Turán problems for expanded hypergraphs

Peter Keevash* Noam Lifshitz† Eoin Long‡ Dor Minzer§

Abstract

We obtain new results on the Turán number of any bounded degree uniform hypergraph obtained as the expansion of a hypergraph of bounded uniformity. These are asymptotically sharp over an essentially optimal regime for both the uniformity and the number of edges and solve a number of open problems in Extremal Combinatorics.

Firstly, we give general conditions under which the crosscut parameter asymptotically determines the Turán number, thus answering a question of Mubayi and Verstraëte. Secondly, we refine our asymptotic results to obtain several exact results, including proofs of the Huang–Loh–Sudakov conjecture on cross matchings and the Füredi–Jiang–Seiver conjecture on path expansions.

We have introduced two major new tools for the proofs of these results. The first of these, Global Hypercontractivity, is used as a 'black box' (we present it in a separate paper with several other applications). The second tool, presented in this paper, is a far-reaching extension of the Junta Method, which we develop from a powerful and general technique for finding matchings in hypergraphs under certain pseudorandomness conditions.

1 Introduction

A longstanding and challenging direction of research in Extremal Combinatorics, initiated by Turán in the 1940's, is that of determining the maximum size of a k-graph (k-uniform hypergraph) $\mathcal{H} \subset \binom{[n]}{k}$ on n vertices not containing some fixed k-graph F; this is the Turán number, denoted $\mathrm{ex}(n, F)$. Turán numbers of graphs (the case $k = 2$) are quite well-understood (if F is not bipartite), but there are very few results even for specific hypergraphs, let alone general results for families of hypergraphs (see the survey [27]).

In this paper we prove a number of general results on Turán numbers for the family of bounded degree expanded hypergraphs (see Section 1.2), thus solving several open problems: a question of Mubayi and Verstraëte relating asymptotics of the Turán number to the crosscut (see Theorem 1.4), the Huang–Loh–Sudakov conjecture on cross matchings (see Theorem 1.2) and the Füredi–Jiang–Seiver conjecture on path expansions (see Corollary 1.7).

A striking feature of our results is their applicability across an essentially optimal range of uniformities and sizes, which previously seemed entirely out of reach. This is achieved via two new methods. The first is a new sharp threshold theorem (see Theorem 3.4) derived from our theory of Global Hypercontractivity, which was presented in the first version of this paper (arXiv:1906.05568); that method is now split off into a separate paper [29] with several other applications unrelated to the questions of Extremal Combinatorics considered here. The second method is a far-reaching extension of the Junta Method of Keller and Lifshitz [30] (which itself greatly extended the applications of an approach initiated by Dinur and Friedgut [4]). A large part of the technical work in this paper goes

*Mathematical Institute, University of Oxford, Oxford, UK. Email: keevash@maths.ox.ac.uk.
Research supported in part by ERC Consolidator Grant 647678.
†Einstein Institute for Mathematics, Hebrew University of Jerusalem. Email: noamlifshitz@gmail.com.
Research supported in part by ERC advanced grant 834735.
‡University of Birmingham, Birmingham, UK. Email: e.long@bham.ac.uk.
Research supported in part by EPSRC New Investigator Award [EP/Y006399/1].
§Department of Mathematics, Massachusetts Institute of Technology. Email: minzer.dor@gmail.com.
Some of the work was done while the author was a postdoc in the Institute for Advanced Study, Princeton, supported NSF grant CCF-1412958 and Rothschild Fellowship.

into developing a powerful and general machinery for finding matchings in hypergraphs under certain pseudorandomness conditions.

1.1 Cross matchings

Before introducing the general setting of expanded hypergraphs, we first consider an important case, which is in itself a source of many significant problems, namely the problem of finding matchings. In both theory and application, a wide range of significant questions can be recast as existence questions for matchings (see e.g. the books [35, 39] and the survey [28]).

Perhaps the most well-known open question concerning matchings, due to Erdős [11], asks how large a family $\mathcal{F} \subset \binom{[n]}{k}$ can be if it does not contain an s-matching, i.e. sets $\{A_1, \dots, A_s\}$ with $A_i \cap A_j = \emptyset$ for all distinct $i, j \in [s]$. Two natural families of such $\mathcal{F}$ are stars $\mathcal{S}_{n,k,s-1} := \left\{A \in \binom{[n]}{k} : A \cap [s-1] \neq \emptyset\right\}$ and cliques $\mathcal{C}_{k,s-1} := \binom{[ks-1]}{k}$. Erdős conjectured that one of these families is always extremal.

Conjecture 1.1 (Erdős Matching Conjecture)**.** *Let $n \geq ks$ and suppose that $\mathcal{F} \subset \binom{[n]}{k}$ does not contain an s-matching. Then $|\mathcal{F}| \leq \max\left\{|\mathcal{S}_{n,k,s-1}|, |\mathcal{C}_{k,s-1}|\right\}$.*

This conjecture remains open, despite an extensive literature, of which we will mention a few highlights. The case $s = 2$ is the classical Erdős–Ko–Rado theorem [12]. Erdős and Gallai [10] confirmed the conjecture for $k = 2$. The case $k = 3$ was proven by Łuczak and Mieczkowska [37] for large s and by Frankl [17] for all s. Bollobás, Daykin and Erdős [1] proved the conjecture provided $n = \Omega(k^3 s)$, which was reduced to $n = \Omega(k^2 s)$ by Huang, Loh and Sudakov [24] and finally to $n = \Omega(ks)$ by Frankl [13] (in fact to $n \gtrsim 2ks$, improved by Frankl and Kupavskii [16] to $n \geq 5ks/3$ for large s), which is the optimal order of magnitude for the extremal family to be a star rather than a clique – or even to just contain s disjoint k-sets.

Our first result in this context is a cross version of that of Frankl, which proves (a strengthened form of) a conjecture of Huang, Loh and Sudakov [24]. Here we say that families $\mathcal{F}_1, \dots, \mathcal{F}_s$ cross contains a hypergraph $\{A_1, \dots, A_s\}$ (e.g. an s-matching) if $A_i \in \mathcal{F}_i$ for each $i \in [s]$.

Theorem 1.2. *There is a constant $C > 0$ so that if $n, s, k_1, \dots, k_s \in \mathbb{N}$ with $k_i \leq \frac{n}{Cs}$ and $\mathcal{F}_i \subset \binom{[n]}{k_i}$ with $|\mathcal{F}_i| \geq |\mathcal{S}_{n,k_i,s-1}|$ for all $i \in [s]$, either $\mathcal{F}_1, \dots, \mathcal{F}_s$ cross contain an s-matching, or there is $J \subset [n]$ with $|J| = s - 1$ such that each $\mathcal{F}_i = \mathcal{S}_{n,k_i,J} := \{A \in \binom{[n]}{k_i} : A \cap J \neq \emptyset\}$.*

Remark 1.3. Theorem 1.2 in the case that all $k_i = k$ was proved by Huang, Loh and Sudakov [24] for $n = \Omega(k^2 s)$ and recently by Frankl and Kupavskii [15] for $n = \Omega(ks \log s)$; our result applies to $n = \Omega(ks)$, which is the optimal order of magnitude. Subsequent to our work, a very different proof of the Huang-Loh-Sudakov Conjecture has been given by Lu, Wang and Yu [36]. We also obtain a strong stability result (see Theorem 6.1 below) which gives structural information even if we only assume that the size of each family is within a constant factor of that of a star: either there is a cross matching or some family correlates strongly with a star. Besides having independent interest, this stability result will play a key role in the proof of our general Turán results.

1.2 Expanded hypergraphs

As mentioned above, there are very few general results on Turán numbers for a family of hypergraphs. One family for which there has been substantial progress is that of expanded graphs (see the survey [38]). Given an r-graph G and $k \geq r$, the k-expansion $G^+ = G^+(k)$ is the k-uniform hypergraph obtained from G by adding $k - r$ new vertices to each edge, i.e. G^+ has edge set $\{e \cup S_e : e \in E(G)\}$ where $|S_e| = k - r$, $S_e \cap V(G) = \emptyset$ and $S_e \cap S_{e'} = \emptyset$ for all distinct $e, e' \in E(G)$. In particular, a k-graph s-matching is the k-expansion of a graph s-matching.

When G is a graph (the case $r = 2$), in the non-degenerate case when k is less than the chromatic number $\chi(G)$ the Turán numbers $\mathrm{ex}(n, G^+(k))$ are well-understood (see [38, Section 2]), so the main focus for ongoing research is the degenerate case $k \geq \chi(G)$. Here Frankl and Füredi [14] introduced the following important parameter and corresponding construction that seems to often determine the asymptotics of the Turán number. For any r-graph G, we call $S \subset V(G^+)$ a *crosscut* if $|E \cap S| = 1$ for all $E \in G^+$. The crosscut $\sigma(G)$ of G is the size of the minimal such set, i.e.

$$\sigma(G) := \min\big\{|S| : S \subset V(G^+) \text{ with } |E \cap S| = 1 \text{ for all } E \in G^+\big\}.$$

It is easy to see that $\sigma(G)$ exists for $k \geq r+1$ and that in this regime the parameter does not depend on k. Clearly,

$$\mathcal{S}^{(1)}_{n,k,\sigma(G)-1} := \big\{A \in \tbinom{[n]}{k} : |A \cap [\sigma(G)-1]| = 1\big\}$$

is G^+-free. Moreover, this simple construction determines the asymptotics of $\mathrm{ex}(n, G^+(k))$ for $n > n_0(k, G)$ for several graphs G, including paths [22, 31], cycles [21, 31] and trees [20, 32]. Given this phenomenon, according to Mubayi and Verstraëte [38], one of the major open problems on expansions is to decide when the Turán number is asymptotically determined by the crosscut construction. Our next result resolves this problem for all bounded degree r-graphs (so in particular for graphs) in a range of parameters that is optimal up to constant factors. Moreover, we also obtain a strong structural approximation for any family that is close to extremal (see Theorem 1.8 below).

