J. Eur. Math. Soc. 27, 801–875 (2025) DOI 10.4171/JEMS/1388

Patrick Morris

Clique factors in pseudorandom graphs

Received July 2, 2021; revised July 1, 2022

Abstract. An n-vertex graph is said to to be (p, β)-*bijumbled* if for any vertex sets $A, B \subseteq V(G)$, we have

$$e(A, B) = p|A|\,|B| \pm \beta\sqrt{|A|\,|B|}.$$

We prove that for any $r \in \mathbb{N}_{\geq 3}$ and $c > 0$ there exists an $\varepsilon > 0$ such that any n-vertex (p, β)-bijumbled graph with $n \in r\mathbb{N}$, $p > 0$, $\delta(G) \geq cpn$ and $\beta \leq \varepsilon p^{r-1}n$ contains a K_r-factor. This implies a corresponding result for the stronger pseudorandom notion of (n, d, λ)-graphs.

For the case of triangle factors, that is, when $r = 3$, this result resolves a conjecture of Krivelevich, Sudakov and Szabó from 2004 and it is tight due to a pseudorandom triangle-free construction of Alon. In fact, in this case even more is true: as a corollary to this result and a result of Han, Kohayakawa, Person and the author, we can conclude that the same condition of $\beta = o(p^2 n)$ actually guarantees that a (p, β)-bijumbled graph G contains every graph on n vertices with maximum degree at most 2.

Keywords. Pseudorandom graphs, clique factors, extremal graph theory

Contents

1. Introduction . . . 802
2. Proof of main theorem . . . 806
3. Preliminaries . . . 817
   3.1. Notation . . . 817
   3.2. Properties of bijumbled graphs . . . 818
   3.3. Concentration of random variables . . . 823
   3.4. Perfect fractional matchings . . . 823
   3.5. Almost perfect matchings in hypergraphs . . . 825
   3.6. Templates . . . 828
4. Diamond trees . . . 829
   4.1. Scattered diamond trees . . . 831
5. Cascading absorption through orchards . . . 834
   5.1. Absorbing orchards . . . 835
   5.2. Shrinkable orchards . . . 838
6. Shrinkable orchards of small order . . . 840
   6.1. From orchards to systems . . . 840

Patrick Morris: Department of Mathematics, Universitat Politècnica de Catalunya, Barcelona, Spain; pmorrismaths@gmail.com

Mathematics Subject Classification (2020): Primary 05C35; Secondary 05C48, 05C70

6.2. Sufficient conditions for shrinkability . . . 841
6.3. The existence of shrinkable orchards of small order . . . 844
6.4. Controlling degrees relative to removable sets of vertices . . . 845
6.5. The existence of shrinkable orchards of larger order . . . 847
7. Shrinkable orchards of large order . . . 848
7.1. A density condition which guarantees shrinkability . . . 848
7.2. Popular diamond trees . . . 850
7.3. The existence of shrinkable orchards of large order . . . 852
7.4. Preprocessing the orchard . . . 854
7.5. Completing the orchard . . . 857
7.6. The existence of shrinkable orchards of smaller order . . . 859
8. The final absorption . . . 860
8.1. Defining an absorbing structure . . . 861
8.2. Finding an absorbing structure . . . 863
9. Concluding remarks . . . 871
References . . . 872

1. Introduction

We say a graph G contains a K_r*-factor* if there is a collection of vertex disjoint copies of K_r that completely cover the vertex set of G. When $r = 3$, we often refer to a K_3-factor as a *triangle factor*. As a natural generalisation of a perfect matching in a graph, K_r-factors are a fundamental object in graph theory with a wealth of results studying various aspects and variants, in particular exploring probabilistic [11, 36, 39, 45, 51], extremal [4, 12, 62, 68], and algorithmic [18, 42, 44, 46] considerations. However, unlike perfect matchings, it is not easy to verify whether a graph G contains a K_r-factor or not. Certainly it is necessary that the number of vertices of G must be divisible by r but given this, it was proved by Schaeffer [43] (in the case $r = 3$) and by Kirkpatrick and Hell [46] (in general) that determining whether a graph on $n \in r\mathbb{N}$ vertices contains a K_r-factor is an NP-complete problem. Given that we cannot hope for a nice characterisation of graphs which contain K_r-factors, there has been a focus on providing sufficient conditions which are computationally easy to verify. One classical such theorem is due to Hajnal and Szemerédi [29] who showed that a K_r-factor is guaranteed if the host graph is sufficiently dense. The case of triangle factors was previously shown by Corrádi and Hajnal [24].

Theorem 1.1. *If $r \in \mathbb{N}_{\geq 3}$ and G is a graph on $n \in r\mathbb{N}$ vertices with minimum degree $\delta(G) \geq (1 - 1/r)n$, then G contains a K_r-factor.*

This theorem is tight, as can be seen, for example, by taking G to be a complete graph with a clique of size $n/r + 1$ removed to leave an independent set of vertices, say I. One then has $\delta(G) = (1 - 1/r)n - 1$ and G does not have a K_r-factor. Indeed, any copy of K_r in a family of vertex disjoint K_rs can use at most one vertex of I but a K_r-factor should contain $n/r < |I|$ copies of K_r. All examples verifying the tightness of Theorem 1.1 share some features with the graph given here. For example they contain much larger independent sets than almost all graphs of this density. Therefore, one might hope to capture more graphs having a K_r-factor by adding a condition that precludes the atypical behaviour of the extremal examples.

This naturally leads us to the notion of *pseudorandom graphs*, which are, roughly speaking, graphs which imitate random graphs of the same density. The study of pseudorandom graphs, initiated in the 1980s by Thomason [65, 66], has become a central and vibrant field at the intersection of combinatorics and theoretical computer science. We refer to the excellent survey of Krivelevich and Sudakov [53] for an introduction to the topic. One way of imposing pseudorandomness is through the spectral notion of the eigenvalue gap. This then leads to the study of (n, d, λ)-graphs G which are d-regular n-vertex graphs with *second eigenvalue* λ. By second eigenvalue, what is actually meant is the second largest eigenvalue in absolute value, as follows. Given an n-vertex d-regular graph G, we can look at the eigenvalues of the adjacency matrix A of G which, as A is a symmetric 0/1-matrix, are real and can be ordered as $\lambda_1 \geq \cdots \geq \lambda_n$. The second eigenvalue is then defined to be $\lambda := \max\{|\lambda_2|, |\lambda_n|\}$. It turns out that this parameter λ controls the pseudorandomness of the graph G, with smaller values of λ giving graphs that have stronger pseudorandom properties. More concretely, the relation is given by the following property of (n, d, λ)-graphs (see e.g. [53, Theorem 2.1]), which is known as the *Expander Mixing Lemma* and shows that λ controls the edge distribution between vertex sets. For any vertex subsets A, B of an (n, d, λ)-graph G, one has

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

where $e(A, B) := |\{uv \in E(G) : u \in A,\ b \in B\}|$ denotes the number[1] of edges in G with one endpoint in A and the other in B. Note that d/n is the density of the graph G, and hence one would expect to see $\frac{d}{n}|A|\,|B|$ edges between the vertex sets A and B in a random graph G. The pseudorandom parameter λ then controls the discrepancy from this paradigm.

It follows from simple linear algebra (see e.g. [53]) that for an (n, d, λ)-graph, one has $\lambda \leq d$ always and moreover, as long as d is not too close to n, say $d \leq 2n/3$, one has $\lambda = \Omega(\sqrt{d})$. Thus, we think of (n, d, λ)-graphs with $\lambda = \Theta(\sqrt{d})$ as being *optimally pseudorandom*. For example, it is known that random regular graphs are optimally pseudorandom (n, d, λ)-graphs with high probability[2] [16, 67].

A prominent theme in the study of pseudorandom graphs has been to give conditions on the parameters, n, d and λ that guarantee certain properties of an (n, d, λ)-graph. For example, it follows easily from (1.1) that any (n, d, λ)-graph G with $\lambda < d^2/n$ contains a triangle as there is an edge in the neighbourhood of every vertex. In particular, any optimally pseudorandom graph with $d = \omega(n^{2/3})$ must contain a triangle. Moreover, this condition is tight due to a triangle-free construction of an (n, d, λ)-graph due to Alon [5] with $d = \Theta(n^{2/3})$ and $\lambda = \Theta(n^{1/3})$. Alon's construction is optimally pseudorandom and Krivelevich, Sudakov and Szabó [54] generalised it to the whole possible

[1] Note that edges that lie in $A \cap B$ are counted twice.

[2] Here, and throughout, we say that a property holds *with high probability* if the probability that it holds tends to 1 as the number n of vertices tends to infinity.

range of densities. That is, for any $d = d(n)$ such that $\Omega(n^{2/3}) = d \leq n$, they gave a sequence of infinitely many n and triangle-free (n', d, λ)-graphs with $n' = \Theta(n)$ and $\lambda = \Theta(d^2/n)$. In general, finding optimal conditions for subgraph appearance in (n, d, λ)-graphs seems hard. Indeed, the only tight conditions that are known are those for fixed size odd cycles [9, 53]. With respect to spanning structures, it is only perfect matchings that have been well understood [17, 20, 53]. Whilst such questions are interesting in their own right, they also have implications in other areas of mathematics. As an example, we mention the beautiful connection given by Alon and Bourgain [6] (see also [2]) who used the existence of certain subgraphs in pseudorandom graphs to prove the existence of additive patterns in large multiplicative subgroups of finite fields.

The purpose of this paper is to answer what has become one of the central problems in this area, by giving a tight condition for an (n, d, λ)-graph to contain a triangle factor.

Theorem 1.2. *There exists $\varepsilon > 0$ such that any (n, d, λ)-graph with $n \in 3\mathbb{N}$, $d > 0$ and $\lambda \leq \varepsilon d^2/n$ contains a triangle factor.*

Theorem 1.2 was conjectured by Krivelevich, Sudakov and Szabó [54] in 2004. Focusing solely on optimally pseudorandom graphs, that is, setting $\lambda = \Theta(\sqrt{d})$, Theorem 1.4 implies that any optimally pseudorandom graph with $d = \omega(n^{2/3})$ contains a triangle factor. Comparing this to Theorem 1.1, we see that imposing pseudorandomness, which is easy to compute via the second eigenvalue, allows us to capture much sparser graphs which are guaranteed to contain a triangle factor.

Theorem 1.2 (and the more general Theorem 1.4 below) conclude a body of work towards the conjecture of Krivelevich, Sudakov and Szabó, and the proof of the theorem, discussed in Section 2, builds upon the many beautiful ideas of various authors, which have arisen in this study. The first step towards the conjecture was given by Krivelevich, Sudakov and Szabó [54] themselves, who showed that $\lambda \leq \varepsilon d^3/(n^2 \log n)$ for some sufficiently small ε is enough to guarantee a triangle factor. This was improved to $\lambda \leq \varepsilon d^{5/2}/n^{3/2}$ by Allen, Böttcher, Hàn, Kohayakawa and Person [3] who also proved that the same condition guarantees the appearance of the square of a Hamilton cycle, a supergraph of a triangle factor. Recently, Nenadov [61] got very close to the conjecture, showing that $\lambda \leq \varepsilon d^2/(n \log n)$ guarantees a triangle factor. Concentrating solely on optimally pseudorandom graphs, these results imply that having degree $d = \omega(n^{4/5}(\log n)^{2/5})$, $\omega(n^{3/4})$ and $\omega((n \log n)^{2/3})$ respectively, guarantees the existence of a triangle factor.

In a different direction, one can fix the condition that $\lambda \leq \varepsilon d^2/n$ for some small $\varepsilon > 0$ and prove the existence of other structures giving evidence for a triangle factor. Again, this was initiated by Krivelevich, Sudakov and Szabó [54] who proved that with this condition, one can guarantee the existence of a *fractional triangle factor*. That is, they showed that there is a function w which assigns a weight $w(T) \in [0, 1]$ to each triangle T in a pseudorandom graph G and is such that for every vertex $v \in V(G)$, the sum $\sum_{v \in T} w(T)$ of the weights of triangles containing v is precisely equal to 1. Imposing $\{0, 1\}$-weights recovers the notion of a triangle factor and a fractional triangle factor is thus a natural relaxation. Another interesting result of Sudakov, Szabó and Vu [64] showed that when $\lambda \leq \varepsilon d^2/n$, we have many triangles and these are well distributed in the (n, d, λ)-graph G.

Indeed, they proved a Turán-type result showing that any triangle-free subgraph of such a graph G must contain at most half the edges of G. A more recent result due to Han, Kohayakawa and Person [34, 35] shows that $\lambda \leq \varepsilon d^2/n$ guarantees the existence of a *near triangle factor*: there are vertex disjoint triangles covering all but $n^{647/648}$ vertices of such an (n, d, λ)-graph.

We will deduce Theorem 1.2 from a more general theorem (Theorem 1.4 below) which deals with K_r-factors for all $r \geq 3$ and works with a larger class of pseudorandom graphs where we do not restrict solely to regular graphs. Indeed, we will work with the notion of *bijumbledness*, whose usage dates back to the original works of Thomason [65, 66], and whose definition captures the key property of edge distribution, given for (n, d, λ)-graphs by (1.1).

Definition 1.3. Let $n \in \mathbb{N}$, $p = p(n) \in [0, 1]$ and $\beta = \beta(n, p) > 0$. An n-vertex graph $G = (V, E)$ is (p, β)-*bijumbled* if for every pair of vertex subsets $A, B \subseteq V$, one has

$$\Big|e(A, B) - p|A|\,|B|\Big| \leq \beta\sqrt{|A|\,|B|}. \tag{1.2}$$

Note that, due to (1.1), (n, d, λ)-graphs are $(d/n, \lambda)$-bijumbled. As with (n, d, λ)-graphs, we are interested in finding conditions on the parameters n, p and β that guarantee the existence of certain subgraphs in n-vertex (p, β)-bijumbled graphs. Our main theorem gives conditions for the existence of K_r-factors for all $r \geq 3$ in this setting.

Theorem 1.4. *For every $r \in \mathbb{N}_{\geq 3}$ and $c > 0$ there exists an $\varepsilon > 0$ such that any n-vertex (p, β)-bijumbled graph with $n \in r\mathbb{N}$, $p > 0$, $\delta(G) \geq cpn$ and $\beta \leq \varepsilon p^{r-1} n$ contains a K_r-factor.*

We remark that the condition that $\delta(G) \geq cpn$ is natural. Indeed, Definition 1.3 implies that almost all vertices will have degree at least cpn, and some lower bound on minimum degree is necessary to avoid isolated vertices. Theorem 1.2 follows directly from Theorem 1.4, and much of the context and past results discussed above have analogous statements when $r \geq 4$ with many authors also working in the more general setting of (p, β)-bijumbled graphs. In particular, for all $r \geq 3$, a condition of $\beta = o(p^{r-1}n)$ guarantees a copy of K_r, and before Theorem 1.4 the best condition known for ensuring a K_r-factor was $\beta = o(p^{r-1}n/\log n)$ due to Nenadov [61]. Another result due to Han, Kohayakawa, Person and the author [32] appeared at roughly the same time as that of Nenadov and gave a condition of $\beta = o(p^r n)$ for a K_r-factor, which for $r \geq 4$ gives a stronger result than the previously best known condition of Allen, Böttcher, Hàn, Kohayakawa and Person [3]. Although this condition is weaker than Nenadov's only when the bijumbled graph is very dense, it turns out that the proof methods of both results will be useful in proving Theorem 1.4.

There is one key difference between the pictures for $r = 3$ and for $r \geq 4$: the tightness of the condition $\beta = o(p^{r-1}n)$ for *both* the clique and the clique factor when $r \geq 4$ is unknown. We defer a more in-depth discussion of this to our concluding remarks (Section 9) and conclude this introduction by again focusing on the most interesting case of triangle factors where we know that Theorems 1.4 and 1.2 are tight due to the construction of Alon (and its generalisation to the whole range of densities by Krivelevich, Sudakov and Szabó) discussed above. Indeed, one of the reasons that the Krivelevich–Sudakov–Szabó conjecture (Theorem 1.2) has attracted so much attention is that it marks a distinct difference between the behaviour of random graphs and that of (optimally) pseudorandom graphs. In random graphs, we know that triangles appear at density roughly $p = n^{-1}$, whilst for triangle factors the threshold is considerably denser, namely $p = n^{-2/3}(\log n)^{1/3}$ [39] (see also recent results [37,40,41,63] that imply that this threshold is sharp). On the other hand, there exist triangle-free, optimally pseudorandom graphs with density roughly $n^{-1/3}$, but Theorem 1.4 asserts that any pseudorandom graph whose density is a constant factor larger than this is guaranteed to have not only a triangle but a triangle factor. Furthermore, it follows from Theorem 1.4 and (the proof of) a result of Han, Kohayakawa, Person and the author [33] that even more is true.

Corollary 1.5. *For every $c > 0$ there exists an $\varepsilon > 0$ such that any n-vertex (p, β)-bijumbled graph with $\delta(G) \geq cpn$, $p > 0$ and $\beta \leq \varepsilon p^2 n$ is* 2-universal. *That is, given any graph F on at most n vertices, with maximum degree 2, G contains a copy of F. In particular, any (n, d, λ)-graph G with $\lambda \leq \varepsilon d^2/n$ is 2-universal.*

Our proof of Theorem 1.4 incorporates discrete algorithmic techniques, probabilistic methods, fractional relaxations and linear programming duality, and the method of absorption. In the next section we discuss the proof in detail and reduce the problem to proving two intermediate propositions and a lemma. These will then be proven in what follows after developing the necessary theory.

Remark. An accompanying conference version [59] of this work deals solely with the setting of Theorem 1.2. More technical parts of the proof are omitted there and we hope that it serves as a gentle introduction to the present paper.

2. Proof of main theorem

The proof of Theorem 1.4 rests on the shoulders of the previous results [3,32–35,54,61] working towards the conjecture of Krivelevich, Sudakov and Szabó. Indeed, it is fair to say that the solution of the conjecture would not have been possible without the insights and ideas of the many authors who tackled this problem. In this section, we discuss these as well as our novel ideas and lay out the key concepts and scheme of the proof. In doing so, we will reduce the theorem to several intermediate results, whose proofs will be the subject of the rest of the paper.

Our proof, like some of its predecessors [3,32,61], works by the method of absorption. It turns out that finding many vertex disjoint copies of K_r in a (p, β)-bijumbled graph G as in Theorem 1.4 is easy. This follows from a simple consequence of Definition 1.3 which guarantees that any small linear sized set of vertices contains a copy of K_r; see e.g. Corollary 3.5 (2) for a precise statement. Therefore we can greedily choose copies of K_r to be in our K_r-factor and continue this process until we are left with some small leftover set of vertices L, where small means of size at most εrn, say. However, at this point we get stuck: we have no way of guaranteeing the existence of a K_r in L and so we do not know how to get a larger set of vertex disjoint copies of K_r. The idea of absorption is to put aside an *absorbing set* of vertices which can *absorb* the leftover vertices L into a K_r-factor. That is, before running this greedy process to build a K_r-factor, we find some special set of vertices $X \subset V(G)$ which has the property that for *any* small set of vertices $L \subset V(G) \setminus X$, there is a K_r-factor in $G[X \cup L]$ (under the trivial divisibility constraint that $r \mid (|X| + |L|)$). If we can find such an X in G, then we can put it to one side and run the greedy argument to cover almost all the vertices which do not lie in X, with vertex disjoint copies of K_r. We can then use the absorbing property to *absorb* the leftover vertices L and get a full K_r-factor.

This leaves the challenge of defining some structure in G which has this absorbing property and finding such a structure (on some vertex set X) in G. The building blocks of our absorbing structure will be subgraphs that we call *K_r-diamond trees*. In words, a K_r-diamond tree $\mathcal{D} = (T, R, \Sigma)$ is the graph obtained by taking a tree T and replacing each edge $e \in E(T)$ by a copy of K_{r+1}^- whose degree $r-1$ vertices are the vertices of e and whose degree r vertices are new and distinct from previous choices; see Figure 1 for an example. The following definition formalises this notion.

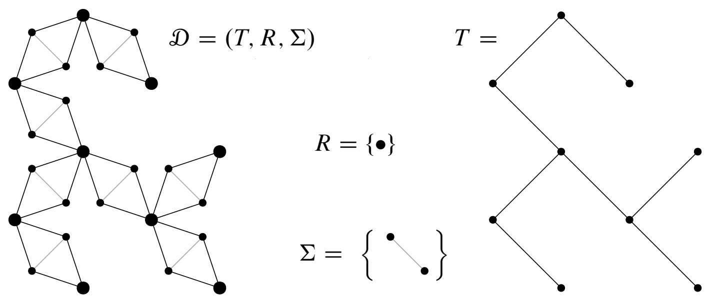

Fig. 1. An example of a K_3-diamond tree $\mathcal{D} = (T, R, \Sigma)$ of order 9 shown on the left. The removable vertices R are the larger vertices of $\mathcal{D}$ and the interior cliques Σ are the edges given in grey. The auxiliary tree T is depicted on the right.

Definition 2.1. A *K_r-diamond tree* $\mathcal{D}$ of *order* m in a graph G is a tuple $\mathcal{D} = (T, R, \Sigma)$ where T is an (auxiliary) tree of order m (i.e. with m vertices), $R \subset V(G)$ is a subset of m vertices of G and[3] $\Sigma \subset K_{r-1}(G)$ is a set of $m-1$ copies of K_{r-1} in G with the following property. There are bijective maps $\rho : V(T) \to R$ and $\sigma : E(T) \to \Sigma$ such that

- the copies of K_{r-1} in Σ are pairwise vertex disjoint in G and they are also disjoint from R, i.e. $V(S) \cap V(S') = \emptyset$ and $V(S) \cap R = \emptyset$ for all $S, S' \in \Sigma$;
- for all $e = uv \in E(T)$, we have $V(\sigma(e)) \subseteq N^G(\rho(u)) \cap N^G(\rho(v))$, that is, the $r-1$-clique $\sigma(e) \in K_{r-1}(G)$ can be extended to a copy of K_r in G by adding the vertex $\rho(u)$ and likewise with $\rho(v)$.

[3] Here and throughout, we use $K_{r-1}(G)$ to denote the family of $(r-1)$-cliques in G.

We refer to R as the set of *removable vertices* of $\mathcal{D}$ and to Σ as the set of *interior cliques* of $\mathcal{D}$. We define the vertices of $\mathcal{D}$ to be all the removable vertices and the vertices in interior cliques. That is, $V(\mathcal{D}) := (\bigcup_{S\in\Sigma} V(S)) \cup R$. Finally, we define the *leaves* of the diamond tree to be the vertices which are images of leaves in T under ρ.

Note that a K_r-diamond tree of order m has exactly $(m-1)r+1$ vertices. Krivelevich [51] used K_3-diamond trees in an absorption argument for triangle factors in random graphs which is often cited as one of the first appearances of the absorption method. Nenadov [61] also used this idea in his result that got within a log-factor of Theorem 1.4. The utility of these subgraphs in absorption arguments comes from the following key observation which shows that they can contribute to a K_r-factor in many ways.

Observation 2.2. Given a K_r-diamond tree $\mathcal{D} = (T, R, \Sigma)$ in G, for *any* removable vertex $v \in R$ there is a K_r-factor of $G[V(\mathcal{D}) \setminus \{v\}]$. Indeed, consider $u = \rho^{-1}(v)$ in $V(T)$ and the map $\varphi : E(T) \to V(T) \setminus \{u\}$ which maps each edge e of T to the vertex in e which has the larger distance from u in T. Then φ is a bijection and taking the copies of K_r on $\sigma(e) \cup \rho(\varphi(e))$ for each edge $e \in E(T)$ gives the required K_r-factor. See Figure 2 for some examples.

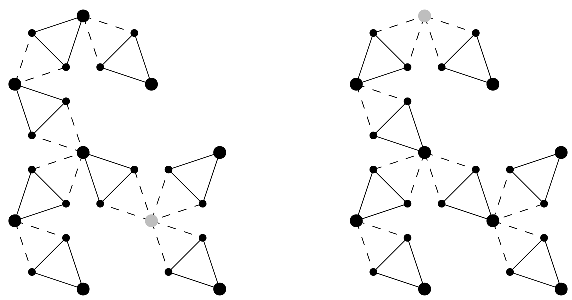

Fig. 2. Some examples of K_3-factors found after removing a removable vertex from the K_3-diamond tree in Figure 1 (see Observation 2.2).

Observation 2.2 works for any underlying auxiliary tree T. It turns out that in the (p, β)-bijumbled graphs G we are interested in, one can find K_r-diamond trees of any order up to linear size. Indeed, one can use the argument of Krivelevich [51] to construct these or a different argument due to Nenadov [61]. The method of Nenadov gives diamond trees whose auxiliary tree is a path, whilst the argument of Krivelevich gives no control over the underlying auxiliary tree which defines the diamond tree found. As a key part of our argument, we will need to prove the existence of diamond trees which have extra structure, as we discuss shortly.

In order to utilise the absorbing power of diamond trees, we need to group them together in collections. The following definition of an *orchard* captures how we do this.

Definition 2.3. We say a collection $\mathcal{O} = \{\mathcal{D}_1, \ldots, \mathcal{D}_k\}$ of pairwise vertex disjoint K_r-diamond trees in a graph G is a $(k, m)_r$*-orchard* if there are k diamond trees in the collection and each has order at least m and at most $2m$. We refer to k as the *size* of the orchard, and to m as its *order*.[4] We denote by $V(\mathcal{O})$ the vertices featuring in diamond trees in $\mathcal{O}$, that is, $V(\mathcal{O}) = \bigcup_{i\in[k]} V(\mathcal{D}_i)$. Finally, if $\mathcal{O}' \subseteq \mathcal{O}$ is a subset of diamond trees in an orchard $\mathcal{O}$, we call $\mathcal{O}'$ a *suborchard* of $\mathcal{O}$.

The term orchard here is supposed to be instructive, indicating that this is a 'neat' collection of diamond trees that all have a similar order and are completely disjoint from one another. As noted in Observation 2.2, a K_r-diamond tree can contribute to a K_r-factor in many ways. By grouping together many vertex disjoint K_r-diamond trees into a $(k, m)_r$-orchard such that $km = \Omega(n)$, we get a structure with a strong absorbing property, as the following lemma shows. We say a $(K, M)_r$-orchard $\mathcal{O}$ *absorbs* a $(k, m)_r$-orchard $\mathcal{R}$ if there is an $((r-1)k, M)_r$-suborchard $\mathcal{O}' \subset \mathcal{O}$ such that there is a K_r-factor in $G[V(\mathcal{R}) \cup V(\mathcal{O}')]$.

Lemma 2.4. *For any $r \in \mathbb{N}_{\geq 3}$ and $0 < \zeta, \eta < 1$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$. Let $\mathcal{O}$ be a $(K, M)_r$-orchard in G such that $KM \geq \zeta n$. Then there exists a set $B \subset V(G)$ such that $|B| \leq \eta p^{2r-4} n$ and $\mathcal{O}$ absorbs any $(k, m)_r$-orchard $\mathcal{R}$ in G with*

$$V(\mathcal{R}) \cap (B \cup V(\mathcal{O})) = \emptyset, \quad k \leq K/(8r) \quad \textit{and} \quad kM \leq mK. \tag{2.1}$$

Morally, Lemma 2.4 says that *large orchards absorb small orchards*. Here, by large we refer to both the size and the order of the orchards. Indeed, the second condition in (2.1) shows that the larger orchard has to have a larger size than the smaller orchard. This is the critical condition when we want absorption between orchards of similar order. The third condition shows that the ratio between the orders of the orchards is constrained by the ratio of the sizes. That is, the larger $\mathcal{O}$ is compared to $\mathcal{R}$ with respect to their sizes, the smaller $\mathcal{R}$ can be than $\mathcal{O}$ with respect to their orders. This will be the critical condition when we want absorption between orchards of (polynomially) different orders. The first condition in (2.1) simply states that in order for $\mathcal{O}$ to absorb $\mathcal{R}$, we need that $\mathcal{R}$ avoids some small set B of bad vertices. This will be easy to implement in applications.

Lemma 2.4 will be proven in Section 5.1. It provides us with an absorption property between two distinct orchards. We will also need an absorption property within orchards themselves, showing that we can find a large suborchard which hosts a K_r-factor in G. Given Observation 2.2, in order to find K_r-factors on suborchards it suffices to find copies of K_r which traverse sets of removable vertices. We therefore make the following definition.

Definition 2.5. Given a $(k, m)_r$-orchard $\mathcal{O} = \{\mathcal{D}_1, \ldots, \mathcal{D}_k\}$ in a graph G, the K_r-*hypergraph* generated by $\mathcal{O}$, denoted $H = H(\mathcal{O})$, is the r-uniform hypergraph with

[4] Note that we abuse notation slightly here. Indeed, we refer to *the* order of an orchard although this may not be uniquely defined by the orchard. We take the convention that when we refer to the order of an orchard, we simply fix one of the possible orders arbitrarily, noting that these possible orders differ by a factor of at most 2.

vertex set $V(H) = \mathcal{O}$ and with $\{\mathcal{D}_{i_1}, \ldots, \mathcal{D}_{i_r}\}$ for distinct $i_1, \ldots, i_r \in [k]$ forming a hyperedge in H if and only if there is a copy of K_r traversing[5] the sets $R_{i_1}, \ldots, R_{i_r}$ in G, where R_{i_j} is the set of removable vertices of $\mathcal{D}_{i_j}$ for all j.

Appealing to Observation 2.2 then gives the following, as finding copy of K_r traversing r sets of removable vertices removes exactly one vertex from each set.

Observation 2.6. If $\mathcal{O}$ is an orchard of K_r-diamond trees in a graph G and $H(\mathcal{O})$ contains a perfect matching, then $G[V(\mathcal{O})]$ contains a K_r-factor.

We will be particularly interested in orchards which contain near K_r-factors in a robust way. This gives us the notion of a shrinkable orchard.

Definition 2.7. Given $0 < \gamma < 1$, we say a $(k,m)_r$-orchard $\mathcal{O}$ in a graph G is γ*-shrinkable* if there exists a suborchard $\mathcal{Q} \subset \mathcal{O}$ of size at least γk such for *any* suborchard $\mathcal{Q}' \subseteq \mathcal{Q}$, there is a matching in $H := H(\mathcal{O} \setminus \mathcal{Q}')$ covering all but $k^{1-\gamma}$ of the vertices of H.

Our first key proposition gives the existence of shrinkable orchards. It will be discussed in Section 5.2 and proven in Sections 6 and 7.

Proposition 2.8. *For any $r \in \mathbb{N}_{\geq 3}$ and $0 < \alpha, \gamma < 1/2^{12r}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$ and any vertex subset $U \subseteq V(G)$ with $|U| \geq n/2$. For any $m \in \mathbb{N}$ with $1 \leq m \leq n^{7/8}$ there exists a γ-shrinkable $(k,m)_r$-orchard $\mathcal{O}$ in $G[U]$ with $k \in \mathbb{N}$ such that $\alpha n \leq km \leq 2\alpha n$.*

Given Lemma 2.4 and Proposition 2.8, we know that we can find orchards which contain large K_r-factors and that large orchards can absorb smaller orchards. This suggests the following approach for giving an absorbing structure which can absorb leftover vertices in our (p, β)-bijumbled graph G (we keep the discussion at a high level here to highlight the key idea; the details of this scheme are elaborated within the proof of Theorem 1.4). Find a sequence of vertex disjoint shrinkable orchards, each on a linear number of vertices. Each orchard in the sequence will have a larger order than that of the previous orchard and the first orchard in the sequence will be composed of linearly many K_r-diamond trees of constant size. We can then run a *cascading absorption* through the sequence of orchards. That is, given some small leftover set of vertices L (which is itself a $(|L|, 1)$-orchard), we use the first orchard in the sequence to absorb L. We then use the fact that the first orchard is shrinkable and so we can cover most of what remains of the first orchard with vertex disjoint copies of K_r. There will be some K_r-diamond trees of the first orchard left at the end of this and for these we appeal to Lemma 2.4 to absorb this small suborchard using the second orchard. Then again, the second orchard is shrinkable and so the remainder of the second orchard can be almost fully covered with vertex disjoint K_rs, leaving some small leftover suborchard uncovered. We then repeat to absorb this leftover with the third orchard and continue in this fashion. In this way we cascade

[5]Here and throughout, we say a copy of K_r *traverses* r disjoint sets of vertices if it contains one vertex from each set.

the absorption through the orchards and each time we do this, we increase the order of the orchard which we need to absorb.

This approach is promising but we need to cut this process off at some point and find a *full* K_r-factor on the vertices which have not already been covered by vertex disjoint copies of K_r. The next proposition states that once the orchard has a large enough order, we can find a structure that can *fully absorb* any leftover.

Proposition 2.9. *For any $r \in \mathbb{N}_{\geq 3}$ and $0 < \alpha, \eta < 1/2^{3r}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1}n$ and any vertex subset $W \subseteq V(G)$ with $|W| \geq n/2$.*

There exist vertex subsets $A, B \subset V(G)$ such that $A \subset W$, $|A| \leq \alpha n$, $|B| \leq \eta p^{2r-4}n$ and for any $(k,m)_r$-orchard $\mathcal{R}$ whose vertices lie in $V(G) \setminus (A \cup B)$, if $|A| + |V(\mathcal{R})| \in r\mathbb{N}$, $k \leq \alpha^2 n^{1/8}$ and $m \geq n^{7/8}$ then $G[A \cup V(\mathcal{R})]$ has a K_r-factor.

In our absorption scheme sketched above, $\mathcal{R}$ will be the leftover of the last orchard (the one with the largest order) after having cascaded the absorption through the sequence of orchards. Proposition 2.9 states that the vertex set A can fully absorb this $\mathcal{R}$. Hence when constructing our absorbing structure, we will first find A and then construct our sequence of shrinkable orchards so that they avoid A and also the small set B of bad vertices given by Proposition 2.9. Finally, we remark that the necessity of W in Proposition 2.9 comes from the fact that before we find our absorbing structure, we will put aside some small set Y which will be used later in the proof to help with bad vertices, and so need to find A in $W = V(G) \setminus Y$.

We are now in a position to prove Theorem 1.4, using only Lemma 2.4, Propositions 2.8 and 2.9, some simple properties of (p, β)-bijumbled graphs and Chernoff's Theorem (Theorem 3.6), a well-known result which gives concentration of binomial random variables.

Proof of Theorem 1.4. For convenient reference throughout the proof, let us fix our constants

$$\gamma := \frac{c}{2^{24r}}, \quad \lambda := \gamma^2, \quad \alpha := \lambda^2, \quad \zeta = \alpha^2 \quad \eta := \zeta^2 \quad \text{and} \quad t := \frac{7}{8\lambda}. \tag{2.2}$$

We further fix $\varepsilon > 0$ much smaller than all these constants and small enough to apply Lemma 2.4 and Propositions 2.8 and 2.9 with these parameters. We also use some simple consequences of Definition 1.3 which imply that, by choosing $\varepsilon > 0$ sufficiently small, we guarantee that any vertex subset of size ζpn contains a copy of K_{r-1}, whilst any vertex set of size ζn contains a copy of K_r; see e.g. Corollary 3.5. Finally, we note that if $\delta(G) \geq (1 - 1/r)n$ then it follows from Theorem 1.1 that G has a K_r-factor and so we can assume that $\delta(G) < (1 - 1/r)n$. For such n-vertex (p, β)-bijumbled graphs G with $\beta \leq \varepsilon p^{r-1}n$, a well-known fact (see Fact 3.1) implies that by choosing $\varepsilon > 0$ sufficiently small, we can assume that n is sufficiently large in what follows, because otherwise no n-vertex (p, β)-bijumbled graphs with $\beta \leq \varepsilon p^{r-1}n$ exist and the theorem is vacuously true. Moreover, another well-known fact (see Fact 3.2) implies that (p, β)-bijumbled graphs cannot be too sparse. In particular, with our condition on β, by choosing $\varepsilon > 0$ sufficiently small, we can also assume that $p \geq n^{-1/3}$ in what follows.

Before finding our K_r-factor in G we need to do some preparation. We begin by setting aside a randomly chosen subset $Y \subset V(G)$. We let each vertex be in Y with probability α. It follows from Chernoff's Theorem (see Theorem 3.6) and a union bound that with high probability, as $n \to \infty$, we have $|Y| \leq 2\alpha n$ and $\deg_Y(v) \geq c\alpha pn/2$ for all $v \in V(G)$. Indeed, this follows because $\mathbb{E}[\deg_Y(v)] \geq c\alpha pn = \Omega(n^{2/3})$ for each $v \in V(G)$. Therefore, as n is large, we can fix such an instance of Y. We will use the vertices of Y to find copies of K_r containing 'bad' vertices later in the argument.

