(H)| \geq (n^{m_1(F)}/C)^{e_F n/r} = (\varepsilon n)^{\frac{e_F}{r m_1(F)}n}$, as desired. \square

8. CONCLUDING REMARKS

In this paper, we established an algorithmic result (Theorem ??) that guarantees, with high probability, the existence of a perfect matching in a random subgraph of H when the algorithm accepts, and the non-existence of such a matching when it rejects.

While our algorithm relies on probabilistic techniques, it would be of great interest to develop a direct, deterministic algorithm for this problem. However, the random subgraph H_p with $p \approx C \log n/n^{k-1}$ is extremely sparse, with minimum degree $\Omega(\log n)$, and most pairs of vertices have collective degree zero. This sparsity presents significant challenges for deterministic decision procedures. To the best of our knowledge, no result of this kind is currently known, and we leave this as an open question.

In contrast to the case of perfect matching, we are unable to obtain a similar algorithm as in Theorem ?? for the setting of F -packings. If we follow the proof idea of Theorem ??, one may reveal many copies of F in the random subgraph to testify specific divisibility conditions. These revealed copies are non-robust (i.e. at most ηn^r many). Although this exposure affects the degree and partition conditions only mildly, it may critically interfere with the third condition required by Theorem ??, namely, that each vertex may lose too many robust copies of F . The reason is that even though the revealed copies of F are non-robust, their edges may still overlap with robust ones, thereby reducing the total number of robust copies. In contrast, in the perfect matching setting, the exposure process only reveals individual edges and does not affect the collection of robust edges at any vertex.

We note that while our main approach follows the randomized algorithm of Kelly, Müyesser, and Pokrovskiy, we also explored alternative methods based on iterative absorption, as developed by Pham, Sah, Sawhney, and Simkin [?]. While we think both approaches work, we chose to present the proof using the method of Kelly et al. as it is conceptually simpler.