Theorem 1.4. *For any $r, \Delta \geq 2$ and $\varepsilon > 0$ there is $C > 0$ so that the following holds for any r-graph G with s edges, maximum degree $\Delta(G) \leq \Delta$ and $\sigma(G) \geq 2$. For any $k, n \in \mathbb{N}$ with $C \leq k \leq n/Cs$ we have $ex(n, G^+(k)) = (1 \pm \varepsilon)|\mathcal{S}^{(1)}_{n,k,\sigma(G)-1}|$.*

Remark 1.5. Some lower bound on k is necessary to obtain the conclusion in Theorem 1.4. Indeed, we have already mentioned that the non-degenerate case $k < \chi(G)$ when G is a graph exhibits different behaviour (a complete partite k-graph shows that $\mathrm{ex}(n, G^+) = \Omega(n/k)^k$), and moreover, examples in [38] show that some lower bound on k may be necessary even if G is bipartite (e.g. if $G = K_{9,9}$ then consider the 3-graph of triangles in a suitably dense random graph made G-free by edge deletions). The upper bound on k in our result is also necessary up to the constant factor by space considerations, as even the complete k-graph $\binom{[n]}{k}$ can only contain $G^+(k)$ if $n \geq |V(G^+)| = |V(G)| + (k-2)s$. With the exception of Frankl's matching theorem [13], Theorem 1.4 appears to be the only known Turán result in which both the uniformity k and the size s can vary over such a wide range.

Next we consider conditions under which we can refine the asymptotic result of Theorem 1.4 and determine the Turán number $\mathrm{ex}(n, G^+)$ exactly. One complication here is that crosscuts may be beaten by stars $\mathcal{S}_{n,k,\tau(G)-1}$, where

$$\tau(G) := \min\big\{|S| : |S \cap e| \geq 1 \text{ for all } e \in E(G)\big\}$$

is the transversal number of G. Clearly $\tau(G) \leq \sigma(G)$. For fixed s, crosscuts cannot be beaten by smaller stars, but this may not hold when s grows with n, as then edges with more than one vertex in the base of the star are significant. Another complication is that lower order correction terms are necessary for certain G, e.g. for k-expanded paths $P^+_\ell(k)$ of length ℓ for $n > n_0(k, \ell)$ we have $\mathrm{ex}(n, P^+_3(k)) = \binom{n-1}{k-1} = |\mathcal{S}_{n,k,1}|$, as predicted by the crosscut/star construction, but $\mathrm{ex}(n, P^+_4(k)) = \binom{n-1}{k-1} + \binom{n-3}{k-2}$, as we can add all sets containing some fixed pair of vertices. This is analogous to the familiar situation in extremal graph theory where we only expect exact results for graphs that are critical with respect to the key parameter of the extremal construction. Accordingly, we introduce the following analogous concept of criticality for expanded hypergraphs with respect to crosscuts and stars: we say that G is *critical* if it has an edge e such that

$$\sigma(G \setminus e) = \tau(G \setminus e) < \tau(G) = \sigma(G).$$

We obtain the following general exact result for Turán numbers.

Theorem 1.6. *For any $r, \Delta \geq 2$ there is $C > 0$ such that for any critical r-graph G with s edges, maximum degree $\Delta(G) \leq \Delta$ and $C \leq k \leq n/Cs$ we have $ex(n, G^+(k)) = |\mathcal{S}_{n,k,\sigma(G)-1}|$.*

This result applies to many graphs considered in the previous literature, such as paths of odd length. Paths of even length are not critical, but satisfy a generalised criticality property: deleting one edge does not reduce the transversal number, but deleting two edges (whether disjoint or intersecting) does reduce the crosscut number. Thus we have the following natural construction for excluding any expanded path P_ℓ^+ of length ℓ. Let $\mathcal{F}^*_{n,k,\ell} = \mathcal{S}_{n,k,J}$ with $|J| = \sigma(P_\ell) - 1$ if ℓ is odd, or if ℓ is even obtain $\mathcal{F}^*_{n,k,\ell}$ from $\mathcal{S}_{n,k,J}$ by adding $\{A \in \binom{[n]}{k} : T \subset A\}$ for some $T \in \binom{[n]\setminus J}{2}$. Clearly $\mathcal{F}^*_{n,k,\ell}$ is P_ℓ^+-free. Füredi, Jiang and Seiver [22] showed that $\mathrm{ex}(n, P_\ell^+) = |\mathcal{F}^*_{n,k,\ell}|$ provided $n \gg n_0(k,\ell)$, and conjectured that this holds provided $n \geq Ck\ell$. We prove this conjecture.

Corollary 1.7. *There is $C > 0$ so that if $n, k, \ell \in \mathbb{N}$ and $C \leq k \leq n/C\ell$ then $ex(n, P_\ell^+) = |\mathcal{F}^*_{n,k,\ell}|$.*

1.3 Junta Approximation

In recent years, the Analysis of Boolean functions has found significant application in Extremal Combinatorics, via the connection provided by the Margulis-Russo formula between the sharp threshold phenomenon and influences of Boolean functions. This approach was initiated by Dinur and Friedgut [4], who applied a theorem of Friedgut [19] on Boolean functions of small influence to prove that large uniform intersecting families can be approximated by juntas, i.e. families that depend only on a few coordinates. This connection has since played a key role in intersection theorems for a variety of settings, including graphs [6], permutations [7] and sets [8, 9].

The approach of Dinur and Friedgut was substantially generalised by Keller and Lifshitz [30] to apply to a variety of Turán problems on expanded hypergraphs. At a very high level, their Junta Method is a version of the Stability Method in Extremal Combinatorics, in that it consists of two steps: an approximate step that determines the rough structure of families that are close to optimal, and an exact step that refines the structure and determines the optimal construction. Their approximate step consisted of showing that any G^+-free family is approximately contained in a G^+-free junta.

The crucial new difficulty that we need to address in this paper is allowing the number of edges in G to grow as a function of n, whereas the previous works needed it to be a fixed constant. Friedgut's theorem can no longer be applied in this setting, as we require a threshold result for Boolean functions $f : \{0,1\}^n \to \{0,1\}$ according to the p-biased measure μ_p in the sparse regime where both p and $\mu_p(f)$ may be functions of n that approach zero.

Our new sharp threshold result (see Theorem 3.4) provides the necessary improvement on the analytic side which, when combined with a number of additional combinatorial ideas, allow us to obtain the following junta approximation theorem. For the statement, we introduce the notation $\mathcal{G}(r, s, \Delta)$ for the family of all r-graphs G with s edges and maximum degree $\Delta(G) \leq \Delta$. We also recall that $S \subset V(G^+)$ is a *crosscut* if $|E \cap S| = 1$ for all $E \in G^+$, and that $\sigma(G)$ denotes the minimum size of a crosscut.

Theorem 1.8. *Let $G \in \mathcal{G}(r, s, \Delta)$ and $C \gg r\Delta\varepsilon^{-1}$. Then for any G^+-free $\mathcal{F} \subset \binom{[n]}{k}$ with $C \leq k \leq n/Cs$, there is $J \subset V(G)$ with $|J| \leq \sigma(G) - 1$ and $|\mathcal{F} \setminus \mathcal{S}_{n,k,J}| \leq \varepsilon |\mathcal{S}_{n,k,\sigma(G)-1}|$.*

We note that Theorem 1.4 is immediate from Theorem 1.8, as for $k \geq C \gg \varepsilon^{-1}$ we have

$$\mathrm{ex}(n, G^+) \geq |\mathcal{S}^{(1)}_{n,k,\sigma(G)-1}| \geq (1-\varepsilon)|\mathcal{S}_{n,k,\sigma(G)-1}|.$$

The set J in Theorem 1.8 will consist of all vertices of suitably large degree. Thus $\mathcal{F}^{\emptyset}_J := \mathcal{F} \setminus \mathcal{S}_{n,k,J}$ does not have any vertices of large degree, which we will think of a pseudorandomness property.

While Theorem 1.8 suffices for asymptotic results, for our exact results we will require the following refined junta approximation result proved in Section 5, in which we improve the bound on $|\mathcal{F}^{\emptyset}_J|$.

Theorem 1.9. *Let $G \in \mathcal{G}(r, s, \Delta)$, $0 < C^{-1} \ll \delta \ll \varepsilon \ll (r\Delta)^{-1}$ and $C \leq k \leq n/Cs$. Then for any G^+-free $\mathcal{F} \subset \binom{[n]}{k}$ with $|\mathcal{F}| > |\mathcal{S}_{n,k,\sigma(G)-1}| - \delta\binom{n-1}{k-1}$ there is $J \in \binom{[n]}{\sigma(G)-1}$ with $|\mathcal{F} \setminus \mathcal{S}_{n,k,J}| \leq \varepsilon\binom{n-1}{k-1}$.*

1.4 Structure, strategy and techniques

To introduce our new techniques, we will first provide some context by indicating the overall structure and where new ingredients are needed. In the proof of Theorem 1.8 we will consider separately the two steps of showing $|J| \leq \sigma(G) - 1$ and $|\mathcal{F} \setminus \mathcal{S}_{n,k,J}| \leq \varepsilon |S_{n,k,\sigma(G)-1}|$. For both steps we consider a two step embedding strategy for G^+, where in the first step we embed[1] G in the 'fat shadow' of $\mathcal{F}$ (meaning that the image of every edge has many extensions to an edge of $\mathcal{F}$) and in the second step we 'lift' edges from the fat shadow to the original family.

This proof strategy is implemented in the next section, assuming results that will be proved in later sections. The analysis of fat shadows and the embedding steps will be carried out in Section 4. The lifting step requires results on cross matchings presented in Section 3, which will also be used for the proof of the Huang–Loh–Sudakov Conjecture in Section 6.

These cross matching results in Section 3 and their further refinements in Section 5 are where we need the new techniques, arising from the interplay of two combinatorial pseudorandomness notions with sharp threshold results from global hypercontractivity. After developing these techniques, the final three sections of the paper apply them in conjunction with some additional combinatorial ideas to prove our exact results on the Turán numbers of expanded hypergraphs.

Pseudorandomness

An important theme throughout this paper will be the interplay between two pseudorandomness notions: globalness and uncapturability. Informally, a hypergraph is 'uncapturable' if there is no small set that hits most of its edges and 'global' if one cannot obtain a significant density increment by restricting to those edges that contain some small fixed set. We will see that globalness implies uncapturability, and that uncapturability can be 'upgraded' to globalness by taking appropriate restrictions.

Here we highlight an important new phenomenon for cross matchings with the following result (a simplified form of Lemma 5.7). Whereas an extremal existence result requires minimum density of order sk/n, we see that a pseudorandom existence result only requires a density parameter of order $(sk/n)^d$ for any fixed constant d (see the next section for the precise definition of uncapturability).