Next, we apply Proposition 2.9 (with $W = V(G) \setminus Y$) to obtain vertex sets $A \subset V(G) \setminus Y$ and B such that $|A| \leq \alpha n$, $|B| \leq \eta p^{2r-4} n$ and we have the following key absorption property. For any $(k, m)_r$-orchard $\mathcal{R}$ whose vertices lie in $V(G) \setminus (A \cup B)$, if $|A| + |V(\mathcal{R})| \in r\mathbb{N}$, $k \leq \zeta n^{1/8}$ and $m \geq n^{7/8}$ then $G[A \cup V(\mathcal{R})]$ has a K_r-factor. That is, A can absorb orchards whose order is sufficiently large.

As sketched above, the idea is now to provide constantly many (namely, $t+1$) vertex disjoint shrinkable orchards $\mathcal{O}_0, \mathcal{O}_1, \ldots \mathcal{O}_t$, each on a linear number of vertices and whose vertices are disjoint from A. The order of these orchards will increase slightly (namely, by a factor of n^{λ}) at each step in the sequence. Due to our definition (2.2) of t, $\mathcal{O}_0$ has $\Omega(n)$ diamond trees of constant order, while $\mathcal{O}_t$ has diamond trees of order $\Omega(n^{7/8})$. The point is that we will be able to repeatedly apply Lemma 2.4 and the fact that each orchard is shrinkable to create a cascading absorption through the shrinkable orchards. Indeed, $\mathcal{O}_0$ will be able to absorb leftover vertices and each $\mathcal{O}_i$ will be able to absorb any leftover K_r-diamond trees in $\mathcal{O}_{i-1}$, after using the fact that $\mathcal{O}_{i-1}$ is shrinkable to cover almost all of the vertices of $\mathcal{O}_{i-1}$ with disjoint copies of K_r. Once this absorption reaches $\mathcal{O}_t$, we will be able to use A to absorb the leftover K_r-diamond trees in $\mathcal{O}_t$ and complete a K_r-factor. In fact, when absorbing between orchards we do not use *all* of $\mathcal{O}_i$ to absorb leftover diamond trees in $\mathcal{O}_{i-1}$ but rather a suborchard $\mathcal{Q}_i \subset \mathcal{O}_i$ which contains a γ-proportion of the K_r-diamond trees in $\mathcal{O}_i$. Indeed, this $\mathcal{Q}_i$ is provided by the fact that $\mathcal{O}_i$ is shrinkable (see Definition 2.7) and guarantees that removing diamond trees from $\mathcal{Q}_i$ will not prevent us from covering almost all of what remains of $\mathcal{O}_i$ with vertex disjoint copies of K_r.

In detail, we collect what we require in the following claim.

Claim 2.10. *There exist vertex disjoint orchards $\mathcal{O}_0, \mathcal{O}_1, \ldots, \mathcal{O}_t$ in G such that the following properties hold.*

(i) *For all $0 \leq i \leq t$, we have $V(\mathcal{O}_i) \cap (A \cup B \cup Y) = \emptyset$.*

(ii) *For each $0 \leq i \leq t$, fixing $m_i := n^{i\lambda}$, the orchard $\mathcal{O}_i$ is a $(k_i, m_i)_r$-orchard for some k_i such that $\alpha n \leq k_i m_i \leq 2\alpha n$.*

(iii) *Each $\mathcal{O}_i$ is γ-shrinkable with respect to some suborchard $\mathcal{Q}_i \subset \mathcal{O}_i$ such that*

$$k_i^* := |\mathcal{Q}_i| \geq \gamma k_i.$$

(iv) *For $1 \leq i \leq t$, given any suborchard $\mathcal{P} \subset \mathcal{O}_{i-1}$ such that $|\mathcal{P}| \leq k_{i-1}^{1-\gamma}$, the orchard $\mathcal{Q}_i$ absorbs $\mathcal{P}$.*

Before verifying the claim, let us see how we can derive the theorem using the claim. So suppose we have found such orchards $\mathcal{O}_0, \ldots, \mathcal{O}_t$ and fix

$$X := A \cup \bigcup_{i=0}^{t} V(\mathcal{O}_i).$$

Furthermore, note that as $k_0^* m_0 = k_0^* \geq \gamma \alpha n \geq \zeta n$, by Lemma 2.4 there exists some set $B_0 \subset V(G)$ such that $|B_0| \leq \eta p^{2r-4} n$ and $\mathcal{Q}_0$ absorbs any $(k, 1)$-orchard[6] $\mathcal{R}$ such that

$$k \leq \zeta n \leq \frac{\gamma \alpha n}{8r} \leq \frac{k_0^*}{8r} \tag{2.3}$$

and $V(\mathcal{R}) \cap (B_0 \cup V(\mathcal{Q}_0)) = \emptyset$. Indeed, the condition on k comes from (2.1), using $m_0 = 1$ and our lower bound on k_0^*. Fix $Z := B_0 \setminus X$ and note that $z := |Z| \leq \eta p^{2r-4} n$ as Z is a subset of B_0. Note also that $X \cap Y = \emptyset$ due to Claim 2.10 (i) and how we defined A.

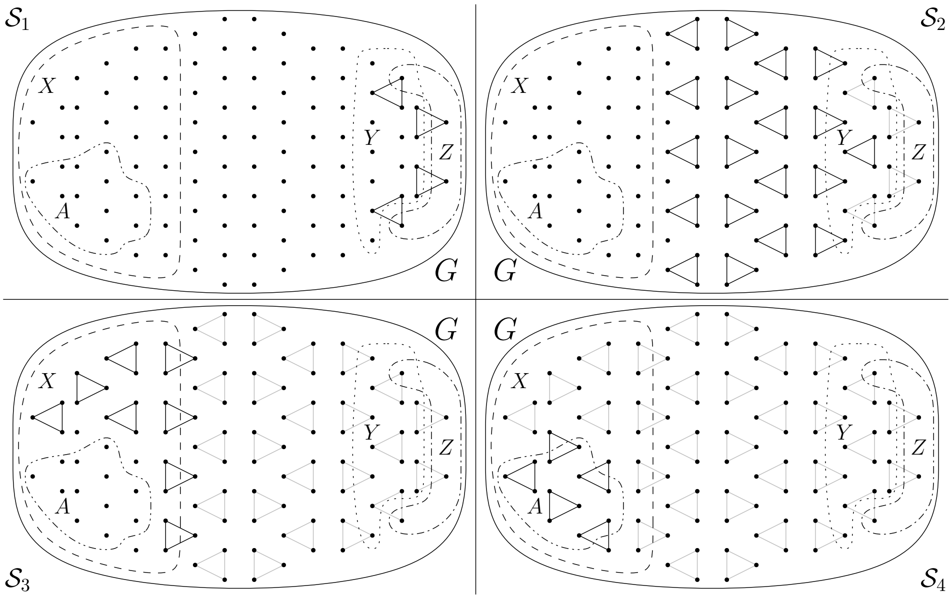

Fig. 3. A schematic to demonstrate the triangles found (and the vertex sets they cover) by our four-phase algorithm that finds a K_3-factor in G.

We are now ready to find our K_r-factor $\mathcal{S}$, which we do algorithmically in four phases. See Figure 3 for a visual guide to the cliques found in each phase. So let us initiate with

[6]Note that a $(k, m)_r$-orchard with $m = 1$ is simply a set of vertices. Each K_r-diamond tree in the orchard has order 1 and so is a single isolated vertex.

$\mathcal{S}_1 = \emptyset$. In the first phase we find copies of K_r containing the vertices in Z, using some vertices in Y. So let us order the vertices of Z arbitrarily as $Z := \{b_1, \ldots, b_z\}$ and fix $Y_1 := Y \setminus Z$. Now for $1 \le j \le z$, we find an r-clique S_j containing b_j and $r-1$ vertices of Y_j. We add S_j to $\mathcal{S}_1$, fix $Y_{j+1} := Y_j \setminus V(S_j)$ and move to step $j+1$. To see that we can always find such a clique, note that for each $j \in [z]$ we have

$$\deg_{Y_j}(b_j) \ge \deg_Y(b_j) - |Z| - r(j-1) \ge c\alpha pn/2 - r\eta p^{2r-4} n \ge \zeta pn,$$

recalling our key property of Y and using the definitions of our constants (2.2). A simple consequence of (1.2) (see e.g. Corollary 3.5 (1) (i)) implies that there is a copy of K_{r-1} in $N(b_j) \cap Y_j$ and so this forms an r-clique S_j with b_j. In this way, we see that we succeed at every step j and at the end of the first phase we have a set of vertex disjoint r-cliques $\mathcal{S}_1$ in G of size z such that every vertex in Z is contained in a clique in $\mathcal{S}_1$.

In the second phase we find the majority of the K_r-factor which we do greedily. We initiate with $\mathcal{S}_2 = \emptyset$ and $W = V(G) \setminus (X \cup V(\mathcal{S}_1))$. Now, whilst $|W| \ge \zeta n$, we can find an r-clique S in W. Again, this is a simple consequence of (1.2); see Corollary 3.5 (2). We add S to $\mathcal{S}_2$ and delete its vertices from W. Therefore at the end of the second phase, we are left with some vertex set $L \subset V(G) \setminus X$ such that $|L| \le \zeta n$ and $\mathcal{S}_1 \cup \mathcal{S}_2$ form a K_r-factor in $G[V(G) \setminus (X \cup L)]$.

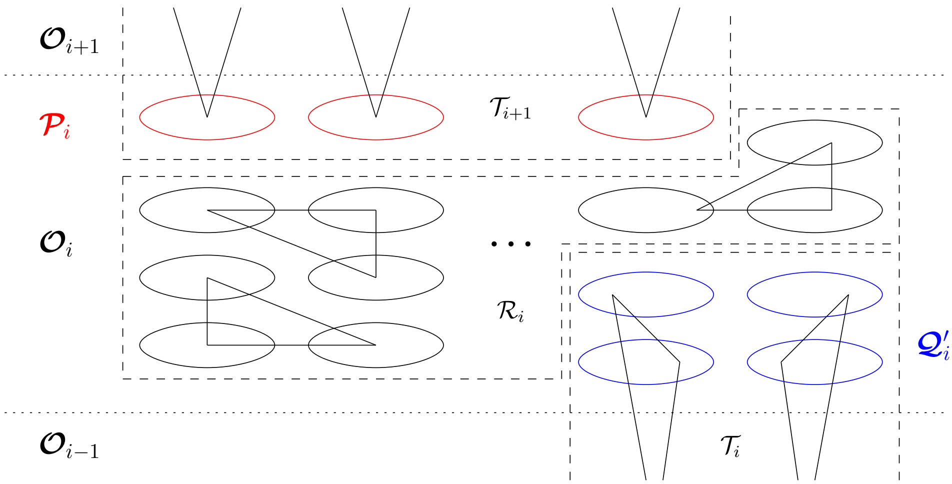

Fig. 4. A closer look at phase 3 of the algorithm in the case $r = 3$.

In our third phase, we will find vertex disjoint r-cliques $\mathcal{S}_3$ which cover L and use almost all the vertices of $X \setminus A$. We begin by fixing $\ell := |L|$ and noting that L is an $(\ell, 1)$-orchard which we relabel as $\mathcal{P}_{-1}$. Now we run the following procedure for $0 \le i \le t$ (see Figure 4). We first absorb $\mathcal{P}_{i-1}$ using $\mathcal{Q}_i$. That is, we find a suborchard $\mathcal{Q}'_i \subset \mathcal{Q}_i$ such that there is a K_r-factor $\mathcal{T}_i$ in $G[V(\mathcal{P}_{i-1}) \cup V(\mathcal{Q}'_i)]$. We add the r-cliques in $\mathcal{T}_i$ to $\mathcal{S}_3$. Then, using the fact that $\mathcal{O}_i$ is γ-shrinkable (Claim 2.10 (iii)), we can define some

$\mathcal{P}_i \subset \mathcal{O}_i \setminus \mathcal{Q}_i'$ such that $|\mathcal{P}_i| \leq k_i^{1-\gamma}$ and there is a perfect matching in the K_r-hypergraph $H(\mathcal{O}_i \setminus (\mathcal{Q}_i' \cup \mathcal{P}_i))$. By Observation 2.6, this perfect matching gives a K_r-factor $\mathcal{R}_i$ in $G[V(\mathcal{O}_i \setminus (\mathcal{Q}_i' \cup \mathcal{P}_i))]$. We add $\mathcal{R}_i$ to $\mathcal{S}_3$ and move to step $i+1$ or finish if $i = t$. Note that in order to find $\mathcal{T}_i$ and $\mathcal{Q}_i'$ in each step $i \geq 1$, we appeal to Claim 2.10 (iv), whilst when $i = 0$, the existence of $\mathcal{T}_1$ and $\mathcal{Q}_1'$ is guaranteed by the fact that L is an $(\ell, 1)$-orchard with $\ell \leq \zeta n$ as in (2.3) and L is disjoint from Z and hence B_0.

Let $\mathcal{R} := \mathcal{P}_t \subset \mathcal{O}_t$. Then $\mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$ is a K_r-factor in $G[V(G) \setminus (A \cup V(\mathcal{R}))]$. Hence as $r \mid n$, we must have $r \mid (|A| + |V(\mathcal{R})|)$. Moreover, $\mathcal{R}$ is a $(k,m)_r$-orchard with $k \leq k_t^{1-\gamma} \leq (2\alpha n^{1/8})^{1-\gamma} < \alpha^2 n^{1/8}$ and $m = n^{7/8}$. Finally, note that $V(\mathcal{R}) \cap B = \emptyset$ due to Claim 2.10 (i). Therefore, by the key property of the absorbing vertex set A in Proposition 2.9, there is a K_r-factor $\mathcal{S}_4$ in $G[A \cup V(\mathcal{R})]$. It follows that $\mathcal{S} := \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3 \cup \mathcal{S}_4$ is a K_r-factor in G, completing the proof.

It remains to establish Claim 2.10 and find the shrinkable orchards as stated. We will do this algorithmically in decreasing order. The reason for this is that in order for (iv) to hold we will appeal to Lemma 2.4 and therefore there will be some set B_i of bad vertices which we want $\mathcal{O}_{i-1}$ to avoid. In fact, we will ensure that $\mathcal{O}_{i-1}$ avoids B_j for all $i \leq j \leq t$. This is not necessary but eases our definitions (as we do not have to reintroduce vertices into the pool U_i of available vertices); the important condition in what follows is that $\mathcal{O}_{i-1}$ avoids B_i for all i.

We start by fixing $U_{t+1} := V(G) \setminus (A \cup B \cup Y)$. Now for $t \geq i \geq 0$ in descending order, we apply Proposition 2.8 to find a γ-shrinkable $(k_i, m_i)_r$-orchard $\mathcal{O}_i$ such that $\alpha n \leq k_i m_i \leq 2\alpha n$ and $V(\mathcal{O}_i) \subset U_{i+1}$. We then define U_i as follows. As $\mathcal{O}_i$ is γ-shrinkable, it defines some suborchard $\mathcal{Q}_i \subset \mathcal{O}_i$ as in condition (iii) of the claim. Now as $k_i^* m_i \geq \gamma\alpha n \geq \zeta n$, it follows from Lemma 2.4 that there exists some $B_i \subset V(G)$ with $|B_i| \leq \eta p^{r-1} n$ such that if k and m satisfy $k \leq k_i^*/(8r)$ and $k m_i \leq m k_i^*$ and $\mathcal{R}$ is a $(k,m)_r$-orchard with $V(\mathcal{R}) \subset V(G) \setminus (B_i \cup V(\mathcal{Q}_i))$ then $\mathcal{Q}_i$ absorbs $\mathcal{R}$. We fix $U_i := U_{i+1} \setminus (V(\mathcal{O}_i) \cup B_i)$ and move onto the next index $i-1$.

Let us first check that the process succeeds in finding the shrinkable orchards $\mathcal{O}_t, \ldots, \mathcal{O}_0$ at each step. Note that we start with $|U_{t+1}| \geq n - 3\alpha n - \eta p^{2r-4} n \geq n - 4\alpha n$. Moreover at each step i, we remove at most $\eta p^{2r-4} n \leq \alpha n$ vertices which lie in B_i and at most $4r\alpha n$ vertices from U_{i+1} which lie in the orchard $\mathcal{O}_i$. Indeed, the orchard is composed of k_i vertex disjoint K_r-diamond trees of order at most $2m_i$, the number of vertices in each diamond tree is less than r times its order, and $k_i m_i \leq 2\alpha n$. Hence for all $t \geq i \geq 0$, we have

$$|U_i| \geq n - (t+2) \cdot 5r\alpha n \geq n/2,$$

using $t \leq 1/\lambda$, $\alpha = \lambda^2$ and the definition of λ (see (2.2)). Hence Proposition 2.8 gives the existence of $\mathcal{O}_i$ at each step and verifies part (iii) of the claim. Note that conditions (i) and (ii) also hold simply from how we defined the $\mathcal{O}_i$ and the fact that we found them in the sets U_i, each of which is a subset of U_{t+1}.

Thus it remains to verify the absorption property between orchards, namely (iv). For each $1 \leq i \leq t$, we chose $\mathcal{O}_{i-1}$ to have vertices in U_i and hence $V(\mathcal{O}_{i-1}) \cap (B_i \cup V(\mathcal{Q}_i))$

$= \emptyset$. Therefore, by Lemma 2.4, $\mathcal{Q}_i$ absorbs any suborchard $\mathcal{P} \subset \mathcal{O}_{i-1}$ with $|\mathcal{P}| \le k_{i-1}^{1-\gamma}$ if $k_{i-1}^{1-\gamma} \le k_i^*/(8r)$ and $k_{i-1}^{1-\gamma} m_i \le k_i^* m_{i-1}$.

Now as $m_i = n^\lambda m_{i-1}$ and $n^{-\lambda} \le 1/(8r)$ for sufficiently large n, it suffices to show that $k_{i-1}^{1-\gamma} \le k_i^* n^{-\lambda}$. To see this, note that since $\alpha n \le k_{i-1} m_{i-1}$ and $k_i m_i \le 2\alpha n$, we have

$$k_{i-1} \le \frac{2\alpha n}{m_{i-1}} = \frac{2\alpha n^{1+\lambda}}{m_i} \le 2k_i n^\lambda \le \frac{2k_i^* n^\lambda}{\gamma},$$

and using this as a lower bound for k_i^*, it suffices to show that

$$k_{i-1}^{\gamma} \ge \frac{2n^{2\lambda}}{\gamma}.$$

This is certainly true as $k_{i-1} \ge k_t \ge \alpha n^{1/8} > n^{4\lambda/\gamma}$, recalling that $4\lambda/\gamma = 4\gamma$ from (2.2). This shows that (iv) holds for all i and concludes the proof of the claim and hence the proof of Theorem 1.4. ■

We remark that this proof scheme builds on that of Nenadov [61] (which in turn is influenced by that of Krivelevich [51]), who proved that $\beta \le \varepsilon p^2 n/\log n$ suffices for a triangle factor in an n-vertex (p, β)-bijumbled graph. Indeed, Nenadov also uses a result akin to Lemma 2.4, albeit between orchards whose orders only differ by a constant factor. His absorbing structure then contains a sequence of $\Theta(\log n)$ orchards whose order increases by a constant factor along the sequence. Therefore the last orchard in the sequence contains constantly many diamond trees of large order (of order $\Theta(n/\log n)$). These can be fully absorbed because any three large sets host a transversal triangle and so transversal triangles between removable sets can be greedily found, completing a triangle factor in the last step. Similarly, the $(k, m)_3$-orchards used in his argument are not imposed to be shrinkable but can be seen to host a triangle factor on all but $o(k)$ of the diamond trees by again applying a greedy approach of finding transversal triangles. The necessity of the $\log n$ in the condition of Nenadov is thus due to needing $\Theta(\log n)$ orchards in the absorbing structure and thus requiring slightly stronger properties of the (p, β)-bijumbled graph, for example the existence of triangles on sets of $\Omega(n/\log n)$ vertices.

The key challenge in this paper is then to prove Propositions 2.8 and 2.9. Both results rely heavily on a technique we develop to provide the existence of K_r-diamond trees in which we have some control over the set of removable vertices. This control is rather weak; we cannot guarantee that any fixed vertices appear as removable vertices but we can give some flexibility over the choice of removable vertices. See Proposition 4.1 for the technical statement of what we prove.

In order to prove Proposition 2.8, we build on the approach of Han, Kohayakawa and Person [34, 35]. Indeed, their result showing the existence of a near K_r-factor (covering all but some $n^{1-\varepsilon'}$ vertices) in (n, d, λ)-graphs can be seen as a step towards proving the existence of shrinkable orchards of order 1. The approach involves showing the existence of a near-perfect matching in a subhypergraph H' of the K_r-hypergraph generated by $V(G)$. In order to do this, one needs to carefully choose H' and this is done by finding many fractional K_r-factors in G which do not put too much weight on (copies of K_r containing) any given edge. Therefore, the methods of Krivelevich, Sudakov and Szabó [54], who proved the existence of singular fractional K_r-factors, become pertinent. They use the power of linear programming duality to prove that certain expansion properties guarantee the existence of fractional factors. In our setting, it turns out that we need several distinct arguments to prove the existence of shrinkable orchards of different orders. We follow the scheme of using fractional factors (in fact, fractional perfect matchings in K_r-hypergraphs) but need to adapt the method for different applications and we rely crucially on probabilistic methods to actually prove the existence of orchards which satisfy the necessary expansion properties.

It can be seen that Proposition 2.8 alone (for all orders of orchards) would lead via the same proof scheme to a condition of $\beta \leq \varepsilon p^{r-1} n/(\log \log n)$. In order to close the gap and achieve Theorem 1.4, Proposition 2.9 is necessary. To prove this, we appeal to a different absorption argument whose roots go back to an ingenious argument of Montgomery [57, 58] in his work on spanning trees in random graphs. The approach, sometimes called the absorption-reservoir method, uses a bipartite graph, which we call a template (see Section 3.6) as an auxiliary graph to define an absorbing structure. This idea was previously used by Han, Kohayakawa, Person and the author [32] to find clique factors in pseudorandom graphs, and we used this approach again in our result on 2-universality [33]. Here we combine this idea with the absorbing power of orchards and prove Proposition 2.9 with a three-stage algorithm which finds the absorbing structure necessary.

The rest of this paper is organised as follows. In the next section, we run through the necessary preliminaries, providing the background theory that we will use. This includes properties of bijumbled graphs, the study of perfect fractional matchings via linear programs, probabilistic methods and the absorption-reservoir method of Montgomery [57, 58]. In Section 4 we then study what kinds of diamond trees we can guarantee in our bijumbled graph. The key result here is Proposition 4.1, which will be crucial at various points in our proof. We then turn to addressing the necessary results for the cascading absorption through the orchards in Section 5. We prove Lemma 2.4 in Section 5.1 and discuss Proposition 2.8 in Section 5.2, reducing it to two intermediate propositions which tackle small and large order shrinkable orchards separately. We go on to prove the existence of shrinkable orchards of small order in Section 6 and large order in Section 7. Finally, we prove Proposition 2.9 which provides the final absorption in the proof of Theorem 1.4, in Section 8.

3. Preliminaries

3.1. Notation

For a graph G and $r \in \mathbb{N}$, we define $K_r(G)$ to be the set of copies of K_r in G. When referring to (a copy of) a clique $S \in K_r(G)$, we will identify the copy with the set of vertices that hosts it. That is, we think of $S \in K_r(G)$ as a set of r vertices which host a clique in G rather than the copy of the clique itself. Given a set $\Sigma \subseteq K_r(G)$ of r-cliques, we use the notation $V(\Sigma)$ to denote all vertices that feature in cliques in Σ, i.e. $V(\Sigma) := \bigcup_{S\in\Sigma} S$. We call $\Sigma \subset K_r(G)$ a *matching* of cliques if it is composed of pairwise vertex disjoint cliques, that is, $S \cap S' = \emptyset$ for any $S \neq S' \in \Sigma$. Now given subsets $S, W \subset V(G)$ of vertices, we let $N^G_W(S)$ denote the common neighbours of the vertices in S which lie in W. That is, $N^G_W(S) := (\bigcap_{v\in S} N^G(v)) \cap W$. Likewise, we define $\deg^G_W(S) := |N^G_W(S)|$ to be the cardinality of this neighbourhood. If the graph G is clear from context then we drop the superscripts. Also if $S = \{u\}$ is a single vertex, we will drop the set brackets. We say that a clique $S \in K_r(G)$ *traverses* vertex subsets $U_1, \dots, U_r \subseteq V(G)$ if there exists some ordering of S as $S = \{u_1, \dots, u_r\}$ such that $u_i \in U_i$ for all $i \in [r]$. Note that when the U_i are pairwise disjoint, this simplifies to requiring that S contains one vertex from each U_i. However, at times we will deal with not necessarily disjoint sets U_i and so this more delicate definition is needed.

If H is an r-uniform hypergraph for some $r \in \mathbb{N}$ and $v, u \in V(H)$, then $\deg^H(v)$ denotes the number of edges in H containing v, and $\mathrm{codeg}^H(u, v)$ denotes the number of edges of H which contain both u and v. If the hypergraph H is clear from context, we drop the superscripts. If H is an r-uniform hypergraph with $r \geq 3$ and J is a 2-*uniform* graph on the same vertex set $V(H)$, then H_J denotes the subhypergraph of H given by all edges of H that contain some edge of J.

For graphs $\tilde{G}$ and G on the same vertex set with $\tilde{G}$ a subgraph of G, we let $G \setminus \tilde{G}$ denote the graph on $V(G)$ given by the set of edges that feature in G but not in $\tilde{G}$. If H' and H are r-uniform hypergraphs with H' a subgraph of H, then $H \setminus H'$ is defined similarly.

We use $x = y \pm z$ to denote that $x \leq y + z$ and $x \geq y - z$, and we say a property holds *with high probability* (whp, for short) if the probability that it holds tends to 1 with some parameter n (usually the number of vertices of a graph). Finally, we drop ceilings and floors unless necessary, so as not to clutter the arguments.

3.2. *Properties of bijumbled graphs*

Here we collect some properties of bijumbled graphs. These range from simple consequences of Definition 1.3 to more involved statements catered to our purposes. We begin by showing that we can assume that the graphs we consider have an arbitrarily large number of vertices.

Fact 3.1. *Given any $r \in \mathbb{N}_{\geq 3}$ and $n_0 \in \mathbb{N}$, there exists $\varepsilon > 0$ such that any n-vertex (p, β)-bijumbled graph G with $n \in r\mathbb{N}$, $p > 0$, $\delta(G) < (1 - 1/r)n$ and $\beta \leq \varepsilon p^{r-1} n$ must have $n \geq n_0$.*

Proof. Let $\varepsilon > 0$ be such that $\varepsilon < 1/(2n_0 r)$. Suppose for a contradiction that there exists an n-vertex (p, β)-bijumbled graph with $\delta(G) < (1 - 1/r)n$, $\beta \leq \varepsilon p^{r-1} n$ and $n < n_0$. Then due to the upper bound on the minimum degree of G, there exists a vertex $u \in V(G)$

and a set $W \in V(G) \setminus \{u\}$ such that $|W| = n/r$ and $\deg_W^G(u) = 0$. However, from Definition 1.3, we have

$$e(\{u\}, W) \geq p|W| - \varepsilon p^{r-1} n \sqrt{\frac{n}{r}} \geq \frac{pn}{r}(1 - \varepsilon\sqrt{nr}) \geq \frac{pn}{2r} > 0,$$

a contradiction. ■

Fact 3.1 shows that by choosing $\varepsilon > 0$ sufficiently small, we guarantee that any bijumbled graph G we are interested in either has a large number of vertices or has $\delta(G) \geq (1 - 1/r)n$, in which case Theorem 1.1 implies the existence of a K_r-factor and we are done. We will use this at various points in our argument and simply state that we choose $\varepsilon > 0$ sufficiently small to force n to be sufficiently large.

The following well known fact states that bijumbled graphs cannot to be too sparse.

Fact 3.2. *For any $r \in \mathbb{N}_{\geq 3}$ and any $C > 0$, there exists an $\varepsilon > 0$ such that if G is an n-vertex (p, β)-bijumbled graph with $p > 0$ and $\beta \leq \varepsilon p^{r-1} n$, then $p \geq Cn^{-1/(2r-3)} \geq Cn^{-1/3}$.*

Proof. Let $\varepsilon > 0$ be such that $\varepsilon^2 \leq 1/(32C^{2r-3})$ and small enough that we can assume that

(i) $n \geq 9$;

(ii) $p \leq 1/16$.

Indeed, from Fact 3.1, we can choose ε so that (i) holds and $Cn^{-1/(2r-3)} < 1/16$ and so we are done if we are not in case (ii). We will also restrict to the case that

(iii) $p \geq 1/(2n)$.

To see that we can do this, suppose for a contradiction that there exists a (p, β)-bijumbled graph $G = (V, E)$ with $pn < 1/2$. We appeal to Definition 1.3 and upper bound $2e(G) = e(V, V)$ by $pn^2 + \varepsilon p^{r-1} n^2 < n - 1$. Hence there must be some vertex $u \in V$ which is isolated in G. But then defining $W := V \setminus \{u\}$, the lower bound of Definition 1.3 gives $e(\{u\}, W) \geq p(n-1) - \varepsilon p^{r-1} n \sqrt{n-1} \geq pn(1/2 - \varepsilon pn) > 0$, a contradiction.

We now turn to proving the statement in full generality. Our aim is to construct large (disjoint) vertex subsets U and W such that $e(U, W) = 0$. We do this in the following greedy fashion. We initiate the process by setting $U = \emptyset$ and $W = V(G)$. Now, whilst $|W| \geq 3n/4$, there exists some $u \in W$ with $\deg_W(u) \leq 2p|W| \leq 2pn$. Indeed, this follows from Definition 1.3 because

$$\sum_{w \in W} \deg_W(w) = e(W, W) \leq p|W|^2 + \varepsilon p^{r-1}|W| \leq 2p|W|^2.$$

We then choose such a u, delete it from W and add it to U and also remove $N_W(u)$ from W.

Let U and W be the resulting sets after this process terminates. It is clear that $e(U, W) = 0$ as we have removed all the neighbours of each vertex $u \in U$ from W during the process. We also claim that $|W| \geq n/2$ and $U \geq 1/(16p)$. Indeed, the last step removed at most $1 + 2pn$ vertices from W. Due to our assumptions (i) and (ii), we see that $1 + 2pn < n/4$ and so as W had size greater than $3n/4$ before this step, we indeed have $|W| \geq n/2$ as the process terminates. To see the lower bound on the size of U, note that if this was not the case, then

$$|V(G) \setminus W| = \left|\bigcup_{u \in U} (\{u\} \cup N^G(u))\right| \leq \sum_{u \in U} |\{u\} \cup N^G(u)| \leq |U|(1 + 2pn)$$
$$\leq \frac{1}{16p} + \frac{n}{8} \leq \frac{n}{4},$$

using assumption (iii) in the last inequality. This implies that $|W| \geq 3n/4$, a contradiction as the process terminated.

Thus $|W| \geq n/2$, $|U| \geq 1/(16p)$ and from Definition 1.3, we have

$$0 = e(U, W) \geq p|U|\,|W| - \varepsilon p^{r-1} n \sqrt{|U|\,|W|},$$

implying that $p^{2r-3} \geq 1/(32\varepsilon^2 n)$. Given our upper bound on ε, this implies that $p \geq Cn^{-1/(2r-3)}$ as required. ∎

Our first lemma shows that few vertices have degree much smaller or much larger than expected with respect to a given set.

Lemma 3.3. *For any $r \in \mathbb{N}_{\geq 3}$ and $\eta > 0$ there exists an $\varepsilon > 0$ such that if G is an n-vertex (p, β)-bijumbled graph with $\beta \leq \varepsilon p^{r-1} n$ then for $W \subseteq V(G)$:*

(i) *The number of vertices $v \in V(G)$ such that $\deg_W(v) < p|W|/2$ is less than*

$$\frac{\eta p^{2r-4} n^2}{|W|}.$$

(ii) *For any q such that $2p \leq q \leq 1$, the number of vertices $v \in V(G)$ such that $\deg_W(v) > q|W|$ is less than*

$$\frac{\eta p^{2r-2} n^2}{q^2 |W|}.$$

Proof. Fix $\varepsilon > 0$ such that $4\varepsilon^2 < \eta$. We prove only (ii); the proof of (i) is both similar and simpler. We set B to be the set of 'bad' vertices, i.e. vertices v such that $\deg_W(v) > q|W|$. Thus we have

$$q|B|\,|W| < e(B, W) \leq p|B|\,|W| + \varepsilon p^{r-1} n \sqrt{|B|\,|W|},$$

using the definition of B and (1.2). Rearranging gives

$$|B| < \frac{\varepsilon^2 p^{2r-2} n^2}{(q - p)^2 |W|},$$

and using $p \leq q/2$ gives the desired conclusion with our choice of ε. ∎

Next, we state some further consequences of Definition 1.3, showing that we can find cliques traversing large enough subsets of vertices. The following lemma is very general and will be used at various points in our argument. Due to its generality, there are some technical features. Whilst these are all necessary for certain parts of our argument, we do not need all of these at once. In fact, for easy reference, we list the consequences of Lemma 3.4 that we will use in Corollary 3.5. This may also serve to digest the statement of Lemma 3.4, seeing how it is applied in practice.

Lemma 3.4. *For any $r \in \mathbb{N}_{\geq 3}$ and $0 < \alpha < 1/2^{2r}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1}n$. Suppose that there are integers x_i, $i \in [r+1]$, such that $x_1 \geq \cdots \geq x_{r+1} \geq 0$ and for some $r^* \in [r]$, one has*

$$x_i + x_{i+1} + 2i \leq 2r - 2 \quad \text{for all } 1 \leq i \leq r^*. \tag{3.1}$$

Define $y := \max\{x_{i+1} + i : i \in [r^*]\}$. Then for any collection of subsets $U_i \subseteq V(G)$ such that $|U_i| \geq \alpha p^{x_i} n$ for all $i \in [r+1]$ and for any subgraph $\tilde{G}$ of G with maximum degree less than $\alpha^2 p^y n$, defining $G' := G \setminus \tilde{G}$, there exists a clique $S \in K_{r^*}(G')$ traversing $U_1, \ldots, U_{r^*}$ such that

$$\deg_{U_j}^{G'}(S) \geq \alpha p^{r^*}|U_j| \quad \text{for } r^* + 1 \leq j \leq r + 1.$$

Proof. Fix $\varepsilon > 0$ small enough to apply Lemma 3.3 (i) with $\eta := \alpha^2/(2^{4r}r)$. Further, fix y and $\tilde{G}$ as in the statement, setting $G' := G \setminus \tilde{G}$. We will prove inductively that for $i = 1, \ldots, r^*$, there exists an i-clique $S_i \in K_i(G')$ traversing $U_1, \ldots, U_i$ such that $\deg_{U_j}^{G'}(S_i) \geq (p/4)^i |U_j|$ for all j with $i + 1 \leq j \leq r + 1$. Note that S_{r^*} is the desired copy of K_{r^*} in the statement, using $\alpha \leq 1/4^{r^*}$ here.

So fix some $i \in [r^*]$. If $i \geq 2$, by induction we deduce the existence of S_{i-1} as claimed and for $i \leq j \leq r+1$, define $W_j \subseteq U_j$ so that $W_j := N_{U_j}^{G'}(S_i)$. If $i = 1$, we simply set $W_j := U_j$ for all j. We thus have

$$|W_j| \geq \left(\frac{p}{4}\right)^{i-1} |U_j| \geq \alpha 4^{1-i} p^{x_j+i-1} n \tag{3.2}$$

for $i \leq j \leq r+1$. Now we appeal to Lemma 3.3 (i) and conclude that for each j with $i + 1 \leq j \leq r + 1$, there is some set $B_j \subset V(G)$ such that $\deg_{W_j}^{G}(v) \geq p|W_j|/2$ for all $v \in V(G) \setminus B_j$ and

$$|B_j| \leq \frac{\eta p^{2r-4} n^2}{|W_j|} \leq \frac{\eta 4^{i-1} p^{2r-3-i-x_j} n}{\alpha} \leq \frac{\alpha p^{2r-3-i-x_{i+1}} n}{4^i r} \leq \frac{\alpha p^{x_i+i-1} n}{4^i r} \leq \frac{|W_i|}{2r}. \tag{3.3}$$

Here, we used (3.2) in the second inequality, the definition of η and the fact the $x_j \leq x_{i+1}$ in the third, (3.1) in the fourth and (3.2) once again in the final inequality. We can thus conclude from (3.3) that there exists a vertex $w_i \in W_i$ such that $w_i \notin B_j$ for all $i + 1 \leq j \leq r - 1$. We claim that choosing $S_i = S_{i-1} \cup \{w_i\}$ completes the inductive step. Indeed, $S_i \in K_i(G')$ as w_i was chosen from the common neighbourhood of S_{i-1} in G'. Also, fixing some $i + 1 \leq j \leq r - 1$, we see that $N^G(w_i)$ intersects $W_j = N_{U_j}^{G'}(S_{i-1})$ in at least $p|W_j|/2$ vertices. Furthermore, at most

$$\alpha^2 p^y n \leq \alpha^2 p^{x_{i+1}+i} n \leq \frac{\alpha}{2^{2r}} p^{x_j+i} n \leq p|W_j|/4$$

edges adjacent to w_i lie in $\tilde{G}$, using the definition of y, the upper bound on α, the fact that $x_j \le x_{i+1}$ and (3.2). Therefore we can conclude that for all $i+1 \le j \le r$, we have $\deg^{G'}_{U_j}(S_i) \ge \deg^{G'}_{W_j}(S_i) \ge p|W_j|/4 \ge (p/4)^i |U_j|$, as required. This completes the induction and the proof. ∎

We now collect some easy consequences of Lemma 3.4 for reference later in the proof.