REFERENCES

- [1] P. Allen, J. Böttcher, J. Corsten, E. Davies, M. Jenssen, P. Morris, B. Roberts, and J. Skokan. A robust Corrádi–Hajnal theorem. *Random Structures & Algorithms*, 65(1):61–130, 2024.
- [2] J. Alman and V. V. Williams. A refined laser method and faster matrix multiplication. *TheoretCS*, 3, 2024.
- [3] N. Alon and R. Yuster. H -factors in dense graphs. *Journal of Combinatorial Theory, Series B*, 66(2):269–282, 1996.
- [4] A. Asadpour, U. Feige, and A. Saberi. Santa Claus meets hypergraph matchings. In *Approximation, randomization and combinatorial optimization*, volume 5171 of *Lecture Notes in Comput. Sci.*, pages 10–20. Springer, Berlin, 2008.
- [5] O. Cooley, D. Kühn, and D. Osthus. Perfect packings with complete graphs minus an edge. *European Journal of Combinatorics*, 28(8):2143–2155, 2007.
- [6] B. Cuckler and J. Kahn. Hamiltonian cycles in Dirac graphs. *Combinatorica*, 29:299–326, 2009.
- [7] J. Edmonds. Paths, trees, and flowers. *Canadian Journal of mathematics*, 17:449–467, 1965.
- [8] P. Erdős. On extremal problems of graphs and generalized graphs. *Israel Journal of Mathematics*, 2(3):183–190, 1964.
- [9] A. Ferber, L. Hardiman, and A. Mond. Counting Hamilton cycles in Dirac hypergraphs. *Combinatorica*, 43(4):665–680, 2023.
- [10] A. Ferber and M. Kwan. Dirac-type theorems in random hypergraphs. *Journal of Combinatorial Theory, Series B*, 155:318–357, 2022.
- [11] K. Frankston, J. Kahn, B. Narayanan, and J. Park. Thresholds versus fractional expectation-thresholds. *Annals of Mathematics*, 194(2):475–495, 2021.
- [12] L. Gan and J. Han. On the Keevash-Knox-Mycroft conjecture. *Journal of Combinatorial Theory, Series B*, 174:214–242, 2025.
- [13] A. Ghouila-Houri. Une condition suffisante d’existence d’un circuit hamiltonien. *Comptes Rendus de l’Académie des Sciences Paris*, 251:495–497, 1960.
- [14] S. Glock, S. Gould, F. Joos, D. Kühn, and D. Osthus. Counting Hamilton cycles in Dirac hypergraphs. *Combinatorics, Probability and Computing*, 30(4):631–653, 2021.
- [15] S. Glock, D. Kühn, A. Lo, and D. Osthus. The existence of designs via iterative absorption: hypergraph F -designs for arbitrary F . *Mem. Amer. Math. Soc., to appear*, 2022.
- [16] J. Han. Near perfect matchings in k -uniform hypergraphs. *Combinatorics, Probability and Computing*, 24(5):723–732, 2015.
- [17] J. Han. Decision problem for perfect matchings in dense k -uniform hypergraphs. *Transactions of the American Mathematical Society*, 369(7):5197–5218, 2017.
- [18] J. Han and P. Keevash. Finding perfect matchings in dense hypergraphs. In *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2366–2377, 2020.
- [19] J. Han and A. Treglown. The complexity of perfect matchings and packings in dense hypergraphs. *Journal of Combinatorial Theory, Series B*, 141:72–104, 2020.
- [20] A. Johansson, J. Kahn, and V. Vu. Factors in random graphs. *Random Structures & Algorithms*, 33(1):1–28, 2008.
- [21] F. Joos, R. Lang, and N. Sanhueza-Matamala. Robust Hamiltonicity. *arXiv preprint arXiv:2312.15262*, 2023.
- [22] D. Y. Kang, T. Kelly, D. Kühn, D. Osthus, and V. Pfenninger. Perfect matchings in random sparsifications of Dirac hypergraphs. *Combinatorica*, pages 1–34, 2024.
- [23] R. M. Karp. Reducibility among combinatorial problems. In *50 Years of Integer Programming 1958-2008: from the Early Years to the State-of-the-Art*, pages 219–241. Springer, 2009.
- [24] M. Karpiński, A. Ruciński, and E. Szymańska. Computational complexity of the perfect matching problem in hypergraphs with subcritical density. *International Journal of Foundations of Computer Science*, 21(06):905–924, 2010.
- [25] P. Keevash. The existence of designs. *arXiv e-prints*, page arXiv:1401.3665, Jan. 2014.
- [26] P. Keevash, F. Knox, and R. Mycroft. Polynomial-time perfect matchings in dense hypergraphs. In *Proceedings of the forty-fifth annual ACM symposium on Theory of Computing*, pages 311–320, 2013.
- [27] P. Keevash and R. Mycroft. A geometric theory for hypergraph matching. *Memoirs of the American Mathematical Society*, 233(1098), 2015.
- [28] T. Kelly, A. Müyesser, and A. Pokrovskiy. Optimal spread for spanning subgraphs of Dirac hypergraphs. *Journal of Combinatorial Theory, Series B*, 169:507–541, 2024.
- [29] D. G. Kirkpatrick and P. Hell. On the complexity of general graph factor problems. *SIAM Journal on Computing*, 12(3):601–609, 1983.
- [30] J. Komlós, G. Sárközy, and E. Szemerédi. Proof of the Alon–Yuster conjecture. *Discrete Mathematics*, 235(1-3):255–269, 2001.