Lemma 1.10. *Let $\mathcal{F}_i \subset \binom{[n]}{k}$ for $i \in [s]$, where $2d \leq k \leq n/Cs$ with $C \gg d \geq 1$. If each $\mathcal{F}_i$ is $(2ds, (2sk/n)^d)$-uncapturable then $\mathcal{F}_1, \ldots, \mathcal{F}_s$ cross contain a matching.*

Sharp thresholds

A classical theorem of Bollobás and Thomason [2] shows that any monotone property (i.e. hypergraph) $\mathcal{F}_n \subset \{0,1\}^n$ has a threshold. Writing $p_{\mathcal{F}_n}(t) = \inf\{p : \mu_p(\mathcal{D}) \geq t\}$, this means that for any $\varepsilon > 0$ there is $C > 0$ such that $p_{\mathcal{F}_n}(1-\varepsilon) \leq C p_{\mathcal{F}_n}(\varepsilon)$. Many natural properties exhibit the 'sharp threshold phenomenon' that $C = 1 + o(1)$ as $n \to \infty$. In particular, our results on Global Hypercontractivity give such a result for global properties (see Theorem 3.4). Any hypergraph $\mathcal{F}$ has a global restriction $\mathcal{F}'$ obtained by taking those edges containing some small fixed set, so our sharp threshold result enables us to find $\mu_{p'}(\mathcal{F}') \gg \mu_p(\mathcal{F})$ for some p' close to p.

We can now give a rough indication (omitting many details) of how this sharp threshold result can be used to prove a result in the direction of Lemma 1.10 (weakening $(sk/n)^d$ to sk/Cn as in Lemma 3.1). Given uncapturable families $\mathcal{F}_1, \ldots, \mathcal{F}_s$, we can upgrade to global families $\mathcal{F}'_1, \ldots, \mathcal{F}'_s$, where we find a small set R partitioned into $(R_1, \ldots, R_s)$ and each $\mathcal{F}'_i = \{A \setminus R_i : A \in \mathcal{F}_i, A \cap R = R_i\}$. Via the sharp threshold result we can then find further restrictions to pass to families $\mathcal{F}''_1, \ldots, \mathcal{F}''_s$ with $\mu_{2p}(\mathcal{F}''_i) \gg \mu_p(\mathcal{F}_i)$, where $p = k/n$. This increase in density is sufficient to find a cross matching by a weak form of the extremal result (translated to the product measure setting), which can then be extended to a cross matching in the original families.

[1] For simplicity we are only describing the embedding strategy used to bound $|\mathcal{F} \setminus \mathcal{S}_{n,k,J}|$; the strategy for bounding $|J|$ is similar, but adapted so that J can play the role of a crosscut in G.

2 Globalness and uncapturability

This section introduces the two key pseudorandomness concepts that will be fundamental throughout this paper. After some basic definitions in the first subsection, we will define and analyse these pseudorandomness notions in the second subsection. We conclude in the third section by proving our junta approximation theorem, assuming two embedding lemmas that will be proved in Section 4.

2.1 Definitions

Given $m, n \in \mathbb{N}$ with $m \le n$ we let $[n] = \{1, 2, \dots, n\}$ and $[m, n] = \{m, m+1, \dots, n\}$. We write $\{0,1\}^X$ for the power set (set of subsets) of a set X (identifying sets with their characteristic 0/1 vectors) and $\binom{X}{k} = X^{(k)} = \{A \subset X : |A| = k\}$. We call $\mathcal{F} \subset \{0,1\}^X$ a family or a hypergraph on the vertex set X, and the elements of $\mathcal{F}$ are called edges. We say $\mathcal{F}$ is k-uniform if $\mathcal{F} \subset \binom{X}{k}$; we also call $\mathcal{F}$ a k-graph on X.

Given a family $\mathcal{F} \subset \{0,1\}^X$ and $B \subset J \subset X$ we write $\mathcal{F}_J^B$ for the family

$$\mathcal{F}_J^B := \{A \in \{0,1\}^{X \setminus J} : A \cup B \in \mathcal{F}\} \subset \{0,1\}^{X \setminus J}.$$

Clearly $\mathcal{F}_J^B$ is $(k - |B|)$-uniform if $\mathcal{F}$ is k-uniform. If either B or J has a single element $\{j\}$ then we will often suppress the bracket, e.g. $\mathcal{F}_v^v = \mathcal{F}_{\{v\}}^{\{v\}}$.

We refer to $\mathcal{F}_v^v$ as the *exclusive link* of v in $\mathcal{F}$. The *inclusive link* of v in $\mathcal{F}$ is $\mathcal{F} * v := \{E \in \mathcal{F} : v \in E\}$. The *degree* of a vertex v in $\mathcal{F}$ is $d_{\mathcal{F}}(v) = |\mathcal{F}_v^v| = |\mathcal{F} * v|$. The minimum and maximum degrees of $\mathcal{F}$ are $\delta(\mathcal{F}) = \min_{v \in V(\mathcal{F})} d_{\mathcal{F}}(v)$ and $\Delta(\mathcal{F}) = \max_{v \in V(\mathcal{F})} d_{\mathcal{F}}(v)$.

Let $\mathcal{H}_1, \dots, \mathcal{H}_s \subset \{0,1\}^V$. We say that $\mathcal{F}_1, \dots, \mathcal{F}_s \subset \{0,1\}^X$ *cross contain* $\mathcal{H}_1, \dots, \mathcal{H}_s$ if there is an injection $\phi : V \to X$ such that $\phi(\mathcal{H}_i) \subset \mathcal{F}_i$ for all $i \in [s]$. Here we write $\phi(\mathcal{H}_i) = \{\phi(e) : e \in \mathcal{H}_i\}$ with each $\phi(e) = \{\phi(x) : x \in e\}$.

We simply say that $\mathcal{F}_1, \dots, \mathcal{F}_s$ cross contain $\mathcal{H}$ if $e(\mathcal{H}) = s$ and $\mathcal{F}_1, \dots, \mathcal{F}_s$ cross contain an ordering of the edges of $\mathcal{H}$, i.e. if $\mathcal{H} = \{e_i : i \in [s]\}$ then there is a permutation $\sigma \in S_s$ such that the hypergraphs $\mathcal{F}_1, \dots, \mathcal{F}_s$ cross contain $\{e_{\sigma(1)}\}, \dots, \{e_{\sigma(s)}\}$. Thus a single hypergraph $\mathcal{F}$ contains $\mathcal{H}$ if $\mathcal{F}_1, \dots, \mathcal{F}_s$ cross contain $\mathcal{H}$, where $\mathcal{F}_i = \mathcal{F}$ for all $i \in [s]$.

Given an r-graph G and $k \ge r$, we recall that the k-expansion $G^+ = G^+(k)$ is the k-uniform hypergraph obtained from G by adding $k - r$ new vertices to each edge, i.e. G^+ has edge set $\{e \cup S_e : e \in E(G)\}$ where $|S_e| = k - r$, $S_e \cap V(G) = \emptyset$ and $S_e \cap S_{e'} = \emptyset$ for all distinct $e, e' \in E(G)$.

When embedding expanded hypergraphs in uniform families, we may allow the uniformity of our families to vary, defining cross containment of G^+ in the obvious way: the edge of G^+ embedded in the family $\mathcal{F}_i \subset \binom{[n]}{k_i}$ is obtained from an edge of G by adding $k_i - r$ new vertices.

A family $\mathcal{F} \subset \{0,1\}^X$ is said to be *monotone* if given $F \in \mathcal{F}$ and $F \subset F' \subset X$ we also have $F' \in \mathcal{F}$. Given $\mathcal{F} \subset \{0,1\}^X$ the *up closure* of $\mathcal{F}$ is the monotone family $\mathcal{F}^{\uparrow} = \{B \subset X : A \subset B \text{ for some } A \in \mathcal{F}\} \subset \{0,1\}^X$. The ℓ-shadow of $\mathcal{F}$ is $\partial^{\ell}(\mathcal{F}) := \{F \in \binom{X}{\ell} : F \subset G \text{ for some } G \in \mathcal{F}\}$. We usually simply write $\partial(\mathcal{F})$ for $\partial^1(\mathcal{F})$.

Given $\mathcal{F} \subset \binom{X}{k}$ we will write $\mu(\mathcal{F}) = |\mathcal{F}| / \binom{|X|}{k}$. Some of our results are more naturally stated with $|\mathcal{F}|$ and others with $\mu(\mathcal{F})$, so we will freely move between these settings. Given $p \in [0,1]$ we will use μ_p to denote the p-biased measure on $\{0,1\}^n$, where a set $\boldsymbol{A} \sim \mu_p$ is selected by including each $i \in [n]$ independently with probability p. We extend this notation to families $\mathcal{F} \subset \{0,1\}^n$ by $\mu_p(\mathcal{F}) := \Pr_{\boldsymbol{A} \sim \mu_p}[\boldsymbol{A} \in \mathcal{F}]$. We often identify a family $\mathcal{F}$ with its characteristic Boolean function $f : \{0,1\}^n \to \{0,1\}$ and apply the above terminology freely in either setting, e.g. we call f monotone if $\mathcal{F}$ is monotone and write $\mu_p(f)$ for the expectation of f under μ_p.

To pass between these measures we note the following simple properties that will be henceforth used without further comment. For any $\mathcal{F} \subset \{0,1\}^n$ and $J \subset [n]$, we have the union bound estimate

$$\mu_p(\mathcal{F}) \le \mu_p(\mathcal{F}_J^{\emptyset}) + p \sum_{j \in J} \mu_p(\mathcal{F}_j^j) \le \mu_p(\mathcal{F}_J^{\emptyset}) + |J| p,$$

and in the opposite direction

$$\mu_p(\mathcal{F}) \geq (1-p)^{|J|}\mu_p(\mathcal{F}_J^\emptyset).$$

Similar estimates hold replacing μ_p by uniform measures μ for $\mathcal{F} \subset \binom{[n]}{k}$ with $k = pn$, remembering to use the correct normalisations: we have $\mu(\mathcal{F}) = |\mathcal{F}|\binom{n}{k}^{-1}$ and $\mu(\mathcal{F}_j^j) = |\mathcal{F}_j^j|\binom{n-1}{k-1}^{-1}$. This gives

$$\mu(\mathcal{F}) \leq \mu(\mathcal{F}_J^\emptyset) + \left(\tfrac{k}{n}\right)\sum_{j\in J}\mu(\mathcal{F}_j^j) \leq \mu(\mathcal{F}_J^\emptyset) + \tfrac{|J|k}{n}, \text{ and}$$

$$\mu(\mathcal{F}) \geq \binom{n}{k}^{-1}\binom{n-|J|}{k}\mu(\mathcal{F}_J^\emptyset) \geq \left(1 - \tfrac{|J|}{n-k}\right)^k \mu(\mathcal{F}_J^\emptyset).$$

Throughout $a \ll b$ or $a^{-1} \gg b^{-1}$ will mean that the following statement holds provided a is sufficiently small as a function of b.