Corollary 3.5. *For any $r \in \mathbb{N}_{\ge 3}$ and $0 < \alpha < 1/2^{2r}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \le \varepsilon p^{r-1} n$:*

(1) *Let $\tilde{G}$ be any subgraph $\tilde{G}$ of G with maximum degree less than $\alpha^2 p^{r-1} n$.*
   - (i) *For any $U_1, \dots, U_{r-1} \subseteq V(G)$ such that $|U_i| \ge \alpha p n$ for $i \in [r-1]$, there exists an $(r-1)$-clique $S \in K_{r-1}(G \setminus \tilde{G})$ traversing the U_i.*
   - (ii) *For any $U_1, \dots, U_r \subseteq V(G)$ such that $|U_1| \ge \alpha p^{2r-4} n$ and $|U_i| \ge \alpha n$ for $2 \le i \le r$, there exists an r-clique $S \in K_r(G \setminus \tilde{G})$ traversing the U_i.*

(2) *For any $U_1, \dots, U_r \subseteq V(G)$ such that $|U_1| \ge \alpha p^{r-1} n$, $|U_i| \ge \alpha p n$ for $2 \le i \le r-2$ and $|U_{r-1}|, |U_r| \ge \alpha n$, there exists an r-clique $S \in K_r(G)$, traversing the U_i.*

(3) *For any $W_0, W_1, W_2 \subseteq V(G)$ such that $|W_0|, |W_1|, |W_2| \ge \alpha n$, there exists an $S \in K_{r-1}(G[W_0])$ such that $\deg_{W_j}(S) \ge \alpha^2 p^{r-1} n$ for $j = 1, 2$.*

Proof. Fix $\varepsilon > 0$ small enough to apply Lemma 3.4. This is predominantly a case of plugging in the values and checking the conditions of Lemma 3.4. For part (1), we let $G' = G \setminus \tilde{G}$. Then for (1) (i), we take $r^* = r-2$, $x_i = 1$ for $1 \le i \le r+1$ and $y = r-1$. We thus see that for $i \in [r^*]$, $x_i + x_{i+1} + 2i = 2 + 2i \le 2r - 2$ and $x_{i+1} + i = 1 + i \le r - 1 = y$. Therefore taking U_i for $1 \le i \le r-1$ with $|U_i| \ge \alpha p n$ (and defining $U_{r+1} = U_r = U_{r-1}$), Lemma 3.4 gives us an $(r-2)$-clique $S' \in K_{r-1}(G')$ traversing $U_1, \dots, U_{r-2}$ such that $\deg^{G'}_{U_{r-1}}(S') \ge \alpha^2 p^{r-1} n > 0$ (here Fact 3.2 shows positivity). Therefore choosing any vertex $v \in N^{G'}_{U_{r-1}}(S')$ and fixing $S = S' \cup \{v\}$ gives the required clique.

The other cases are similar. For part (1) (ii), we fix $r^* = r-1$, $x_1 = 2r-4$, $x_i = 0$ for $2 \le i \le r+1$ and $y = r-1$. Again, it is easily checked that the conditions on the x_i are all satisfied and so applying Lemma 3.4 (fixing $U_{r+1} = U_r$) gives an $(r-1)$-clique S' in G' traversing $U_1, \dots, U_{r-1}$ such that S' has a nonempty G'-neighbourhood in U_r. Therefore adding any vertex in this neighbourhood to S' gives the required r-clique $S \in K_r(G')$.

For part (2), we fix $r^* = r-1$, $x_1 = r-1$, $x_i = 1$ for all i such that $2 \le i \le r-2$ and $x_{r-1} = x_r = x_{r+1} = 0$. We also let $\tilde{G}$ be the empty graph and so $G = G'$. Now note that for $r = 3$, we have $x_1 = 2$ and $x_2 = 0$ and so $x_1 + x_2 + 2 = 4 = 2r - 2$, whilst for $r \ge 4$, we have $x_1 + x_2 + 2 = r + 2 \le 2r - 2$. Conditions (3.1) for $2 \le i \le r^* = r-1$ can be similarly checked. Therefore Lemma 3.4 gives an $(r-1)$-clique $S' \in K_{r-1}(G)$ traversing $U_1, \dots, U_{r-1}$ such that $N^G_{U_r}(S') \ne \emptyset$ and so as above, we extend S' to the required r-clique S.

Finally, for part (3) we fix $r^* = r-1$, $x_i = 0$ for all $1 \le i \le r$ and define our sets as $U_i = W_0$ for $i \in [r-1]$ and $U_r = W_1$, $U_{r+1} = W_2$. Applying Lemma 3.4 then directly gives the required $(r-1)$-clique $S \in K_{r-1}(G[W_0])$ (again here $\tilde{G}$ is taken to be empty). ∎

3.3. Concentration of random variables

We will use the following well-known concentration bounds (see e.g. [38, Theorem 2.1, Corollary 2.4 and Theorem 2.8]).

Theorem 3.6 (Chernoff bounds). *Let X be the sum of a set of mutually independent Bernoulli random variables and let $\lambda = \mathbb{E}[X]$. Then for any $0 < \delta < 3/2$, we have*

$$\mathbb{P}[X \geq (1+\delta)\lambda] \leq e^{-\delta^2\lambda/3} \quad \text{and} \quad \mathbb{P}[X \leq (1-\delta)\lambda] \leq e^{-\delta^2\lambda/2}.$$

Furthermore, if $x \geq 7\lambda$, then $\mathbb{P}[X \geq x] \leq e^{-x}$.

3.4. Perfect fractional matchings

Given an r-uniform hypergraph H, a *fractional matching* in H is a function $f : E(H) \to \mathbb{R}_{\geq 0}$ such that $\sum_{e:\, v\in e} f(e) \leq 1$ for all $v \in V(H)$. We say the fractional matching is *perfect* if $\sum_{e:\, v\in e} f(e) = 1$ for all $v \in V(H)$. The *value* of a fractional matching f is $|f| := \sum_{e\in E(H)} f(e)$. The maximum value $|f|$ over all choices of fractional matching f of H, we call the *fractional matching number* of H, which we denote by $\nu^*(H)$.

A *fractional cover* of H is a function $g : V(H) \to \mathbb{R}_{\geq 0}$ such that for all $e \in E(H)$, one has $\sum_{v\in e} g(v) \geq 1$. The *value* of a fractional cover g is $|g| := \sum_{v\in V(H)} g(v)$. The *fractional cover number* of H, denoted $\tau^*(H)$, is then the minimum value of a fractional cover g of H.

For an r-uniform hypergraph H, the fractional matching number of H can be encoded as the optimal solution of a linear program. Taking the dual of this linear program gives another linear program which outputs the fractional cover number as an optimal solution. The duality theorem from linear programming thus tells us that $\nu^*(H) = \tau^*(H)$ for any hypergraph H. Using this, as well as the so called ‘complementary slackness conditions’ that follow from the duality theorem, one can derive the following simple consequences (see e.g. [50, Proposition 2] or [35, Proposition 2.4]).

Proposition 3.7. *For any r-uniform hypergraph H on N vertices, the following hold:*

(1) *$\nu^*(H) \leq N/r$, with equality if and only if there exists a perfect fractional matching in H.*

(2) *$\nu^*(H) \geq \nu(H)$ where $\nu(H)$ denotes the size of the largest matching in H.*

(3) *If $g : V(H) \to \mathbb{R}_{\geq 0}$ is a fractional cover and $U \subset V(H)$, then $g' := g|_U : U \to \mathbb{R}_{\geq 0}$ is a fractional cover of $H[U]$ and hence $|g'| = \sum_{u\in U} g(u) \geq \tau^*(H[U]) = \nu^*(H[U])$.*

(4) *If $g : V(H) \to \mathbb{R}_{\geq 0}$ is an optimal fractional cover, i.e. $|g| = \tau^*(H)$, then $\nu^*(H) \geq |W|/r$ where $W := \{v \in V(H) : g(v) > 0\}$.*

We now give two lemmas, exploring some simple conditions which guarantee the existence of a perfect fractional matching.

Lemma 3.8. *Suppose H is an N-vertex, r-uniform hypergraph such that given any vertex $v \in V(H)$ and any subset $W \subseteq V(H) \setminus \{v\}$ of at least $N/(2r)$ vertices, there exists* *an edge in H containing v and $r-1$ vertices of W. Then H has a perfect fractional matching.*

Proof. Suppose for a contradiction that H does *not* have a perfect fractional matching. Thus, by Proposition 3.7 (1), if we take $g : V(H) \to \mathbb{R}_{\geq 0}$ to be an optimal fractional cover of H so that $|g| = \nu^*(H) = \tau^*(H)$ we have $|g| < N/r$. Hence, if we order the vertices in decreasing weight order according to g, we see by Proposition 3.7 (4) that $g(w) = 0$, where w is the final vertex in this order. Take $W \subset V(H) \setminus \{w\}$ to be the set of $N/(2r)$ vertices preceding w in the order. Then by the condition of the lemma, there exists an edge using w and $r-1$ vertices of W. Since $g(w) = 0$, there is some vertex w' in W with $g(w') \geq 1/(r-1)$. Therefore all vertices preceding W in the order (as well as w') have at least this weight and in total

$$|g| \geq \frac{N - N/r}{r-1} \geq N/r,$$

a contradiction. ∎

Given a vertex subset $U \subset V := V(H)$ in a hypergraph H, a *fan focused at* U in H is a subset $F \subset E(H)$ of edges of H such that $|e \cap U| = 1$ for all $e \in F$ and $e \cap e' \cap (V \setminus U) = \emptyset$ for all $e \neq e' \in F$. In words, each edge of a fan intersects U in exactly one vertex and outside of U, the edges in a fan are pairwise disjoint. The *size* of a fan is simply the number of edges in the fan. If $U = \{u\}$ is a single vertex, we simply refer to a fan focused at u.

Lemma 3.8 shows that if H has the property that the link of every vertex v has no large independent sets, then it must have a perfect fractional matching. In fact, we do not necessarily need such an expansion property to hold locally at every vertex and can instead focus on subsets of vertices, if we have an added condition that every vertex has a large enough fan focused at it. This is the content of the following lemma.

Lemma 3.9. *Suppose H is an N-vertex, r-uniform hypergraph and there exists $r \leq M \leq N/(2r)$ such that*

(i) *for all $v \in V(H)$ there is a fan focused at v in H of size M;*

(ii) *for every subset $W_0 \subset V(G)$ with $|W_0| = M$ and every subset $W_1 \subset V(G) \setminus W_0$ with $|W_1| \geq N/(2r)$, there exists an edge of H with one vertex in W_0 and the other $r-1$ vertices in W_1.*

Then H has a perfect fractional matching.

Proof. We start by noticing that (ii) leads to the following two consequences:

(a) For all $U \subset V(H)$ with $|U| = (r-1)M$, fixing $V' := V(H) \setminus U$ we find that for all $U' \subset V'$ such that $|U'| = M$, there is a fan of size $N/r - M$ focused at U' in $H[V']$.

(b) Every subset of at least N/r vertices of H induces an edge in H.

Indeed, for U' as in (a) we can build the fan $F_{U'}$ focused at U' greedily. Whilst $|F_{U'}| \leq N/r - M$, we see that $W := V(G) \setminus (V(F_{U'}) \cup U' \cup U)$ has size at least

$$N - (N/r - M)(r-1) - M - (r-1)M \geq N - (2r-1)N/(2r) \geq N/(2r),$$

using $M \le N/(2r)$ here. Hence we can find an edge using one vertex of U' and $r-1$ vertices of W which extends the fan $F_{U'}$. Condition (b) also follows easily because taking W' to be a set with N/r vertices, we find that for any $W'' \subset W'$ with $|W''| = M$, there is an edge containing a vertex in W'' and $r-1$ vertices of $W' \setminus W''$ from (ii).

Now we turn to the main proof. We fix $g : V(H) \to \mathbb{R}_{\ge 0}$ to be an optimal fractional cover and suppose for a contradiction that $|g| < N/r$. We deduce the existence of a vertex $w \in V(H)$ with $g(w) = 0$ and a fan F_w focused at w of size M. Taking $U_1 := \bigcup\{e \setminus \{w\} : e \in F_w\}$, we have $|U_1| = (r-1)M$ and $\sum_{u \in U_1} g(u) \ge M$.

Now consider $V' := V(H) \setminus U_1$. If $\nu^*(H[V']) \ge N/r - M$ then we can conclude that $\sum_{v \in V'} g(v) \ge N/r - M$ from Proposition 3.7 (3), which implies that $|g| \ge N/r$, a contradiction. Hence

$$\nu^*(H[V']) < \frac{N}{r} - M = \frac{N' - M}{r}, \tag{3.4}$$

where $N' := |V'| = N - (r-1)M$. We fix $g' : V' \to \mathbb{R}_{\ge 0}$ to be some optimal fractional cover of $H[V']$ with $|g'| = \nu^*(H[V'])$. By Proposition 3.7 (4), we therefore deduce that there is some set $U_2 \subset V'$ with $|U_2| = M$ and $g'(u') = 0$ for all $u' \in U_2$. By (a) there exists a fan F_{U_2} of size $N/r - M$ focused at U_2 in $H[V']$. Taking $Z := \bigcup\{e : e \in F_{U_2}\} \setminus U_2$, we have $|Z| = (r-1)(N/r - M)$ and similarly to before, using the fact that for each edge $e \in F_{U_2}$ we have $\sum_{v \in e} g'(v) \ge 1$ and $g'(u') = 0$ for all $u' \in U_2$, we can conclude that $\sum_{z \in Z} g'(z) \ge |F_{U_2}| = N/r - M$.

Finally, we look at $V'' := V' \setminus Z$. We have $N'' := |V''| = N' - (r-1)(N/r) + (r-1)M$ and using (b) and Proposition 3.7 (2), we find that

$$\nu^*(H[V'']) \ge \frac{N'' - N/r}{r} = \frac{N' + (r-1)M}{r} - \frac{N}{r}.$$

Hence, by Proposition 3.7 (3), we deduce that $\sum_{v'' \in V''} g'(v'') \ge (N' + (r-1)M)/r - N/r$. Combining this with the lower bound on the sum of g' values on Z implies that $|g'| = \nu^*(H[V']) \ge (N' - M)/r$, contradicting (3.4). ∎

3.5. *Almost perfect matchings in hypergraphs*

It is well known that hypergraphs that have roughly regular vertex degrees and small codegrees contain large matchings. This is often referred to as Pippenger's Theorem but there are in fact a family of similar results, all following from the "semi-random" or "nibble" method (see e.g. [10, Section 4.7]). Here we use the following explicit version which follows directly from a result of Kostochka and Rödl [49].

Theorem 3.10. *For any integers $r \ge 3$ and $K \ge 4$ there exists $\Delta_0 > 0$ such that for all $\Delta \ge \Delta_0$ the following holds. If H is a r-uniform hypergraph on N vertices such that*

(1) *for all vertices $v \in V(H)$, we have* $\deg(v) = \Delta(1 \pm K\sqrt{(\log \Delta)/\Delta})$;

(2) *for all $u \ne v \in V(H)$, we have* $\operatorname{codeg}(u, v) \le \Delta^{1/(2r-1)}$,

then H has a matching covering all but at most $\Delta^{-1/r} N$ vertices.

Indeed, [49, Theorem 4] states that for all $r \geq 3$, $K_0 \geq 8$ and reals $0 < \delta, \gamma < 1$, there exists a D_0 such that if H is an r-uniform hypergraph on N vertices with

$$D - K_0\sqrt{D \log D} \leq \deg^H(v) \leq D,$$

for all $v \in V(H)$, where $D \geq D_0$, and $\operatorname{codeg}^H(u,v) \leq C < D^{1-\gamma}$ for all pairs of vertices $u \neq v$, then H has a matching covering all but at most $O(N(C/D)^{(1-\delta)/(r-1)})$ vertices. In order to derive Theorem 3.10 from this we fix $K_0 = 2K$, $\delta = \frac{1}{4r}$ and $\gamma = \frac{2r-2}{2r-1}$. Letting D_0 be the resulting constant given by [49, Theorem 4], we fix $\Delta_0 \geq D_0$ to be some large constant. Hence, our conditions (1) and (2) of Theorem 3.10 guarantee that H satisfies the conditions of [49, Theorem 4] with $D = \Delta + K\sqrt{\Delta \log \Delta}$ and $C = \Delta^{1-\gamma}$. Now note that

$$\frac{C}{D} = (1+o(1))\Delta^{-\gamma} = (1+o(1))\Delta^{-(2r-2)/(2r-1)} = o(\Delta^{-(4r-4)/(4r-1)}).$$

Combining this with the fact that $\frac{1-\delta}{r-1} = \frac{4r-1}{4r(r-1)}$ and $\Delta \geq \Delta_0$ is sufficiently large, we conclude from [49, Theorem 4] that the number of vertices uncovered by a largest matching is always less than $\Delta^{-1/r}N$, as required.

Clearly, in order to prove that a hypergraph H has a large matching, it suffices to establish the conditions of Theorem 3.10 for a spanning subgraph $H' \subset H$. An idea introduced by Alon, Frankl, Huang, Rödl, Ruciński and Sudakov [8] is to find such an H' as a random subhypergraph of H and guarantee that the conditions of Theorem 3.10 hold for H' by using perfect fractional matchings to dictate the probability with which we take each edge into H'. This idea was then used in the context of finding almost K_r-factors in pseudorandom graphs by Han, Kohayakawa and Person [34, 35]. We will also adopt this idea and so give the following theorem.

Theorem 3.11. *For all $r \in \mathbb{N}_{\geq 3}$ and $0 < \eta < 1/2$, there exists an N_0 such that the following holds for all $N \geq N_0$. Suppose H is an N-vertex, r-uniform hypergraph such that there exist $t := 2N^\eta$ perfect fractional matchings $f_1, \dots, f_t : E(H) \to \mathbb{R}_{\geq 0}$ in H with the property that*

$$\sum_{i=1}^{t} \sum_{e \in E(H)\colon \{u,v\} \subset e} f_i(e) \leq 2 \tag{3.5}$$

for all pairs of vertices $u \neq v \in V(H)$. Then H has a matching covering all but at most $N^{1-\eta/r}$ vertices.

Proof. We take a random subgraph $H' \subseteq H$ by keeping every edge $e \in E(H)$ independently with probability $p_e = \sum_{i=1}^{t} f_i(e)/2$, noting that $p_e \in [0,1]$ for all $e \in E(H)$ due to (3.5). We fix $\Delta := t/2 = N^\eta$ and $K := 4/\eta$ and claim that H' satisfies the conditions of Theorem 3.10 whp as N tends to infinity.

To check that H' satisfies the conditions of Theorem 3.10, note that for each $v \in V$ we have

$$\mathbb{E}[\deg^{H'}(v)] = \sum_{e\colon v \in e} p_e = \sum_{e\colon v\in e} \sum_{i=1}^{t} f_i(e)/2 = \frac{1}{2}\sum_{i=1}^{t}\Big(\sum_{e\colon v\in e} f_i(e)\Big) = t/2 = \Delta,$$

using the fact that each f_i is a perfect fractional matching. Applying Theorem 3.6 then gives

$$\mathbb{P}\left[\deg^{H'}(v) \neq \Delta\left(1 \pm K\sqrt{\frac{\log\Delta}{\Delta}}\right)\right] \leq 2\exp\left(-\frac{K^2\log\Delta}{3}\right) \\ \leq 2\exp\left(-\frac{K^2\eta\log N}{3}\right) \leq \frac{1}{N^2} \tag{3.6}$$

for N sufficiently large. Similarly, for $u \neq v \in V(H)$, we find that $\mathbb{E}[\operatorname{codeg}^{H'}(u,v)] = \sum_{e:\{u,v\}\subset e} p_e \leq 1$ by (3.5) and applying Theorem 3.6 gives

$$\mathbb{P}[\operatorname{codeg}^{H'}(u,v) \geq \Delta^{1/(2r-1)}] \leq \exp(-\Delta^{1/(2r-1)}) \leq \frac{1}{N^3} \tag{3.7}$$

for large N. Hence taking a union bound over all vertices and pairs of vertices and upper bounding the failure probabilities with (3.6) and (3.7) shows that H' satisfies the conditions of Theorem 3.10 whp. Therefore for N (and hence Δ) sufficiently large, we can fix such an instance of H' and apply Theorem 3.10 which gives the large matching in H' and hence in H, concluding the proof. ∎

It will be useful for us to work with the following corollary to Theorem 3.11 which gives us a sufficient condition for us to be able to generate the perfect fractional matchings needed in Theorem 3.11 via a greedy process. Recall that for a 2-uniform graph J on $V(H)$, H_J denotes the subhypergraph of H given by all edges of H which contain some edge of J.

Theorem 3.12. *For all $r \in \mathbb{N}_{\geq 3}$ and $0 < \gamma < 1/(2r^2)$, there exists an N_0 such that the following holds for all $N \geq N_0$. Suppose H is an N-vertex, r-uniform hypergraph such that given any* graph *J on $V(H)$ of maximum degree at most $N^{r^2\gamma}$, the set $H \setminus H_J$ contains a perfect fractional matching. Then H has a matching covering all but at most $N^{1-\gamma}$ vertices.*

Proof. We will prove this by appealing to Theorem 3.11 with $\eta := r\gamma$ and so we set out to find $t := 2N^\eta$ perfect fractional matchings $f_1, \dots, f_t$ such that (3.5) holds. We do this algorithmically, finding the f_i one at a time. We begin by defining J_1 to be the empty (2-uniform) graph on $V(H)$ and for $1 \leq i \leq t$ we do the following. We find a perfect fractional matching f_i in $H \setminus H_{J_i}$ and add this to our family of perfect fractional matchings. We then define a graph G_i with vertex set $V(H)$ and a pair of vertices $\rho \in \binom{V(H)}{2}$ forming an edge in G_i if

$$\sum_{\rho\subset e\in E(H)} f_i(e) \geq \frac{N^{-\eta}}{2}.$$

Finally, we define $J_{i+1} := J_i \cup G_i$ and move to step $i+1$.

We claim that this algorithm does not stall and we complete our collection of t perfect fractional matchings. In order to check this, we need to verify that we can find a perfect fractional matching in $H \setminus H_{J_j}$ for each $j \in [t]$. This follows because at each step i we have, for any $v \in V(H)$,

$$\sum_{u \in V(H)\setminus\{v\}} \Big(\sum_{\{u,v\} \subset e \in E(H)} f_i(e) \Big) = (r-1) \sum_{v \in e \in E(H)} f_i(e) = r-1,$$

as f_i is a perfect fractional matching. Hence the number of pairs $\rho \in \binom{V(H)}{2}$ which contain v and form an edge of G_i is at most $2(r-1)N^{\eta}$. As this holds for all choices of $v \in V(H)$ we see that G_i has maximum degree less than $2(r-1)N^{\eta}$. Thus for each $j \in [t]$, $J_j := \bigcup_{i=1}^{j-1} G_i$ has maximum degree less than

$$2(r-1)N^{\eta} \cdot (j-1) \le 2(r-1)N^{\eta} t = 4(r-1)N^{2\eta} \le N^{r\eta} = N^{r^2\gamma}$$

for N sufficiently large. So $H \setminus H_{J_i}$ does indeed host a perfect fractional matching by assumption.

Finally, we need to check condition (3.5) for each pair of vertices $\rho = \{u, v\} \in \binom{V(H)}{2}$. Note that for any pair $\rho \in \binom{V(H)}{2}$ of vertices of H we have

$$\sum_{i=1}^{t} \Big(\sum_{\rho \subset e \in E(H)} f_i(e) \Big) \le \frac{tN^{-\eta}}{2} \le 1$$

if ρ does not feature as an edge in any of the G_i. On the other hand, if $\rho = \{u, v\} \in E(G_j)$ for some $j \in [t]$, then because we forbid the edges of H containing $\{u, v\}$ from being used again we have

$$\sum_{i=j+1}^{t} \Big(\sum_{\rho \subset e \in E(H)} f_i(e) \Big) = 0.$$

Also $\rho \notin E(G_i)$ for $i < j$ as otherwise there could be no weight on (edges containing) ρ in f_j. Hence

$$\sum_{i=1}^{j-1} \Big(\sum_{\rho \subset e \in E(H)} f_i(e) \Big) \le \frac{(j-1)N^{-\eta}}{2} \le 1,$$

and using the fact that

$$\sum_{\rho \subset e \in E(H)} f_j(e) \le \sum_{u \in e \in E(H)} f_j(e) = 1$$

shows that (3.5) holds for all $\rho \in \binom{V(H)}{2}$ as required. So by Theorem 3.11, H contains a matching covering all but at most $N^{1-\eta/r} = N^{1-\gamma}$ vertices, concluding the proof. ■

3.6. Templates

In this section we concentrate on a powerful new approach introduced by Montgomery [57, 58], in his work on spanning trees in random graphs. The general idea is to use the following key notion as an auxiliary graph to define absorbing structures in the host graph of interest.

Definition 3.13. A *template* $\mathcal{T}$ with *flexibility* $t \in \mathbb{N}$ is a bipartite graph on $7t$ vertices with vertex classes I and $J_1 \cup J_2$ such that $|I| = 3t$, $|J_1| = |J_2| = 2t$, and for any $\bar{J} \subset J_2$ with $|\bar{J}| = t$, the induced graph $\mathcal{T}[V(\mathcal{T}) \setminus \bar{J}]$ has a perfect matching. We call J_2 the *flexible* set of vertices for the template.

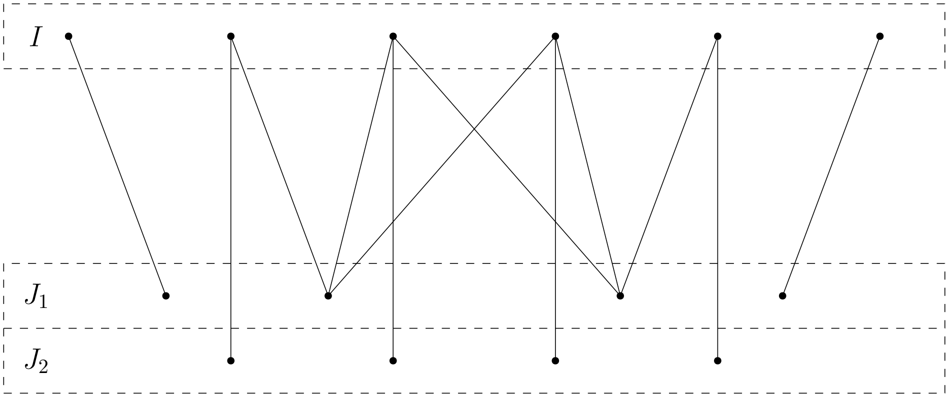

Fig. 5. A template $\mathcal{T}$ of flexibility 2. One can check that the key property is indeed satisfied.

See Figure 5 for an example of a template. The definition implies that a template is *robust* with respect to having a perfect matching. It is not hard to come up with examples of templates, indeed a complete bipartite graph certainly satisfies the condition. The utility of the notion for defining absorbing structures that are possible to find in the desired host graphs comes with the fact that *sparse templates* exist. Indeed, Montgomery [57,58] proved the following using a probabilistic argument.

Theorem 3.14. *For all sufficiently large t, there exists a template of flexibility t and maximum degree* 40.

Han, Kohayakawa, Person and the author [33] then showed how to derandomise the argument for the existence of templates and find templates with bounded maximum degree efficiently in polynomial time. We will use the method of template absorption in proving Proposition 2.9.

4. Diamond trees

Recall the definition of diamond trees from Section 2, namely Definition 2.1. In this section we prove the existence of diamond trees in our bijumbled graphs. The main aim is to prove the following proposition which gives us some flexibility over which vertices feature as removable vertices of our diamond tree. This will turn out to be very valuable at various points in our proof.

Proposition 4.1. *For any $r \in \mathbb{N}_{\geq 3}$ and $0 < \alpha < 1/2^{2r}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$. For any $2 \leq z \leq \alpha n$ and any pair of disjoint vertex subsets $U, W \subset V(G)$ such that $|U|, |W| \geq 4\alpha r n$, there exist disjoint vertex subsets $X, Y \subset U$ such that*

(1) $|X| + |Y| = z$;

(2) $|X| \leq \max\{1, 2z/d_*\}$ *with* $d_* = \alpha^2 p^{r-1} n$;

(3) *for* any *subset $Y' \subset Y$, there exists a K_r-diamond tree $\mathcal{D} = (T, R, \Sigma)$ such that $R = X \cup Y'$ and $\Sigma \subset K_{r-1}(G[W])$ is a matching of $(r-1)$-cliques in W.*

Let us pause to digest the proposition. Firstly, note that by choosing $Y' = Y$ in (3) and varying z, we can guarantee the existence of K_r-diamond trees of any order up to linear in our bijumbled graph G. However, the proposition is much more powerful than just this. The vertex set Y and property (3) allow us *flexibility* in which vertices appear in the removable set of vertices of the diamond tree we take from the proposition. We can start with z much larger than the desired order of the diamond tree we want and then remove unwanted vertices from Y to end up with some Y' that we include in the removable vertices of the diamond tree. The point is that by starting with a larger z (and hence larger $|Y|$), we can deduce stronger properties about the vertices in Y, allowing us to then ensure properties of the set of removable vertices R that we would otherwise have no hope in guaranteeing. There is a catch, as we are forced to include the set X in any diamond tree we produce, but note that due to property (2), the size of X is negligible compared to the size of Y. Indeed, due to Fact 3.2, d_* is polynomial in n (of order at least $\Omega(n^{(r-2)/(2r-3)})$, to be precise). Thus we can choose Y' to be much smaller than Y and still have the vertices in Y' contribute a significant subset (at least half, say) of the removable vertices of the diamond tree we obtain. We delay applications of Proposition 4.1 to later in the proof but refer the reader to Lemmas 6.6, 7.2 and 8.4 for a flavour of the consequences of the proposition.

The rest of this section is concerned with proving Proposition 4.1. The idea behind the proof is simple: we look to find a large (order z) K_r-diamond tree in G with the property that many of the removable vertices are leaves (the set Y). This allows us to pick and choose which leaves (the set Y') we include in our desired diamond tree, as we can simply remove the other leaves and their corresponding interior cliques;[7] see Figure 6 for an example. In order to find diamond trees with many leaves, we introduce the notion of a scattered diamond tree and deduce the existence of such diamond trees in a suitably pseudorandom graph.

[7]That is, for each unwanted leaf $v \in R$ in the diamond tree $\mathcal{D} = (T, R, \Sigma)$, we remove the interior clique in Σ which corresponds to the edge adjacent to the (preimage of) v in the defining tree T, as well as the leaf v itself.

4.1. Scattered diamond trees

One way to ensure a large set of leaves in a tree is to impose a minimum degree on all non-leaf vertices. This leads to the following definition.

Definition 4.2. We say a tree T (of order at least 2) is *d-scattered* if every vertex in $V(T)$ which is not a leaf in T has degree at least d. As a convention we will also say that a tree of order 1 (a single vertex) is d-scattered for all d. We say a diamond tree $\mathcal{D} = (T, R, \Sigma)$ is d-scattered if its underlying auxiliary tree T is d-scattered.

See Figure 6 for an example of a scattered K_3-diamond tree.The following simple lemma shows that most of the vertices in a scattered tree (and hence most of the removable vertices in a scattered diamond tree) are leaves.

Lemma 4.3. *Let $d \geq 2$ and suppose that T is a d-scattered tree of order $m \geq 3$. Then defining $X \subset V(T)$ to be the vertices*[8] *which have degree greater than 1 in T, we have $|X| \leq \frac{m-2}{d-1}$.*

Proof. By the definition of d-scattered trees, every vertex in X has degree at least d. We define $x := |X|$. Note that $T[X]$ is a connected subtree of T. Indeed, the interior vertices of a path between any two vertices of T must lie in X (as they have degree at least 2). Hence $T[X]$ has exactly $x - 1$ edges and we can estimate the number of edges in T as follows:

$$e(T) = m - 1 = \sum_{v \in X} \deg(v) - e(T[X]) \geq xd - (x - 1).$$

Rearranging, one obtains $x \leq \frac{m-2}{d-1}$, as required. ∎

We will show that we can find large scattered diamond trees in our bijumbled graph. To begin with, we focus on diamond trees for which the auxiliary tree is a star, which we call *diamond stars*. The next lemma shows that we can find large diamond stars in a suitably pseudorandom graph.

Lemma 4.4. *For any $r \in \mathbb{N}_{\geq 3}$ and $0 < \alpha < 1/2^{2r}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$, fixing $d_* := \alpha^2 p^{r-1} n$. Let $U_0, U_1, U_2 \subseteq V(G)$ be disjoint vertex subsets such that $|U_i| \geq \alpha n$ for $i = 0, 1$ and $|U_2| \geq \alpha r n$. Then there exists a K_r-diamond tree $\mathcal{D}^* = (T^*, R^*, \Sigma^*)$ in G such that T^* is a star of order $1 + d_*$ centred at x, say, with*[9] *$\rho^*(x) \in U_0$, $R^* \setminus \{\rho^*(x)\} \subset U_1$ and $\Sigma^* \subset K_{r-1}(G[U_2])$ a matching of $(r-1)$-cliques in U_2.*

Proof. Fix $\varepsilon > 0$ small enough to apply Corollary 3.5 (3). Shrink U_0 (if necessary) to be a set of exactly αn vertices. We claim that there is a matching $M \subset K_{r-1}(G[U_2])$ of $(r-1)$-cliques such that $|M| = \alpha n$ and each clique $S \in M$ has $\deg_{U_i}(S) \geq d_*$ for

[8] That is, X is the set of vertices of T which are not leaves.

[9] Here $\rho^* : V(T^*) \to R^*$ is the associated bijection in the definition of $\mathcal{D}^*$.

$i = 0, 1$. Indeed, we can find M greedily by applying Corollary 3.5 (3) (with $W_i = U_{2-i}$ for $i = 0, 1, 2$) repeatedly, adding an $(r-1)$-clique S to M and removing its vertices from U_2 after each application. While $|M| \leq \alpha n$, we have $|U_2| \geq \alpha n$ and so we are indeed in a position to apply Corollary 3.5 (3) throughout the process.

Now once we have found M, for each $S \in M$ and for $i = 0, 1$, let $N_i(S) := N_{U_i}(S)$, that is, the set of vertices in U_i which form a K_r with S. By construction we have $|N_0(S)| \geq d_*$ for each S in M and so

$$|\{(v, S) \in U_0 \times M : v \in N_0(S)\}| \geq |M| d_* = \alpha n d_*.$$

Hence, as $|U_0|$ has size αn (as we imposed at the start of the proof), by averaging, there exists a vertex $v_0 \in U_0$ and a subset Σ^* of d_* cliques in M such that v_0 is in $N_0(S)$ for all $S \in \Sigma^*$. We can now construct our diamond star greedily, with v_0 as the image of the large degree vertex. Sequentially, for each clique S in Σ^*, choose a vertex u in $N_1(S)$ which has not been previously chosen and add the copy of K_{r+1}^- on S, v_0 and u to the diamond star (adding u to R^*). As $N_1(S) \geq d_*$ for all $S \in \Sigma^* \subseteq M$, there is always an option for u and so this process succeeds in building the required diamond star. ∎

Our next lemma follows the scheme of Krivelevich [51] to construct large diamond trees. We adapt his proof to guarantee that the diamond tree obtained is scattered.