- [31] M. Krivelevich, C. Lee, and B. Sudakov. Robust Hamiltonicity of Dirac graphs. *Transactions of the American Mathematical Society*, 366(6):3095–3130, 2014.
- [32] D. Kühn and D. Osthus. Critical chromatic number and complexity of perfect packings in graphs. In *Proceedings of the 17th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 851–859, 2006.
- [33] A. Liebenau and N. Wormald. Asymptotic enumeration of graphs by degree sequence, and the degree sequence of a random graph. *Journal of the European Mathematical Society*, 26(1):1–40, 2023.
- [34] A. Lo and K. Markström. F -factors in hypergraphs via absorption. *Graphs and Combinatorics*, 31:679–712, 2015.
- [35] C. McDiarmid et al. On the method of bounded differences. *Surveys in combinatorics*, 141(1):148–188, 1989.
- [36] M. Molloy and B. Reed. *Graph colouring and the probabilistic method*, volume 23. Springer Science & Business Media, 2002.
- [37] H. T. Pham, A. Sah, M. Sawhney, and M. Simkin. A toolkit for robust thresholds. *arXiv preprint arXiv:2210.03064*, 2022.
- [38] O. Riordan. Random cliques in random graphs and sharp thresholds for F -factors. *Random Structures & Algorithms*, 61(4):619–637, 2022.
- [39] V. Rödl, A. Ruciński, and E. Szemerédi. Perfect matchings in large uniform hypergraphs with large minimum collective degree. *Journal of Combinatorial Theory, Series A*, 116(3):613–636, 2009.
- [40] H. Ryser. Neuere probleme in der kombinatorik. *Vorträge über Kombinatorik, Oberwolfach*, pages 69–91, 1967.
- [41] G. N. Sárközy, S. M. Selkow, and E. Szemerédi. On the number of Hamiltonian cycles in Dirac graphs. *Discrete Mathematics*, 265(1-3):237–250, 2003.
- [42] A. Shokoufandeh and Y. Zhao. Proof of a tiling conjecture of Komlós. *Random Structures & Algorithms*, 23(2):180–205, 2003.
- [43] E. Szymańska. The complexity of almost perfect matchings and other packing problems in uniform hypergraphs with high codegree. *European Journal of Combinatorics*, 34(3):632–646, 2013.
- [44] W. T. Tutte. The factorization of linear graphs. *Journal of the London Mathematical Society*, 1(2):107–111, 1947.

APPENDIX A. PROOF OF LEMMA ??

Here we provide the full proof of Lemma ?. We first state a lemma analogous to Lemma 6.1 from [?]. For a vector \mathbf{v} , we write $\mathbf{v}(i)$ to denote its i th coordinate.

Lemma A.1. *Suppose $k \geq 3$, and let L be a lattice in \mathbb{Z}^P , where \mathcal{P} is a partition of a set V with $|\mathcal{P}| = d$. If L contains a full set of k -vectors, then*

- (A1) *for any $i_1, i_2, i_3 \in [d]$, there exists $i_4 \in [d]$ such that $\mathbf{u}_{i_1} + \mathbf{u}_{i_2} - \mathbf{u}_{i_3} - \mathbf{u}_{i_4} \in L$,*
- (A2) *for any $\mathbf{v} \in L_{\max}^d$ and $i \in [d]$, there exists $j \in [d]$ such that $\mathbf{v} - \mathbf{u}_i + \mathbf{u}_j \in L$.*

Proof. We first verify ?. Fix arbitrary $i_1, i_2, i_3 \in [d]$, and choose an arbitrary $(k-3)$ -vector \mathbf{w} . Since $\mathbf{w} + \mathbf{u}_{i_1} + \mathbf{u}_{i_2}$ is a $(k-1)$ -vector and L contains a full set of k -vectors, there exists $j \in [d]$ such that $\mathbf{w} + \mathbf{u}_{i_1} + \mathbf{u}_{i_2} + \mathbf{u}_j \in L$. Similarly, there is $i_4 \in [d]$ such that $\mathbf{w} + \mathbf{u}_{i_3} + \mathbf{u}_j + \mathbf{u}_{i_4} \in L$. Then $\mathbf{u}_{i_1} + \mathbf{u}_{i_2} - \mathbf{u}_{i_3} - \mathbf{u}_{i_4}$ is the difference of these two index vectors and hence lies in L .