Recall that $\mathcal{G}(r,s,\Delta)$ denotes the family of all r-graphs G with s edges and maximum degree $\Delta(G) \leq \Delta$. Throughout the remainder of the paper it will often be convenient to assume that G belongs to the subset $\mathcal{G}'(r,s,\Delta)$ of $\mathcal{G}(r,s,\Delta)$ consisting of its r-partite r-graphs. There is no loss of generality in this assumption, as $G^+(r\Delta)$ is $r\Delta$-partite for any $G \in \mathcal{G}(r,s,\Delta)$. To see this, consider a greedy algorithm in which we assign vertices of G sequentially to $r\Delta$ parts, ensuring for every edge that all of its vertices are in distinct parts. Clearly this algorithm can be completed. Then the expansion vertices can be assigned so that each edge of G^+ has one vertex in each part.

2.2 Pseudorandomness

Here we define our two key notions of pseudorandomness for set systems, namely uncapturability and globalness, and explore some of their basic properties.

Definition 2.1. Let $\mathcal{F} \subset \{0,1\}^n$ and μ be a measure on $\{0,1\}^n$.

We say $\mathcal{F}$ is (μ,a,ε)-uncapturable if $\mu(\mathcal{F}_J^\emptyset) \geq \varepsilon$ whenever $J \subset [n]$ with $|J| \leq a$.

We say $\mathcal{F}$ is (μ,a,ε)-global if $\mu(\mathcal{F}_J^J) \leq \varepsilon$ whenever $J \subset [n]$ with $|J| \leq a$.

We say $\mathcal{F}$ is (μ,a,ε)-capturable if it is not (μ,a,ε)-uncapturable, or (μ,a,ε)-local if it is not (μ,a,ε)-global. We omit μ from the notation if it is clear from the context, i.e. if $\mathcal{F} \subset \binom{[n]}{k}$ with uniform measure or $\mathcal{F} \subset \{0,1\}^n$ with p-biased measure μ_p, where p is clear from the context.

We now establish some basic properties of these definitions. For each property we state two lemmas that apply when μ is uniform or $\mu = \mu_p$. We only give proofs in the uniform setting, as those in the p-biased setting are essentially the same. The following pair of lemmas shows that globalness is preserved by restrictions.

Lemma 2.2. *If $\mathcal{F} \subset \binom{[n]}{k}$ is (a,ε)-global and $I \subset J \subset [n]$ with $|I| < a$ and $|J| < n/2k$ then $\mathcal{F}_J^I$ is $(a-|I|, 2\varepsilon)$-global.*

Lemma 2.3. *If $\mathcal{F} \subset \{0,1\}^n$ under μ_p is (a,ε)-global and $I \subset J \subset [n]$ with $|I| < a$ and $|J| < 1/2p$ then $\mathcal{F}_J^I$ is $(a-|I|, 2\varepsilon)$-global.*

Proof of Lemma 2.2. Let $K \subset [n] \setminus J$ with $|K| \leq a - |I|$. Then we have $\mu((\mathcal{F}_J^I)_K^K) = \mu((\mathcal{F}_{I\cup K}^{I\cup K})_{J\setminus I}^\emptyset) \leq \left(1 - \frac{|J\setminus I|}{n-k}\right)^{-k}\mu(\mathcal{F}_{I\cup K}^{I\cup K}) \leq 2\mu(\mathcal{F}_{I\cup K}^{I\cup K}) < 2\varepsilon$, using that $|I \cup K| \leq a$ and that $\mathcal{F}$ is (a,ε)-global. □

The next pair shows that globalness implies uncapturability.

Lemma 2.4. *If $\mathcal{F} \subset \binom{[n]}{k}$ is $(1,\varepsilon)$-global with $\varepsilon = \mu(\mathcal{F})n/2ak$ then $\mathcal{F}$ is $(a, \mu(\mathcal{F})/2)$-uncapturable.*

Lemma 2.5. *If $\mathcal{F} \subset \{0,1\}^n$ under μ_p is $(1,\varepsilon)$-global with $\varepsilon = \mu_p(\mathcal{F})/2ap$ then $\mathcal{F}$ is $(a, \mu_p(\mathcal{F})/2)$-uncapturable.*

Proof of Lemma 2.4. If $|J| \leq a$ then $\mu(\mathcal{F}_J^\emptyset) \geq \mu(\mathcal{F}) - \left(\frac{k}{n}\right)\sum_{j\in J}\mu(\mathcal{F}_j^j) \geq \mu(\mathcal{F}) - \left(\frac{k}{n}\right)|J|\varepsilon \geq \mu(\mathcal{F})/2$. □

Uncapturability does not imply globalness, but we do have a partial converse: by taking restrictions we can upgrade uncapturable families to families that are global or large.

Lemma 2.6. *Suppose $\beta \in (0, .1)$ and $\mathcal{F}_i \subset \binom{[n]}{k_i}$ with $2r < k_i < \beta n/2rm$ are (rm, δ_i)-uncapturable for $i \in [m]$. Then there are pairwise disjoint $S_1, \dots, S_m$ with each $|S_i| \le r$ such that, setting $\mathcal{G}_i = (\mathcal{F}_i)^{S_i}_S$ where $S = \bigcup_i S_i$, whenever $\mu(\mathcal{G}_i) < \beta$ we have $S_i = \emptyset$ and $\mathcal{G}_i$ is $(r, 2\beta)$-global with $\mu(\mathcal{G}_i) > \delta_i$.*

Lemma 2.7. *Suppose $\beta \in (0, .1)$ and $\mathcal{F}_i \subset \binom{[n]}{k_i}$ with $k_i < \beta n/2rm$ are (rm, δ_i)-uncapturable for $i \in [m]$. Then there are pairwise disjoint $S_1, \dots, S_m$ with each $|S_i| \le r$ such that, setting $\mathcal{G}_i = (\mathcal{F}_i^{\uparrow})^{S_i}_S$ where $S = \bigcup_i S_i$ and $p_i = k_i/(n - |S|)$, whenever $\mu_{p_i}(\mathcal{G}_i) < \beta$ we have $S_i = \emptyset$ and $\mathcal{G}_i$ is $(r, 2\beta)$-global with $\mu_{p_i}(\mathcal{G}_i) > \delta_i/4$.*

Proof of Lemma 2.6. Let $I \subset [m]$ be maximal such that there exists a collection of pairwise disjoint sets $(S_i : i \in I)$ with $|S_i| \le r$ and $\mu((\mathcal{F}_i)^{S_i}_{S_i}) > 1.5\beta$. Let $S = \bigcup_{i \in I} S_i$ and $\mathcal{G}_i = (\mathcal{F}_i)^{S_i}_S$ for each $i \in [m]$, where $S_i = \emptyset$ for $i \in [m] \setminus I$. For any $i \in I$ we have $\mu(\mathcal{G}_i) > \mu((\mathcal{F}_i)^{S_i}_{S_i}) - |S \setminus S_i| k_i/n > \beta$. Now consider i with $\mu(\mathcal{G}_i) < \beta$. Then $i \notin I$, so $S_i = \emptyset$ and $\mu(\mathcal{G}_i) > \delta_i$ by uncapturability. Furthermore, for any $R \subset [n] \setminus S$ with $|R| \le r$ we have $\mu((\mathcal{F}_i)^R_R) \le 1.5\beta$, so $(\mathcal{G}_i)^R_R = ((\mathcal{F}_i)^R_R)^{\emptyset}_S$ has $\mu((\mathcal{G}_i)^R_R) \le \left(1 - \frac{|S|}{n-k_i}\right)^{-k_i} \mu((\mathcal{F}_i)^R_R) < 2\beta$. □

We conclude this subsection with a lemma on decomposing any family according to its vertex degrees, where to make an analogy with the regularity method we think of high degree vertex links as 'structured' and the low degree remainder as 'pseudorandom'.

Lemma 2.8. *Let $\mathcal{F} \subset \binom{[n]}{k}$ and $J = \{i : \mu(\mathcal{F}^i_i) > \varepsilon\}$. If $|J| < n/2k$ then $\mathcal{G} = \mathcal{F}^{\emptyset}_J$ is $(1, 2\varepsilon)$-global, and so $(a, \mu(\mathcal{G})/2)$-uncapturable with $a = \mu(\mathcal{G})n/4k\varepsilon$.*

Proof. If $j \in [n] \setminus J$ then $\mu(\mathcal{F}^j_j) \le \varepsilon$ by definition of J, so $\mu(\mathcal{G}^j_j) = \mu((\mathcal{F}^j_j)^{\emptyset}_J) \le \left(1 - \frac{|J|}{n-k}\right)^{-k} \mu(\mathcal{F}^j_j) < 2\varepsilon$. The lemma follows by Definition 2.1 and Lemma 2.4. □

2.3 Embeddings

Here we will prove Theorem 1.8 assuming two fundamental embedding results, which will be proved in Section 4. The first of these shows that sufficiently large families contain a cross copy of any expanded hypergraph G^+. Our bound on $\mu(\mathcal{F}_i)$ is sharper for larger k_i: when $k_i = O(1)$ it is a constant, which is relatively weak (but still useful), whereas when $k_i \gg \log n$ it is $O(sk_i/n) = O(\sigma(G)k_i/n)$, which is tight up to the constant factor.

Lemma 2.9. *Let $G \in \mathcal{G}(r, s, \Delta)$, $C \gg r\Delta$ and $C \le k_i \le n/Cs$ for all $i \in [s]$. Then any $\mathcal{F}_i \subset \binom{[n]}{k_i}$ with $\mu(\mathcal{F}_i) \ge e^{-k_i/C} + Csk_i/n$ for all $i \in [s]$ cross contain G^+.*

When the uniformities k_i are small we cannot improve this cross containment result, as below density $e^{-\Omega(k_i)}$ the families $\mathcal{F}_i$ may have disjoint supports. However, when finding G^+ in a single family $\mathcal{F}$ we can get a much better bound on the density, and moreover it suffices to assume that $\mathcal{F}$ is sufficiently uncapturable, as follows.

Lemma 2.10. *Given $G \in \mathcal{G}(r, s, \Delta)$, $C \gg C_1 \gg C_2 \gg r\Delta$ and $C \le k \le n/Cs$, any $(C_1 s, sk/C_2 n)$-uncapturable $\mathcal{F} \subset \binom{[n]}{k}$ contains G^+.*

We conclude this section by deducing our junta approximation theorem from the above lemmas.