Lemma 4.5. *For any $r \in \mathbb{N}_{\geq 3}$ and $0 < \alpha < 1/2^{2r}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$, fixing $d_* := \alpha^2 p^{r-1} n$. For any $2 \leq z \leq \alpha n$ and any pair of disjoint vertex subsets $U, W \subset V(G)$ such that $|U|, |W| \geq 4\alpha r n$, there exists a d_*-scattered K_r-diamond tree $\mathcal{D}_{\mathrm{sc}} = (T_{\mathrm{sc}}, R_{\mathrm{sc}}, \Sigma_{\mathrm{sc}})$ of order m such that $z \leq m \leq z + d_*$, $R_{\mathrm{sc}} \subset U$ and $\Sigma_{\mathrm{sc}} \subset K_{r-1}(G[W])$ is a matching of $(r-1)$-cliques in $G[W]$.*

Proof. Our proof is algorithmic and works by building a diamond tree forest, that is, a set of pairwise vertex disjoint diamond trees. At each step of the algorithm, we will add to one of the trees in our forest, boosting the degree of a vertex in the underlying auxiliary tree by d_*, using Lemma 4.4. By discarding trees when the sum of the orders of the trees gets too large, we will show that one of the trees in our forest will eventually obtain the desired order after finitely many steps of the algorithm. The details follow.

Initiate the process by fixing $U_0 \subset U$ to be an arbitrary subset of αn vertices, $W_0 = \emptyset \subset W$ to be empty and $\mathcal{D}_1, \ldots, \mathcal{D}_\ell$ with $\ell = \alpha n$ to be the diamond trees which are defined to be the single vertices in U_0. That is, for $i \in [\ell]$, the K_r-diamond tree $\mathcal{D}_i = (T_i, R_i, \Sigma_i)$ corresponds to an auxiliary tree T_i which is just a single vertex and thus R_i is also a single vertex and Σ_i is empty. In general, at each step of the process we will have a family $\mathcal{D}_1, \ldots, \mathcal{D}_\ell$ (for some $\ell \in \mathbb{N}$) of vertex disjoint K_r-diamond trees such that for each i, the diamond tree $\mathcal{D}_i = (T_i, R_i, \Sigma_i)$ is d_*-scattered, has $R_i \subset U_0$ and $\Sigma_i \subset G[W_0]$. Furthermore, we will have $U_0 = \bigcup_{i \in [\ell]} R_i$ and $W_0 = \bigcup_{i \in \ell} V(\Sigma_i) \subset W$ and maintain throughout $\alpha n \leq |U_0| \leq 2\alpha n$ and $|W_0| \leq 2(r-1)\alpha n$.

Now at each step, given such a set U_0 and family $\mathcal{D}_1, \ldots, \mathcal{D}_\ell$, we apply Lemma 4.4 with $U_1 = U \setminus U_0$ and $U_2 = W \setminus W_0$, noting that the conditions on the size of U and W in the statement of the lemma and the imposed conditions on the size of U_0 and W_0 throughout the process indeed allow Lemma 4.4 to be applied. Thus, we find a K_r-diamond star $\mathcal{D}^* = (T^*, R^*, \Sigma^*)$ of order $d_* + 1$ with centre $v_0 \in U_0$, $R^* \setminus \{v_0\} \subset U \setminus U_0$ and $\Sigma^* \subset K_{r-1}(G[U_2])$ a matching of $(r-1)$-cliques. As U_0 is the union of the removable vertices of the family of diamond trees, there is some $i_0 \in [\ell]$ such that $v_0 \in R_{i_0}$. We then update $\mathcal{D}_{i_0}$ by adjoining the diamond star to the tree at v_0, we add all the vertices of R^* to U_0, and all the vertices of the $(r-1)$-cliques in Σ^* to W_0. Now if there is a K_r-diamond tree among the (new) family $\mathcal{D}_1, \dots, \mathcal{D}_\ell$ which has order at least z, we take such a diamond tree as $\mathcal{D}_{\mathrm{sc}}$ and finish the process. If not, then we look at the size of U_0. If $|U_0| < 2\alpha n$, we continue to the next step. If $|U_0| \geq 2\alpha n$, then we sequentially discard arbitrary K_r-diamond trees $\mathcal{D}_j = (T_j, R_j, \Sigma_j)$ from the family. That is, we choose a $\mathcal{D}_j$ in the family, delete R_j from U_0 and delete the vertices that belong to $(r-1)$-cliques in Σ_j from W_0. We continue discarding diamond trees until $|U_0| \leq 2\alpha n$. Note that as $|R_j| \leq z \leq \alpha n$ for all j, the updated U_0 at the end of this discarding process will have size at least αn as required. We then move to the next step.

All the diamond trees in our family are d_*-scattered throughout the process and also W_0, as the set of vertices featuring in interior cliques of a family of K_r-diamond trees whose orders add up to less than $2\alpha n$, has size less than $2(r-1)\alpha n$ throughout. It is also clear that as the order of any diamond tree in our collection grows by at most d_* in each step, the order of the diamond tree which is found by the algorithm will be at most $z + d_*$. It only remains to check that the algorithm terminates but this is guaranteed because the number of diamond trees is decreasing throughout the process. Indeed, we never add new diamond trees to the family and every $\alpha n / d_*$ steps we have to discard at least one diamond tree from the family. If the algorithm does not terminate after finding an appropriate $\mathcal{D}_{\mathrm{sc}}$, then eventually we will be left with just one diamond tree $\mathcal{D}_1$ in the family, but at this point the order of $\mathcal{D}_1$ would be at least $\alpha n \geq z$, contradicting that the algorithm is still running. ■

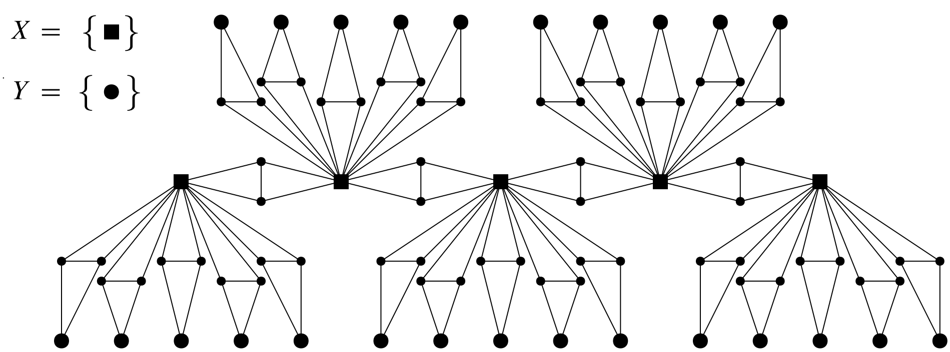

Fig. 6. A 6-scattered K_3-diamond tree.

Using Lemmas 4.4 and 4.5 we can now deduce Proposition 4.1.

Proof of Proposition 4.1. Fix $\varepsilon > 0$ small enough to apply Lemmas 4.4 and 4.5 and small enough to force n to be sufficiently large in what follows. Let us first deal with the case when $z \le d_* := \alpha^2 p^{r-1} n$. Here, we arbitrarily partition U into U_0 and U_1 of size at least αn, fix $U_2 = W$ and apply Lemma 4.4 to get a K_r-diamond star $\mathcal{D}^* = (T^*, R^*, \Sigma^*)$ of order $1 + d_*$ with $R^* \subset U$ and $\Sigma^* \subset K_{r-1}(G[W])$ a matching of $(r-1)$-cliques in W. Let $x \in R^*$ be the only non-leaf vertex in R^* and define $X = \{x\}$. Further, let $Y \subset R^* \setminus X$ be an arbitrary subset of $z - 1$ vertices. Now taking $\rho^* : V(T^*) \to R^*$ and $\sigma^* : E(T^*) \to \Sigma^*$ to be the defining bijective maps for $\mathcal{D}^*$, note that for any $Y' \subset Y$, the set $\{\rho^{*-1}(v) : v \in Y' \cup X\} \subset V(T^*)$ spans a subtree (or rather a substar) of T^*, say T. Therefore, taking $\mathcal{D} = (T, X \cup Y', \Sigma)$ where $\Sigma := \{\sigma^*(e) : e \in E(T)\}$ defines a K_r-diamond tree with removable vertices $Y' \cup X$. Therefore (1)–(3) of the proposition are all satisfied.

When $d_* < z \le \alpha n$, the proof is similar. We apply Lemma 4.5 to get a d_*-scattered K_r-diamond tree $\mathcal{D}_{\rm sc} = (T_{\rm sc}, R_{\rm sc}, \Sigma_{\rm sc})$ as given by the lemma and define $X \subset R_{\rm sc}$ to be the non-leaves of $\mathcal{D}_{\rm sc}$. See Figure 6 for an example. In order to bound $|X|$ and prove property (2), we appeal to Lemma 4.3 which gives

$$|X| \le \frac{|R_{\rm sc}| - 2}{d_* - 1} \le \frac{z + d_* - 2}{d_* - 1} \le \frac{2z}{d_*},$$

using $z \ge d_*$ in the final inequality.

We note that for n large (using Fact 3.2) we have $d_* \ge 4$, implying that $|X| \le z/2$. We fix $Y \subset R_{\rm sc} \setminus X$ to be an arbitrary subset of size $z - |X|$ and claim that conditions (1)–(3) of the proposition are all satisfied. Indeed, it remains only to prove (3) and this follows similarly to above, by taking sub-diamond-trees of $\mathcal{D}_{\rm sc}$. In detail, fix some $Y' \subset Y$ and let $R = Y' \cup X$. Then if $\rho_{\rm sc} : V(T_{\rm sc}) \to R_{\rm sc}$ and $\sigma_{\rm sc} : E(T_{\rm sc}) \to \Sigma_{\rm sc}$ are the defining bijective maps for $\mathcal{D}_{\rm sc}$, then the set $\{\rho_{\rm sc}^{-1}(v) : v \in R\}$ of vertices spans a subtree $T \subset T_{\rm sc}$. Indeed, we simply deleted leaves from $T_{\rm sc}$, namely $\rho_{\rm sc}^{-1}(x)$ for $x \in R_{\rm sc} \setminus Y'$. Taking $\Sigma = \{\sigma_{\rm sc}(e) : e \in E(T)\}$, we conclude that $\mathcal{D} = (T, R, \Sigma)$ is the desired diamond tree. ∎

5. Cascading absorption through orchards

In this section we discuss orchards in our (p, β)-bijumbled graphs. We begin in Section 5.1 by proving Lemma 2.4 which details conditions for when one orchard absorbs another. In Section 5.2, we then discuss the existence of shrinkable orchards, addressing Proposition 2.8 which tells us that we can find shrinkable orchards of all desired orders in the graphs we are interested in. The proof of Proposition 2.8 requires many ideas and two distinct approaches. Therefore, we defer the majority of the work to later sections and simply reduce the proposition here, splitting it into two 'subpropositions' which will be tackled separately. Recall that Lemma 2.4 and Proposition 2.8 were the two ingredients we needed to prove the cascading absorption through constantly many orchards in the proof of Theorem 1.4.

5.1. Absorbing orchards

Recall the definition (Definition 2.3) of an orchard and that we say a $(K, M)_r$-orchard $\mathcal{O}$ *absorbs* a $(k, m)_r$-orchard $\mathcal{R}$ if there is an $((r-1)k, M)_r$-suborchard $\mathcal{O}' \subset \mathcal{O}$ such that there is a K_r-factor in $G[V(\mathcal{R}) \cup V(\mathcal{O}')]$.

In this section we prove Lemma 2.4, restated below for convenience, which is a generalisation of [61, Lemma 3.5]. The lemma gives some sufficient conditions for an orchard to be able to absorb another orchard.

Lemma 2.4 (restated). *For any $r \in \mathbb{N}_{\geq 3}$ and $0 < \zeta, \eta < 1$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$. Let $\mathcal{O}$ be a $(K, M)_r$-orchard in G such that $KM \geq \zeta n$. Then there exists a set $B \subset V(G)$ such that $|B| \leq \eta p^{2r-4} n$ and $\mathcal{O}$ absorbs any $(k, m)_r$-orchard $\mathcal{R}$ in G with*

$$V(\mathcal{R}) \cap (B \cup V(\mathcal{O})) = \emptyset, \quad k \leq K/(8r) \quad \text{and} \quad kM \leq mK. \tag{5.1}$$

Our proof scheme follows that of [33] which gives a polynomial time two-phase algorithm for finding the necessary K_r-factor. The algorithm is a simple greedy algorithm and works by absorbing each diamond tree $\mathcal{B}$ in the small orchard $\mathcal{R}$, one at a time. In more detail, for each diamond tree $\mathcal{B}$ in $\mathcal{R}$, we find $r-1$ diamond trees $\mathcal{D}_1, \ldots, \mathcal{D}_{r-1} \in \mathcal{O}$ such that there is a copy of K_r traversing the sets of removable vertices of $\mathcal{B}$ and the diamond trees $\mathcal{D}_1, \ldots, \mathcal{D}_{r-1}$. This implies that there is a K_r-factor in $G[V(\mathcal{B}) \cup V(\mathcal{D}_1) \cup \cdots \cup V(\mathcal{D}_{r-1})]$ (see Observation 2.2) and so we can add the $\mathcal{D}_i$ to the suborchard $\mathcal{O}'$, forbid them from being used again, and move to the next diamond tree $\mathcal{B}' \in \mathcal{R}$. Note that typically, we expect to succeed with this process. Indeed, the set of removable vertices of diamond trees in $\mathcal{O}$ is linear in size (and remains linear even after forbidding diamond trees $\mathcal{D} \in \mathcal{O}$ used for previous $\mathcal{B} \in \mathcal{R}$) and so a typical vertex has $\Omega(pn)$ neighbours among this set of removable vertices. Hence, appealing to Corollary 3.5 (1) (i) which states that sets of size $\Omega(pn)$ host copies of K_{r-1}, we can expect to find a copy of K_{r-1} in the neighbourhood of a typical removable vertex of $\mathcal{B} \in \mathcal{R}$ which lies on the removable vertices of diamond trees in $\mathcal{O}$. As long as this copy of K_{r-1} traverses sets of removable vertices of distinct diamond trees in $\mathcal{O}$, we will succeed. With a few extra ideas and a bit of preprocessing (for example partitioning $\mathcal{O}$ into $r-1$ suborchards at the start), this intuition holds true and we can successfully greedily start to build $\mathcal{O}'$.

In fact, if kM is small compared to pn, we can fully form $\mathcal{O}'$ in this way and no second phase is necessary. However, if kM is large compared to pn we may run into trouble as with this greedy approach, it may be the case that the neighbourhood of a removable vertex v of a diamond tree $\mathcal{B} \in \mathcal{R}$ has too small a size by the time we come to considering $\mathcal{B}$. Indeed, as we run this greedy process, we forbid the diamond trees (and their removable vertices) which we add to $\mathcal{O}'$ from being used again. This could result in v having much fewer than pn neighbours in the removable vertices of diamond trees in (the remainder of) $\mathcal{O}$ and so we have no guarantee of finding a copy of K_{r-1} in this neighbourhood.

We resolve this issue by running a two-phase algorithm and reserving half of $\mathcal{O}$ for the second phase. The key point is that if a diamond tree $\mathcal{B}$ fails in the first round then it must be the case that *all* of the removable vertices of $\mathcal{B}$ have small neighbourhoods amongst the removable vertices of diamond trees in $\mathcal{O}$. Given that throughout the process, many diamond trees in $\mathcal{O}$ will remain available to use, pseudorandomness (more precisely, Corollary 3.5 (1) (ii)) tells us that the number of vertices that do not have large enough neighbourhoods is relatively small. Hence, as each diamond tree $\mathcal{B} \in \mathcal{R}$ which failed in the first phase has a set of removable vertices which are atypical in this way, we can upper bound the number of diamond trees in $\mathcal{R}$ that fail in the first round. This upper bound will then be used to show that in the second phase, we are successful with each diamond tree, as throughout the second round, the number of removable vertices being forbidden (due to being used to absorb other diamond trees in $\mathcal{R}$) will be negligible and so the neighbourhoods of vertices amongst the removable vertices of diamond trees in the half of $\mathcal{O}$ reserved for this second phase will remain large.

Proof of Lemma 2.4. We fix $\alpha, \eta' < \frac{\eta\zeta^2}{2^{3r}r^2}$ and choose $\varepsilon > 0$ small enough to apply Lemma 3.3 with $\eta_{3.3} = \eta'$ and Corollary 3.5 with $\alpha_{3.5} = \alpha$. Let $\mathcal{O} = \{\mathcal{D}_1, \dots, \mathcal{D}_K\}$ be the $(K, M)_r$-orchard with each $\mathcal{D}_i = (T_i, R_i, \Sigma_i)$ being a K_r-diamond tree of order between M and $2M$. We start by arbitrarily partitioning $\mathcal{O}$ into $2(r-1)$ suborchards of size as equal as possible so that $\mathcal{O} = \bigcup_{j=1}^{2(r-1)} \mathcal{O}_j$ and each $\mathcal{O}_j$ is a $(K_j, M)_r$-orchard with $K_j = \frac{K}{2(r-1)} \pm 1 \geq \frac{K}{2r}$. For $j \in [2(r-1)]$, we let

$$Y_j := \bigcup_{i:\, \mathcal{D}_i \in \mathcal{O}_j} R_i$$

be the set of removable vertices of the diamond trees which feature in the jth suborchard. Note that $|Y_j| \geq K_j M \geq KM/(2r) \geq \zeta n/(2r)$ for each $j \in [2(r-1)]$. We define B to be the set of vertices $v \in V \setminus V(\mathcal{O})$ such that for some $j \in [2(r-1)]$, $\deg_{Y_j}(v) < p|Y_j|/2$. By Lemma 3.3 (i), we have

$$|B| < \frac{2(r-1)\eta' p^{2r-4} n^2}{\min_j |Y_j|} \leq \eta p^{2r-4} n,$$

due to our lower bound on the size of the $|Y_j|$ and our upper bound on η'.

Now as in the statement of the lemma, consider a $(k, m)_r$-orchard $\mathcal{R} = \{\mathcal{B}_1, \dots, \mathcal{B}_k\}$ of diamond trees whose vertices lie in $V \setminus (B \cup V(\mathcal{O}))$. For $i' \in [k]$, let $Q_{i'}$ be the set of removable vertices of the diamond tree $\mathcal{B}_{i'}$. We will show that for each $i' \in [k]$, there exist distinct indices $i_1 = i_1(i'), \dots, i_{r-1} = i_{r-1}(i') \in [K]$ such that there is a copy of K_r which traverses the sets $Q_{i'}$ and $R_{i_1}, \dots, R_{i_{r-1}}$, where R_{i_1} is the set of removable vertices of $\mathcal{D}_{i_1}$ and likewise for $i_2, \dots, i_{r-1}$. Now, from Observation 2.2, for such an r-tuple $\mathcal{B}_{i'}$, $\mathcal{D}_{i_1}, \dots, \mathcal{D}_{i_{r-1}}$, there is a K_r-factor in $G[V(\mathcal{B}_{i'}) \cup V(\mathcal{D}_{i_1}) \cup \cdots \cup V(\mathcal{D}_{i_{r-1}})]$. We will prove that one can choose such indices $i_1, \dots, i_{r-1}$ for each $i' \in [k]$ in such a way that no $i \in [K]$ is chosen more than once. That is, for $i' \neq j' \in [k]$, the sets $\{i_1(i'), \dots, i_{r-1}(i')\}$ and $\{i_1(j'), \dots, i_{r-1}(j')\}$ are disjoint. Therefore our suborchard $\mathcal{O}' \subset \mathcal{O}$ can simply be defined to be the union of all the choices of $\mathcal{D}_{i_j(i')}$ for $i' \in [k]$ and $j \in [r-1]$.

We now show how to find the indices $i_1(i'), \dots, i_{r-1}(i')$ for each $i' \in [k]$. We will achieve this via the following simple algorithm. We initiate the first round of the algorithm with $\mathcal{O}' = \emptyset$, $I = [k]$, $\mathcal{P}_j = \mathcal{O}_j$ and $Z_j = Y_j$ for $1 \le j \le r-1$. Note that the $\mathcal{O}_j$ for $r \le j \le 2(r-1)$ do not feature in these definitions. This is because we will not use any diamond trees that lie in $\bigcup_{j=r}^{2(r-1)} \mathcal{O}_j$ in this first round. Now the algorithm runs as follows. For $i' = 1, \dots, k$, we check if there exists some set $\{\mathcal{D}_{i_j} \in \mathcal{P}_j : j \in [r-1]\}$ such that there is a K_r traversing $Q_{i'}$ and the sets of removable vertices $R_{i_1}, \dots, R_{i_{r-1}}$. If this is the case then we delete $\mathcal{D}_{i_j}$ from $\mathcal{P}_j$ and add it to $\mathcal{O}'$ for $j \in [r-1]$ and we also delete R_{i_j} from Z_j for all $j \in [r-1]$. Furthermore, we delete i' from I and move to the next index $i'+1$ (or finish this round if $i' = k$). If it is not the case that such diamond trees exist in the orchards $\mathcal{P}_j$, then we simply leave i' as a member of I and move on to the next index.

At the end of the first round, we have some set I of indices remaining. We define $t := |I|$ at this point. We will now use diamond trees in the orchards $\mathcal{O}_j$ with $r \le j \le 2(r-1)$ to absorb these remaining diamond trees $\mathcal{B}_{i'}$ with $i' \in I$. Thus we reset the process, setting $\mathcal{P}_j = \mathcal{O}_{j+r-1}$ and $Z_j = Y_{j+r-1}$ for all $j \in [r-1]$. We then follow the same simple process in the second round as we did in the first, running through the (remaining) $i' \in I$ in order and trying to find an appropriate set $\{\mathcal{D}_{i_j} \in \mathcal{P}_j : j \in [r-1]\}$ of diamond trees at each step. We claim that in this second round, we can find such a set for every $i' \in I$ and so by the end of the second round, the set I is empty and $\mathcal{O}'$ is such that $G[V(\mathcal{R}) \cup V(\mathcal{O}')]$ hosts a K_r-factor.

In order to prove this, our analysis splits into two cases. First consider when $kM < \frac{\zeta pn}{16r}$. In this case, the second round is not even necessary as all indices succeeded in the first round. Indeed, note that every time we are successful for an index i', we delete at most $2M$ vertices from each of the Z_j. Therefore, at any instance in the first round of the process, any vertex v which is not in B has

$$\deg_{Z_j}(v) \ge \frac{p|Y_j|}{2} - 2kM \ge \frac{\zeta pn}{4r} - 2kM \ge \frac{\zeta pn}{8r}$$

for all $j \in [r-1]$, using our lower bound on the $|Y_j|$ and our upper bound on kM. But then, by Corollary 3.5 (1) (i) (applied in this instance with $\tilde{G}$ being the empty graph and $G' = G$), there exists a copy of K_{r-1} traversing the sets $N_{Z_j}(v)$ for $1 \le j \le r-1$. When v is any vertex in the removable set of vertices $Q_{i'}$ for some diamond tree $\mathcal{B}_{i'}$ in the process, this gives a copy of K_r traversing $Q_{i'}$ and some sets of removable vertices R_{i_j} for diamond trees $\mathcal{D}_{i_j} \in \mathcal{P}_j$, $j \in [r-1]$, as desired. In this way, we see that the process succeeds in every step of the first round to find a suitable $\{i_j(i') : j \in [r]\}$ for each $i' \in [k]$ and I is empty (i.e. $t = 0$) at the end of the round. Note that we have used here the fact that the vertices of $Q_{i'}$ are not in B.

When $\frac{\zeta pn}{16r} \le kM \le mK$, the second round may be neccesary and we start with estimating t, the size of I after the first round. Now note that at the end of the first round, *before* we reassign the sets Z_j to removable vertices in diamond trees in $\mathcal{O}_{j+r-1}$ for $j \in [r-1]$, if we take $Q = \bigcup_{i' \in I} Q_{i'}$, then there is no K_r traversing Q and the sets $Z_1, \dots, Z_{r-1}$. Indeed, otherwise there would be an $i' \in I$ and a vertex $v \in Q_{i'} \subseteq Q$

which is contained in a K_r with a set of vertices $\{v_{i_j} \in Z_j : j \in [r-1]\}$. This contradicts that for the index i' we failed to find a suitable set of i_j in the first round. Thus, at the end of the first round, there is no K_r traversing Q, and the Z_j, $j \in [r-1]$. Moreover,

$$|Z_j| \geq \frac{KM}{2r} - 2kM \geq \frac{KM}{4r} \geq \frac{\zeta n}{4r},$$

using the upper bound on k from (5.1) and the fact that at most $2M$ vertices are deleted from Z_j every time we are successful with an index $i' \in I$. Thus, we can conclude from Corollary 3.5 (1) (ii) that at the end of the first round, $tm < |Q| < \alpha p^{2r-4} n$. Therefore

$$t < \frac{\alpha p^{2r-4} n}{m} \leq \frac{\alpha p^2 n}{m} \leq \frac{16\alpha rpK}{\zeta} \leq \frac{16\alpha rpn}{\zeta M} \leq \frac{\zeta pn}{16rM},$$

where we have used our lower and upper bounds on kM to give an upper bound on pn/m in the third inequality, the fact that $KM \leq n$ in the fourth inequality and our upper bound on α in the final inequality.

We now turn to analyse the second round. Using our upper bound on t, we can upper bound the number of vertices deleted in each Z_j throughout the second round, and using this we find that for any vertex v not in B, any $j \in [r-1]$ and at any point in the second round,

$$\deg_{Z_j}(v) \geq \frac{p|Y_{j+r-1}|}{2} - 2tM \geq \frac{\zeta pn}{4r} - \frac{\zeta pn}{8r} \geq \frac{\zeta pn}{8r}.$$

Thus we can repeat the argument used for the case when kM was small, seeing that at every step in the second round we are successful in finding an appropriate set of i_j for $j \in [r-1]$ for each $i' \in I$. This completes the proof. ■

5.2. *Shrinkable orchards*

Here we are concerned with the existence of shrinkable orchards in pseudorandom graphs and verifying Proposition 2.8, which we restate below for the convenience of the reader. We also encourage the reader to remind themselves of Definitions 2.5 and 2.7 as well as Observation 2.6.

Proposition 2.8 (restated). *For any $r \in \mathbb{N}_{\geq 3}$ and $0 < \alpha, \gamma < 1/2^{12r}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$ and any vertex subset $U \subseteq V(G)$ with $|U| \geq n/2$. For any $m \in \mathbb{N}$ with $1 \leq m \leq n^{7/8}$ there exists a γ-shrinkable $(k, m)_r$-orchard $\mathcal{O}$ in $G[U]$ with $k \in \mathbb{N}$ such that $\alpha n \leq km \leq 2\alpha n$.*

In order to prove Proposition 2.8, we will appeal to the methods of Sections 3.4 and 3.5. We will use Theorem 3.12 to reduce the problem to establishing the existence of perfect fractional matchings in the appropriate K_r-hypergraphs and we will then employ Lemmas 3.8 and 3.9 to find these perfect fractional matchings. In order that our hypergraph has the desired properties to apply these lemmas, we need to choose the diamond trees which define our orchard carefully.

It turns out that different arguments are needed for finding shrinkable orchards of different orders. In Section 6 we show how to find shrinkable orchards of small order, establishing the following intermediate proposition.

Proposition 5.1. *For any $r \in \mathbb{N}_{\geq 3}$ and $0 < \alpha, \gamma < 1/2^{3r}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$ and any vertex subset $U \subseteq V(G)$ with $|U| \geq n/2$. For any $m \in \mathbb{N}$ with*

$$1 \leq m \leq \min\{p^{r-2} n^{1-2r^3\gamma}, n^{7/8}\},$$

there exists a γ-shrinkable $(k, m)_r$-orchard $\mathcal{O}$ in $G[U]$ with $k \in \mathbb{N}$ such that $\alpha n \leq km \leq 2\alpha n$.

In Section 7 we then address shrinkable orchards with large order, which results in the following.

Proposition 5.2. *For any $r \in \mathbb{N}_{\geq 3}$ and $0 < \alpha, \gamma < 1/2^{12r}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$ and any vertex subset $U \subseteq V(G)$ with $|U| \geq n/2$. For any $m \in \mathbb{N}$ with*

$$p^{r-1} n \leq m \leq n^{7/8},$$

there exists a γ-shrinkable $(k, m)_r$-orchard $\mathcal{O}$ in $G[U]$ with $k \in \mathbb{N}$ such that $\alpha n \leq km \leq 2\alpha n$.

The proof of Proposition 2.8 is basically immediate from Propositions 5.1 and 5.2 but we spell it out nonetheless.

Proof of Proposition 2.8. We split into a case analysis based on the density p of our graph G. First consider $p \geq n^{-1/(10r)}$. Then we claim that $p^{r-2} n^{1-2r^3\gamma} \geq n^{7/8}$ and so the desired γ-shrinkable orchard of all orders up to $n^{7/8}$ can be derived from Proposition 5.1. Indeed, we have $p^{r-2} n^{1-2r^3\gamma} \geq n^{1-\frac{r-2}{10r}-2r^3\gamma}$ and

$$1 - \frac{r-2}{10r} - 2r^3\gamma > 1 - \frac{1}{10} - \frac{1}{40} = 7/8,$$

due to our upper bound on γ (and lower bound on r).

When $p < n^{-1/(10r)}$, we have $p \leq n^{-2r^3\gamma}$ again due to our upper bound on γ. Hence we can apply Proposition 5.1 to find γ-shrinkable orchards of orders $m \leq n^{7/8}$ such that $m < p^{r-1} n \leq p^{r-2} n^{1-2r^3\gamma}$ and apply Proposition 5.2 to find γ-shrinkable orchards with orders m such that $p^{r-1} n \leq m \leq n^{7/8}$. This settles all cases, giving the proposition. ∎

In both cases, a simpler argument works for the extreme cases, that is, when the order is small in Proposition 5.1 or when the order is large in Proposition 5.2. Extra ideas are then needed to push the approaches, extending the ranges of the two propositions so that they meet and cover all desired orders. In more detail, an easier form of Proposition 5.1 can cover orders which get close to $p^{r-1} n$ (see Proposition 6.5). Again the separation required depends on γ, explicitly $m \leq p^{r-1} n^{1-r^3\gamma}$. This is already enough to cover all desired orchard orders when p is large. On the other hand, a basic form of the argument for large order orchards gives shrinkable orchards of order at least $p^{r-1}n$ when p is large and of order at least p^{1-r} when p is smaller (see Proposition 7.3). Interestingly, Fact 3.2 implies exactly that $p^{1-r} = \Omega(p^{r-2}n)$ always and so proves that when p is small (close to the lower bound of $\Omega(n^{-1/(2r-3)})$) and our bijumbled graph is sparse, both the simpler arguments for small orders and large orders *as well as* their extensions are needed. Indeed, using the simpler version, Proposition 6.5, for small orders and the full power of Proposition 5.2 leaves a small gap in the orders, and so does using Proposition 5.1 in conjunction with the easier Proposition 7.3. In order to help the reader through the next two sections, in both cases we begin by presenting the easier weaker versions of the statements we need. This then lays the foundation for the full proofs and allows us to discuss the more technical aspects needed to push the ranges for which we can prove the existence of shrinkable orchards.

6. Shrinkable orchards of small order

Our first argument for proving the existence of shrinkable orchards works provided the order of the orchard is not too large, establishing Proposition 5.1. Before embarking on this we have to go through several steps. Firstly, in Section 6.1, we generalise the theory of shrinkable orchards built up in Section 2, allowing slightly more flexibility for our consequent proofs. In Section 6.2, we then use the theory of perfect fractional matchings to give conditions that guarantee an orchard is shrinkable. In Section 6.3, we show how this immediately implies the existence of shrinkable orchards of small order. However, this falls short of Proposition 5.1 and in the rest of this section we push the ideas to extend the range of orders we can cover, showing how to cleverly choose diamond trees of our orchard in Section 6.4, which allows us to prove the full Proposition 5.1 in Section 6.5.

6.1. *From orchards to systems*

We begin by generalising our definitions slightly, allowing us to work not just with orchards but also with set systems.

Definition 6.1. Given a graph G we say a set $\Lambda \subset 2^{V(G)}$ of pairwise disjoint subsets is a (k,m)-*system* if $m \leq |Q| \leq 2m$ for each $Q \in \Lambda$ and $|\Lambda| = k$. That is, a (k,m)-system is just a family of k disjoint vertex sets of size between m and $2m$.

Now given a (k,m)-system Λ in a graph G, the K_r-*hypergraph* generated by Λ, denoted $H = H(\Lambda; r)$, is the r-uniform hypergraph with vertex set $V(H) = \Lambda$ and with $\{Q_{i_1}, \dots, Q_{i_r}\} \in \binom{\Lambda}{r}$ forming a hyperedge in H if and only if there is a copy of K_r traversing the sets $Q_{i_1}, \dots, Q_{i_r}$ in G.

Finally, for $0 < \gamma < 1$, we say a (k,m)-system Λ in a graph G is γ-*shrinkable* (with respect to r) if there exists a subsystem $\Gamma \subset \Lambda$ of size at least γk such for *any* subsystem $\Gamma' \subseteq \Gamma$, there is a matching in $H := H(\Lambda \setminus \Gamma'; r)$ covering all but $k^{1-\gamma}$ of the vertices of H.

Note that given a $(k,m)_r$-orchard $\mathcal{O}$ we can define a (k,m)-system Λ as the sets of removable vertices of diamond trees in $\mathcal{O}$. That is, $\Lambda := \{R_{\mathcal{D}} : \mathcal{D} \in \mathcal{O}\}$. Then the K_r-hypergraphs generated by $\mathcal{O}$ and Λ coincide, i.e. $H(\Lambda; r) = H(\mathcal{O})$, and $\mathcal{O}$ is γ-shrinkable if and only if Λ is γ-shrinkable. However, Definition 6.1 allows us slightly more flexibility, giving us the ability to focus on subsets of removable vertices. The next observation highlights this and although the result is trivial, it will be important for our proofs.

Observation 6.2. Suppose $r \geq 3$, $0 < \gamma < 1$ and $\mathcal{O} = \{\mathcal{D}_1, \ldots, \mathcal{D}_k\}$ is a $(k,m)_r$-orchard in a graph G with R_i being the set of removable vertices of $\mathcal{D}_i$ for $i \in [k]$. Then if $\Lambda = \{Q_1, \ldots, Q_k\}$ is some (k,m')-system (for some m') such that $Q_i \subseteq R_i$ for $i \in [k]$ and Λ is γ-shrinkable (with respect to r), then $\mathcal{O}$ is also γ-shrinkable.

It will become clear why such a relaxation is useful for us and thus why we make this switch to working with set systems.

6.2. *Sufficient conditions for shrinkability*

We now explore the conditions on set systems which guarantee shrinkability. We begin by giving some local conditions on a set system which guarantee that it is shrinkable given that it lies in the pseudorandom graphs we are interested in.

Lemma 6.3. *For any $r \in \mathbb{N}_{\geq 3}$ and $0 < \alpha, \gamma < 1/(2^r r^2)$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p,β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$. Suppose $\Lambda \subset 2^{V(G)}$ is a (k,m)-system such that $m \leq n^{7/8}$, $km \geq \alpha n$, $pk \geq n^{\gamma}$ and*

(1) *there exists a subsystem $\Gamma \subset \Lambda$ such that $|\Gamma| \geq \gamma k$ and $Y := \bigcup\{P : P \in \Lambda \setminus \Gamma\}$, for every $Q \in \Lambda$ there exists a vertex $v \in Q$ such that*

$$\deg^G_Y(v) \geq \alpha p k m;$$

(2) *for any $u \in \bigcup_{P \in \Lambda} P$ and $Q \in \Lambda$, we have $\deg^G_Q(u) \leq p^{r-1} n^{1-r^3\gamma}$.*

Then Λ is γ-shrinkable with respect to r.

Let us make a few remarks before proving the lemma. Firstly, note that condition (1), despite the slight technicality necessary to avoid dependence on sets in Γ, is a natural condition. Indeed, we are requiring that at least one vertex in each set is well connected to the other sets and has a constant fraction of the degree that we would expect on average. Condition (2) is perhaps more mysterious as it is unclear why having an *upper* bound on the degree of a vertex relative to another set in the system is advantageous. The point is that this guarantees that each of the vertices has a neighbourhood that is well-spread across the other sets of the set system, without being too concentrated on any other single set. Within the proof this necessity manifests itself as we appeal to Theorem 3.12 and so will need that when we disallow edges between certain pairs of sets from being used (dictated by the graph J), we do not significantly alter the graph in which we work. The details follow in the proof.

Proof of Lemma 6.3. Fix $\varepsilon > 0$ small enough to apply Corollary 3.5 with $\alpha_{3.5} = \alpha' < \alpha^3/(2r)$ and small enough to force n to be sufficiently large in what follows. Fixing $\Gamma \subset \Lambda$ as in condition (1), we have to show that for any $\Gamma' \subset \Gamma$, the K_r-hypergraph $H = H(\Lambda \setminus \Gamma'; r)$ has a matching covering all but $k^{1-\gamma}$ vertices of H. So fix such a Γ', let $\Lambda^* := \Lambda \setminus \Gamma'$ and let $H := H(\Lambda^*; r)$.