Now, we prove ?. For any $\mathbf{v} \in L_{\max}^d$, consider $\mathbf{v}' \in L$ that minimises $\sum_{i \in [d]} |\mathbf{v}(i) - \mathbf{v}'(i)|$ subject to $\sum_{i \in [d]} \mathbf{v}(i) = \sum_{i \in [d]} \mathbf{v}'(i)$. We claim that $\sum_{i \in [d]} |\mathbf{v}(i) - \mathbf{v}'(i)| \leq 2$. Indeed, if not there exist $i_1, i_2, i_3 \in [d]$ satisfying one of the following:

- $i_1 \neq i_2$, $\mathbf{v}(i_1) - \mathbf{v}'(i_1) > 0$, $\mathbf{v}(i_2) - \mathbf{v}'(i_2) > 0$ and $\mathbf{v}(i_3) - \mathbf{v}'(i_3) < 0$,
- $i_1 = i_2$, $\mathbf{v}(i_1) - \mathbf{v}'(i_1) > 1$ and $\mathbf{v}(i_3) - \mathbf{v}'(i_3) < 0$,
- $i_1 \neq i_2$, $\mathbf{v}(i_1) - \mathbf{v}'(i_1) < 0$, $\mathbf{v}(i_2) - \mathbf{v}'(i_2) < 0$ and $\mathbf{v}(i_3) - \mathbf{v}'(i_3) > 0$,
- $i_1 = i_2$, $\mathbf{v}(i_1) - \mathbf{v}'(i_1) < -1$ and $\mathbf{v}(i_3) - \mathbf{v}'(i_3) > 0$.

In each case, by ?, we obtain $\mathbf{w} := \mathbf{u}_{i_1} + \mathbf{u}_{i_2} - \mathbf{u}_{i_3} - \mathbf{u}_{i_4} \in L$ for some $i_4 \in [d]$. Then either $\mathbf{v}' + \mathbf{w}$ or $\mathbf{v}' - \mathbf{w}$ contradicts the minimality of \mathbf{v}' . Thus the claim holds, and we have $\mathbf{v}' = \mathbf{v} - \mathbf{u}_{j_1} + \mathbf{u}_{j_2}$ for some $j_1, j_2 \in [d]$. For any $i \in [d]$, by ? again, there is some $j \in [d]$ such that $\mathbf{u}_i + \mathbf{u}_{j_1} - \mathbf{u}_{j_2} - \mathbf{u}_j \in L$, denoted by \mathbf{w}' . Then $\mathbf{v} + \mathbf{u}_i - \mathbf{u}_j = \mathbf{v}' + \mathbf{w}' \in L$, as desired. \square

Now we present the proof of Lemma ?.

Proof. Fix any $(k-1)$ -vector \mathbf{w} , and let $d := |\mathcal{P}|$. We claim that every coset $\mathbf{v} + L$ of L in L_{\max}^P contains an index vector $\mathbf{w} + \mathbf{u}_i$ for some $i \in [d]$. Indeed, since L contains a full set of k -vectors,

there exists $j \in [d]$ such that $\mathbf{w} + \mathbf{u}_j \in L$. Moreover, by Lemma ????, we obtain $-\mathbf{v} - \mathbf{u}_j + \mathbf{u}_i \in L$ for some $i \in [d]$. Thus, $\mathbf{w} + \mathbf{u}_i = \mathbf{v} + (-\mathbf{v} - \mathbf{u}_j + \mathbf{u}_i) + (\mathbf{w} + \mathbf{u}_j) \in \mathbf{v} + L$, as claimed. Therefore, $|Q(\mathcal{P}, L)| \leq |\mathcal{P}|$. \square

SCHOOL OF MATHEMATICS AND STATISTICS AND CENTER OF APPLIED MATHEMATICS, BEIJING INSTITUTE OF TECHNOLOGY, BEIJING, CHINA, 100081. EMAIL: {han.jie|jingwen.zhao}@bit.edu.cn.