Proof of Theorem 1.8. Let $G \in \mathcal{G}(r, s, \Delta)$ and $C \gg C_1 \gg C_2 \gg r\Delta\varepsilon^{-1}$. Consider any G^+-free $\mathcal{F} \subset \binom{[n]}{k}$ with $C \le k \le \frac{n}{Cs}$. Let $J = \{i \in [n] : \mu(\mathcal{F}^i_i) \ge \beta\}$, where $\beta := e^{-k/C_1} + C_1 sk/n$. We need to show $|J| \le \sigma(G) - 1$ and $|\mathcal{F}^{\emptyset}_J| \le \varepsilon |\mathcal{S}_{n,k,\sigma(G)-1}|$.

The bound on $|J|$ follows from Lemma 2.9. Indeed, supposing for a contradiction $|J| \ge \sigma(G)$, we may fix a minimal crosscut S of G^+ and distinct $i_s \in J$ for each $s \in S$. Let $I = \{i_s : s \in S\}$ and

$\mathcal{F}_s := \mathcal{F}_I^{i_s}$ for $s \in S$. By definition of J, for each $s \in S$ we have $\mu(\mathcal{F}_s) > \beta - |I|k/n > \beta/2$, so by Lemma 2.9 the families $(\mathcal{F}_s : s \in S)$ cross contain the exclusive links $((G^+)^s_s : s \in S)$. However, this contradicts $\mathcal{F}$ being G^+-free.

As $|J| < s \le n/Ck$ we can apply Lemma 2.8 to see that $\mathcal{G} = \mathcal{F}_J^\emptyset$ is $(a, \mu(\mathcal{G})/2)$-uncapturable with $a = \mu(\mathcal{G})n/4k\beta$. However, by Lemma 2.10 $\mathcal{G}$ is $(C_1 s, sk/C_2 n)$-capturable, so we must have $\mu(\mathcal{G})/2 < sk/C_2 n$, or $a < C_1 s$, so again $\mu(\mathcal{G}) < 4\beta C_1 sk/n < sk/C_2 n$. As $\mu(\mathcal{S}_{n,k,\sigma(G)-1}) > .9(\sigma(G)-1)k/n$ and $s \le \Delta\sigma(G)$ we deduce $|\mathcal{F}_J^\emptyset| = |\mathcal{G}| < \varepsilon|\mathcal{S}_{n,k,\sigma(G)-1}|$. □

3 Matchings

The main result of this section is the following lemma on cross containment of matchings in uncapturable families, which will be used for 'lifting' (as described in Section 1.4) and also in the proof of the Huang–Loh–Sudakov Conjecture.

Lemma 3.1. *Let $C \gg C_1 \gg C_2 \gg 1$ and $\mathcal{F}_i \subset \binom{[n]}{k_i}$ with $k_i \le n/Cs$ for $i \in [s]$. Suppose $\mathcal{F}_i$ is $(C_1 m, mk_i/C_2 n)$-uncapturable for $i \in [m]$ and $\mu(\mathcal{F}_i) > C_1 sk_i/n$ for $i > m$. Then $\mathcal{F}_1, \dots, \mathcal{F}_s$ cross contain a matching.*

We start in the first subsection by recalling some basic probabilistic tools, and also our new sharp threshold result from [29]. Next we present some extremal results on cross matchings in the second subsection. We conclude by proving the uncapturability result in the third subsection.

3.1 Probabilistic tools and sharp thresholds

We start with the following lemma that will be used to pass between the uniform and p-biased measures.

Lemma 3.2. *Let $n, k \in \mathbb{N}$ with $k = pn \le n$. Then $\mathbb{P}\big(\mathrm{Bin}(n,p) \ge k\big) \ge 1/4$. Thus if $\mathcal{A} \subset \binom{[n]}{k}$ we have $\mu_p(\mathcal{A}^\uparrow) \ge \mu(\mathcal{A})/4$.*

Proof. The first statement appears in [23]. With $\alpha := \mu(\mathcal{A})$, the second holds as $\big|\mathcal{A}^\uparrow \cap \binom{[n]}{j}\big| \ge \alpha\binom{n}{j}$ for $j \ge k$ by the LYM inequality, and so we have $\mu_p(\mathcal{A}^\uparrow) \ge \sum_{j=k}^n \mathbb{P}\big(\mathrm{Bin}(n,p) = j\big)\mu\big(\mathcal{A}^\uparrow \cap \binom{[n]}{j}\big) \ge \mathbb{P}\big(\mathrm{Bin}(n,p) \ge k\big)\alpha \ge \alpha/4$. □

We will also need the following well-known Chernoff bound (see [25, Theorem 2.8]), as applied to sums of Bernoulli random variables, i.e. random variables which take values in $\{0,1\}$; if these are identically distributed then we obtain a binomial variable. The inequality can also be applied to a hypergeometric random variable (see [25, Remark 2.11]), i.e. $|S \cap T|$ with $S \in \binom{X}{s}$ and uniformly random $T \in \binom{X}{t}$ for some X, s and t.

Lemma 3.3. *Let X be a sum of independent Bernoulli random variables and $0 < a < 3/2$. Then $\mathbb{P}\big[|X - \mathbb{E}X| \ge a\mathbb{E}X\big] \le 2e^{-\frac{a^2}{3}\mathbb{E}X}$.*

Next we state our sharp threshold result for global functions which will play a crucial role in this section, and so for all subsequent applications of Lemma 3.1.

Theorem 3.4. *[29, Theorem 1.9] For any $\zeta > 0$ there is $C_0 > 0$ such that for any $\varepsilon, p, q \in (0, 1/2)$ with $q \ge (1+\zeta)p$ and $C > C_0$, writing $r = C\log\varepsilon^{-1}$ and $\delta = C^{-r}$, any monotone (μ_p, r, δ)-global $\mathcal{F} \subset \{0,1\}^n$ with $\mu_p(\mathcal{F}) \le \delta$ satisfies $\mu_q(\mathcal{F}) \ge \mu_p(\mathcal{F})/\varepsilon$.*

We will apply the following two consequences of this result.

Theorem 3.5. *Suppose $\mathcal{F} \subset \{0,1\}^n$ is monotone with $\mu_p(\mathcal{F}) = \mu$.*

1. *If $\mu \ll r^{-1} \ll \varepsilon$ then there is $R \subset [n]$ with $|R| \le r$ and $\mu_{2p}(\mathcal{F}_R^R) \ge \mu/\varepsilon$.*

2. If $p \ll K^{-1} \ll \eta \ll 1$ then there is $R \subset [n]$ with $|R| \le K \log \mu^{-1}$ and $\mu_{Kp}(\mathcal{F}^R_R) \ge \mu^\eta$.

Proof. For (1) we apply Theorem 3.4 with $\zeta = 1$ and the same ε. If $\mathcal{F}$ is not (r,δ)-global then for some R with $|R| \ge r$ we have $\mu_{2p}(\mathcal{F}^R_R) \ge \mu_p(\mathcal{F}^R_R) \ge \delta \ge \mu/\varepsilon$. On the other hand, if $\mathcal{F}$ is (r,δ)-global then we can take $R = \emptyset$, as Theorem 3.4 gives $\mu_{2p}(\mathcal{F}) \ge \mu/\varepsilon$.

For (2), we repeatedly apply Theorem 3.4 with $\zeta = 1$ and $\varepsilon = \mu^{\eta^2}$, so $r = C \log \varepsilon^{-1} = C\eta^2 \log \mu^{-1}$ and $\delta = C^{-r} = \mu^{\eta^2 C \log C} \ge \mu^\eta$, as we may assume $\eta \ll C^{-1}$. We can assume that $\mathcal{F}$ is (r,δ)-global, otherwise we immediately obtain R as required, so $\mu_{2p}(\mathcal{F}) \ge \mu/\varepsilon = \mu^{1-\eta^2}$. Repeating the argument, if we do not find R then after $t \le \eta^{-2}$ iterations we reach $\mu_{2^t p}(\mathcal{F}) \ge \delta \ge \mu^\eta$, so we can take $R = \emptyset$. $\square$

3.2 Extremal results

In this subsection we adapt the method of [24, Lemma 3.1] to prove a variant form of the following result of Huang, Loh and Sudakov [24].

Lemma 3.6. *Let $k_1, \dots, k_s, n \in \mathbb{N}$ with $\sum_{i \in [s]} k_i \le n$. Suppose $\mathcal{F}_i \subset \binom{[n]}{k_i}$ for all $i \in [s]$ and that $\mathcal{F}_1, \dots, \mathcal{F}_s$ do not cross contain a matching. Then $\mu(\mathcal{F}_i) \le k_i(s-1)/n$ for some $i \in [s]$.*

We will prove the following variant that allows a few families to be significantly smaller.

Lemma 3.7. *Let $1 \le m \le s$, $k_1, \dots, k_s \ge 0$ and $n \ge \sum_{i \in [s]} k_i$. Suppose $\mathcal{F}_i \subset \binom{[n]}{k_i}$ with $\mu(\mathcal{F}_i) > 2k_i m/n$ for $i \in [m]$ and $\mu(\mathcal{F}_i) > 2k_i s/n$ for $i \in [m+1, s]$. Then $\{\mathcal{F}_i\}_{i \in [s]}$ cross contain a matching.*

We also require the following version for the p-biased measure, which we will deduce from Lemma 3.7 by a limit argument similar to those in [5, 18].

Lemma 3.8. *Let $m \le s$ and $p_1, \dots, p_s > 0$ with $\sum_{i \in [s]} p_i \le 1/2$. Suppose that $\mathcal{F}_1, \dots \mathcal{F}_s \subset \{0,1\}^n$ are monotone families with $\mu_{p_i}(\mathcal{F}_i) \ge 3mp_i$ for $i \in [m]$ and $\mu_{p_i}(\mathcal{F}_i) \ge 3sp_i$ for $i \in [m+1, s]$. Then $\{\mathcal{F}_i\}_{i \in [s]}$ cross contain a matching.*

We introduce the following terminology. Given $\mathbf{a} = (a_1, \dots, a_s) \in \mathbb{R}^s$ and $n, k_1, \dots, k_s \ge 0$ we say $\mathbf{a}$ is *forcing* for $(n, k_1, \dots, k_s)$ if any families $\mathcal{F}_1, \dots, \mathcal{F}_s$ with $\mathcal{F}_i \subset \binom{[n]}{k_i}$ and $\mu(\mathcal{F}_i) > \frac{a_i k_i}{n}$ for all $i \in [s]$ cross contain an s-matching. We say $\mathbf{a} = (a_1, \dots, a_s) \in \mathbb{R}^s$ is *forcing* if it is forcing for $(n, k_1, \dots, k_s)$ whenever $n \ge \sum_{i \in [s]} k_i$ and *exactly forcing* if it is forcing for $(n, k_1, \dots, k_s)$ whenever $n = \sum_{i \in [s]} k_i$. Any forcing sequence is clearly exactly forcing; we establish the converse.