In order to show the existence of a large matching in H, we appeal to Theorem 3.12. So let us fix $N = |V(H)|$ and note that as $N \geq (1-\gamma)k$ and $k \geq \alpha n^{1/8}$ due to our conditions on k and m, we can assume that N is sufficiently large in what follows. Now fix some 2-uniform graph J on $V(H)$ of maximum degree at most $N^{r^2\gamma}$. If we can show that $H \setminus H_J$ contains a perfect fractional matching, then we are done by Theorem 3.12 because, J being arbitrary, the theorem guarantees a matching covering all but at most $N^{1-\gamma} \leq k^{1-\gamma}$ vertices of H.

In order to study $H \setminus H_J$, we look at the forbidden edges of G which J imposes. That is, we define

$$\tilde{G}_J := \bigcup_{\{Q_1,Q_2\}\in E(J)} G[Q_1, Q_2] \cup \bigcup_{Q\in\Lambda^*} G[Q]$$

where $G[Q_1, Q_2]$ denotes the set of all edges in G between the sets Q_1 and Q_2 and $G[Q]$ denotes all the edges induced by G in the set Q. Then for any $v \in V(G)$, we have $\deg^{\tilde{G}_J}(v) = 0$ if $v \notin \bigcup_{P\in\Lambda^*} P$, while if $v \in Q \in \Lambda^*$ then

$$\deg^{\tilde{G}_J}(v) \leq \sum_{P\in N^J(Q)\cup\{Q\}} \deg_P^G(u) \leq (N^{r^2\gamma}+1)p^{r-1}n^{1-r^3\gamma} \leq p^{r-1}n^{1-\gamma}, \qquad (6.1)$$

using (2), the upper bound on the degrees in J and the fact that $N \leq n$.

Now defining $G'_J := G \setminus \tilde{G}_J$, we see that $H \setminus H_J$ is precisely the hypergraph obtained by viewing Λ^* as an (N, m)-system in G'_J and taking the K_r-hypergraph $H^* = H(\Lambda^*; r)$ in G'_J. Indeed, as there are no edges of G'_J between two sets, Q_1 and Q_2 say, which form an edge in J, there can be no edge of H in H^* which contains both Q_1 and Q_2. We therefore switch from now on to considering H^* as the K_r-hypergraph generated by Λ^* in G'_J.

In order to prove the existence of a perfect fractional matching in H^*, we will appeal to Lemma 3.9, fixing $M = \alpha^2 pk$. Note that due to our lower bound on pk, we certainly have $M_1 \geq r$. We thus need to check that conditions (i) and (ii) of that lemma hold. For (i), fix some $Q \in \Lambda^*$. From (1) we know that there exists some vertex v in Q such that $\deg_W^G(v) \geq \alpha pkm$ where $W := \bigcup_{P\in\Lambda^*} P$, and so taking $U := N_W^{G'_J}(v)$ we have

$$|U| \geq \alpha pkm - \deg^{\tilde{G}_J}(v) \geq \alpha pkm/2,$$

using (6.1). Moreover, due to (2), we can spilt U into disjoint sets $U_1, \ldots, U_{r-1}$ such that $|U_i| \geq \alpha pkm/(2r)$ for each i, and for any $P \in \Lambda^*$, there exists an $i \in [r-1]$ such that $|P \cap U| \subset U_i$. That is, we simply partition U into $r-1$ roughly equal size parts such that vertices which lie in the same P end up in the same part. Condition (2) of the lemma guarantees that $U \cap P$ is small enough for each $P \in \Lambda^*$ and so we can do this partition in such a way that all the U_i are roughly equal in size. We will now repeatedly find $(r-1)$-cliques in G'_J traversing the U_i and build a fan F_Q of size M in H^* focused at Q. We start with F_Q being empty and each time we find a copy $S=\{u_1,\dots,u_{r-1}\}\in K_{r-1}(G'_J)$ in G' with $u_i\in U_i$ for $i\in[r-1]$, there $P_1,\dots,P_{r-1}\in\Lambda^*$ such that $u_i\in P_i$. We add the hyperedge between $P_1,\dots,P_{r-1}$ and Q to the fan F_Q and delete any vertices in P_i from U_i for $i\in[r-1]$. We repeat this process and note that we are successful in every step until F_Q has size M. Indeed, this follows from Corollary 3.5 (1) (i) because while $|F_Q|<M$, we have deleted at most[10] $M2m\le\alpha pkm/(4r)$ vertices from each U_i and so $U_i\ge\alpha pkm/(4r)\ge\alpha' pn$, using our upper bound on α' and our lower bound on km.

We now turn to verifying (ii) of Lemma 3.9. We will show that given any r-tuple of disjoint subsystems $\Gamma_1,\dots,\Gamma_r\subset\Lambda^*$ such that $|\Gamma_1|=M$ and $|\Gamma_i|\ge\alpha k/r$ for $2\le i\le r$, there exists a hyperedge of H^* with one endpoint in each of the Γ_i. Indeed, this follows from Corollary 3.5 (1) (ii) because taking $U_i:=\bigcup_{P\in\Gamma_i}P$ for $i\in[r]$, there exists an r-clique $S\in K_r(G'_J)$ traversing the U_i which in turn gives the hyperedge. Condition (ii) then clearly follows since any subsystem of Γ^* of size $N/(2r)\ge\alpha k$ can be split into $r-1$ subsystems of size at least $\alpha k/r$. Note that in both applications of Corollary 3.5 (1) above we have used (6.1) to show that we could find cliques that avoid using edges of $\tilde{G}_J$. The lemma now follows from Lemma 3.9. ■

As previously noted, condition (1) of Lemma 6.3 is somewhat weak and just requires that each set in the set system contains a vertex that acts typically. We now show how we can 'clean up' a set system, losing sets which do not have a typical vertex in order to recover condition (1). This allows us to focus on finding systems which satisfy condition (2) of Lemma 6.3.

Lemma 6.4. *For any $r\in\mathbb{N}_{\ge 3}$ and $0<\alpha,\gamma<1/(2^r r^2)$ there exists an $\varepsilon>0$ such that the following holds for any n-vertex (p,β)-bijumbled graph G with $\beta\le\varepsilon p^{r-1}n$. Suppose $\Lambda'\subset 2^{V(G)}$ is a $((1+\gamma)k,m)$-system such that $m\le n^{7/8}$, $\alpha n\le km\le 2\alpha n$, $pk\ge n^{\gamma}$ and for any $u\in\bigcup_{P\in\Lambda'}P$ and $Q\in\Lambda'$, we have $\deg^G_Q(u)\le p^{r-1}n^{1-r^3\gamma}$. Then there exists a (k,m)-system $\Lambda\subset\Lambda'$ which is γ-shrinkable with respect to r.*

Proof. We fix $\varepsilon>0$ small enough to apply Lemma 6.3 and to apply Lemma 3.3 with $\eta<\gamma\alpha^2/2$. The method of the proof is simple; we aim to apply Lemma 6.3 and so obtain Λ from Λ' by losing the sets which violate condition (1) of that lemma. By Lemma 3.3 (i), there are few vertices which have small degree ($\le p|Y|/2$) relative to any set Y which is large enough and so we can expect that we do not lose many sets when transitioning from Λ' to Λ. One complication is that the definition of Y in condition (1) of Lemma 6.3 *depends* on the sets in the system and so we cannot guarantee that a set satisfying (1) continues to satisfy the condition once other sets have been removed. In order to handle this, we delete sets in the system one by one, creating a process which will terminate with a system which has the desired minimum degree condition. The details now follow.

[10] Here we use the fact that every set in Λ has size at most $2m$.

We begin by fixing $\Gamma_0 \subset \Lambda'$ to be some arbitrary subsystem of size $(1-\gamma)k$ and we initiate the process by setting $\Theta = \Lambda'$ and setting a 'bin' system Φ which we initiate as being empty, that is, we set $\Phi = \emptyset$. Throughout the process we also define

$$W := \bigcup_{Q \in \Gamma_0 \cap \Theta} Q$$

to be the subset of vertices that lie in (sets that belong to) the current system Θ as well as the system Γ_0. Now the process runs as follows. If there is a set P in Θ such that $\deg_W(v) < \alpha pkm$ for all $v \in P$, then we delete P from Θ and add it to Φ. Hence if $P \in \Gamma_0$ then we also delete P from W. We claim that this process terminates with $|\Phi| \leq \gamma k$. Indeed, if this were not the case then consider the process at the point where $|\Phi| = \gamma k$. At this point we have

$$|W| = \Big| \bigcup_{Q \in \Gamma_0 \setminus \Phi} Q \Big| = (|\Gamma_0| - |\Phi|)m = (k - 2\gamma k)m \geq km/2 \geq \alpha n/2.$$

Now Lemma 3.3 (i) implies that at most

$$\frac{\eta p^{2r-4} n^2}{|W|} \leq \frac{2\eta p^{2r-4} n}{\alpha} < \gamma \alpha n \leq \gamma k m = |\Phi| m$$

vertices can have degree less than $p|W|/2$ relative to W. This leads to a contradiction. Indeed, it follows from how Φ is defined that at this point in the process, $|\bigcup_{P \in \Phi} P| \geq |\Phi| m$ and for all $v \in \bigcup_{P \in \Phi} P$, we have $\deg_W(v) < \alpha pkm < p|W|/2$. Indeed, if a vertex $v \in P \in \Phi$ had a larger degree relative to W then P would not have been added to Φ in the process.

Hence when the process terminates we have $|\Phi| \leq \gamma k$ and $|\Theta| = |\Lambda' \setminus \Phi| \geq k$. We fix $\Lambda \subseteq \Theta$ of size k so that $\Gamma_0 \cap \Theta \subseteq \Lambda$. We also fix $\Gamma \subseteq \Lambda \setminus (\Gamma_0 \cap \Theta)$ of size γk (which is possible as $|\Gamma_0 \cap \Theta| \leq (1-\gamma)k$). We claim that Λ is γ-shrinkable with respect to r. Indeed, this follows directly from Lemma 6.3 noting that condition (1) is satisfied in Λ with respect to Γ due to how we constructed Λ. ∎

6.3. *The existence of shrinkable orchards of small order*

We are now ready to prove the existence of shrinkable orchards by appealing to Lemma 6.4. Indeed, we simply need to find orchards which satisfy the maximum degree condition given there. This condition is immediate when the order of the orchards which we aim for is sufficiently small, leading to the following easy consequence.

Proposition 6.5. *For any $r \in \mathbb{N}_{\geq 3}$ and $0 < \alpha, \gamma < 1/2^{3r}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$ and any vertex subset $U \subseteq V(G)$ with $|U| \geq n/2$. For any $m \in \mathbb{N}$ with*

$$1 \leq m \leq \min\{p^{r-1} n^{1-r^3\gamma}, n^{7/8}\},$$

there exists a γ-shrinkable $(k,m)_r$-orchard $\mathcal{O}$ in $G[U]$ with k such that $\alpha n \leq km \leq 2\alpha n$.

Proof. We fix $\varepsilon > 0$ small enough to apply Proposition 4.1 and Lemma 6.4 with α, γ as defined here and $k \in \mathbb{N}$ such that $\alpha n \leq km \leq 2\alpha n$. Note that due to our upper bound on m, we certainly have $pk \geq n^{\gamma}$. We begin by finding a $((1+\gamma)k, m)_r$-orchard $\mathcal{O}'$ in $G[U]$. This can be done by repeated applications of Proposition 4.1. Indeed, we initiate a process by fixing $U' = U$ and $\mathcal{O}' = \emptyset$ and at each step we find some K_r-diamond tree $\mathcal{D}$ of order m in U', add it to $\mathcal{O}'$ and delete its vertices from U'. We claim that we can do this until $\mathcal{O}'$ has size $(1+\gamma)k$. Indeed, this follows because at any point in the process, $|V(\mathcal{O}')| \leq (1+\gamma)kmr = (1+\gamma)2\alpha rn \leq n/4$ due to our upper bounds on α and γ. Therefore, throughout the process, we have $|U'| \geq n/4$ and so U' can be split into two disjoint sets of size at least $4\alpha rn$. Therefore applying Proposition 4.1 with $z = m$ (and taking $Y' = Y$ in (3)) gives us the existence of the diamond tree at each step of this process.

Now defining $\Lambda' = \{R_{\mathcal{D}} : \mathcal{D} \in \mathcal{O}'\}$ to be the $((1+\gamma)k, m)$-system generated by taking the sets of removable vertices of diamond trees that lie in $\mathcal{O}'$, we see that Λ' satisfies the hypothesis of Lemma 6.4 since $m \leq p^{r-1} n^{1-r^3\gamma}$. Hence Lemma 6.4 implies the existence of a subsystem $\Lambda \subset \Lambda'$ of size k which is γ-shrinkable with respect to r. Finally, we conclude that $\mathcal{O} := \{\mathcal{D} \in \mathcal{O}' : R_{\mathcal{D}} \in \Lambda\}$ is the required γ-shrinkable $(k, m)_r$-orchard by Observation 6.2. ■

For dense graphs (that is, when p is large), Proposition 6.5 is already enough to establish Proposition 2.8. On the other hand, for sparse graphs Proposition 6.5 can only be used for orchards of very small order and becomes redundant as the order m approaches $p^{r-1}n$. However, in deriving Proposition 6.5, we were quite naive in our application of Lemma 6.4, using the order of a diamond tree as an upper bound on the degrees of vertices to the removable set of vertices of the diamond tree. For a set Q we expect a typical vertex $v \in V(G)$ to have $\deg_Q(v) \leq p|Q|$ and so we can hope that Lemma 6.4 can be applied to imply the existence of shrinkable orchards whose orders approach $p^{r-2}n$, gaining an extra power of p over Proposition 6.5. This is the content of the rest of this section.

6.4. Controlling degrees relative to removable sets of vertices

A reasonable approach when trying to apply Lemma 6.4 to deduce the existence of larger shrinkable orchards is to start with a larger (in size) orchard than we desire and crop diamond trees which fail the bounded degree condition. This approach is reminiscent of how we derived Lemma 6.4 from Lemma 6.3, where we greedily lost diamond trees which violated condition (1) of Lemma 6.3. In this case, though, our condition is harder to satisfy. Indeed, we require that *all* vertices in a set in our system satisfy the degree condition and not just a single vertex. In order to achieve this, we will need to appeal to (the full power of) Proposition 4.1 to choose our diamond trees. As Proposition 4.1 does not give full control over the set of vertices which end up as the set of removable vertices, we have to settle with being able to conclude our desired upper bound on the degrees of vertices relative to a *subset* of the removable vertices. The detailed statement is as follows.

Lemma 6.6. *For any $r \in \mathbb{N}_{\geq 3}$ and $0 < \gamma, \eta < 1/r^2$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$ and any vertex subset $U \subseteq V(G)$ with $|U| \geq n/4$. For any $m \in \mathbb{N}$ with $p^{r-1} n^{1-\gamma} \leq m \leq p^{r-2} n^{1-2\gamma}$, there exists a K_r-diamond tree $\mathcal{D} = (T, R, \Sigma)$ of order at most $2m$ such that $V(\mathcal{D}) \subset U$ and there exists a subset $Q \subseteq R$ of removable vertices such that $|Q| = m$ and all but at most ηm vertices $v \in V(G)$ have $\deg_Q(v) \leq p^{r-1} n^{1-\gamma}$.*

Proof. Fix $\varepsilon > 0$ small enough to apply Proposition 4.1 with $\alpha = 1/(2^{2r} r)$, small enough to apply Lemma 3.3 with $\eta_{3.3} = \eta' < \alpha^2 \eta / 2^6$ and small enough to force n to be sufficiently large in what follows. We begin by splitting U into disjoint subsets U' and W' arbitrarily so that $|U'|, |W'| \geq n/8 \geq 4\alpha r n$, noting that this is possible due to our definition of α.

Now fix some m with $p^{r-1} n^{1-\gamma} \leq m \leq p^{r-2} n^{1-2\gamma}$ and define $q := p^{r-1} n^{1-\gamma} m^{-1}$. Note that $8p \leq p n^{\gamma} \leq q \leq 1$ due to our conditions on m. As we aim to find a set Q of size m, the condition that $\deg_Q(v) > p^{r-1} n^{1-\gamma}$ is equivalent to $\deg_Q(v) > q|Q|$. As discussed above, given a diamond tree of the correct order and a subset Q of m removable vertices, we can appeal to Lemma 3.3 (ii) to bound the number of vertices which have high degree relative to Q. However, the bound is not strong enough for our purposes so we instead appeal to the full power of Proposition 4.1. The idea is to take Y of size much bigger than m. Therefore applying Lemma 3.3 (ii) with respect to Y gives a much stronger upper bound on the number of vertices which have large degree (at least $q|Y|$, say) relative to Y. If we then take Q to be a *random subset* of Y then we expect the density of the neighbourhood of a vertex in Q to have roughly the same density as the neighbourhood of that vertex in Y. Hence, we can bound the number of vertices which have large degree relative to Q by 'carrying over' the bound on the number of vertices which had large degree relative to Y. The details follow.

First we fix $d_* := \alpha^2 p^{r-1} n$ and $z := \min\{\alpha n, d_* m/2\}$. Note that for n large, due to Fact 3.2, d_* will also be large. Now apply Proposition 4.1 to obtain disjoint subsets $X, Y \subset U'$ as in the statement of Proposition 4.1. Note that $|X| \leq 2z/d_* \leq m$ and $|Y| \geq z - |X| \geq z/2$ for n sufficiently large. Fix a subset $Z \subset Y$ of size $z/2$ and let $B \subset V(G)$ be the set of vertices $v \in V(G)$ such that $\deg_Z(v) > q|Z|/4$. We claim that $|B| \leq \eta m$. Indeed, noting that $q/4 \geq 2p$, Lemma 3.3 (ii) gives

$$|B| \leq \frac{2^4 \eta' p^{2r-2} n^2}{q^2 |Z|} = \frac{2^5 \eta' n^{2\gamma} m^2}{z} \leq \begin{cases} \eta n^{2\gamma-1} p^{1-r} m & \text{if } z = d_* m/2, \\ \eta n^{2\gamma-1} m^2 & \text{if } z = \alpha n. \end{cases} \tag{6.2}$$

In the case that $z = d_* m/2$, the estimate in (6.2) is less than ηm for large n due to the condition that $\gamma < 1/r^2$ and the fact that $p \geq n^{-1/(2r-3)}$ (Fact 3.2). In the case that $z = \alpha n$, the estimate in (6.2) is less than ηm since $m \leq p^{r-2} n^{1-2\gamma} \leq n^{1-2\gamma}$.

For each $v \notin B$, we have $\deg_Z(v) \leq q|Z|/4$ and so we let $N_v \subset Z$ be a subset of exactly $q|Z|/4$ vertices in Z such that N_v contains all the neighbours of v which lie in Z. Now consider a random subset $Q_1 \subset Z$ where we keep each vertex independently with probability $p' = 4m/z$, noting that $0 \leq p' \leq 1$ for large enough n. Clearly $\mathbb{E}[|Q_1|] = p'|Z| = 2m$ and for each $v \in V(G) \setminus B$, we have $\mathbb{E}[|Q_1 \cap N_v|] = p'|N_v| = qm/2$. We get concentration for these random variables from Theorem 3.6 which is strong enough to do a union bound and conclude that whp as n (and hence m and qm) tend to infinity, we have $|Q_1| \geq m$ and $|Q_1 \cap N_v| \leq qm$ for all $v \in V(G) \setminus B$. Therefore, for sufficiently large n, we can fix such an instance of Q_1 and take Q to be a subset of Q_1 such that $|Q| = m$. Therefore for all vertices $v \in V(G) \setminus B$, we have

$$\deg_Q(v) \leq |Q \cap N_v| \leq |Q_1 \cap N_v| \leq qm = p^{r-1}n^{1-\gamma}.$$

We know that $|B| \leq \eta m$ from above and we use the conclusion of Proposition 4.1 to give a K_r-diamond tree $\mathcal{D} = (T, R, \Sigma)$ with removable vertices $R := X \cup Q \subset U' \subset U$ and Σ a matching of $(r-1)$-cliques in $W' \subset U$. We thus have $V(\mathcal{D}) \subset U$ as required and the order of $\mathcal{D}$ is $|Q| + |X| \leq 2m$. ∎

6.5. *The existence of shrinkable orchards of larger order*

Lemma 6.6 gives us the key to being able to push the methods above (which culminated in Lemma 6.4) to be able to handle orchards with larger order. We remark that the flexibility given by dealing with (k, m)-systems and Observation 6.2 is necessary in order to handle this extension. Indeed, this is due to Lemma 6.6 only giving control over the degree relative to a subset of the removable vertices of the diamond tree generated.

Proof of Proposition 5.1. By Proposition 6.5 we can focus on the case that

$$p^{r-1}n^{1-r^3\gamma} \leq m \leq \min\{p^{r-2}n^{1-2r^3\gamma}, n^{7/8}\}.$$

We fix $\varepsilon > 0$ small enough to apply Lemma 6.4 with $\alpha_{6.4} = \alpha' := \alpha/4$ and $\gamma_{6.4} = \gamma$ and small enough to apply Lemma 6.6 with $\gamma_{6.6} = \gamma' := r^3\gamma$ and $\eta_{6.6} = \eta < \alpha/8$. Finally, we fix some $k \in \mathbb{N}$ such that $\alpha n \leq km \leq 2\alpha n$ and note that $pk \geq n^\gamma$ due to our upper bound on m. By repeatedly applying Lemma 6.6, we find a $(2k, m)$-orchard $\mathcal{O}_0$ with $V(\mathcal{O}_0) \subset U$ and each $\mathcal{D} = (T, R, \Sigma) \in \mathcal{O}_0$ has the property that there exists some distinguished subset $Q_\mathcal{D} \subset R$ of removable vertices such that $|Q_\mathcal{D}| = m$ and all but at most ηm vertices v in $V(G)$ have $\deg_{Q_\mathcal{D}}(v) \leq p^{r-1}n^{1-\gamma'}$. Indeed, we can find $\mathcal{O}_0$ by sequentially choosing diamond trees and deleting their vertices from U, using the fact that $|V(\mathcal{O}_0)| \leq 4r\alpha n$ at all times in this process and so $|U \setminus V(\mathcal{O}_0)| \geq n/4$ and we can apply Lemma 6.6.

Now we will crop our orchard $\mathcal{O}_0$ to arrive at an orchard for which we can apply Lemma 6.4 to subsets of removable vertices. Similarly to the proof of Lemma 6.4, we do this by a process of 'cleaning up', losing diamond trees in the orchard which have lots of removable vertices which are atypical. So let $B_1 \subset V(G)$ be the set of vertices $v \in V(G)$ such that $\deg_{Q_\mathcal{D}}(v) > p^{r-1}n^{1-\gamma'}$ for some $\mathcal{D} \in \mathcal{O}_0$ as above. It follows that $|B_1| \leq \eta m \cdot 2k \leq \alpha n/4$. Next we delete $\mathcal{D}'$ from $\mathcal{O}_0$ if $|B_1 \cap Q_{\mathcal{D}'}| \geq m/2$. Due to our upper bound on $|B_1|$, we delete at most $k/2$ of the diamond trees $\mathcal{D}'$ from $\mathcal{O}_0$. Let the resulting suborchard be $\mathcal{O}_1 \subseteq \mathcal{O}_0$ and for each diamond tree $\mathcal{D} = (T, R, \Sigma) \in \mathcal{O}_1$ define a distinguished subset $S_\mathcal{D} \subset Q_\mathcal{D} \subseteq R$ of removable vertices such that $|S_\mathcal{D}| = m/2$ and

$$\deg_{S_\mathcal{D}}(v) \leq p^{r-1}n^{1-\gamma'} = p^{r-1}n^{1-r^3\gamma} \text{ for all } \mathcal{D} \in \mathcal{O}_1 \text{ and all } v \in \bigcup_{\mathcal{D}' \in \mathcal{O}_1} S_{\mathcal{D}'}. \tag{6.3}$$

Let $\mathcal{O}_2$ be an arbitrary suborchard of $\mathcal{O}_1$ with $|\mathcal{O}_2| = (1+\gamma)k$. Moreover, let $\Lambda' = \{S_{\mathcal{D}} : \mathcal{D} \in \mathcal{O}_2\}$ be the $((1+\gamma)k, m/4)$-system defined by the distinguished subsets of removable vertices for the K_r-diamond trees in $\mathcal{O}_2$. Now due to (6.3), Lemma 6.4 gives the existence of some γ-shrinkable (with respect to r) subsystem $\Lambda \subset \Lambda'$. Taking $\mathcal{O} := \{\mathcal{D} \in \mathcal{O}_2 : S_{\mathcal{D}} \in \Lambda\}$ thus gives a γ-shrinkable $(k,m)_r$-orchard as required, appealing to Observation 6.2. ∎

7. Shrinkable orchards of large order

In this section, we establish the existence of shrinkable orchards with large order, proving Proposition 5.2. Our approach is to find an orchard such that the K_r-hypergraph H generated by the orchard is very dense. This allows us to apply Lemma 3.8 in many subhypergraphs of H. Coupled with Theorem 3.12, this will imply that the orchard is shrinkable. As in the previous section, we begin in Section 7.1 by using these results on fractional matchings to deduce conditions on an orchard which guarantee shrinkability. We will then show in Section 7.2 that we can appeal to Proposition 4.1 to generate diamond trees whose removable vertices are contained in many copies of K_r. This will then allow us to prove the existence of shrinkable orchards of large order in Section 7.3. As in Section 6, however, this first argument will fall short of the range of orders needed in Proposition 5.2. The rest of the section is thus concerned with extending our methods to capture more orders. This leads us to a process which generates an orchard in two rounds. The outcome of the first round is discussed in Section 7.4, and building on this, in Section 7.5 we detail properties of the orchard after a second round of generation. Finally, in Section 7.6, we show that by generating orchards via this two-phase process, we end up with orchards which are shrinkable. This allows us to complete the proof of Proposition 5.2.

7.1. A density condition which guarantees shrinkability

We begin by applying Lemma 3.8 and Theorem 3.12 to give a density condition which we can use to show that an orchard is shrinkable. This transforms our problem into finding orchards which satisfy this condition.

Lemma 7.1. *For all $r \in \mathbb{N}_{\geq 3}$ and $0 < \gamma < 1/(2r^3)$, there exists a $k_0 \in \mathbb{N}$ such that the following holds. Suppose that $\mathcal{O}$ is a $(k,m)_r$-orchard in a graph G with $k \in \mathbb{N}_{\geq k_0}$ and $m \in \mathbb{N}$. For a diamond tree $\mathcal{D} \in \mathcal{O}$, let $R_{\mathcal{D}}$ denote its removable vertices and for a suborchard $\mathcal{O}' \subset \mathcal{O}$, let $R(\mathcal{O}') := \bigcup_{\mathcal{D}\in\mathcal{O}'} R_{\mathcal{D}}$ denote the union of the sets of removable vertices of diamond trees in $\mathcal{O}'$. Suppose that the following condition holds:*

For any $\mathcal{D} \in \mathcal{O}$ and $\mathcal{P} \subset \mathcal{O} \setminus \{\mathcal{D}\}$ such that $|\mathcal{P}| \geq k/(4r)$, there exists a suborchard $\mathcal{P}^* = \mathcal{P}^*(\mathcal{D}, \mathcal{P}) \subset \mathcal{P}$ such that $|\mathcal{P}^*| \leq k^{1-r^3\gamma}$ and for any *disjoint suborchards $\mathcal{O}_1, \dots, \mathcal{O}_{r-2} \subset \mathcal{P} \setminus \mathcal{P}^*$, with $|\mathcal{O}_i| \geq k^{1-r^3\gamma}$ for $i \in [r-3]$ and $|\mathcal{O}_{r-2}| \geq \gamma k$, there is a copy of K_r in G traversing $R_{\mathcal{D}}$, $R(\mathcal{P}^*)$ and $R(\mathcal{O}_i)$ for $i \in [r-2]$.* (7.1)

Then $\mathcal{O}$ is γ-shrinkable.

Let us take a moment to digest the density condition (7.1). For simplicity, one can think of $\mathcal{P}^*$ being a single diamond tree $\mathcal{D}^* = \mathcal{D}^*(\mathcal{D}, \mathcal{P})$. Indeed, this is the setting that we will work in first when applying Lemma 7.1. Simplifying further and just focusing on the case that $r = 3$, condition (7.1) translates as stating that for any K_3-diamond tree $\mathcal{D}$ in the orchard and large suborchard $\mathcal{P} \subset \mathcal{O}$, there is some diamond tree $\mathcal{D}^* \in \mathcal{P}$ such that the pair $\{\mathcal{D}, \mathcal{D}^*\}$ has high degree in the K_3-hypergraph generated by $\mathcal{P}$. Indeed, for any small linear sized $\mathcal{O}_1 \subset \mathcal{P}$, there is a hyperedge in $H(\mathcal{O})$ containing $\mathcal{D}, \mathcal{D}^*$ and a diamond tree in $\mathcal{O}_1$. In general, when $r \geq 4$, we need to guarantee traversing K_rs when some of the sets we look to traverse are smaller than linear (size $k^{1-r^3\gamma}$). Also later on we will need the full power of Lemma 7.1 which allows us to choose the $\mathcal{P}^*$ as a small suborchard as opposed to a single diamond tree. We now prove the lemma.

Proof of Lemma 7.1. Let $\mathcal{Q} \subset \mathcal{O}$ be an arbitrary suborchard of $\mathcal{O}$ of size γk. We will show that $\mathcal{O}$ is shrinkable with respect to $\mathcal{Q}$. So fix some arbitrary suborchard $\mathcal{Q}' \subset \mathcal{Q}$ and let $H := H(\mathcal{O} \setminus \mathcal{Q}')$ be the K_r-hypergraph generated by $\mathcal{O} \setminus \mathcal{Q}'$. We have to show that H has a matching covering all but at most $k^{1-\gamma}$ vertices of H.

In order to show the existence of a large matching in H, as we did in Lemma 6.3, we appeal to Theorem 3.12. So let us fix $N = |V(H)|$ and note that as $N \geq (1-\gamma)k$, by choosing k_0 to be large, we can assume that N is sufficiently large in what follows. Now fix some 2-uniform graph J on $V(H)$ of maximum degree at most $N^{r^2\gamma}$. If we can show that $H \setminus H_J$ contains a perfect fractional matching, then we are done by Theorem 3.12 because, J being arbitrary, the theorem guarantees a matching covering all but at most $N^{1-\gamma} \leq k^{1-\gamma}$ vertices of H.

In order to prove the existence of a perfect fractional matching in $H \setminus H_J$, we appeal to Lemma 3.8, fixing $M := N/(2r)$. Thus, we need to show that given any K_r-diamond tree $\mathcal{D} \in V(H) = \mathcal{O} \setminus \mathcal{Q}$ and suborchard $\mathcal{P}_0 \subset V(H) \setminus \{\mathcal{D}\}$ with $|\mathcal{P}_0| \geq M$, there is an edge in $H \setminus H_J$ containing $\mathcal{D}$ and $r-1$ K_r-diamond trees in $\mathcal{P}_0$. So fix such a $\mathcal{D}$ and $\mathcal{P}_0$. Let $\mathcal{P} := \mathcal{P}_0 \setminus N^J(\mathcal{D})$. Then

$$|\mathcal{P}| \geq |\mathcal{P}_0| - |N^J(\mathcal{D})| \geq \frac{N}{2r} - N^{r^2\gamma} \geq \frac{(1-\gamma)k}{2r} - k^{r^2\gamma} \geq \frac{k}{4r}$$

for k sufficiently large. Hence by condition (7.1), we have the existence of some $\mathcal{P}^* = \mathcal{P}^*(\mathcal{D}, \mathcal{P}) \subset \mathcal{P}$ as in the hypothesis. Now we will iteratively define $\mathcal{O}_i$ for $1 \leq i \leq r-2$ as follows. We begin by fixing $\mathcal{P}' = \mathcal{P}$ and defining $\mathcal{Q}_0 := \bigcup_{\mathcal{C} \in \mathcal{P}^*}(N^J(\mathcal{C}) \cup \{\mathcal{C}\})$. For $1 \leq i \leq r-2$, we update $\mathcal{P}'$ by removing any diamond trees in $\mathcal{Q}_{i-1}$ from $\mathcal{P}'$ and then define $\mathcal{O}_i$ to be an arbitrary suborchard of $\mathcal{P}'$ of size $k^{1-r^3\gamma}$ if $i \in [r-3]$, and of size γk if $i = r-2$. If $i = r-2$ we end this process. If $i < r-2$, we define $\mathcal{Q}_i := \bigcup_{\mathcal{C} \in \mathcal{O}_i}(N^J(\mathcal{C}) \cup \{\mathcal{C}\})$ and move to the next index.

Let us check that we are successful in each round. Indeed, this follows because at the beginning of step i in the process, $\mathcal{P}'$ has size

$$|\mathcal{P}'| \geq \frac{k}{4r} - ik^{1-r^3\gamma}(1 + N^{r^2\gamma}) \geq \frac{k}{4r} - rk^{1-r^3\gamma+r^2\gamma} \geq \gamma k \geq k^{1-r^3\gamma}$$

for large k. Therefore there is always space in $\mathcal{P}'$ to choose our suborchard $\mathcal{O}_i$ at each step i. Now condition (7.1) gives a copy of K_r in G traversing $R_{\mathcal{D}}$, $R(\mathcal{P}^*)$ and $R(\mathcal{O}_i)$ for $i \in [r-2]$. This gives a hyperedge e in the K_r-hypergraph $H = H(\mathcal{O} \setminus \mathcal{Q}')$ which has one vertex $\mathcal{D}$, one vertex in $\mathcal{P}^* \subset \mathcal{P}_0$ and one vertex in each of the $\mathcal{O}_i \subset \mathcal{P}_0$. Moreover, this edge e lies in $H \setminus H_J$. Indeed, by our construction of $\mathcal{P}^*$ and the $\mathcal{O}_i$, there is no edge in J between any pair of distinct sets in the family $\{\{\mathcal{D}\}, \mathcal{P}^*, \mathcal{O}_1, \ldots, \mathcal{O}_{r-2}\}$. We have therefore established the existence of a perfect fractional matching in $H \setminus H_J$ due to Lemma 3.8, which implies that $\mathcal{O}$ is γ-shrinkable as detailed above. ∎

Lemma 7.1 gives a route to proving the existence of shrinkable orchards. Indeed, if the sets of vertices which arise as pools of removable vertices of suborchards are sufficiently large, then appealing to Corollary 3.5 can give the required transversal copy of K_r in G, so that (7.1) is satisfied. However, we cannot immediately derive such results because the sizes of the sets required in (7.1) are too small. In particular, (7.1) forces only one set (namely $R(\mathcal{O}_{r-2})$) to be linear in size, whilst all other sets that feature can have sublinear size. This is troublesome because the examples we have from Corollary 3.5 to generate transversal copies of K_r require at least two of the sets involved to be linear. Indeed, it can be seen from the more general Lemma 3.4 that we cannot do any better. That is, in order to use Definition 1.3 and our condition on β to derive the existence of a copy of K_r that traverses a family of sets, at least two of the sets in the family must be linear in size. Therefore in order to apply Lemma 7.1 and derive the existence of shrinkable orchards, we have to obtain orchards with some additional structure. We start by exploring properties of singular diamond tress that we can guarantee.

7.2. *Popular diamond trees*

As was the case when we were interested in proving the existence of shrinkable orchards with small order, Proposition 4.1 gives a powerful tool for proving the existence of diamond trees with additional desired properties. Here we show that we can choose a diamond tree so that there are many copies of K_r formed with its removable vertices.

Lemma 7.2. *For any $r \in \mathbb{N}_{\geq 3}$ and $0 < \alpha < 1/2^{12r}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$ and any vertex subset $U \subseteq V(G)$ with $|U| \geq n/4$. Suppose that $m \in \mathbb{N}$ with*

$$\max\{p^{1-r}, p^{r-1}n\} \leq m \leq n^{7/8}$$

and we have set families $\mathcal{W}_0, \mathcal{W}_1, \ldots, \mathcal{W}_{r-1} \subset 2^{V(G)}$ such that

(1) *$|W_0| \geq \alpha p^{r-1} n$ for all $W_0 \in \mathcal{W}_0$;*

(2) *$|W_i| \geq \alpha p n$ for all $W_i \in \mathcal{W}_i$, $1 \leq i \leq r-3$;*

(3) *$|W_{r-2}| \geq \alpha n$ for all $W_{r-2} \in \mathcal{W}_{r-2}$;*

(4) *$\prod_{i=0}^{r-2} |\mathcal{W}_i| \leq 2^{m/4}$.*

Then there exists a K_r-diamond tree $\mathcal{D} = (T, R, \Sigma)$ in $G[U]$ of order at least m and at most $2m$ such that for any *choice of sets $\mathbf{W} = (W_0, \dots, W_{r-2}) \in \mathcal{W}_0 \times \cdots \times \mathcal{W}_{r-2}$, there is a copy of K_r in G traversing R and the sets $W_0, \dots, W_{r-2}$.*

Proof. Let us fix $\varepsilon > 0$ small enough to apply Proposition 4.1 with $\alpha_{4.1} = \alpha' := 1/2^{3r}$ and Corollary 3.5 with $\alpha_{3.5} = \alpha$, as well as being small enough to force n to be sufficiently large. Note that our lower bound of $\Omega(p^{r-1}n)$ on m and Fact 3.2 imply that $m \to \infty$ as $n \to \infty$ and so we can also assume m is sufficiently large in what follows. We begin by splitting U into disjoint subsets U' and W' arbitrarily so that $|U'|, |W'| \geq n/8 = 4\alpha' rn$, noting that this is possible by our definition of α'. We further fix $d_* := \alpha'^2 p^{r-1}n$.