Lemma 3.9. *A sequence $\mathbf{a} \in \mathbb{R}^s$ is forcing if and only if it is exactly forcing.*

We require the following compression operators. Given distinct $i, j \in [n]$ and $F \subset [n]$, we let

$$C_{i,j}(F) := \begin{cases} (F \setminus \{j\}) \cup \{i\} & \text{if } j \in F, i \notin F; \\ F & \text{otherwise.} \end{cases}$$

Given $\mathcal{F} \subset \{0,1\}^n$, we let $C_{i,j}(\mathcal{F}) = \{C_{i,j}(F) : F \in \mathcal{F}\} \cup \{F \in \mathcal{F} : C_{i,j}(F) \in \mathcal{F}\}$. We say $\mathcal{F}$ is $C_{i,j}$-compressed if $C_{i,j}(\mathcal{F}) = \mathcal{F}$.

Proof of Lemma 3.9. A forcing sequence is clearly exactly forcing, so it remains to prove the converse. We argue by induction on s; the base case $s = 1$ is clear. Suppose that $\mathbf{a} \in \mathbb{R}^s$ is exactly forcing. We fix $k_1, \dots, k_s \ge 0$ and show by induction on $n \ge \sum_{i \in [s]} k_i$ that $\mathbf{a}$ is forcing for $(n, k_1, \dots, k_s)$, i.e. any families $\mathcal{F}_1, \dots, \mathcal{F}_s$ with $\mathcal{F}_i \subset \binom{[n]}{k_i}$ and $\mu(\mathcal{F}_i) > \frac{a_i k_i}{n}$ for all $i \in [s]$ cross contain an s-matching. The base case $n = \sum_{i \in [s]} k_i$ holds as $\mathbf{a}$ is exactly forcing.

First suppose $k_i = 0$ for some $i \in [s]$; without loss of generality $i = s$. Then $\mathbf{a}' = (a_1, \dots, a_{s-1})$ is exactly forcing, and so forcing by induction on s. Thus $\mathcal{F}_1, \dots, \mathcal{F}_{s-1}$ cross contain an $(s-1)$-matching. Combined with $\emptyset \in \mathcal{F}_s$ we find a cross s-matching in $\mathcal{F}_1, \dots, \mathcal{F}_s$, as required.

We may now assume $k_i \geq 1$ for all $i \in [s]$. We suppose for contradiction that $\mathcal{F}_1, \ldots, \mathcal{F}_s$ do not cross contain an s-matching. Let $\mathcal{G}_1, \ldots, \mathcal{G}_s$ be obtained from $\mathcal{F}_1, \ldots, \mathcal{F}_s$ by successively applying the compression operators $C_{1,n}, C_{2,n}, \ldots, C_{n-1,n}$. As is well-known (e.g. see [24, Lemma 2.1 (iii)]), $\mathcal{G}_1, \ldots, \mathcal{G}_s$ do not cross contain an s-matching and are $C_{j,n}$-compressed for all $j \in [n-1]$. For each $i \in [s]$ let

$$\mathcal{G}_i(n) := \big\{A \subset [n-1] : A \cup \{n\} \in \mathcal{G}_i\big\} \subset \tbinom{[n-1]}{k_i-1};$$
$$\mathcal{G}_i(\overline{n}) := \big\{A \subset [n-1] : A \in \mathcal{G}_i\big\} \subset \tbinom{[n-1]}{k_i}.$$

We now claim that if $I \subset [s]$ then $\{\mathcal{H}_i\}_{i\in[s]}$ are cross free of an s-matching, where $\mathcal{H}_i = \mathcal{G}_i(n)$ for $i \in I$ and $\mathcal{H}_i = \mathcal{G}_i(\overline{n})$ for $i \notin I$. For contradiction, suppose $\{A_i\}_{i\in[s]}$ is such a cross matching in $\{\mathcal{H}_i\}_{i\in[s]}$. Then $A_i \cup \{n\} \in \mathcal{G}_i$ for all $i \in I$ and $A_i \in \mathcal{G}_i$ for $i \notin I$. However, as $\mathcal{G}_i$ is $C_{j,n}$-compressed for all $j \in [n-1]$ and $n > \sum_{i\in[s]} k_i$, there are distinct $j_i \in [n] \setminus \big(\cup_{i\in[s]} A_i\big)$ for all $i \in I$ such that $A_i \cup \{j_i\} \in \mathcal{G}_i$. Then $\{A_i \cup \{j_i\}\}_{i\in I} \cup \{A_i\}_{i\in[s]\setminus I}$ is a cross s-matching in $\{\mathcal{G}_i\}_{i\in[s]}$, a contradiction. Thus the claim holds.

By induction on n, it now suffices to show that for each $i \in [s]$ either $\mu(\mathcal{G}_i(n)) > a_i(k_i-1)/(n-1)$ or $\mu(\mathcal{G}_i(\overline{n})) > a_i k_i/(n-1)$; indeed, we then obtain the required contradiction by setting $I = \{i \in [s] : \mu(\mathcal{G}_i(n)) > a_i(k_i-1)/(n-1)\}$ in the above claim. But this is clear, as otherwise

$$\frac{a_i k_i}{n} < \mu(\mathcal{G}_i) = \Big(\frac{n-k_i}{n}\Big)\mu(\mathcal{G}_i(\overline{n})) + \Big(\frac{k_i}{n}\Big)\mu(\mathcal{G}_i(n)) \leq \Big(\frac{n-k_i}{n}\Big)\Big(\frac{a_i k_i}{n-1}\Big) + \Big(\frac{k_i}{n}\Big)\Big(\frac{a_i(k_i-1)}{n-1}\Big) = \frac{a_i k_i}{n},$$

a contradiction. This completes the proof. □

We conclude this subsection by deducing Lemmas 3.7 and 3.8.

Proof of Lemma 3.7. By Lemma 3.9 it suffices to prove the statement under the assumption $n = \sum_{i\in[s]} k_i$. Note first that if $n = 0$ then $\mathcal{F}_i = \{\emptyset\}$ for all $i \in [s]$ which clearly cross contain an s-matching. Thus we may assume $n > 0$. For any $i \in [m]$ we have $2k_i m/n < \mu(\mathcal{F}_i) \leq 1$, so $k_i < n/2m$, and similarly $k_i < n/2s$ for $i \in [m+1, s]$. But now $n = \sum_{i\in[s]} k_i < m \cdot n/2m + (s-m) \cdot n/2s < n$ is a contradiction. □

Proof of Lemma 3.8. Let $N^{-1} \ll \varepsilon \ll \min_{i\in[s]} p_i$ and $\mathcal{G}_i = \mathcal{F}_i \times \{0,1\}^{[N]\setminus[n]} \subset \{0,1\}^N$ for each $i \in [s]$. Then each $\mu_{p_i}(\mathcal{G}_i) = \mu_{p_i}(\mathcal{F}_i)$. Writing $I_i = \big[(1-\varepsilon)Np_i, (1+\varepsilon)Np_i\big]$, by Lemma 3.3 each $\mu_{p_i}\big(\cup_{k\notin I_i} \binom{[N]}{k}\big) < \varepsilon$, so there are $k_i \in I_i$ such that each $\mu\big(\mathcal{G}_i \cap \binom{[N]}{k_i}\big) > \mu_{p_i}(\mathcal{F}_i) - \varepsilon$, which is at least $2mk_i/N$ for $i \in [m]$ and $2sk_i/N$ for $i \in [m+1, s]$. The result now follows from Lemma 3.7. □

3.3 Capturability

In this subsection we conclude this section by proving its main lemma on cross matchings in uncapturable families. The idea of the proof is to take suitable restrictions that boost the measure of the families so that we can apply the extremal result from the previous subsection. However, uncapturability is not preserved by restrictions, so we first upgrade to globalness, which is preserved by restrictions. We also pass from the setting of uniform families to that of biased measures, which allows us to apply our sharp threshold result, and also has the technical advantage that we do not need to assume any lower bound on the uniformity of our families.

Proof of Lemma 3.1. Let $C \gg C_1 \gg C_2 \gg 1$ and $\mathcal{F}_i \subset \binom{[n]}{k_i}$ with $k_i \leq n/Cs$ for $i \in [s]$. Suppose $\mathcal{F}_i$ is $(C_1 m, mk_i/C_2 n)$-uncapturable for $i \in [m]$ and $\mu(\mathcal{F}_i) > C_1 sk_i/n$ for $i > m$. We need to show that $\mathcal{F}_1, \ldots, \mathcal{F}_s$ cross contain a matching.

We start by upgrading uncapturability to globalness and moving to biased measures. By Lemma 2.7 with $r = C_1$ and $\beta = C_1^{-2}$ there are pairwise disjoint $S_1, \ldots, S_m$ with each $|S_i| \leq r$ such that, setting $\mathcal{G}_i = (\mathcal{F}_i^{\uparrow})^{S_i}_S$ where $S = \bigcup_i S_i$ and $p_i = k_i/(n-|S|)$, whenever $\mu_{p_i}(\mathcal{G}_i) < C_1^{-2}$ we have $S_i = \emptyset$

and $\mathcal{G}_i$ is $(C_1, 2C_1^{-2})$-global with $\mu_{p_i}(\mathcal{G}_i) > mk_i/4C_2n > mp_i/5C_2$. We note by Lemma 2.5 that $\mathcal{G}_i$ is $(a, mp_i/10C_2)$-uncapturable, where $a = (mp_i/5C_2)/(4p_iC_1^{-2}) > C_1m$.

Next we will choose pairwise disjoint $R_1, \dots, R_m \subset [n] \setminus S$ with each $|R_i| < C_1/8$, write $R_{<j} = \bigcup_{i<j} R_i$, and define families $\mathcal{G}_i^j$ by $\mathcal{G}_i^j = (\mathcal{G}_i)^{\emptyset}_{R_{<j}}$ for $i \geq j$ or $\mathcal{G}_i^j = (\mathcal{G}_i)^{R_i}_{R_{<j}}$ for $i < j$.