Now we apply Proposition 4.1 with $z := \alpha'^2 n/4 = n/2^{6r+2}$ and fix the sets $X \subset U'$ and $Y \subset U'$ which are output. Note that

$$|X| \leq \max \begin{Bmatrix} \frac{2z}{d_*} = \frac{1}{2p^{r-1}} \\ 1 \end{Bmatrix} \leq \frac{m}{2} \quad \text{and} \quad |Y| = z - |X| \geq \frac{z}{2} = \frac{n}{2^{6r+3}},$$

for n large.

Now for each choice of $\mathbf{W} = (W_0, \dots, W_{r-2}) \in \mathcal{W}_0 \times \cdots \times \mathcal{W}_{r-2}$, we find some subset $Y(\mathbf{W}) \subset Y$ of size $|Y|/2$ such that for every $v \in Y(\mathbf{W})$, there is a copy of K_{r-1} in the neighbourhood of v which traverses $W_0, \dots, W_{r-2}$. In other words, for every $v \in Y(\mathbf{W})$, there is a copy of K_r traversing $W_0, \dots, W_{r-2}$ and $\{v\}$. We can find $Y(\mathbf{W})$ by repeated applications of Corollary 3.5 (2). In more detail, we initiate with $Y_0 = Y$ and $Y(\mathbf{W})$ empty and in each step we find a copy of K_r traversing $W_0, \dots, W_{r-2}$ and Y_0. Taking v to be the[11] vertex of this K_r that lies in Y_0, we add v to $Y(\mathbf{W})$, delete it from Y_0 and move to the next step. We continue for $|Y|/2$ steps using the fact that the conditions of Corollary 3.5 (2) are satisfied at each step. Indeed, this is due to the lower bounds on the sizes of W_i in conditions (1)–(3) of this lemma and the fact that $|Y_0| \geq |Y| - |Y(\mathbf{W})| \geq |Y|/2 \geq \alpha n$ throughout, using our upper bound on α and our lower bound on $|Y|$ here.

Similarly to the proof of Lemma 6.6, we now take Q to be a random subset of Y by taking each vertex of Y into Q independently with probability $p' := \frac{5m}{4|Y|}$. Thus $\mathbb{E}[|Q|] = 5m/4$ and by Theorem 3.6, we have $m \leq |Q| \leq 3m/2$ with probability at least $1 - 2e^{-m/60}$. Furthermore, for any fixed $\mathbf{W} \in \mathcal{W}_0 \times \cdots \times \mathcal{W}_{r-2}$,

$$\mathbb{E}[|Q \cap Y(\mathbf{W})|] = p'|Y(\mathbf{W})| = 5m/8.$$

Applying Theorem 3.6 again implies that the probability that $|Q \cap Y(\mathbf{W})| = 0$ is less than $e^{-5m/16}$. Therefore using the inequality $\prod_{i=0}^{r-2} |\mathcal{W}_i| \leq 2^{m/4}$ and appealing to a union bound, we can conclude that whp as n (and hence m) tends to infinity, we see that $m \leq |Q| \leq 3m/2$ and $Q \cap Y(\mathbf{W}) \neq \emptyset$ for *all* choices of $\mathbf{W} \in \mathcal{W}_0 \times \cdots \times \mathcal{W}_{r-2}$. So for sufficiently large n we can fix such an instance $Q \subset Y$ and taking $R := X \cup Q$ we have

[11] Here we refer to *the* vertex that lies in Y_0 although there may be several (if the W_i intersect the Y_0). What we mean here is the vertex v in the copy of K_r which is assigned to Y_0 by virtue of the copy being traversing.

that a K_r-diamond tree $\mathcal{D} = (T, R, \Sigma)$ with removable set of vertices R is guaranteed by Proposition 4.1. We claim that $\mathcal{D}$ satisfies all the necessary conditions. Indeed, the fact that the order of $\mathcal{D}$ lies between m and $2m$ follows from the fact that $m \le |Q| \le 3m/2$ and $|X| \le m/2$, whilst the fact that $Q \cap Y(\mathbf{W}) \ne \emptyset$ for each choice of $\mathbf{W} = (W_0, \dots, W_{r-2})$ guarantees that we have a copy of K_r traversing $Q \subset R$ and the sets $W_0, \dots, W_{r-2}$. ■

7.3. The existence of shrinkable orchards of large order

Using Lemma 7.2 to generate the diamond trees that form our orchard, we can prove that the orchard generated satisfies the condition of Lemma 7.1 and hence is shrinkable. This gives the following.

Proposition 7.3. *For any $r \in \mathbb{N}_{\ge 3}$ and $0 < \alpha, \gamma < 1/2^{12r}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \le \varepsilon p^{r-1} n$ and any vertex subset $U \subseteq V(G)$ with $|U| \ge n/2$. For any $m \in \mathbb{N}$ with*

$$\max\{p^{1-r}, p^{r-1}n\} \le m \le n^{7/8},$$

there exists a γ-shrinkable $(k, m)_r$-orchard $\mathcal{O}$ in $G[U]$ with $k \in \mathbb{N}$ such that $\alpha n \le km \le 2\alpha n$.

Proof. Fix $\varepsilon > 0$ small enough to apply Lemma 7.1 with $\gamma_{7.1} = \gamma$ and Lemma 7.2 with $\alpha_{7.2} = \alpha' = \alpha\gamma$. Fix some $k \in \mathbb{N}$ such that $\alpha n \le k \le 2\alpha n$. We also ensure that ε is small enough to force n (and hence k, due to our upper bound on m) to be sufficiently large in what follows. Now we begin by noticing that $k \le m/(8r)$. Indeed, if $p \ge n^{-1/(2r-2)}$, then

$$p^{1-r} \le \sqrt{n} \le p^{r-1} n \le m,$$

while if $p \le n^{-1/(2r-2)}$, then

$$p^{r-1} n \le \sqrt{n} \le p^{1-r} \le m.$$

Therefore, for any p we have $m \ge \sqrt{n}$ and $k \le 2\alpha n/m \le 2^{-11r}\sqrt{n} \le m/(8r)$.

Now we turn to finding our $(k, m)_r$-orchard in $G[U]$. We do this by finding one diamond tree at a time as follows. For $1 \le i \le k$, fix $U_i := U \setminus \bigcup_{i' < i} V(\mathcal{D}_{i'})$ and note that $|U_i| \ge |U| - 2\alpha r n \ge n/4$ throughout due to our condition on α. We then apply Lemma 7.2 to find a diamond tree $\mathcal{D}_i = (T_i, R_i, \Sigma_i)$ such that $V(\mathcal{D}_i) \subset U_i$ and for any choice of $i' \in [i-1]$ and disjoint subsets $I_1, \dots, I_{r-2} \subset [i-1] \setminus \{i'\}$ with $|I_j| \ge pk$ for $1 \le j \le r-3$, and $|I_{r-2}| \ge \gamma k$, there is a copy of K_r traversing R_i, $R_{i'}$ and the sets $\bigcup_{\ell \in I_j} R_\ell$ for $j \in [r-2]$. The existence of such a $\mathcal{D}_i$ follows from Lemma 7.2. Indeed, we define $\mathcal{W}_0 = \{R_{i'} : i' \in [i-1]\}$, $\mathcal{W}_j = \{\bigcup_{\ell \in I'} R_\ell : I' \subset [i-1], |I'| \ge pk\}$ for $1 \le j \le r-3$ and finally $\mathcal{W}_{r-2} = \{\bigcup_{\ell \in I'} R_\ell : I' \subset [i-1], |I'| \ge \gamma k\}$. We need to check that conditions (1)–(4) of Lemma 7.2 are satisfied. Indeed, (1) follows from our lower bound on m, whilst (2) and (3) follow from the fact that $km \ge \alpha n$ and our definition of α'. Finally, note that each choice of a set in any of the $\mathcal{W}_j$ comes from a subset of $[i-1]$. Hence we can upper bound $\prod_{j=0}^{r-2} |\mathcal{W}_j|$ by $(2^i)^{r-1} \le 2^{rk}$. As discussed in the opening paragraph, we have $k \le m/(8r)$ and so condition (4) of Lemma 7.2 is also satisfied. Thus Lemma 7.2 succeeds in finding the necessary K_r-diamond tree at every step of this process.

Let $\mathcal{O} = \{\mathcal{D}_1, \dots, \mathcal{D}_k\}$ be the orchard obtained by this process. We claim that $\mathcal{O}$ is γ-shrinkable and to show this we appeal to Lemma 7.1 and so need to show that the density condition (7.1) is satisfied by $\mathcal{O}$. So fix $\mathcal{D}_i \in \mathcal{O}$ and $\mathcal{P} \subset \mathcal{O} \setminus \{\mathcal{D}_i\}$ with $|\mathcal{P}| \ge k/(4r)$. We then define $\mathcal{D}^* = \mathcal{D}^*(\mathcal{D}_i, \mathcal{P})$ (this plays the role of $\mathcal{P}^*$ in (7.1)) to be the diamond tree in $\mathcal{P}$ with the highest index. That is, we define $i^* := \max\{i' : \mathcal{D}_{i'} \in \mathcal{P}^*\}$ and set $\mathcal{D}^* = \mathcal{D}_{i^*}$. Note that we may have $i^* < i$ but this will not be a problem. We claim that condition (7.1) is satisfied with this choice of $\mathcal{D}^*$. Indeed, let $\mathcal{O}_1, \dots, \mathcal{O}_{r-2} \subset \mathcal{P}^* \setminus \{\mathcal{D}^*\}$ be disjoint suborchards satisfying the lower bounds on the sizes given by (7.1). For each $j \in [r-2]$, define $I_j := \{i' : \mathcal{D}_{i'} \in \mathcal{O}_j\}$. Then $|I_{r-2}| \ge \gamma k$. For $1 \le j \le r-3$ we have $|I_j| \ge k^{1-r^3\gamma} \ge pk$. This follows from the fact that

$$k^{-r^3\gamma} \ge k^{-1/2^r} \ge n^{-1/2^{r+1}} \ge n^{-\frac{1}{8(r-1)}} \ge p,$$

where we have used the upper bound on γ in the first inequality, the fact that $k \le \sqrt{n}$ in the second inequality (see the opening paragraph of the proof), and $p^{r-1}n \le m \le n^{7/8}$ in the last inequality. Now relabelling $\{i, i^*\}$ as $\{\ell_0, \ell_1\}$ so that $\ell_0 < \ell_1$, we see that at the point of choosing $\mathcal{D}_{\ell_1}$, we guaranteed that there was a K_r traversing R_{ℓ_1}, R_{ℓ_0} and the sets $R(\mathcal{O}_j) = \bigcup_{i' \in I_j} R_{i'}$ for $j \in [r-2]$. By Lemma 7.1 this completes the proof that $\mathcal{O}$ is γ-shrinkable. ∎

Proposition 7.3 establishes Proposition 5.2 when G is very dense. However, when G is sparse (when $p \le n^{-1/(2r-2)}$ to be specific), the lower bound of $m \ge p^{1-r}$ takes over and we are left with a gap between the range covered by Proposition 7.3 and the desired range of Proposition 5.2. Tracing the condition that $m = \Omega(p^{1-r})$ back through the proof, we can see that this was necessary in order to prove Lemma 7.2. There, we used our key Proposition 4.1 to generate a diamond tree where we had a large pool Y of vertices which were candidates for being removable vertices. In order to establish the existence of the cliques we need in Lemma 7.2, we needed Y to be linear in size. The sticking point then comes from the fact that Proposition 4.1 can only guarantee a maximum factor of $O(p^{r-1}n)$ between the size of the pool of vertices Y and the order of the diamond tree that we generate. Indeed, in Proposition 4.1 we are forced to include the set X in the removable vertices of the diamond tree we generate and when Y is linear in size, X could have size as large as $\Omega(p^{1-r})$. It is unclear how one would improve on this and find diamond trees with smaller order that are still contained in sufficiently many copies of K_r.

Thankfully, there is a way to circumvent this issue and apply our methods to close the gap in the range of orders nonetheless. The key idea is to replace the diamond tree generated by Lemma 7.2 with a *set of diamond trees*, that is, a small suborchard. Indeed, by grouping together diamond trees, we can decrease their order but guarantee that the collective pool of potential removable vertices for the group is still linear in size. Through following a similar proof to that of Lemma 7.2, this has the outcome of being able to guarantee many copies of K_r which contain a vertex in the removable vertices of *one of* the diamond trees in the group. Moreover, in the proof of Proposition 7.3, we crucially used the fact that we could generate diamond trees from Lemma 7.2 to establish the density condition (7.1) of Lemma 7.1. We chose an appropriate $\mathcal{D}^*$ and used the fact that it had been generated by Lemma 7.2 to prove the required existence of transversal copies of K_r. However, Lemma 7.1 allows us to use a much larger suborchard $\mathcal{P}^*$ for this condition as opposed to a single diamond tree. Therefore there is hope to incorporate the idea of using a suborchard instead of a single diamond tree in Lemma 7.2, whilst maintaining the overall, scheme of the proof. There are some further difficulties to overcome, but on a high level, this is the approach we follow in the next sections to establish Proposition 5.2.

7.4. *Preprocessing the orchard*

As discussed above, in order to prove Proposition 5.2 and remove the condition that $m = \Omega(p^{1-r})$ from Proposition 7.3, we need to replace the role played by $\mathcal{D}^*$ in the proof by a small suborchard $\mathcal{P}^*$. This allows us to prove an analogue of Lemma 7.2, where one now finds an orchard whose collective set of removable vertices lies in many copies of K_r. Our shrinkable orchard will then be formed as the union of many of these smaller orchards. Indeed, in what follows we will split k as $k = \ell t$ and will aim to have t smaller $(\ell, m)_r$-orchards contributing to our shrinkable orchard $\mathcal{O}$. Each of the $(\ell, m)_r$-orchards will have strong connectivity to the rest of the orchard $\mathcal{O}$.

In order to work with the fact that we are splitting k into t sets of size ℓ, we introduce a two-coordinate index system, with $(i, j) \in [t] \times [\ell]$ indicating that we are referring to the jth object in the ith subset and we will work through these indices lexicographically. In more detail, we let $<_L$ denote the lexicographic order on the pairs $(i, j) \in [t] \times [\ell]$. That is, $(i', j') <_L (i, j)$ if and only if either $1 \le i' \le i - 1$ and $1 \le j' \le \ell$ or $i' = i$ and $1 \le j' \le j - 1$. Furthermore, for each $1 \le i \le t$ and $1 \le j \le \ell$, we define

$$I_{<ij} := \{(i', j') \in [t] \times [\ell] : (i', j') <_L (i, j)\}$$

to be the indices (i', j') which come before (i, j) in the lexicographic order.

A hurdle that arises with our new approach is that we lose the symmetry provided by the fact that both $\mathcal{D}$ and $\mathcal{D}^*$ in our applications of Lemma 7.1 were given by singular diamond trees. Indeed, in our proof of Proposition 7.3, when verifying condition (7.1) of Lemma 7.1, we use the fact that both the arbitrary diamond tree $\mathcal{D} = \mathcal{D}_i$ and the diamond tree $\mathcal{D}^* = \mathcal{D}^*(\mathcal{D}_i, \mathcal{P})$ that we can choose were generated using Lemma 7.2. We now hope to generate our suborchards $\mathcal{P}^*$ using an equivalent to Lemma 7.2, and this will mean that we can no longer switch the roles of $\mathcal{D}$ and $\mathcal{P}^*$ when appealing to the conclusion of (the proof method of) Lemma 7.2. In particular, this places a higher demand on the properties we need to conclude of our $(\ell, m)_r$-suborchards.

In more detail, we need to generate suborchards which are highly connected to *all* the other vertices of the K_r-hypergraph $H(\mathcal{O})$. Therefore it no longer suffices to build our orchard in a linear fashion, choosing diamond trees (or indeed suborchards) to be well connected (in terms of the K_r-hypergraph) with previously chosen diamond trees. We will instead generate our orchard in two rounds. In the first round we fix a part of each diamond tree and using Proposition 4.1, provide large pools of vertices which can extend the parts of the diamond trees chosen so far, which we will then do in the second round. Lemma 7.4 details the outcome we draw from this preprocessing first round.

Lemma 7.4. *For any $r \in \mathbb{N}_{\geq 3}$ and $0 < \alpha < 1/2^{12r}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$, any vertex subset $U \subseteq V(G)$ with $|U| \geq n/2$ and any $k, m, t, \ell \in \mathbb{N}$ such that*

$$k = t\ell, \quad \alpha n \leq km \leq 2\alpha n \quad \text{and} \quad \ell m \geq p^{1-r}.$$

There exist vertex sets $Z_{ij}, Y_{ij} \subset U$ and matchings $\Pi_{ij}, \Upsilon_{ij} \subset K_{r-1}(G[U])$ of $(r-1)$-cliques for each $i \in [t]$ and $j \in [\ell]$ such that the copies of K_{r-1} in each $\Upsilon_{ij} := \{S_v : v \in Y_{ij}\}$ are indexed by the vertices in Y_{ij} and conditions (1_{ij}) through (5_{ij}) below are satisfied for all $1 \leq i \leq t$ and $1 \leq j \leq \ell$:

(1_{ij}) *$|Z_{ij}| = m$ and $|\Pi_{ij}| = |Z_{ij}| - 1$.*

(2_{ij}) *$|Y_{ij}| = |\Upsilon_{ij}| = \sqrt{\alpha}\, n/\ell$.*

(3_{ij}) *The vertex sets Z_{ij}, Y_{ij}, $V(\Pi_{ij})$ and $V(\Upsilon_{ij})$ are all disjoint from each other.*

(4_{ij}) *$A \cap A' = \emptyset$ for any choice of $A \in \{Z_{ij}, V(\Pi_{ij}), Y_{ij}, V(\Upsilon_{ij})\}$ and*[12]

$$A' \in \{Z_{i'j'}, V(\Pi_{i'j'}) : (i', j') \in I_{<ij}\} \cup \{Y_{ij'}, V(\Upsilon_{ij'}) : 1 \leq j' \leq j - 1\}.$$

(5_{ij}) *For any choice of $\tilde{Y}$ such that $\tilde{Y} \subseteq Y_{ij}$, there exists a K_r-diamond tree $\mathcal{D} = (T, R, \Sigma)$ such that $R = Z_{ij} \cup \tilde{Y}$ and $\Sigma = \Pi_{ij} \cup \tilde{\Upsilon}_{ij}$, where $\tilde{\Upsilon}_{ij} \subset \Upsilon_{ij}$ is defined to be*

$$\tilde{\Upsilon}_{ij} := \{S_{\tilde{v}} : \tilde{v} \in \tilde{Y} \subset Y_{ij}\}.$$

As mentioned above, in this first round we put aside part of every single diamond tree in the $(k, m)_r$-orchard we are going to generate, thus partially defining the orchard. We also put aside large pools of vertices which will be used to extend these diamond trees in the second round of generating our orchard. The fixed parts of the diamond trees chosen in Lemma 7.4 are the sets Z_{ij} and the interior cliques Π_{ij}, whilst the pools of potential removable vertices and interior cliques that can be used to extend the diamond trees chosen are given by the sets Y_{ij} and Υ_{ij}, respectively. We make sure through conditions (1_{ij}) that these fixed diamond subtrees contribute a substantial portion of the final diamond trees that we are shooting for (which will have order between m and $2m$). We also guarantee through conditions (4_{ij}) that the parts of the diamond trees that we put aside in this preprocessing round do not interfere with each other, in that they are vertex disjoint. Notice also that if we fix $i \in [t]$, then conditions (4_{ij}) for all $j \in [\ell]$ guarantee that the

[12]Crucially, we do not require that A is disjoint from *all* $Y_{i'j'}$ and $V(\Upsilon_{i'j'})$, only those that are in the same subfamily indexed by i.

sets Y_{ij}, $V(\Upsilon_{ij})$, $j \in [\ell]$, do not intersect each other. This is important because in the second round of generating our orchard, we will want to extend all the diamond trees in the ith $(\ell, m)_r$-suborchard simultaneously and so we do not want any interference between the choices of the extensions within such a suborchard. Also note that conditions (2_{ij}), for fixed $i \in [t]$ and all $j \in [\ell]$, guarantee that the collective pool of potential removable vertices for the ith $(\ell, m)_r$-suborchard (the set $\bigcup_{j\in[\ell]} Y_{ij}$) is linear in size, as required. Finally, conditions (5_{ij}) contain the heart of Proposition 4.1, allowing us to arbitrarily extend any of the diamond trees we have so far using any subsets of the pools (the Y_{ij}) of potential removable vertices and interior cliques (the Υ_{ij}) we have put aside.

Our final remark on the statement of Lemma 7.4 is that we do not require e.g. Y_{ij} and $Y_{i'j'}$ for $i \neq i'$, to be disjoint. Indeed, as we have t suborchards and each has a linear collective pool of potential removable vertices, there would not be enough space in the graph to keep these pools disjoint. However, by requiring that the collective pool is much larger than all the vertices in our orchard (that is, much larger than km), we guarantee that we will be able to proceed greedily in our second round (Lemma 7.5) of defining the orchard, always having a large enough set of potential removable vertices at each step.

Proof of Lemma 7.4. Let us fix $\varepsilon > 0$ small enough to apply Proposition 4.1 with $\alpha_{4.1} = \alpha' := 1/2^{2r+1}$. We will find these vertex sets and matchings of $(r-1)$-cliques algorithmically working through the pairs $(i, j) \in [t] \times [\ell]$ in lexicographic order. So let us fix some $(i^*, j^*) \in [t] \times [\ell]$ and suppose that we have already found Z_{ij}, Y_{ij}, Π_{ij} and Υ_{ij} such that conditions (1_{ij}) through (5_{ij}) are satisfied for all $(i, j) \in I_{<i^*j^*}$. We fix $W^* \subset U$ to be

$$W^* := \Big(\bigcup\{Z_{ij} \cup V(\Pi_{ij}) : (i, j) \in I_{<i^*j^*}\}\Big) \\ \cup \Big(\bigcup\{Y_{i^*j} \cup V(\Upsilon_{i^*j}) : 1 \leq j \leq j^* - 1\}\Big),$$

and let $U^* := U \setminus W^*$. We use conditions (1_{ij}) and (2_{ij}) to upper bound the size of W^* as follows. We have

$$|W^*| \leq rm((i^* - 1)\ell + j^* - 1) + \frac{\sqrt{\alpha}\, rn}{\ell}(j^* - 1) \leq rmt\ell + \sqrt{\alpha}\, rn \leq (2\alpha + \sqrt{\alpha})rn,$$

using $mt\ell = mk \leq 2\alpha n$. Hence $|U^*| \geq n/4$ from our upper bound on α. We will find $Z_{i^*j^*}, Y_{i^*j^*} \subset U^*$ and $\Pi_{i^*j^*}, \Upsilon_{i^*j^*} \subset K_{r-1}(G[U^*])$ and so condition $(4_{i^*j^*})$ will be satisfied. The required vertex sets $Z_{i^*j^*}$ and $Y_{i^*j^*}$ are found by an application of Proposition 4.1. So let us split U^* into disjoint subsets U' and W' arbitrarily so that $|U'|, |W'| \geq n/8 \geq 4\alpha' rn$, noting that this is possible by our definition of α'. We further fix $d_* := \alpha'^2 p^{r-1} n$ and $z := m + \sqrt{\alpha}\, n/\ell$ and note that $z \leq \alpha' n$ due to the fact that $m \leq 2\alpha n/k \leq 2\alpha n$ and our upper bound on α.

So Proposition 4.1 shows that there exists disjoint vertex subsets $X, Y \subset U' \subset U^*$ such that $|X| + |Y| = z$ and $|X| = 1 \leq m$ or

$$|X| \leq 2z/d_* \leq \frac{2m}{d_*} + \frac{2\sqrt{\alpha}\, n}{d_* \ell} \leq \frac{m}{2} + \frac{2\sqrt{\alpha}}{\alpha'^2 p^{r-1} \ell} \leq m,$$

using our upper bound on α and lower bound on ℓm in the last inequality. As $|X| \le m$, we can fix some $Z_{i^*j^*} \subset X \cup Y$ such that $X \subseteq Z_{i^*j^*}$ and $|Z_{i^*j^*}| = m$. Therefore letting $Y_{i^*j^*} := Y \setminus Z_{i^*j^*}$, we have $|Y_{i^*j^*}| = z - m = \sqrt{\alpha}\, n/\ell$ and so the size requirements on $Z_{i^*j^*}$ in $(1_{i^*j^*})$ and on $Y_{i^*j^*}$ in $(2_{i^*j^*})$ are both satisfied. Moreover, part of $(5_{i^*j^*})$ is also satisfied. Indeed, for some $\tilde{Y} \subset Y_{i^*j^*}$, taking $Y' = \tilde{Y} \cup (Z_{i^*j^*} \setminus X)$, Proposition 4.1 implies that there is a diamond tree $\mathcal{D} = (T, R, \Sigma)$ with removable vertices $R = X \cup Y' = Z_{i^*j^*} \cup \tilde{Y}$ and Σ a matching of $(r-1)$-cliques in $G[U^*]$.

Now in order to complete the proof of the lemma, we need to define the matchings of $(r-1)$-cliques $\Pi_{i^*j^*}$ and $\Upsilon_{i^*j^*}$ and reason that the remaining conditions of the lemma are satisfied. This comes from recalling how we proved Proposition 4.1 in Section 4.1 (see also Figure 6). There, we applied Lemma 4.5 to find a large d_*-scattered K_r-diamond tree $\mathcal{D}_{\mathrm{sc}} = (T_{\mathrm{sc}}, R_{\mathrm{sc}}, \Sigma_{\mathrm{sc}})$, where $R_{\mathrm{sc}} = X \cup Y$ was the set of removable vertices of $\mathcal{D}_{\mathrm{sc}}$ and $Y \subset R_{\mathrm{sc}}$ was the set of leaves in $\mathcal{D}_{\mathrm{sc}}$. The conclusion of Proposition 4.1 then followed readily as we could choose which leaves in Y to include in a diamond subtree $\mathcal{D}$ of $\mathcal{D}_{\mathrm{sc}}$. From this proof we see that we can partition Σ_{sc} into $\Sigma_{\mathrm{sc}} =: \Pi_{i^*j^*} \cup \Upsilon_{i^*j^*}$ where the $(r-1)$-cliques $\Pi_{i^*j^*}$ are interior cliques of the K_r-diamond subtree of $\mathcal{D}_{\mathrm{sc}}$ spanned by the removable vertices $Z_{i^*j^*}$. Furthermore, we can label $\Upsilon_{i^*j^*}$ with the vertices in $Y_{i^*j^*}$ so that $(5_{i^*j^*})$ is satisfied. Indeed, each vertex v in $Y_{i^*j^*}$ corresponds to a leaf of the diamond tree $\mathcal{D}_{\mathrm{sc}}$ and so there is an interior clique $S_v \in \Sigma_{\mathrm{sc}}$ such that any diamond subtree which contains the non-leaves X of $\mathcal{D}_{\mathrm{sc}}$ can be extended by adding v to the set of removable vertices and S_v to the set of interior cliques. As $\mathcal{D}_{\mathrm{sc}}$ is a well-defined K_r-diamond tree, condition $(3_{i^*j^*})$ is also satisfied and the size constraints on $\Pi_{i^*j^*}$ and $\Upsilon_{i^*j^*}$ in $(1_{i^*j^*})$ and $(2_{i^*j^*})$ are also immediate, noting that $|\Pi_{i^*j^*}| = |Z_{i^*j^*}| - 1$ as the set of interior cliques of a diamond tree with removable vertices $Z_{i^*j^*}$. ∎

7.5. *Completing the orchard*

We will now use Lemma 7.4 to generate our orchard. This can be thought of as extending the parts of the diamond trees (the Z_{ij} and Π_{ij}) which were fixed in Lemma 7.4. The strategy is very similar to that of Lemma 7.2 and Proposition 7.3. Indeed, we take random subsets of the pools of potential vertices in order to guarantee that the K_r-hypergraph generated by our final orchard is sufficiently dense. The key difference here is that, as opposed to fixing our orchard one diamond tree at a time, we appeal to Lemma 7.4 to fix part of all the diamond trees in our orchard and then carry out the extensions on $(\ell, m)_r$-suborchards. That is, we apply the approach of Lemma 7.2 on the whole suborchard as opposed to a singular K_r-diamond tree. After doing this process for all suborchards we end up with an orchard which generates a dense K_r-hypergraph. This is detailed in the following lemma.

Lemma 7.5. *For any $r \in \mathbb{N}_{\ge 3}$, $0 < \alpha < 1/2^{12r}$ and $0 < \gamma < 1$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \le \varepsilon p^{r-1} n$, any vertex subset $U \subseteq V(G)$ with $|U| \ge n/2$ and any $k, m, t, \ell \in \mathbb{N}$ such that*

$$k = t\ell, \quad m \ge p^{r-1} n, \quad \ell m \ge p^{1-r} \quad \text{and} \quad \alpha n \le km \le 2\alpha n. \tag{7.2}$$

There exists a $(k,m)_r$-orchard $\mathcal{O}$ in G such that $V(\mathcal{O}) \subset U$ and $\mathcal{O}$ can be partitioned into suborchards $\mathcal{Q}_1, \ldots, \mathcal{Q}_t$ such that each $\mathcal{Q}_i$ with $1 \leq i \leq t$ is an $(\ell, m)_r$-orchard and we have the following property. For any $i \in [t]$, any $\mathcal{D}' \in \mathcal{O}$, any suborchard $\mathcal{Q}' \subseteq \mathcal{Q}_i$ with $|\mathcal{Q}'| \geq \ell/(4r)$ and any set of disjoint suborchards $\mathcal{O}'_1, \ldots, \mathcal{O}'_{r-2} \subset \mathcal{O}$ with $|\mathcal{O}'_{i'}| \geq pk$ for $i' \in [r-3]$ and $|\mathcal{O}'_{r-2}| \geq \gamma k$, there exists a copy of K_r traversing[13] $R_{\mathcal{D}'}$, $R(\mathcal{Q}')$ and $R(\mathcal{O}_{i'})$ for $i' \in [r-2]$.

Proof. Fix $\varepsilon > 0$ small enough to apply Corollary 3.5 with $\alpha_{3.5} = \alpha' := \gamma\alpha$, small enough to apply Lemma 7.4 with $\alpha_{7.4} = \alpha$ and small enough to force n to be sufficiently large in what follows. We begin by applying Lemma 7.4 to get vertex sets Z_{ij}, Y_{ij} and matchings of $(r-1)$-cliques Π_{ij} and Υ_{ij} for $(i, j) \in [t] \times [\ell]$ satisfying (1_{ij}) through (5_{ij}) as listed in that lemma. Now we turn to finding the diamond trees $\mathcal{D}_{ij}$ for $(i, j) \in [t] \times [\ell]$ which will form our orchard $\mathcal{O}$, so that the suborchard $\mathcal{Q}_i$ is defined to be $\mathcal{Q}_i := \{\mathcal{D}_{ij} : j \in [\ell]\}$ for each $i \in [t]$. We will appeal in particular to condition (5_{ij}) of Lemma 7.4 to find each $\mathcal{D}_{ij} = (T_{ij}, R_{ij}, \Sigma_{ij})$. In more detail, for each $(i, j) \in [t] \times [\ell]$, we will find $\tilde{Y}_{ij} \subset Y_{ij}$ and apply (5_{ij}) to find a diamond tree with removable set of vertices $R_{ij} := Z_{ij} \cup \tilde{Y}_{ij}$.

Now for a set of indices $I' \subseteq [t] \times [\ell]$, we let $Z(I') = \bigcup_{(i,j)\in I'} Z_{ij}$. In order to guarantee the key property of $\mathcal{O}$ it suffices that for each $i \in [t]$ we have the following. For any choice of $J \subseteq [\ell]$ with $|J| \geq \ell/(4r)$ and any choice of $(i_0, j_0) \in [t] \times [\ell]$ and subsets $I_1, \ldots, I_{r-2} \subset [t] \times [\ell]$ with $|I_{i'}| \geq pk$ for $i' \in [r-2]$ and $|I_{r-2}| \geq \gamma k$, the following holds. There exists a copy of K_r traversing $\bigcup_{j\in J} \tilde{Y}_{ij}$, $Z_{i_0 j_0}$ and the sets $Z(I_{i'})$ for $i' \in [r-2]$. This is what we prove in what follows as we select our sets $\tilde{Y}_{ij}$.

We work through the $i \in [t]$ in order. Let $W_0 := \bigcup_{(i,j)\in[t]\times[\ell]} (Z_{ij} \cup V(\Pi_{ij}))$ and initiate with $U_0 = U \setminus W_0$. Suppose that we are at some step $i^* \in [t]$ and we have fixed $\mathcal{D}_{ij} = (T_{ij}, R_{ij}, \Sigma_{ij})$ for all $i < i^*$. In this step, we will fix $\mathcal{D}_{i^* j}$ for all $j \in [\ell]$. We define $W_{i^*} := (\bigcup_{(i,j): i<i^*} V(\mathcal{D}_{ij})) \cup W_0$. We further define, for each $J \subseteq [\ell]$,

$$Y_J^{i^*} := \{v \in U \setminus W_{i^*} : v \in Y_{i^* j} \text{ for some } j \in [J] \text{ and } S_v \in \Upsilon_{i^* j} \cap K_{r-1}(G[U \setminus W_{i^*}])\}.$$

In words, $Y_J^{i^*}$ keeps track of the vertices v which lie in one of the $Y_{i^* j}$ with $j \in J$ which we can still use, in that the vertex v has not been used in a previous diamond tree and neither have the vertices of its associated copy of K_{r-1}, S_v. Note that $|W_{i^*}| \leq 4\alpha r n$ as a subset of vertices of a (k, m)-orchard with $km \leq 2\alpha n$. Hence if $|J| \geq \ell/(4r)$, then

$$|Y_J^{i^*}| \geq \frac{\ell}{4r}\left(\frac{\sqrt{\alpha}\, n}{\ell}\right) - 4\alpha r n \geq \left(\frac{\sqrt{\alpha}}{4r} - 4\alpha r\right) n \geq 2\alpha n,$$

by conditions (2_{ij}) and (4_{ij}) of Lemma 7.4 for $i = i^*$ and our upper bound on α.

We now define a random subset $\tilde{Y}^{i^*}$ by taking each vertex $v \in Y_{[\ell]}^{i^*}$ into $\tilde{Y}^{i^*}$ independently with probability $q := \frac{\ell m}{2\sqrt{\alpha}\, n}$, noting that $0 < q \leq 1$ since $\ell m \leq km \leq 2\alpha n \leq 2\sqrt{\alpha}\, n$.

[13] Here as before, for a diamond tree $\mathcal{D}$, $R_{\mathcal{D}}$ denotes the set of removable vertices of $\mathcal{D}$ and for a suborchard $\mathcal{O}' \subseteq \mathcal{O}$, $R(\mathcal{O}')$ denotes the union of the sets of removable vertices of diamond trees in $\mathcal{O}'$. That is, $R(\mathcal{O}') := \bigcup_{\mathcal{D}\in\mathcal{O}'} R_{\mathcal{D}}$.