We claim that we can choose each R_i to ensure $\mu_{2p_i}(\mathcal{G}_i^i) \geq 7mp_i$. To see this, first note that $\mathcal{G}_i^{i-1} = (\mathcal{G}_i)^{\emptyset}_{R_{<i}}$ has $\mu_{p_i}(\mathcal{G}_i^{i-1}) \geq mp_i/10C_2$ by uncapturability. If $\mu_{p_i}(\mathcal{G}_i^{i-1}) \geq 7mp_i$ we let $R_i = \emptyset$ to obtain $\mu_{2p_i}(\mathcal{G}_i^i) = \mu_{2p_i}(\mathcal{G}_i^{i-1}) \geq \mu_{p_i}(\mathcal{G}_i^{i-1}) \geq 7mp_i$. Otherwise, as $mp_i < 2C^{-1} \ll C_1^{-1} \ll C_2^{-1}$ we can apply Theorem 3.5.1 with $\varepsilon^{-1} = 70C_2$ and $r = C_1/8$ to choose R_i with $|R_i| \leq r$ so that $\mathcal{G}_i^i = (\mathcal{G}_i^{i-1})^{R_i}_{R_i}$ has $\mu_{2p_i}(\mathcal{G}_i^i) > \mu_{p_i}(\mathcal{G}_i^{i-1})/\varepsilon \geq 7mp_i$. Either way the claim holds.

By Lemma 2.3 each $\mathcal{G}_i^i$ with $i \in [m]$ is $(C_1/2, 4C_1^{-2})$-global, so $\mathcal{G}_i^m = (\mathcal{G}_i^i)^{\emptyset}_{\bigcup_{j>i} R_j}$ has $\mu_{2p_i}(\mathcal{G}_i^m) \geq \mu_{2p_i}(\mathcal{G}_i^i) - m(C_1/8) \cdot 4C_1^{-2} \cdot 2p_i \geq 3m(2p_i)$. For $i > m$ we have $\mu(\mathcal{F}_i) > C_1sk_i/n$, so $\mu_{p_i}(\mathcal{G}_i^i) > \mu_{p_i}(\mathcal{F}_i)/4 - m(C_1/8)p_i > 3sp_i$. By Lemma 3.8, $\mathcal{G}_1^m, \dots, \mathcal{G}_s^m$ cross contain a matching; hence so do $\mathcal{F}_1, \dots, \mathcal{F}_s$. □

4 Shadows and embeddings

In this section we will complete the proof of our junta approximation theorem by implementing the strategy described above of finding embeddings in fat shadows. We start in the first subsection by defining and analysing fat shadows. In the second subsection we find shadow embeddings. We then conclude in the final subsection with lifted embeddings (using the lifting result from the previous section) that prove Lemmas 2.9 and 2.10, thus proving Theorem 1.8.

4.1 Fat shadows

In this subsection we present various lower bounds on the density of fat shadows, defined as follows.

Definition 4.1. The c-fat r-shadow of $\mathcal{F} \subset \binom{[n]}{k}$ is $\partial_c^r \mathcal{F} := \{A \in \binom{[n]}{r} : \mu(\mathcal{F}_A^A) \geq c\}$.
The c-fat shadow of $\mathcal{F}$ is $\partial_c \mathcal{F} := \bigcup_{r \leq k} \partial_c^r \mathcal{F}$.

The following simple 'Markov' bound is useful when $\mathcal{F}$ is nearly complete.

Lemma 4.2. *If $\mu(\mathcal{F}) \geq 1 - cc'$ then $\mu(\partial_{1-c}^r \mathcal{F}) \geq 1 - c'$.*

Proof. Consider uniformly random $A \subset B \subset [n]$ with $|A| = r$ and $|B| = k$. For any $A \notin \partial_{1-c}^r \mathcal{F}$ we have $\mathbb{P}(B \notin \mathcal{F} \mid A) \geq c$, so $cc' \geq \mathbb{P}(B \notin \mathcal{F}) \geq c \cdot \mathbb{P}(A \notin \partial_{1-c}^r \mathcal{F}) = c(1 - \mu(\partial_{1-c}^r \mathcal{F}))$. □

Another bound is given the following Fairness Proposition of Keller and Lifshitz [30].

Proposition 4.3 (Fairness Proposition)**.** *Let $C \gg r/\varepsilon$ and $\mathcal{F} \subset \binom{[n]}{k}$ with $k \geq r$ and $\mu(\mathcal{F}) \geq e^{-k/C}$. For $c = (1-\varepsilon)\mu(\mathcal{F})$ we have $\mu(\partial_c^r \mathcal{F}) \geq 1 - \varepsilon$.*

When the above bounds are not applicable we rely on the following lemma, whose proof will occupy the remainder of this subsection.

Lemma 4.4. *Let $\mathcal{F} \subset \binom{[n]}{k}$, $r < \ell \leq k$ and $\mathcal{H} = \{B \in \binom{[n]}{\ell} : \partial^r B \subset \partial_c^r \mathcal{F}\}$, where $c = \mu(\mathcal{F})/2\binom{\ell}{r}$. Then $\mu(\mathcal{H}) \geq \mu(\mathcal{F})/2$. Thus $\mu(\partial_c^r \mathcal{F}) \geq (\mu(\mathcal{F})/2)^{r/\ell}$. Furthermore, if $G \in \mathcal{G}'(r, s, \Delta)$, $C \gg r\Delta$ and $\partial_c^r \mathcal{F}$ is G-free then $\mu(\partial_c^r \mathcal{F}) \geq \left((\mu(\mathcal{F})/2 - (s/n)^{\ell/C})n/s\ell^2\right)^{r/(\ell-1)}$.*

We require several further lemmas for the proof of Lemma 4.4. We start by stating a consequence of the Lovász form [34] of the Kruskal–Katona theorem [26, 33].

Lemma 4.5. *If $1 \leq \ell \leq k \leq n$ and $\mathcal{A} \subset \binom{[n]}{k}$ then $\mu(\partial^\ell(\mathcal{A})) \geq \mu(\mathcal{A})^{\ell/k}$.*

Proof. Define $\beta \in [0,1]$ by $|\mathcal{A}| = \binom{\beta n}{k}$, so that $\mu(\mathcal{A}) = \prod_{i=0}^{k-1}(\beta - i/n)$. By the Lovász form of Kruskal–Katona (Problem 13.31(b) in [34]), we have $|\partial^\ell \mathcal{A}| \geq \binom{\beta n}{\ell}$, so $\mu(\partial^\ell(\mathcal{A}))^k \geq \prod_{i=0}^{\ell-1}(\beta - i/n)^k \geq \mu(\mathcal{A})^\ell$. □

Next we require an estimate on the Turán numbers of r-partite r-graphs, which follows from [3, Theorem 2] due to Conlon, Fox and Sudakov. (Recall that $\mathcal{G}'(r,s,\Delta)$ is the family of r-partite r-graphs with s edges and maximum degree Δ.)

Theorem 4.6. *Let $F \in \mathcal{G}'(r,s,\Delta)$ and $C \gg r\Delta$. Then any F-free $\mathcal{H} \subset \binom{[n]}{r}$ with $n > Cs$ has $\mu(\mathcal{H}) < (s/n)^{1/C}$.*

We note that the following lemma is immediate from Theorem 4.6 and Lemma 4.5.

Lemma 4.7. *Let $G \in \mathcal{G}'(r,s,\Delta)$, $C \gg r\Delta$, $C \leq k \leq n/Cs$ and $\mathcal{F} \subset \binom{[n]}{k}$. If $\partial^r \mathcal{F}$ is G-free then $\mu(\mathcal{F}) \leq (s/n)^{k/C}$.*

Our next lemma is an adaptation of one due to Kostochka, Mubayi and Verstraëte [31].

Lemma 4.8. *Suppose $G \in \mathcal{G}'(r,s,\Delta)$, $C \gg r\Delta$ and $\mathcal{F}$ is a G^+-free k-graph on $[n]$. Then $\mu(\partial \mathcal{F}) \geq (\mu(\mathcal{F}) - (s/n)^{k/C})n/sk^2$.*

Proof. We define $\mathcal{G} \subset \mathcal{F}$ by starting with $\mathcal{G} = \mathcal{F}$ and then repeating the following procedure: if there is any $A \in \partial \mathcal{G}$ with $|\mathcal{G}_A^A| \leq ks$ then remove from $\mathcal{G}$ all edges containing A. This terminates with some $\mathcal{G}$ such that $|\mathcal{G}_A^A| > ks$ for all $A \in \partial \mathcal{G}$ and $|\mathcal{G}| \geq |\mathcal{F}| - ks|\partial \mathcal{F}|$, so $\mu(\partial \mathcal{F}) \geq (\mu(\mathcal{F}) - \mu(\mathcal{G}))n/sk^2$.

We will now show that $\partial^r \mathcal{G}$ is G-free, which will complete the proof due to Lemma 4.7. To see this, we suppose that $\phi(G)$ is a copy of G in $\partial^r \mathcal{G}$ and will obtain a contradiction by finding a copy of G^+ in $\mathcal{G}$. To do so, we start by fixing for each edge A of G an edge e_A of $\mathcal{G}$ containing $\phi(A)$. Then we repeat the following procedure: while some e_A contains some $\phi(x)$ with $x \notin A$, replace e_A by some edge $(e_A \setminus \{\phi(x)\}) \cup \{v\}$ with $v \notin \operatorname{Im} \phi$. As $|\mathcal{G}_A^A| > ks$ for all $A \in \partial \mathcal{G}$ we can always choose v as required. The procedure terminates with a copy of G^+, so the proof is complete. □

We conclude this subsection with the proof of its main lemma.

Proof of Lemma 4.4. Consider uniformly random (A,B,C) with $C \subset B \subset A \subset [n]$ and $|C| = r$, $|B| = \ell$, $|A| = k$. Write $p = \mathbb{P}(A \in \mathcal{F}, C \notin \partial_c^r \mathcal{F})$ and $q = \mathbb{P}(A \in \mathcal{F}, B \notin \mathcal{H})$.

For any $C \notin \partial_c^r \mathcal{F}$ we have $\mathbb{P}(A \in \mathcal{F} \mid C) = \mu(\mathcal{F}_C^C) \leq c$, so $p \leq c$. On the other hand, $p \geq q\binom{\ell}{r}^{-1}$, as for any $A \in \mathcal{F}$ and $B \notin \mathcal{H}$ we have $\mathbb{P}(C \notin \partial_c^r \mathcal{F} \mid A, B) \geq \binom{\ell}{r}^{-1}$. We deduce $q \leq \binom{\ell}{r}c = \mu(\mathcal{F})/2$.

Thus $\mu(\mathcal{H}) = \mathbb{P}(B \in \mathcal{H}) \geq \mathbb{P}(A \in \mathcal{F}) - q \geq \mu(\mathcal{F})/2$.

As $\partial^r \mathcal{H} \subset \partial_c^r \mathcal{F}$, Lemma 4.5 gives $\mu(\partial_c^r \mathcal{F}) \geq (\mu(\mathcal{F})/2)^{r/\ell}$.