For $j \in [\ell]$, we define $\tilde{Y}_{i^*j} := \tilde{Y}^{i^*} \cap Y_{i^*j}$. It follows from (2_{ij}) that $\mathbb{E}[|\tilde{Y}_{i^*j}|] = q|Y_{i^*j}| \leq m/2$ for all $j \in [\ell]$, and an application of Theorem 3.6 as well as a union bound shows that with high probability, $|\tilde{Y}_{i^*j}| \leq m$ for all $j \in [\ell]$. Note that in order to show that the upper bound on the failure probability given by Theorem 3.6 is strong enough to beat a union bound of the ℓ events, we use our lower bound on m and Fact 3.2. Furthermore, we find that with high probability, for any choice of $J \subseteq [\ell]$ with $|J| \geq \ell/(4r)$ and any choice of $(i_0, j_0) \in [t] \times [\ell]$ and subsets $I_1, \dots, I_{r-2} \subset [t] \times [\ell]$ with $|I_{i'}| \geq pk$ for $i' \in [r-2]$ and $|I_{r-2}| \geq \gamma k$, there exists a copy of K_r traversing $\bigcup_{j \in J} \tilde{Y}_{i^*j}$, $Z_{i_0 j_0}$ and the sets $Z(I_{i'})$ for $i' \in [r-2]$. Indeed, this follows from an application of Theorem 3.6 very similar to the proof of Lemma 7.2. For a fixed J, (i_0, j_0) and $I_{i'}$ for $i' \in [r-2]$ as above, there is some subset X of αn vertices of $Y_J^{i^*}$ such that each vertex in X has a copy of K_{r-1} in its neighbourhood which traverses the sets $Z_{i_0 j_0}$ and $Z(I_{i'})$ for $i' \in [r-2]$. Indeed, X can be found by repeated applications of Corollary 3.5 (2), deleting vertices from $Y_J^{i^*}$ and adding them to X on each application. Therefore $\mathbb{E}[|X \cap \tilde{Y}^{i^*}|] = q|X| = \sqrt{\alpha}\,\ell m/2$ and by Theorem 3.6, the probability that $X \cap \tilde{Y}^{i^*} = \emptyset$ is less than $e^{-\sqrt{\alpha}\,\ell m/4}$. Since $\ell m \geq m \geq p^{r-1}n \geq n^{(r-2)/(2r-3)}$ because of our lower bound on m and Fact 3.2, this probability tends to 0 as $n \to \infty$. Moreover, as there are less than $k \cdot (2^k)^{r-2} \cdot 2^\ell \leq 2^{rk}$ choices for such (i_0, j_0), $I_{i'}$ for $i' \in [r-2]$ and J, a union bound gives the traversing copies of K_r for all choices with high probability. Indeed,

$$rk \leq \frac{2\alpha r n}{m} \leq \frac{2\alpha r}{p^{r-1}} \leq 2\alpha r \ell m \leq \frac{\sqrt{\alpha}\,\ell m}{8},$$

by our conditions on km, m, ℓm and our upper bound on α from the hypotheses.

Therefore, we can fix an instance of Y^{i^*} which satisfies the desired conclusions that we have shown happen with high probability. For each $j \in [\ell]$, taking $\tilde{Y}_{i^*j} = \tilde{Y}^{i^*} \cap Y_{i^*j}$ and defining $\tilde{\Upsilon}_{i^*j} := \{S_v \in \Upsilon_{i^*j} : v \in \tilde{Y}_{i^*j}\}$, we apply condition (5_{ij}) for $i = i^*$ to get a family $\mathcal{Q}_{i^*}$ of $\mathcal{D}_{i^*j} = (T_{i^*j}, R_{i^*j}, \Sigma_{i^*j})$ for $j \in [\ell]$ such that for each j we have $R_{i^*j} = Z_{i^*j} \cup \tilde{Y}_{i^*j}$ and $\Sigma_{i^*j} = \Pi_{i^*j} \cup \tilde{\Upsilon}_{i^*j}$. This completes the step for i^* and we move to $i^* + 1$ and repeat. Doing this for each $1 \leq i^* \leq t$ completes the proof. ∎

7.6. The existence of shrinkable orchards of smaller order

With Lemma 7.5 in hand, we can now complete the proof of Proposition 5.2.

Proof of Proposition 5.2. Fix $\varepsilon > 0$ small enough to apply Lemmas 7.1 and 7.5 and Proposition 7.3 all with the same α and γ and small enough to guarantee that $p \geq Cn^{-1/(2r-3)}$ with $C := (2/\alpha)^{1/r}$ (see Fact 3.2). We also take $\varepsilon > 0$ small enough to ensure that n is sufficiently large in what follows.

Now note that Proposition 7.3 directly implies the existence of the desired shrinkable orchard if $p^{1-r} \leq p^{r-1}n$ or if $p^{r-1}n \leq p^{1-r} \leq m \leq n^{7/8}$ and so we can assume from now on that

$$p^{r-1}n \leq m \leq \min\{p^{1-r}, n^{7/8}\}. \tag{7.3}$$

We are therefore in a position (due to our lower bound on m) to apply Lemma 7.5 but we first need to fix $k, t, \ell \in \mathbb{N}$ so that conditions (7.2) are satisfied. We first fix $\ell \in \mathbb{N}$ so that $p^{1-r} \leq \ell m \leq 2p^{1-r}$. This is possible as $m \leq p^{1-r}$ and so there is a multiple of m in the desired range. Next we fix $t \in \mathbb{N}$ to be any integer such that $\alpha n \leq t\ell m \leq 2\alpha n$. Again, this is possible because $\ell m \leq 2p^{1-r} \leq \alpha n^{(r-1)/(2r-3)} \leq \alpha n$ by Fact 3.2. So there is indeed a multiple t of ℓm in the desired range. Finally, we fix $k = t\ell$ and so all the conditions in (7.2) are satisfied with our choice of parameters. Before analysing the conclusion of Lemma 7.5, let us point out a few further implications of our choices of parameters. Firstly,

$$k^{-r^3\gamma} \geq \left(\frac{1}{k}\right)^{1/2^r} \geq \left(\frac{m}{n}\right)^{1/2^r} \geq (p^{r-1})^{1/2^r} \geq p. \tag{7.4}$$

Moreover,

$$\frac{\ell}{k} \leq \frac{2}{p^{r-1}mk} \leq \frac{2}{\alpha p^{r-1}n} \leq p, \tag{7.5}$$

where we use the upper bound on ℓm in the first inequality, the lower bound on km in the second inequality and the fact that $p \geq Cn^{-1/(2r-3)} \geq Cn^{-1/r}$ from Fact 3.2 in the last inequality. Putting (7.4) and (7.5) together then gives

$$k^{1-r^3\gamma} \geq pk \geq \ell. \tag{7.6}$$

Now we apply Lemma 7.5 and let $\mathcal{O}$ be the resulting $(k, m)_r$-orchard partitioned into $(\ell, m)_r$-suborchards $\mathcal{Q}_1, \ldots, \mathcal{Q}_t$. We will show that $\mathcal{O}$ is γ-shrinkable by appealing to Lemma 7.1. Firstly, due to the upper bound of $m \leq n^{7/8}$ and the fact that $k = \Theta(n/m)$, by forcing n to be sufficiently large, we can assume that k is also sufficiently large to apply Lemma 7.1. We therefore just need to check the density condition (7.1) of the lemma. So fix $\mathcal{D} \in \mathcal{O}$ and a suborchard $\mathcal{P} \subset \mathcal{O} \setminus \{\mathcal{D}\}$ such that $\mathcal{P} \geq k/(4r)$. By a pigeonhole principle, there exists an $i \in [t]$ such that $|\mathcal{P} \cap \mathcal{Q}_i| \geq \ell/(4r)$. So fix such an i and define $\mathcal{P}^* := \mathcal{P} \cap \mathcal{Q}_i$, noting that $|\mathcal{P}^*| \leq k^{1-r^3\gamma}$ due to (7.6). Now we simply need to check that for any choice of suborchards $\mathcal{O}_1, \ldots, \mathcal{O}_{r-2} \subset \mathcal{P} \setminus \mathcal{P}^*$ with $|\mathcal{O}_{i'}| \geq k^{1-r^3\gamma}$ for $i' \in [r-3]$ and $|\mathcal{O}_{r-2}| \geq \gamma k$, there is a copy of K_r in G traversing $R_{\mathcal{D}}$, $R(\mathcal{P}^*)$ and the sets $R(\mathcal{O}_{i'})$ for $i' \in [r-2]$. This is verified by the conclusion of Lemma 7.5, setting $\mathcal{D}' = \mathcal{D}$, $\mathcal{Q}' = \mathcal{P}^*$ and $\mathcal{O}'_{i'} = \mathcal{O}_{i'}$ for $i' \in [r-2]$, noting that the lower bounds on the sizes of the $\mathcal{O}'_{i'}$ are guaranteed by (7.6). Hence $\mathcal{O}$ is indeed γ-shrinkable by Lemma 7.1 and this concludes the proof. ∎

8. The final absorption

The aim of this section is to prove Proposition 2.9 which we restate here for convenience.

Proposition 2.9 (restated). *For any $r \in \mathbb{N}_{\geq 3}$ and $0 < \alpha, \eta < 1/2^{3r}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1}n$ and any vertex subset $W \subseteq V(G)$ with $|W| \geq n/2$.*

There exist vertex subsets $A, B \subset V$ such that $A \subset W$, $|A| \leq \alpha n$, $|B| \leq \eta p^{2r-4} n$ and for any $(k, m)_r$-orchard $\mathcal{R}$ whose vertices lie in $V(G) \setminus (A \cup B)$, if $|A| + |V(\mathcal{R})| \in r\mathbb{N}$, $k \leq \alpha^2 n^{1/8}$ and $m \geq n^{7/8}$ then $G[A \cup V(\mathcal{R})]$ has a K_r-factor.

In order to prove this, in Section 8.1 we first define an absorbing structure whose vertex set will play the role of A in Proposition 2.9. We then prove that it has the required absorbing property. Next, in Section 8.2, we prove that we can find the absorbing structure in a suitably pseudorandom graph and show that this implies Proposition 2.9.

8.1. *Defining an absorbing structure*

Recall from Section 3.6 the definition of a template and the fact that templates of flexibility t with maximum degree 40 exist for all large enough t (Theorem 3.14). We will use a template as an auxiliary graph to define an *absorbing structure* which can contribute to a K_r-factor in many ways.

Definition 8.1. Let $\mathcal{T}$ be a template with flexibility t on vertex sets I and $J := J_1 \cup J_2$ with $|I| = 3t$ and $|J_1| = |J_2| = 2t$. A *K_r-absorbing structure* $\mathbb{A}$ of *order M* with respect to $\mathcal{T}$ in G consists of a labelled matching of $(r-1)$-cliques in G, $\Xi(\mathbb{A}) := \{S_i : i \in I\} \subset K_{r-1}(G)$ and a labelled $(4t, M)_r$-orchard $\mathcal{J}(\mathbb{A}) := \{\mathcal{D}_j : j \in J\}$ such that the following holds for each $i \in I$ and $j \in J$:

- $S_i \cap V(\mathcal{D}_j) = \emptyset$;
- if $ij \in E(\mathcal{T})$ then there is a vertex in the removable set R_j of vertices of $\mathcal{D}_j$ which forms a K_r with S_i in G.

We say that $\mathbb{A}$ has *flexibility* t, which is inherited from the template by which $\mathbb{A}$ is defined. We refer to the vertices of the absorbing structure, denoted $V(\mathbb{A})$, which is all vertices which feature in cliques in $\Xi(\mathbb{A})$ or diamond trees in $\mathcal{J}(\mathbb{A})$.

See Figure 7 for an example of an absorbing structure. Note that a K_r-absorbing structure $\mathbb{A}$ of flexibility t and order M has less than

$$3t(r-1) + 4t((2M-1)r + 1) \leq 8rtM - rt + t \leq 8rtM \tag{8.1}$$

vertices. The absorbing structure is defined in such a way that it inherits the robust property that the template has with respect to perfect matchings but has such a property with respect to K_r-factors. The following lemma demonstrates this and reduces Proposition 2.9 to finding an appropriate absorbing structure in G.

Lemma 8.2. *For any $r \in \mathbb{N}_{\geq 3}$ and $0 < \zeta, \eta < 1/2^{2r}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$. Suppose that $t, M \in \mathbb{N}$ are such that $tM \geq \zeta n$ and there exists a K_r-absorbing structure $\mathbb{A}$ of flexibility t (with respect to some template $\mathcal{T}$) and order M in G. Let $A := V(\mathbb{A})$.*

Then there exists some vertex subset $B \subset V(G)$ such that $|B| \leq \eta p^{2r-4} n$ and for any $(k, m)_r$-orchard $\mathcal{R}$ whose vertices lie in $V(G) \setminus (A \cup B)$, if $|A| + |V(\mathcal{R})| \in r\mathbb{N}$, $k \leq \frac{t}{4r}$ and $m \geq M$ then $G[A \cup V(\mathcal{R})]$ has a K_r-factor.

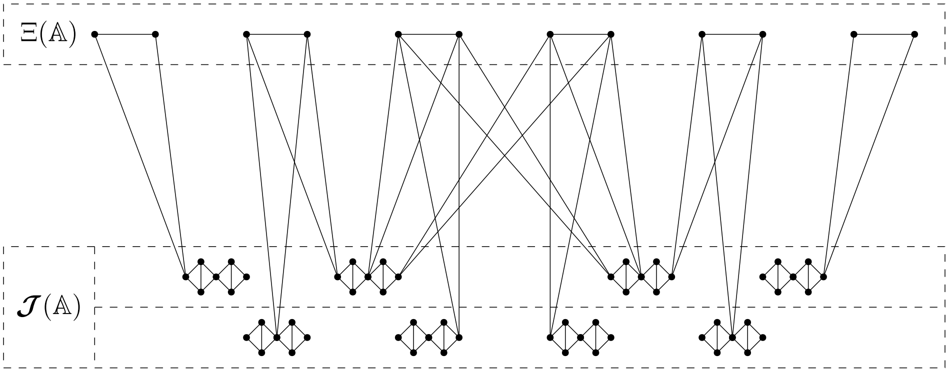

Fig. 7. A K_3-absorbing structure of order 3 and flexibility 2, whose defining template is the template $\mathcal{T}$ displayed in Figure 5.

Proof. Fix $\varepsilon > 0$ small enough to apply Lemma 2.4 with ζ and η as defined here and small enough to apply Corollary 3.5 with $\alpha_{3.5} = \alpha := \zeta/(r+1)$. Let $\mathcal{O} = \{\mathcal{D}_j : j \in J_2\}$ be the suborchard of $\mathcal{J}(\mathbb{A})$ defined by those indices which lie in the flexible set J_2 of the template $\mathcal{T}$ which defines $\mathbb{A}$. Thus $\mathcal{O}$ is a $(2t, M)_r$-orchard. Therefore, by Lemma 2.4, there exists a set $B \subset V(G)$ with $|B| \le \eta p^{2r-4} n$ and for any $(k, m)_r$-orchard $\mathcal{R}$ as in the statement of the lemma, $\mathcal{O}$ absorbs $\mathcal{R}$. Indeed, note that in the notation of Lemma 2.4, k, m and M are as defined here, while $K = 2t$. Hence the condition $k \le K/(8r)$ is precisely the same as our presumption that $k \le t/(4r)$, whilst the condition $kM \le mK$ is guaranteed by the fact that $m \ge M$ and $k \le K/(8r)$. Unpacking the conclusion of Lemma 2.4, we thus find that for any such $\mathcal{R}$ there exists some subfamily $\mathcal{P}_1 \subseteq \mathcal{O}$ such that $|\mathcal{P}_1| = (r-1)k \le t/2$ and $G[V(\mathcal{P}_1) \cup V(\mathcal{R})]$ has a K_r-factor. We will show that $G[A \setminus V(\mathcal{P}_1)]$ also has a K_r-factor, which will complete the proof.

Now note that for any $\mathcal{P} \subset \mathcal{O}$ such that $|\mathcal{P}| = t$, $G[A \setminus V(\mathcal{P})]$ has a K_r-factor. Indeed, let $\bar{J} := \{j \in J_2 : \mathcal{D}_j \in \mathcal{P}\}$ be the indices of diamond trees that feature in $\mathcal{P}$. By the definition of the template $\mathcal{T}$, Definition 3.13, there is a perfect matching $F \subset E(\mathcal{T})$ in $\mathcal{T}[V(\mathcal{T}) \setminus \bar{J}]$. Now for $ij \in F$, we can take a K_r-factor on $S_i \cup V(\mathcal{D}_j)$ in G guaranteed by the fact that S_i forms a copy of K_r with a removable vertex of $\mathcal{D}_j$ (Definition 8.1) and the key property of the removable vertices of a K_r-diamond tree (Observation 2.2). As F is a perfect matching in $\mathcal{T}[V(\mathcal{T}) \setminus \bar{J}]$, we see that by taking these K_r-factors for each $ij \in F$, we obtain a K_r-factor in $G[A \setminus V(\mathcal{P})]$ as required.

If we had $|\mathcal{P}_1| = t$, this would complete the proof. However, $\mathcal{P}_1$ is actually much smaller. Indeed, $|\mathcal{P}_1| \le t/2$. We will proceed by finding some $\mathcal{P}_2 \subset \mathcal{O} \setminus \mathcal{P}_1$ such that $G[V(\mathcal{P}_2)]$ has a K_r-factor and $|\mathcal{P}_1| + |\mathcal{P}_2| = t$. We build $\mathcal{P}_2$ by the following greedy process. We initiate with $\mathcal{O}' = \mathcal{O} \setminus \mathcal{P}_1$ and $\mathcal{P}_2 = \emptyset$. Then at each time step, as long as $|\mathcal{P}_1| + |\mathcal{P}_2| + r \le t$ we partition $\mathcal{O}'$ into r parts $\mathcal{O}_1, \ldots, \mathcal{O}_r$ of sizes of the parts are as equal as possible. We let R_x be the union of the removable vertices of diamond trees in the orchard $\mathcal{O}_x$ for $x \in [r]$. Each R_x has size at least $tM/(r+1) \ge \zeta n/(r+1) = \alpha n$. Therefore, by Corollary 3.5 (2), there is a copy of K_r traversing the R_x, $x \in [r]$. This gives some r-tuple of diamond trees $\mathcal{D}_1, \dots, \mathcal{D}_r$ such that $\mathcal{D}_x \in \mathcal{O}_x$ for all $x \in [r]$ and there is a K_r-factor in $G[V(\mathcal{D}_1) \cup \dots \cup V(\mathcal{D}_r)]$, given by taking the copy of K_r that traverses their sets of removable vertices and applying Observation 2.2. We add $\mathcal{D}_1, \dots, \mathcal{D}_r$ to $\mathcal{P}_2$ and delete these diamond trees from the orchard $\mathcal{O}'$, which completes this time step. Clearly at all points in this process there is a K_r-factor in $G[V(\mathcal{P}_2)]$ and we claim that this process terminates when $|\mathcal{P}_1| + |\mathcal{P}_2|$ is exactly equal to t. Indeed, if this is not the case, as we increase the size of $|\mathcal{P}_2|$ by exactly r in each step, we have $|\mathcal{P}_1| + |\mathcal{P}_2| = t - s$ for some $s \in [r-1]$ at the end of the process. Let $\mathcal{P}_3 \subset \mathcal{O} \setminus (\mathcal{P}_1 \cup \mathcal{P}_2)$ be a set of s K_r-diamond trees. Now as $V_1 := V(\mathcal{R}) \cup V(\mathcal{P}_1) \cup V(\mathcal{P}_2)$ hosts a K_r-factor, it follows that $r \mid |V_1|$. Likewise, we know that $V_2 = A \setminus \bigcup_{i=1}^{3} V(\mathcal{P}_i)$ hosts a K_r-factor and so $r \mid |V_2|$. Since r divides the size of $A \cup V(\mathcal{R})$ and $A \cup V(\mathcal{R}) = V_1 \cup V_2 \cup V(\mathcal{P}_3)$, we can infer that $r \mid |V(\mathcal{P}_3)|$. This is a contradiction as $\mathcal{P}_3$ is a set of s K_r-diamond trees for some $1 \le s \le r-1$ and the number of vertices in any diamond tree is 1 mod r. Therefore we can find a $\mathcal{P}_2$ as claimed.

Finally, taking $\mathcal{P} := \mathcal{P}_1 \cup \mathcal{P}_2$, we are done by taking our K_r-factor in $G[A \cup V(\mathcal{R})]$ to be the union of the K_r-factors in $G[V(\mathcal{R}) \cup V(\mathcal{P}_1)]$, in $G[V(\mathcal{P}_2)]$ and in $G[A \setminus V(\mathcal{P})]$. ∎

8.2. *Finding an absorbing structure*

Lemma 8.2 reduces Proposition 2.9 to finding an appropriate absorbing structure in G. In this section we prove that this is possible by proving the following proposition.

Proposition 8.3. *For any $r \in \mathbb{N}_{\ge 3}$ and $0 < \alpha < 1/2^{3r}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \le \varepsilon p^{r-1} n$ and any vertex subset $W \subseteq V(G)$ with $|W| \ge n/2$. There exists a K_r-absorbing structure $\mathbb{A}$ in G of flexibility $t = \alpha n^{1/8}$ and order $M = n^{7/8}$ such that $V(\mathbb{A}) \subseteq W$.*

With Lemma 8.2 and Proposition 8.3, the proof of Proposition 2.9 follows readily:

Proof of Proposition 2.9. Fix $\zeta := \alpha/(8r)$ and $\varepsilon > 0$ small enough to apply Lemma 8.2 with ζ and η as defined here and small enough to apply Proposition 8.3 with $\alpha_{8.3} = \zeta$. We can therefore apply Proposition 8.3 to get an absorbing structure $\mathbb{A}$ in G with flexibility $t = \zeta n^{1/8}$ and order $M = n^{7/8}$. The set $A = V(\mathbb{A}) \subset W$ has size $|A| \le 8rtM = \alpha n$ (see (8.1)). The conclusion then follows from Lemma 8.2 noting that $k \le \alpha^2 n^{1/8}$ implies $k \le t/(4r) = \zeta n^{1/8}/(4r) = \alpha n^{1/8}/(48r^2)$ due to our upper bound on α. ∎

Now in order to prove the existence of an absorbing structure as in Proposition 8.3, we will first fix some template $\mathcal{T}$ which will define $\mathbb{A}$. Next, we will set aside a large matching $\Pi \subset K_{r-1}(G)$ of $(r-1)$-cliques. These will be candidates for the matching of $(r-1)$-cliques $\Xi(\mathbb{A})$ in our absorbing structure but we start with a much bigger set Π of size $\Omega(n^{2/3})$. Moreover, to each $(r-1)$-clique $S \in \Pi$ we will associate a set of vertices $X_S \subset V(G)$ such that $X_S \subset N^G(S)$, $|X_S| = \Omega(n^{1/3})$ and, crucially, the sets $\{X_S : S \in \Pi\}$ are *disjoint*. We will find this matching Π of $(r-1)$-cliques with a simple greedy procedure, appealing to Corollary 3.5 (3) to find each $S \in \Pi$ (and the corresponding neighbourhood set X_S), one by one.

After finding Π, we then turn to constructing the $(4t, M)$-orchard $\mathcal{J}(\mathbb{A})$ for the absorbing structure $\mathbb{A}$. Again, this will be done greedily, fixing the diamond trees $\mathcal{D} \in \mathcal{J}(\mathbb{A})$ one at a time. Let us consider fixing some diamond tree $\mathcal{D}_j \in \mathcal{J}(\mathbb{A})$. Note that as we fix $\mathcal{D}_j$, we immediately get restrictions on which $S \in \Pi$ remain as candidates to play the rôle of certain $S_i \in \Xi(\mathbb{A})$. Indeed, if the removable vertices of $\mathcal{D}_j$ are disjoint from $N^G(S)$ and ij is an edge in the template $\mathcal{T}$ defining $\mathbb{A}$, then there is no way S can play the role of S_i in $\Xi(\mathbb{A})$. Therefore as we fix our diamond trees, we will aim to have their sets of removable vertices intersect as many of the X_S (and hence neighbourhoods $N^G(S)$) for $S \in \Pi$ as possible.

In order to do this, we will use the following lemma, which shows that we can find diamond trees whose removable vertices intersect many prescribed sets (in our case this will be the sets X_S). The proof of this lemma is a simple application of Proposition 4.1.

Lemma 8.4. *For any $r \in \mathbb{N}_{\geq 3}$ and $0 < \alpha < 1/2^{2r}$, there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$.*

Suppose $\frac{\alpha^2}{2} n^{2/3} \leq \ell \leq \alpha n^{2/3}$ and we have disjoint vertex subsets $W, U_1, \ldots, U_\ell$ such that $|W| \geq n/4$ and $|U_i| \geq n^{1/3}$ for all $i \in [\ell]$. Then there exists a diamond tree $\mathcal{D} = (T, R, \Sigma)$ in G such that

- (i) *$\Sigma \subset K_{r-1}(G[W])$ is a matching of $(r-1)$-cliques in W;*
- (ii) *$R \subset \bigcup_{i=1}^{\ell} U_i$ and R intersects ℓ' of the sets U_i for some $\ell' \geq \ell/(4r)$;*
- (iii) *the order of $\mathcal{D}$ is at most $n^{2/3}$;*
- (iv) *for all but at most $n^{1/2}$ of the indices $i \in [\ell]$, we have $|V(\mathcal{D}) \cap U_i| \leq n^{1/6}$.*

Proof. We begin by fixing $\gamma := \ell/n^{2/3}$ so that $\alpha^2/2 \leq \gamma \leq \alpha$ and we fix ε small enough to apply Proposition 4.1 with $\alpha_{4.1} = \alpha' := \gamma/(4r)$ and small enough to guarantee that $p \geq C n^{-1/(2r-3)}$ with $C = 4/\alpha'$ (see Fact 3.2). Now shrink each set U_i so that it has exactly $n^{1/3}$ vertices and define $U := \bigcup_{i=1}^{\ell} U_i$. Furthermore, fix $d_* := \alpha'^2 p^{r-1} n$ and apply Proposition 4.1 with U, W and $z = \alpha' n$. So we get disjoint subsets $X, Y \subset U$ as in the outcome of Proposition 4.1.

Now firstly note that as $|X| + |Y| = z = \alpha' n$, $|U| = \ell n^{1/3} = \gamma n = 4rz$ and the U_i have equal size, $X \cup Y$ must intersect at least $\ell/(4r)$ of the sets U_i. We will choose our $\mathcal{D} = (T, R, \Sigma)$ so that R intersects all the sets U_i that $X \cup Y$ intersects, thus guaranteeing condition (ii). Indeed, if we let $Y' \subset Y$ be the *minimal* subset of Y such that there exists no $i \in [\ell]$ with $Y \cap U_i \neq \emptyset$ and $Y' \cap U_i = \emptyset$, Proposition 4.1 gives the existence of a diamond tree $\mathcal{D} = (T, R, \Sigma)$ such that $R = X \cup Y'$ and $\Sigma \subset K_{r-1}(G[W])$ and so conditions (i) and (ii) are satisfied.

In order to establish condition (iii), note that $|Y'| \leq \ell \leq n^{2/3}/2$ and if $|X| > 1$ then

$$|X| \leq \frac{2z}{d_*} \leq \frac{2}{\alpha' p^{r-1}} \leq \frac{2n^{(r-1)/(2r-3)}}{\alpha' C^{r-1}} \leq \frac{n^{2/3}}{2},$$

due to our definition of C and the fact that $\frac{r-1}{2r-3} \leq \frac{2}{3}$ for all $r \geq 3$. Finally, (iv) is a simple consequence of (iii). Indeed, if (iv) were not true, then as the U_i are pairwise disjoint, $\mathcal{D}$ would have order greater than $n^{1/2} \cdot n^{1/6} \geq n^{2/3}$, a contradiction. ■

Let us return to sketching the proof of Proposition 8.3, considering now that we can use Lemma 8.4 to find diamond trees $\mathcal{D} \in \mathcal{J}(\mathbb{A})$. As discussed above, the key property of diamond trees generated by Lemma 8.4 is (ii), allowing us to find diamond trees that intersect many of the sets $\{X_S : S \in \Pi\}$ which we begin the proof with. Property (iv) will also be useful as it shows that in the process of building $\mathcal{J}(\mathbb{A})$ one by one, we do not destroy many of the sets X_S and most of them remain large and can be used by other $\mathcal{D} \in \mathcal{J}(\mathbb{A})$.

One potentially troublesome consequence of Lemma 8.4 is that the diamond trees it finds are far too small; see (iii). Indeed, the diamond trees in our orchard $\mathcal{J}(\mathbb{A})$ are supposed to be of order $M = n^{7/8}$. It turns out that this is not such a big hurdle as we can find a large diamond tree disjoint from all the X_S and connect it to the diamond tree $\mathcal{C}$ output by Lemma 8.4. In more detail, we can apply Proposition 4.1 to create a large (linear) pool Y of vertices that can be removable vertices of some diamond tree which will be disjoint from all the vertices in the sets X_S. We also consider the large (linear) pool Z of vertices that lie in some $X_S \setminus V(\mathcal{C})$ with $S \in \Pi$ such that the removable vertices of $\mathcal{C}$ intersect X_S. It is not hard to show (see for example Corollary 3.5 (3)) that there is a copy of K_{r+1}^- with one degree $r-1$ vertex in Y and the other in $X_{S^*} \subset Z$ for some $S^* \in \Pi$. By also taking S^* into $\mathcal{D}$ and choosing an appropriate $Y' \subset Y$ to apply the key property of Proposition 4.1, we can obtain a diamond tree $\mathcal{D}$ of the correct size that contains the diamond tree $\mathcal{C}$ output by Lemma 8.4.

More troublesome is the fact that condition (ii), which means that the removable vertices of $\mathcal{C}$ intersect many of the desired sets X_S, is, in fact, not strong enough. Indeed, consider some fixed $i \in I$ for which we want to find a copy S_i of K_{r-1} to lie in $\Xi(\mathbb{A})$. If $j, j' \in N^{\mathcal{T}}(i)$ and the sets $\{X_S : S \in \Pi,\ R_{\mathcal{D}_j} \cap X_S \neq \emptyset\}$ and $\{X_S : S \in \Pi,\ R_{\mathcal{D}_{j'}} \cap X_S \neq \emptyset\}$ are disjoint (here, as usual, we use $R_{\mathcal{D}}$ to denote the removable vertices of $\mathcal{D}$), then already there are no candidates for S_i in Π. To fix this, we actually need that when we choose a diamond tree $\mathcal{D} \in \mathcal{J}(\mathbb{A})$, the set $R_{\mathcal{D}}$ intersects *almost all* of the sets $\{X_S : S \in \Pi\}$. We achieve this by iterating Lemma 8.4, creating constantly many disjoint diamond trees $\mathcal{C}$ that together hit almost all of the X_S with their removable vertices. We then connect all of these diamond trees $\mathcal{C}$ with a large diamond tree disjoint from the sets X_S to obtain the desired diamond tree $\mathcal{D} \in \mathcal{J}(\mathbb{A})$. This connecting process is similar to (although slightly more involved than) the connecting process outlined in the previous paragraph. We now give the full details for the proof of Proposition 8.3, concluding this section and chapter.

Proof of Proposition 8.3. We begin by fixing $\varepsilon > 0$ small enough to apply Corollary 3.5 with $\alpha_{3.5} = \alpha' := \alpha^2/(16r)$ and to apply Proposition 4.1 and Lemma 8.4 each with $\alpha_{4.1} = \alpha_{8.4} = \alpha$. We also take ε small enough to force n to be sufficiently large and to guarantee that $p \geq C'n^{-1/(2r-3)}$ for $C' := 2/\alpha'^2$ using Fact 3.2. We further fix some template $\mathcal{T}$

with vertex sets I and $J = J_1 \cup J_2$ of flexibility t and maximum degree 40 which we know exists for n (and hence t) sufficiently large by Theorem 3.14 of Montgomery [57].

We will find an absorbing structure with respect to $\mathcal{T}$ and so must prove the existence of a matching $\Xi(\mathbb{A}) = \{S_i : i \in I\} \subset K_{r-1}(G[W])$ of $3t$ copies of K_{r-1}, and a $(4t, M)$-orchard $\mathcal{J} = \mathcal{J}(\mathbb{A}) = \{\mathcal{D}_j : j \in J\}$ such that the conditions of Definition 8.1 are satisfied. We will do this in three stages. In Claim 8.5, we fix some large matching $\Pi \subset K_{r-1}(G[W])$ of $(r-1)$-cliques which will be candidates for the $(r-1)$-cliques which will feature in $\Xi(\mathbb{A})$. We will guarantee that the cliques in Π are contained in many copies of K_r which will help as we proceed to build our absorbing structure. In Claim 8.6, we will fix the K_r-diamond trees which will form our orchard $\mathcal{J}$ for our K_r-absorbing structure. We will carefully control how these diamond trees intersect the cliques in our candidate set Π and their neighbourhoods. Finally, we will show that we can find a suitable $\Xi(\mathbb{A}) \subset \Pi$ so that we obtain the desired absorbing structure.

Claim 8.5. *There exists a matching $\Pi = \{S_1, \ldots, S_\ell\} \subset K_{r-1}(G[W])$ of $\ell := \alpha n^{2/3}$ copies of K_{r-1} and sets $X_h \subset W \setminus V(\Pi)$ for each $h \in [\ell]$ such that the X_h are pairwise disjoint, each has size $2n^{1/3}$ and for all $h \in [\ell]$ we have $X_h \subset N^G_W(S_h)$.*

Proof of Claim. We can do this by way of a simple greedy process choosing such an $(r-1)$-clique S_h and set X_h in order for $h = 1, \ldots, \ell$. When choosing S_h and X_h, we look at the set of vertices $V_h \subset W$ which have not been used in previous choices of $S_{h'}$ or $X_{h'}$. We have

$$|V_h| \geq |W| - \Big|\bigcup_{h'<h} (X_{h'} \cup S_{h'})\Big| \geq n/2 - (\ell-1)(r-1+2n^{1/3}) \geq (1/2-2\alpha)n \geq n/4,$$

and an application of Corollary 3.5 (3) with $W_0 = W_1 = W_2 = V_h$ gives the desired S_h and X_h in V_h since

$$\alpha'^2 p^{r-1} n \geq \alpha'^2 C' n^{1-(r-1)/(2r-3)} \geq 2n^{1/3}$$

due to Fact 3.2. ■

Next we turn to fixing our $(4t, M)$-orchard $\mathcal{J}$.

Claim 8.6. *Let S_h and X_h for $h = 1, \ldots, \ell$ be as in Claim 8.5. Then there exists a $(4t, M)$-orchard $\mathcal{J} = \{\mathcal{D}_1, \ldots, \mathcal{D}_{4t}\}$ such that $V(\mathcal{J}) \subset W$ and the following properties hold for each $\mathcal{D}_j = (T_j, R_j, \Sigma_j)$ with $j \in [4t]$:*

(1) *the set R_j of removable vertices intersects at least $(1-\alpha)\ell$ of the sets X_h with $h \in [\ell]$;*

(2) *$V(\mathcal{D}_j)$ intersects at most $C := \frac{\log(\frac{2}{\alpha})}{\log(\frac{4r}{4r-1})}$ of the S_h with $h \in [\ell]$.*

Before proving the claim, let us see how it implies the proposition. Indeed, taking the $(4t, M)_r$ orchard $\mathcal{J}$ from Claim 8.6 as $\mathcal{J}(\mathbb{A})$, we just need to choose a matching of $(r-1)$-cliques $\Xi(\mathbb{A}) = \{S_i : i \in I\}$ so that $S_i \cap V(\mathcal{J}) = \emptyset$ for all $i \in I$ and whenever $ij \in E(\mathcal{T})$, there is a vertex in R_j which forms a copy of K_r with S_i. We do this greedily, showing that for each $i = 1, \ldots, 3t$ in order, there is a suitable choice for S_i in Π. We initiate by fixing $L \subseteq [\ell]$ to be the indices $h \in [\ell]$ such that $S_h \cap V(\mathcal{J}) = \emptyset$. By condition (2) in Claim 8.6, for large n we have

$$|L| \geq \ell - 4Ct \geq (1-\alpha)\ell$$

at the beginning of this process, recalling that $\ell = \alpha n^{2/3}$ and $t = \alpha n^{1/8}$. Now for $i = 1, \ldots, 3t$, we find an index $h = h(i) \in L$ such that S_h forms a copy of K_r with a vertex in R_j for all j such that $ij \in E(\mathcal{T})$. We fix $S_i = S_h$ and delete h from L. If this process succeeds in finding a suitable $h = h(i)$ for each $i \in I$ then the resulting $\Xi(\mathbb{A}) = \{S_i : i \in I\}$ along with $\mathcal{J}$ form the desired K_r-absorbing structure.

It remains to check that we are successful at each step. So consider step $i^* \in [3t]$. We have that $|L| \geq (1-\alpha)\ell - (i^* - 1) \geq (1-2\alpha)\ell$ at the beginning of the step. Now for each $j \in J$ which is a neighbour of i^* in the template $\mathcal{T}$, by Claim 8.6 (1) there are at most $\alpha\ell$ indices $h \in [\ell]$ such that no vertex of R_j forms a K_r with S_h in G. Indeed, for almost all choices of h, we have $R_j \cap X_h \neq \emptyset$ and $X_h \subset N^G(S_h)$. Given that $\mathcal{T}$ has maximum degree 40, this gives at most $40\alpha\ell$ indices $h \in L$ that would not be a good choice for $h(i^*)$. Therefore there are at least $(1-42\alpha)\ell$ indices $h \in L$ which can be chosen as $h(i^*)$ and we simply choose one arbitrarily.

This shows that the algorithm is successful in generating the desired absorbing structure and so it only remains to prove Claim 8.6, which we do now.

Proof of Claim 8.6. We will find the diamond trees $\mathcal{D}_j$, $j = 1, \ldots, 4t$, one by one so that they are vertex disjoint and satisfy the two conditions in the statement of the claim as well as the further following condition:

(3) $V(\mathcal{D}_j)$ intersects all but at most $Cn^{1/2}$ of the X_h with $h \in [\ell]$ in more than $2Cn^{1/6}$ vertices.

We will initiate the process with $\Lambda := [\ell]$ and $U_h = X_h$ for all $h \in [\ell]$. These sets U_h will keep track of vertices in X_h that we are still allowed to use, that is, those vertices which have not been used in previously chosen diamond trees. Furthermore, the set $\Lambda \subset [\ell]$ will keep track of all indices which are *alive*. When we choose a $\mathcal{D}_j$ for some $j \in [4t]$, we *kill* (and remove from Λ) all the indices $h \in [\ell]$ such that $V(\mathcal{D}_j)$ intersects X_h in more than $2Cn^{1/6}$ vertices. We also kill any index h such that $V(\mathcal{D}_j)$ intersects S_h. Due to our conditions (2) and (3), throughout the process we have

$$|\Lambda| \geq \ell - 4t(C + Cn^{1/2}) \geq (1-\alpha/2)\ell$$

for n large, recalling that $\ell = \alpha n^{2/3}$ and $t = \alpha n^{1/8}$. Moreover, due to condition (3), at any point in the process, for all alive indices h in Λ, the size of $U_h \subset X_h$ is at least

$$|U_h| \geq |X_h| - \sum_j |V(\mathcal{D}_j) \cap X_h| \geq 2n^{1/3} - 8tCn^{1/6} \geq n^{1/3}$$

for n sufficiently large. We remark that it is crucial in the previous two calculations that $t = n^{1/8}$ and so when choosing our diamond trees, we do not kill too many indices or make too many of the sets X_h too small to be used by subsequent diamond trees. In fact, any t polynomially smaller than $n^{1/6}$ would suffice for this.

So let us suppose that we are at step $j^* \in [4t]$ where we look for $\mathcal{D}_{j^*}$ and we have some fixed set Λ of alive indices and subsets $U_h \subset X_h$ for $h \in \Lambda$. We run a subalgorithm that finds $\mathcal{D}_{j^*}$ in two phases. We begin by setting $\Gamma = \Lambda$ and $\boldsymbol{\mathcal{C}} = \emptyset$. The first phase of the subalgorithm works by finding at most C small order diamond trees whose removable vertices intersect many of the U_h for $h \in \Lambda$. The family $\boldsymbol{\mathcal{C}}$ will collect these small order diamond trees and the set Γ will keep track of the indices h in Λ for which we have not yet intersected U_h. In the second phase of the algorithm, we will form $\mathcal{D}_{j^*}$ by joining together the diamond trees in $\boldsymbol{\mathcal{C}}$ so that they form one diamond tree. By guaranteeing that our diamond trees in $\boldsymbol{\mathcal{C}}$ have removable vertices that intersect most of the sets U_h, we will guarantee condition (1) of the claim. Before starting, we also initiate by setting $W' \subset W$ to be

$$W' = W \setminus \Big(\bigcup_{h \in [\ell]} (S_h \cup X_h) \cup \bigcup_{j < j^*} V(\mathcal{D}_j) \Big).$$

In words, W' is the subset of vertices of W that has not been used in any of the structures that we have found so far. Finally, we initiate a counter by setting $s = 1$.

At step s, we apply Lemma 8.4 on the sets W' and $\{U_h : h \in \Gamma\}$. We thus find a K_r-diamond tree $\mathcal{C}_s = (T, R, \Sigma)$ which we add to $\boldsymbol{\mathcal{C}}$, which has the following properties guaranteed by Lemma 8.4:

(i) $\Sigma \subset K_{r-1}(G[W'])$ and we delete $V(\Sigma)$ from W';

(ii) $R \subset \bigcup_{h \in \Gamma} U_h$ and defining $\Gamma_s \subset \Gamma$ to be $\Gamma_s := \{h' : R \cap U_{h'} \neq \emptyset\}$, we have $|\Gamma_s| \geq |\Gamma|/(4r)$; we delete Γ_s from Γ;

(iii) the order of $\mathcal{C}_s$ is at most $n^{2/3}$;

(iv) there is a set $\Phi_s \subset \Gamma_s \subset \Gamma \subset [\ell]$ of at most $n^{1/2}$ indices such that for all $h \in [\ell] \setminus \Phi_s$ we have $|V(\mathcal{C}_s) \cap U_h| \leq n^{1/6}$.

Finding such a $\mathcal{C}_s$ concludes step s. If $|\Gamma| < \alpha\ell/2$, we terminate this phase and move on to the next phase. If $|\Gamma| \geq \alpha\ell/2$, we move to step $s + 1$.

We must check that the conditions for Lemma 8.4 are satisfied throughout this phase in order to find the required diamond trees $\mathcal{C}_s$ at each step. Indeed, this follows because

$$\frac{\alpha^2}{2} n^{2/3} = \frac{\alpha}{2}\ell \leq |\Gamma| \leq \ell = \alpha n^{2/3}$$

throughout, and $|U_h| \geq n^{1/3}$ for all $h \in \Gamma$ since $\Gamma \subset \Lambda$ is a subset of alive indices. Finally, $|W'| \geq n/4$ throughout this process. Indeed, note that due to condition (ii) and the fact that we only continue until $|\Gamma| \leq \alpha\ell/2$, the process runs for a maximum of C steps, recalling the definition of C from condition (2) of the claim. That is, $|\boldsymbol{\mathcal{C}}| \leq C$ throughout and so

$$\begin{aligned} |W'| &\geq |W| - \sum_{h \in [\ell]} (|S_h| + |X_h|) - \sum_{j < j^*} |V(\mathcal{D}_j)| - \sum_{\mathcal{C} \in \boldsymbol{\mathcal{C}}} V(\mathcal{C}) \\ &\geq \frac{n}{2} - \ell \cdot 3n^{1/3} - 8trM - Cn^{2/3} \geq \Big(\frac{1}{2} - (4 + 8r)\alpha \Big) n \geq \frac{n}{4}, \end{aligned} \tag{8.2}$$

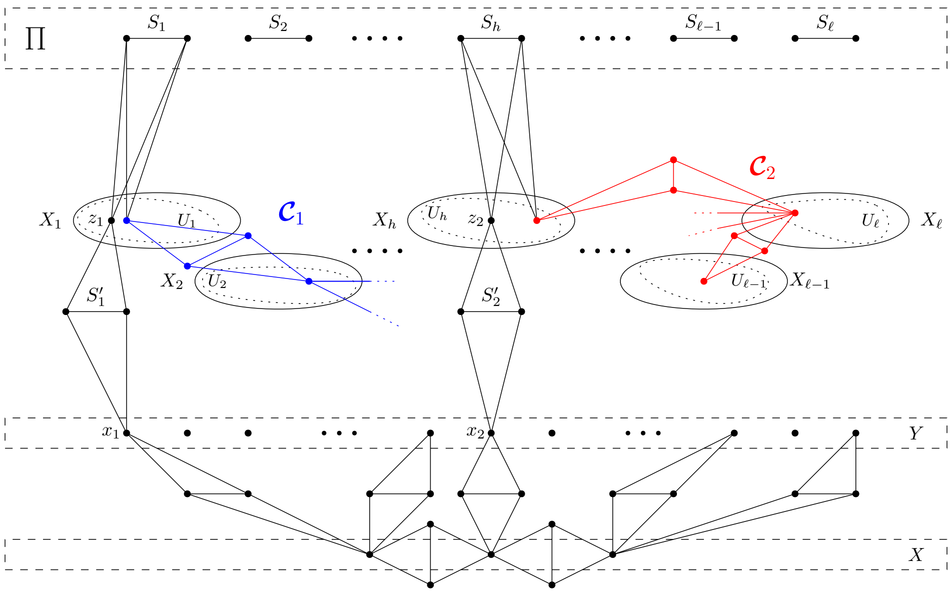

Fig. 8. An example of $\mathcal{D}_{j^*}$ and its components. In this case, we have $c = 2$, $h_1 = 1$ and $h_2 = h$.

due to our upper bound on α, for n sufficiently large. This verifies that we find $\mathcal{C}_s$ at every step s of this process and so we finish this phase with $|\Gamma| < \alpha\ell/2$ and some family $\boldsymbol{\mathcal{C}} = \{\mathcal{C}_1, \dots, \mathcal{C}_c\}$ of $c \le C$ vertex disjoint K_r-diamond trees.

Now we describe how we generate $\mathcal{D}_{j^*}$ which will have all the diamond trees $\mathcal{C}_s \in \boldsymbol{\mathcal{C}}$ as sub-diamond-trees. We refer the reader to Figure 8 to keep track of the many components that contribute to our $\mathcal{D}_{j^*}$. One thing to note is that the sum of the orders of the diamond trees in $\boldsymbol{\mathcal{C}}$ is far too small for us to just build $\mathcal{D}_{j^*}$ from the diamond trees in $\boldsymbol{\mathcal{C}}$. Indeed, the sum of the orders is $O(n^{2/3})$ and we want $\mathcal{D}_{j^*}$ to have order $M = n^{7/8}$. Therefore we will have to find the majority of the K_r-diamond tree $\mathcal{D}_{j^*}$ elsewhere. In order to prepare for this, we first split W' arbitrarily into U_0, W_0 and Z_0 of roughly equal size and note that due to our lower bound (8.2) on $|W'|$, each of these sets has size at least $n/16$. Next we fix $d_* := \alpha^2 p^{r-1} n$ and $z = \alpha^2 n$ and apply Proposition 4.1 with respect to the sets U_0 and W_0 to get disjoint sets $X, Y \subset U_0$ as detailed there. Note that $|X| \le 2n^{2/3}$. Indeed, if $|X| > 1$, then $|X| \le 2z/d_* \le 2p^{1-r} \le 2n^{2/3}$ due to Fact 3.2.

Now for $1 \le s \le c$, define $Z_s := \bigcup_{h \in \Gamma_s \setminus \Phi_s} (U_h \setminus V(\mathcal{C}_s))$. In words, Z_s is the union of the sets U_h which $\mathcal{C}_s$ intersects, after removing the sets $U_{h'}$ which $\mathcal{C}_s$ intersects in too many vertices and then removing the vertices of $\mathcal{C}_s$. Now, for each $s \in [c]$,

$$|Z_s| \ge (|\Gamma_s| - |\Phi_s|)(n^{1/3} - n^{1/6}) \ge \frac{\alpha \ell n^{1/3}}{8r} - 2n^{5/6} \ge \frac{\alpha^2 n}{16r} \ge \alpha' n$$

for n large, as the U_h are pairwise disjoint. Note also that as the Γ_s are pairwise disjoint, so are the Z_s for $s \in [c]$. Now for $1 \le s \le c$, apply Corollary 3.5 (3) to find an $(r-1)$-clique $S'_s \subset K_{r-1}(G[Z_0])$ such that there is a vertex $z_s \in Z_s \cap N^G(S'_s)$ and a vertex

$x_s \in (X \cup Y) \cap N^G(S'_s)$. We delete the S'_s from Z_0 and move to the next index $s+1$ or finish if $s = c$.

Now choose some $Y' \subset Y$ such that $x_s \in X \cup Y'$ for all $s \in [c]$ and

$$|Y'| + |X| + \sum_{s \in [c]} (|R_{\mathcal{C}_s}| + 1) = M.$$

This is easily done as $|X| + |Y| = \alpha^2 n$ is linear and $|X|, |R_{\mathcal{C}_s}| \leq 2n^{2/3}$ for all $s \in [c]$, which is much smaller than $M = n^{7/8}$. By Proposition 4.1, there is a K_r-diamond tree $\tilde{\mathcal{D}} = (\tilde{T}, \tilde{R}, \tilde{\Sigma})$ with $\tilde{R} = X \cup Y'$ and $\tilde{\Sigma} \subset K_{r-1}(G[W_0])$ a matching of $(r-1)$-cliques in $W_0 \subset W'$. Our diamond tree $\mathcal{D}_{j^*}$ is then obtained by connecting $\tilde{\mathcal{D}}$ and all the $\mathcal{C}_s \in \boldsymbol{\mathcal{C}}$. In more detail, for each $s \in [c]$, there exists some $h_s \in \Gamma_s$ such that $z_s \in U_{h_s} \subset X_{h_s}$. We define

$$R_{j^*} := \tilde{R} \cup \bigcup_{s \in [c]} (R_{\mathcal{C}_s} \cup \{z_s\}) \quad \text{and} \quad \Sigma_{j^*} := \tilde{\Sigma} \cup \bigcup_{s \in [c]} (\Sigma_{\mathcal{C}_s} \cup \{S'_s\} \cup \{S_{h_s}\}),$$

where $\Sigma_{\mathcal{C}_s}$ is the set of interior $(r-1)$-cliques of $\mathcal{C}_s$; $S'_s \in K_{r-1}(G[Z_0])$ is the $(r-1)$-clique which forms a clique with both z_s and x_s defined above; and S_{h_s} is the $(r-1)$-clique corresponding to the set X_{h_s} (which contains z_s) in Claim 8.5. We claim that there exists a diamond tree $\mathcal{D}_{j^*}$ of order M which has R_{j^*} as a set of removable vertices and Σ_{j^*} as a set of interior $(r-1)$-cliques. Indeed, we can form the defining auxiliary tree T_{j^*} by starting with the forest of the disjoint union of $\tilde{T}$ and the $T_{\mathcal{C}_s}$ for $s \in [c]$, where $T_{\mathcal{C}_s}$ denotes the defining tree for the K_r-diamond tree $\mathcal{C}_s$. For each $s \in [c]$, we then add a path of length 2 (with two edges) between some vertex in $V(T_{\mathcal{C}_s})$ and $V(\tilde{T})$. The edges of this path correspond exactly to the internal $(r-1)$-cliques S_{h_s} and S'_s and thus the vertices of this path correspond to x_s, z_s and some vertex in $R_{\mathcal{C}_s} \cap U_{h_s}$ for each $s \in [c]$.

This defines $\mathcal{D}_{j^*}$ and so we update all the U_h to be $U_h \setminus V(\mathcal{D}_{j^*})$ for $h \in [\ell]$ and kill any indices $h \in \Lambda$ such that either $V(\mathcal{D}_{j^*})$ intersects S_h or $|X_h \cap V(\mathcal{D}_{j^*})| \geq 2Cn^{1/6}$. We now need to check that conditions (1)–(3) hold for $\mathcal{D}_{j^*}$. To see (1), note that R_{j^*} contains all the $R_{\mathcal{C}_s}$ for $s \in [c]$ and so intersects X_h for all $h \in \bigcup_{s \in [c]} \Gamma_s$. Moreover, taking Γ as defined at the end of finding the $\mathcal{C}_s$, we have $|\Gamma \cup \bigcup_{s \in [c]} \Gamma_s| \geq (1 - \alpha/2)\ell$ and $|\Gamma| \leq \alpha\ell/2$, and so this confirms (1). To see (2), note that the only times we used vertices of the S_h with $h \in [\ell]$ to construct $\mathcal{D}_{j^*}$ was when we added the S_{h_s} for $s \in [c]$ to the set of interior cliques. Thus we intersected exactly $c \leq C$ of these with $V(\mathcal{D}_{j^*})$. Finally, (3) for $\mathcal{D}_{j^*}$ is implied by conditions (iv) when we found the $\mathcal{C}_s$. Indeed, $R_{j^*} \cap \bigcup_{h \in [\ell]} X_h = \bigcup_{s \in [c]} (R_{\mathcal{C}_s} \cup \{z_s\})$ and so for any index h that does not lie in $\bigcup_{s \in [c]} \Lambda_s$ (which has size at most $Cn^{1/2}$), we have

$$|V(\mathcal{D}_{j^*}) \cap X_h| \leq \sum_{s \in [c]} |(V(\mathcal{C}_s) \cup \{z_s\}) \cap X_h| \leq C(n^{1/6} + 1) \leq 2Cn^{1/6}.$$

This concludes the finding of $\mathcal{D}_{j^*}$, and doing this for all $j^* \in [4t]$ gives the desired claim and hence the proposition. ∎

9. Concluding remarks

In this paper, we showed that a condition of $\beta = o(p^k n)$ in an n-vertex (p, β)-bijumbled graph guarantees a K_{k+1}-factor. We conjecture that the same condition in fact guarantees any subgraph with maximum degree k.

Conjecture 9.1. For any $k \in \mathbb{N}_{\geq 2}$ and $c > 0$ there exists an $\varepsilon > 0$ such that any n-vertex (p, β)-bijumbled graph with $\delta(G) \geq cpn$ and $\beta \leq \varepsilon p^k n$ is k-*universal*, that is, given any graph F on at most n vertices, with maximum degree at most k, G contains a copy of F.

Note that Corollary 1.5 settles Conjecture 9.1 for $k = 2$. For $k \geq 3$, the best known result comes from the sparse blow-up lemma of Allen, Böttcher, Hàn, Kohayakawa and Person [2] which gives a condition of $\beta = o(p^{(3k+1)/2} n)$ guaranteeing k-universality in a (p, β)-bijumbled graph.

The conjecture echoes the notion that a K_{k+1}-factor is the 'hardest' maximum degree k graph to find. This idea has manifested itself in various other settings. For example, we know from the theorem of Hajnal and Szemerédi (Theorem 1.1) that any n-vertex graph G with $\delta(G) \geq (k/(k+1))n$ contains a K_{k+1}-factor and that this is tight. Bollobás and Eldridge [15], and independently Catlin [19], conjectured that the same minimum degree condition actually guarantees k-universality. This has been proven for $k = 2, 3$ [1,7,25] but remains open in general. In the case of random graphs, Johansson, Kahn and Vu [39] proved that the threshold for the appearance of a K_{k+1}-factor is of the order of

$$p_k^*(n) := n^{-2/(k+1)} (\log n)^{2/(k^2+k)}.$$

A recent breakthrough result of Frankston, Kahn, Narayanan and Park [28] implies that for any n-vertex graph F with maximum degree k, the threshold for the appearance of F in $G(n, p)$ is at most $p_k^*(n)$. Note that this is *not* implying that $G(n, p)$ is k-universal whp when $p = \omega(p_k^*(n))$ as we can only guarantee that some fixed F appears whp. However, the stronger version that $p_k^*(n)$ is the threshold for k-universality is believed to be true but only verified for $k = 2$ [26].

One thing that sets aside the pseudorandom setting in stark contrast to the other settings discussed above is that it might be possible to replace a K_{k+1}-factor as the benchmark for the 'hardest' graph to find in the host graph, by a single copy of K_{k+1}. Indeed, various authors [22,27,53,64] have stipulated that n-vertex K_{k+1}-free (p, β)-bijumbled graphs exist with $\beta = \Theta(p^k n)$. Such graphs would witness the tightness of both Theorem 1.4 and Conjecture 9.1 for all values of $k \geq 2$ (taking $r = k + 1$ in the setting of Theorem 1.4). Focusing on optimally pseudorandom graphs (that is, fixing $\beta = \Theta(\sqrt{pn})$ in (p, β)-bijumbled graphs), we expect to be able to find K_{k+1}-free optimally pseudorandom graphs with $p = \Omega(n^{-1/(2k-1)})$. These are only known to exist when $k = 2$. Indeed, we discussed the triangle-free construction of Alon in the introduction, and other constructions [21,48] have also been given which are (near-)optimal. For $k \geq 3$, however, this remains a key challenge in the understanding of pseudorandom graphs, with the best known general construction coming from a recent improvement of Bishnoi, Ihringer and Pepe [13] who give K_{k+1}-free optimally pseudorandom graphs of density $p = \Theta(n^{-1/k})$.

Further interest in finding denser such graphs comes from a recent remarkable connection discovered by Mubayi and Verstraëte [60] that shows that if, as we expect, the K_{k+1}-free optimally pseudorandom graphs with density $p = \Omega(n^{-1/(2k-1)})$ do exist, then it is possible to improve the lower bound on the off-diagonal Ramsey numbers to match the upper bound and thus determine the asymptotics of this extremal function. In detail, they show that if these pseudorandom graphs exist, then the off-diagonal Ramsey number is $R(k+1,t) = t^{k+o(1)}$ as t tends to infinity. In fact, even a construction with $p = \omega(n^{-1/(k+1)})$ would improve on the current best known lower bound on off-diagonal Ramsey numbers due to Bohman and Keevash [14].

We conclude by noting that Theorem 1.2 is, in some sense, the first result of its kind, giving a tight condition on pseudorandomness to guarantee the existence of a spanning structure. Indeed, the case of Hamilton cycles remains an intriguing open problem. Krivelevich and Sudakov [52] conjectured that a condition of $\lambda = o(d)$ is sufficient in (n, d, λ)-graphs and proved the currently best known bound of

$$\lambda = o\left(\frac{(\log\log n)^2 d}{\log n(\log\log\log n)}\right).$$

For hypergraphs of higher uniformity, one can easily generalise the notion of bijumbledness in Definition 1.3 but the picture becomes considerably more complex. Indeed, it turns out that the only subgraphs that one can guarantee by imposing conditions on bijumbledness are *linear* subgraphs, those in which pairs of hyperedges intersect in at most one vertex. Building on previous work [23, 47, 55, 56] mainly concerned with dense hypergraphs (the so-called quasirandom regime), Hiệp Hàn, Jie Han and the author [30, 31] recently gave the best-known conditions on pseudorandomness that guarantee different linear subgraphs of hypergraphs. These include all fixed sized linear subgraphs as well as F-factors for linear F (including perfect matchings) and loose Hamilton cycles. The tightness of these results is unclear as no good constructions are known for F-free pseudorandom hypergraphs. In general, the appearance of subgraphs in sparse pseudorandom (hyper-)graphs remains a fascinating area which is far from being understood.

Acknowledgments. Much of this work was done during a visit of the author to IMPA, Rio de Janeiro, Brazil. I am grateful to both Pedro Araújo and Robert Morris with whom I discussed many aspects of this project, for their hospitality, support and enthusiasm. I am also grateful to my coauthors from previous projects, Jie Han, Yoshiharu Kohayakawa and Yury Person, for introducing me to this area and many of the techniques and approaches used to tackle these sorts of problems. Finally, I am grateful to the anonymous referee, to Tibor Szabó and again to Robert Morris for providing many helpful suggestions aiding the presentation and readability of the paper.

Funding. The author was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - The Berlin Mathematics Research Center MATH+ (EXC-2046/1, project ID: 390685689) and by a Walter Benjamin fellowship of the DFG - project number 504502205.

References

[1] Aigner, M., Brandt, S.: Embedding arbitrary graphs of maximum degree two. J. London Math. Soc. (2) **48**, 39–51 (1993) Zbl 0796.05029 MR 1223891

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

[3] Allen, P., Böttcher, J., Hàn, H., Kohayakawa, Y., Person, Y.: Powers of Hamilton cycles in pseudorandom graphs. Combinatorica **37**, 573–616 (2017) Zbl 1399.05118 MR 3694704

[4] Allen, P., Böttcher, J., Hladký, J., Piguet, D.: A density Corrádi–Hajnal theorem. Canad. J. Math. **67**, 721–758 (2015) Zbl 1316.05069 MR 3361011

[5] Alon, N.: Explicit Ramsey graphs and orthonormal labelings. Electron. J. Combin. **1**, art. 12, 8 pp. (1994) Zbl 0814.05056 MR 1302331

[6] Alon, N., Bourgain, J.: Additive patterns in multiplicative subgroups. Geom. Funct. Anal. **24**, 721–739 (2014) Zbl 1377.11013 MR 3213827

[7] Alon, N., Fischer, E.: 2-factors in dense graphs. Discrete Math. **152**, 13–23 (1996) Zbl 0851.05068 MR 1388628

[8] Alon, N., Frankl, P., Huang, H., Rödl, V., Ruciński, A., Sudakov, B.: Large matchings in uniform hypergraphs and the conjecture of Erdős and Samuels. J. Combin. Theory Ser. A **119**, 1200–1215 (2012) Zbl 1242.05189 MR 2915641

[9] Alon, N., Kahale, N.: Approximating the independence number via the ϑ-function. Math. Programming **80**, 253–264 (1998) Zbl 0895.90169 MR 1603356

[10] Alon, N., Spencer, J. H.: The probabilistic method. 4th ed., Wiley Series in Discrete Mathematics and Optimization, Wiley, Hoboken, NJ (2016) Zbl 1333.05001 MR 3524748

[11] Balogh, J., Lee, C., Samotij, W.: Corrádi and Hajnal's theorem for sparse random graphs. Combin. Probab. Comput. **21**, 23–55 (2012) Zbl 1241.05130 MR 2900047

[12] Balogh, J., Molla, T., Sharifzadeh, M.: Triangle factors of graphs without large independent sets and of weighted graphs. Random Structures Algorithms **49**, 669–693 (2016) Zbl 1352.05141 MR 3570984

[13] Bishnoi, A., Ihringer, F., Pepe, V.: A construction for clique-free pseudorandom graphs. Combinatorica **40**, 307–314 (2020) Zbl 1463.05373 MR 4121148

[14] Bohman, T., Keevash, P.: The early evolution of the H-free process. Invent. Math. **181**, 291–336 (2010) Zbl 1223.05270 MR 2657427

[15] Bollobás, B., Eldridge, S. E.: Packings of graphs and applications to computational complexity. J. Combin. Theory Ser. B **25**, 105–124 (1978) Zbl 0387.05020 MR 511983

[16] Broder, A. Z., Frieze, A. M., Suen, S., Upfal, E.: Optimal construction of edge-disjoint paths in random graphs. SIAM J. Comput. **28**, 541–573 (1999) Zbl 0912.05058 MR 1634360

[17] Brouwer, A. E., Haemers, W. H.: Eigenvalues and perfect matchings. Linear Algebra Appl. **395**, 155–162 (2005) Zbl 1056.05097 MR 2112881

[18] Caprara, A., Rizzi, R.: Packing triangles in bounded degree graphs. Inform. Process. Lett. **84**, 175–180 (2002) Zbl 1042.68087 MR 1928953

[19] Catlin, P. A.: Embedding subgraphs and coloring graphs under extremal degree conditions. Ph.D. thesis, Ohio State University (1976) MR 2626461

[20] Cioabă, S. M.: Perfect matchings, eigenvalues and expansion. C. R. Math. Acad. Sci. Soc. R. Canada **27**, 101–104 (2005) Zbl 1110.05058 MR 2204683

[21] Conlon, D.: A sequence of triangle-free pseudorandom graphs. Combin. Probab. Comput. **26**, 195–200 (2017) Zbl 1371.05259 MR 3603964

[22] Conlon, D., Fox, J., Zhao, Y.: Extremal results in sparse pseudorandom graphs. Adv. Math. **256**, 206–290 (2014) Zbl 1285.05096 MR 3177293

[23] Conlon, D., Hàn, H., Person, Y., Schacht, M.: Weak quasi-randomness for uniform hypergraphs. Random Structures Algorithms **40**, 1–38 (2012) Zbl 1236.05137 MR 2864650

[24] Corradi, K., Hajnal, A.: On the maximal number of independent circuits in a graph. Acta Math. Acad. Sci. Hungar. **14**, 423–439 (1963) Zbl 0118.19001 MR 200185

[25] Csaba, B., Shokoufandeh, A., Szemerédi, E.: Proof of a conjecture of Bollobás and Eldridge for graphs of maximum degree three. Combinatorica **23**, 35–72 (2003) Zbl 1046.05040 MR 1996626

[26] Ferber, A., Kronenberg, G., Luh, K.: Optimal threshold for a random graph to be 2-universal. Trans. Amer. Math. Soc. **372**, 4239–4262 (2019) Zbl 1420.05161 MR 4009429

[27] Fox, J., Lee, C., Sudakov, B.: Chromatic number, clique subdivisions, and the conjectures of Hajós and Erdős–Fajtlowicz. Combinatorica **33**, 181–197 (2013) Zbl 1299.05306 MR 3071852

[28] Frankston, K., Kahn, J., Narayanan, B., Park, J.: Thresholds versus fractional expectation-thresholds. Ann. of Math. (2) **194**, 475–495 (2021) Zbl 1472.05132 MR 4298747

[29] Hajnal, A., Szemerédi, E.: Proof of a conjecture of P. Erdős. In: Combinatorial theory and its applications, I–III (Balatonfüred, 1969), Colloquia Mathematica Societatis János Bolyai 4, North-Holland, Amsterdam, 601–623 (1970) Zbl 0217.02601 MR 297607

[30] Hàn, H., Han, J., Morris, P.: Factors and loose Hamilton cycles in sparse pseudo-random hypergraphs. In: Proceedings of the 2020 ACM–SIAM Symposium on Discrete Algorithms, SIAM, Philadelphia, PA, 702–717 (2020) Zbl 07304066 MR 4141225

[31] Hàn, H., Han, J., Morris, P.: Factors and loose Hamilton cycles in sparse pseudo-random hypergraphs. Random Structures Algorithms **61**, 101–125 (2022) Zbl 07304066 MR 4448751

[32] Han, J., Kohayakawa, Y., Morris, P., Person, Y.: Clique-factors in sparse pseudorandom graphs. Eur. J. Combin. **82**, 102999, 8 (2019) Zbl 1419.05116 MR 3989656

[33] Han, J., Kohayakawa, Y., Morris, P., Person, Y.: Finding any given 2-factor in sparse pseudo-random graphs efficiently. J. Graph Theory **96**, 87–108 (2021) MR 4191116

[34] Han, J., Kohayakawa, Y., Person, Y.: Near-perfect clique-factors in sparse pseudorandom graphs. In: Discrete Mathematics Days 2018, Electronic Notes in Discrete Mathematics 68, Elsevier, Amsterdam, 221–226 (2018) Zbl 1397.05093 MR 3844904

[35] Han, J., Kohayakawa, Y., Person, Y.: Near-perfect clique-factors in sparse pseudorandom graphs. Combin. Probab. Comput. **30**, 570–590 (2021) Zbl 1510.05141 MR 4272853

[36] Han, J., Morris, P., Treglown, A.: Tilings in randomly perturbed graphs: bridging the gap between Hajnal–Szemerédi and Johansson–Kahn–Vu. Random Structures Algorithms **58**, 480–516 (2021) MR 4234994

[37] Heckel, A.: Random triangles in random graphs. Random Structures Algorithms **59**, 616–621 (2021) MR 4323311

[38] Janson, S., Łuczak, T., Ruciński, A.: Random graphs. Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York (2000) Zbl 0968.05003 MR 1782847

[39] Johansson, A., Kahn, J., Vu, V.: Factors in random graphs. Random Structures Algorithms **33**, 1–28 (2008) Zbl 1146.05040 MR 2428975

[40] Kahn, J.: Hitting times for Shamir's problem. Trans. Amer. Math. Soc. **375**, 627–668 (2022) Zbl 1479.05326 MR 4358678

[41] Kahn, J.: Asymptotics for Shamir's problem. Adv. Math. **422**, art. 109019, 39 pp. (2023) Zbl 1512.05311 MR 4575035

[42] Kann, V.: Maximum bounded H-matching is MAX SNP-complete. Inform. Process. Lett. **49**, 309–318 (1994) Zbl 0803.68089 MR 1266956

[43] Karp, R. M.: On the computational complexity of combinatorial problems. Networks **5**, 45–68 (1975) Zbl 0324.05003

[44] Kierstead, H. A., Kostochka, A. V., Mydlarz, M., Szemerédi, E.: A fast algorithm for equitable coloring. Combinatorica **30**, 217–224 (2010) Zbl 1224.05176 MR 2676836

[45] Kim, J. H.: Perfect matchings in random uniform hypergraphs. Random Structures Algorithms **23**, 111–132 (2003) Zbl 1028.05088 MR 1995686

[46] Kirkpatrick, D. G., Hell, P.: On the complexity of general graph factor problems. SIAM J. Comput. **12**, 601–609 (1983) Zbl 0525.68023 MR 707416

[47] Kohayakawa, Y., Nagle, B., Rödl, V., Schacht, M.: Weak hypergraph regularity and linear hypergraphs. J. Combin. Theory Ser. B **100**, 151–160 (2010) Zbl 1216.05094 MR 2595699

[48] Kopparty, S.: Cayley graphs (lecture notes). http://sites.math.rutgers.edu/~sk1233/courses/graphtheory-F11/cayley.pdf (2011)

[49] Kostochka, A. V., Rödl, V.: Partial Steiner systems and matchings in hypergraphs. In: Proceedings of the Eighth International Conference “Random Structures and Algorithms” (Poznań, 1997), Random Structures Algorithms **13**, 335–347 (1998) Zbl 0959.05079 MR 1662789

[50] Krivelevich, M.: Perfect fractional matchings in random hypergraphs. Random Structures Algorithms **9**, 317–334 (1996) Zbl 0874.05049 MR 1606849

[51] Krivelevich, M.: Triangle factors in random graphs. Combin. Probab. Comput. **6**, 337–347 (1997) Zbl 0886.05101 MR 1464569

[52] Krivelevich, M., Sudakov, B.: Sparse pseudo-random graphs are Hamiltonian. J. Graph Theory **42**, 17–33 (2003) Zbl 1028.05059 MR 1943104

[53] Krivelevich, M., Sudakov, B.: Pseudo-random graphs. In: More sets, graphs and numbers, Bolyai Society Mathematical Studies 15, Springer, Berlin, 199–262 (2006) Zbl 1098.05075 MR 2223394

[54] Krivelevich, M., Sudakov, B., Szabó, T.: Triangle factors in sparse pseudo-random graphs. Combinatorica **24**, 403–426 (2004) Zbl 1064.05119 MR 2085364

[55] Lenz, J., Mubayi, D.: Perfect packings in quasirandom hypergraphs I. J. Combin. Theory Ser. B **119**, 155–177 (2016) Zbl 1334.05095 MR 3486341

[56] Lenz, J., Mubayi, D., Mycroft, R.: Hamilton cycles in quasirandom hypergraphs. Random Structures Algorithms **49**, 363–378 (2016) Zbl 1346.05153 MR 3536544

[57] Montgomery, R.: Embedding bounded degree spanning trees in random graphs. arXiv:1405.6559 (2014)

[58] Montgomery, R.: Spanning trees in random graphs. Adv. Math. **356**, art. 106793, 92 pp. (2019) Zbl 1421.05080 MR 3998769

[59] Morris, P.: A tight condition for triangle factors in pseudorandom graphs. In: Proceedings of the 2021 ACM–SIAM Symposium on Discrete Algorithms (SODA), SIAM, Philadelphia, PA, 899–918 (2021) MR 4262490

[60] Mubayi, D., Verstraëte, J.: A note on pseudorandom Ramsey graphs. J. Eur. Math. Soc. (online, 2023)

[61] Nenadov, R.: Triangle-factors in pseudorandom graphs. Bull. London Math. Soc. **51**, 421–430 (2019) Zbl 1425.05093 MR 3964496

[62] Nenadov, R., Pehova, Y.: On a Ramsey–Turán variant of the Hajnal–Szemerédi theorem. SIAM J. Discrete Math. **34**, 1001–1010 (2020) Zbl 1436.05087 MR 4080942

[63] Riordan, O.: Random cliques in random graphs and sharp thresholds for F-factors. Random Structures Algorithms **61**, 619–637 (2022) MR 4504837

[64] Sudakov, B., Szabó, T., Vu, V. H.: A generalization of Turán’s theorem. J. Graph Theory **49**, 187–195 (2005) Zbl 1066.05079 MR 2145507

[65] Thomason, A.: Pseudorandom graphs. In: Random graphs ’85 (Poznań, 1985), North-Holland Mathematics Studies 144, North-Holland, Amsterdam, 307–331 (1987) Zbl 0626.00008 MR 930498

[66] Thomason, A.: Random graphs, strongly regular graphs and pseudorandom graphs. In: Surveys in combinatorics 1987 (New Cross, 1987), London Mathematical Society Lecture Note Series 123, Cambridge University Press, Cambridge, 173–195 (1987) Zbl 0672.05068 MR 905280

[67] Tikhomirov, K., Youssef, P.: The spectral gap of dense random regular graphs. Ann. Probab. **47**, 362–419 (2019) Zbl 1407.05212 MR 3909972

[68] Treglown, A.: A degree sequence Hajnal–Szemerédi theorem. J. Combin. Theory Ser. B **118**, 13–43 (2016) Zbl 1332.05117 MR 3471843