Now suppose $G \in \mathcal{G}'(r,s,\Delta)$ and $\partial_c^r \mathcal{F}$ is G-free. Then $\mathcal{H}$ is G^+-free, so Lemma 4.8 gives $\mu(\partial \mathcal{H}) \geq (\mu(\mathcal{H}) - (s/n)^{\ell/C})n/s\ell^2$. As $\partial^r \partial \mathcal{H} \subset \partial_c^r \mathcal{F}$, Lemma 4.5 gives the required bound. □

4.2 Shadow embeddings

The following lemma implements a simple greedy algorithm for cross embedding any bounded degree r-graph in a collection of nearly complete r-graphs (more generally, we also allow smaller edges).

Lemma 4.9. *Let $0 < \eta \ll (r\Delta)^{-1}$ and $G = \{e_1, \dots, e_s\}$ be a hypergraph of maximum degree Δ with each $|e_i| = r_i \leq r$. Suppose for each $i \in [s]$ that $\mathcal{G}_i$ is an r_i-graph on $[n]$, where $n \geq 2rs$ and $\mu(\mathcal{G}_i) > 1 - \eta$. Then $\mathcal{G}_1, \dots, \mathcal{G}_s$ cross contain G.*

Proof. Write $V(G) = \{v_1, \dots, v_m\}$. We may assume that G has no isolated vertices, so $m \leq \sum_i d_G(v_i) \leq rs \leq n/2$. We will construct an injection $\phi : V(G) \to [n]$ such that each $\phi(e_j) \in \mathcal{G}_j$. To do so, we define ϕ sequentially so that, for each $0 \leq t \leq m$ the definition of ϕ on $V_t := \{v_i : i \leq t\}$ is *t-good*, meaning that for each edge e_j we have

$$\phi(e_j \cap V_t) \in \partial_{c_{jt}} \mathcal{G}_j, \text{ where } c_{jt} = 1 - \eta(2\Delta)^{|e_j \cap V_t|}. \tag{1}$$

Note that (1) holds whenever $e_j \cap V_t = \emptyset$, as $\mu(\mathcal{G}_j) > 1-\eta$; in particular, (1) holds when $t=0$.

It remains to show for any $0 \le t < m$ that we can extend any t-good embedding ϕ to a $(t+1)$-good embedding. To see this, first note that we only need to check (1) when e_j is one of at most Δ edges containing v_{t+1}. Fix any such edge e_j, let $f = \phi(e_j \cap V_t)$, and let B_j be the set of $x \in [n]$ such that choosing $\phi(v_{t+1}) = x$ would give $\phi(e_j \cap V_{t+1}) = f \cup \{x\} \notin \partial_{c_{j(t+1)}} \mathcal{G}_j$. Then

$$|B_j|\eta(2\Delta)^{|f|+1} \le \sum_{x \in B} \left(1 - \mu\big((\mathcal{G}_j)^{f\cup\{x\}}_{f\cup\{x\}}\big)\right) \le n(1-\mu((\mathcal{G}_j)^f_f)) < n\eta(2\Delta)^{|f|},$$

so $|B_j| < n/2\Delta$. Summing over at most Δ choices of j forbids fewer than $n/2$ choices of x. The requirement that ϕ be injective also forbids fewer than $n/2$ vertices, so we can extend ϕ as required. □

4.3 Lifted embeddings

We conclude this section by proving the two embedding lemmas assumed above, thus completing the proof of Theorem 1.8.

Proof of Lemma 2.9. Suppose $n, s, k_1, \dots, k_s \in \mathbb{N}$ with $C \le k_i \le \frac{n}{Cs}$ for all $i \in [s]$, and $\mathcal{F}_i \subset \binom{[n]}{k_i}$ with each $\mu(\mathcal{F}_i) \ge e^{-k_i/C} + Csk_i/n$. Let η be as in Lemma 4.9. We can assume C is large enough so that Proposition 4.3 gives $\mu(\mathcal{G}_i) \ge 1-\eta$ for each $i \in [s]$, where $\mathcal{G}_i$ is the r-graph on $[n]$ consisting of all $e \in \binom{[n]}{r}$ with $\mu((\mathcal{F}_i)^e_e) \ge Csk_i/2n$. By Lemma 4.9 we can find $R_1, \dots, R_s$ forming a copy of G with $R_i \in \mathcal{G}_i$ for all $i \in [s]$. Let $R = R_1 \cup \dots \cup R_s$. By the union bound, each $\mu\big((\mathcal{F}_i)^{R_i}_R\big) \ge \mu\big((\mathcal{F}_i)^{R_i}_{R_i}\big) - |R|k_i/n \ge Csk_i/4n$ for $C \ge 8$, so Lemma 3.6 gives a cross matching $E_1, \dots, E_s$ in $(\mathcal{F}_1)^{R_1}_R, \dots, (\mathcal{F}_s)^{R_s}_R$. Now $\mathcal{F}_1, \dots, \mathcal{F}_s$ cross contain a copy of G^+ with edges $R_1 \cup E_1, \dots, R_s \cup E_s$. □

Proof of Lemma 2.10. Let $G \in \mathcal{G}(r,s,\Delta)$ and $C \gg C_1 \gg C_2 \gg r\Delta$. Suppose for a contradiction that $\mathcal{F} \subset \binom{[n]}{k}$ with $C \le k \le n/Cs$ is $(C_1 s, sk/C_2 n)$-uncapturable but G^+-free.

Let $\mathcal{B}$ be a maximal collection of pairwise disjoint sets where each $B \in \mathcal{B}$ has $|B| \le r+1$ and $\mu(\mathcal{F}^B_B) > \beta := e^{-k/C_1} + C_1 sk/n$. We claim that $|\mathcal{B}| < s$. To see this, suppose for a contradiction that we have distinct $B_1, \dots, B_s$ in $\mathcal{B}$. Let $B = \bigcup_{i=1}^s B_i$ and $\mathcal{F}_i = \mathcal{F}^{B_i}_B$ for $i \in [s]$. Then each $\mu(\mathcal{F}_i) > \beta - |B|k/n > e^{-k/C_1} + C_1 sk/2n$. Now Lemma 2.9 gives a cross copy of G^+ in $\mathcal{F}_1, \dots, \mathcal{F}_s$, contradicting $\mathcal{F}$ being G^+-free, so $|\mathcal{B}| < s$, as claimed.

Now let $\mathcal{G} = \mathcal{F}^\emptyset_B$ with $B = \bigcup \mathcal{B}$. Then $\mathcal{G}$ is $(r+1, 2\beta)$-global by definition of $\mathcal{B}$ and $\mu(\mathcal{G}) > sk/C_2 n$ by uncapturability of $\mathcal{F}$. Let $\mathcal{H} = \{B \in \binom{[n]}{C_2} : \partial^r B \subset \partial^r_c \mathcal{G}\}$, where $c = \mu(\mathcal{G})/2\binom{C_2}{r} > sk/nC_2^{2r}$. We have $\mu(\mathcal{H}) \ge \mu(\mathcal{G})/2$ by Lemma 4.4. We will show that $\partial^r \mathcal{H}$ is G-free. Then Lemma 4.7 with $C_2/2 \gg r\Delta$ in place of C will give the contradiction $sk/C_2 n < \mu(\mathcal{G}) \le 2\mu(\mathcal{H}) \le (s/n)^2$.

It remains to show that $\partial^r \mathcal{H}$ is G-free. Suppose for a contradiction that $A_1, \dots, A_s$ is a copy of G in $\partial^r \mathcal{H}$. Let $A = \bigcup_{i=1}^s A_i$ and $\mathcal{G}_i = \mathcal{G}^{A_i}_A$ for $i \in [s]$. Then each $\mathcal{G}_i$ is $(1, 4\beta)$-global by Lemma 2.2 with $\mu(\mathcal{G}_i) > c - |A| \cdot 2\beta k/n > c/2$. Now each $\mathcal{G}_i$ is $(C_1 s, c/4)$-uncapturable by Lemma 2.4, so $\mathcal{G}_1, \dots, \mathcal{G}_s$ cross contain a matching by Lemma 3.1 with $m = s$. However, this contradicts $\mathcal{F}$ being G^+-free. □

5 Refined junta approximation

In this final section of the part we will prove Theorem 1.9, our refined junta approximation result, which will play a key role in the proofs of our results in the next part. We start in the first subsection by setting out the strategy of the proof and implementing it assuming an embedding lemma, whose proof will then occupy the remainder of the section.

5.1 Strategy

Our embedding strategy considers a setup below that blends the two embedding strategies used in the proof of Theorem 1.8: it has elements of Lemma 2.9 (mapping a crosscut to a junta) and of Lemma 2.10 (embedding in the fat shadow and lifting via uncapturability).

Setup 5.1. Let $G \in \mathcal{G}'(r,s,\Delta)$. Let S be a crosscut in $G^+(r+1)$ with $|S| = \sigma := \sigma(G)$. Suppose $S_1 \subset S$ with $|S_1| = \sigma_1 \le \sigma$ and $\{G^x_x : x \in S_1\}$ vertex disjoint. Let $H_1, \dots, H_{\sigma_1}$ be the inclusive links $G * x = \{e \in G : x \in e\}$ for $x \in S_1$ and $H_{\sigma_1+1}, \dots, H_\sigma$ be the exclusive links G^x_x for $x \in S \setminus S_1$. Let $V_1 = \bigcup_{i=1}^{\sigma_1} V(H_i)$ and suppose $\{j : V(H_j) \cap V_1 \neq \emptyset\} = [\sigma_2]$. Let $H'_i = H_i$ for $i \in [\sigma_1]$ and $H'_i = \{e \cap V_1 : e \in H_i\}$ for $i \in [\sigma_1 + 1, \sigma_2]$.

We note that $\sigma \le s \le \Delta\sigma$. To use Setup 5.1 for embedding G^+ in $\mathcal{F} \subset \binom{[n]}{k}$ it suffices to find $J = \{j_{\sigma_1+1}, \dots, j_\sigma\} \subset [n]$ and a cross copy of $H_1^+, \dots, H_\sigma^+$ in $\mathcal{F}_1, \dots, \mathcal{F}_\sigma$, where $\mathcal{F}_i = \mathcal{F}_J^\emptyset$ for $i \in [\sigma_1]$ and $\mathcal{F}_i = \mathcal{F}_J^{j_i}$ for $i \in [\sigma_1+1, \sigma]$. This will be achieved by the following lemma.

Lemma 5.2. *Let $C \gg C_1 \gg \theta^{-1} \gg \varepsilon^{-1} \gg r\Delta$ and $C < k < n/Cs$. Let $G, H_1, \dots, H_\sigma$ be as in Setup 5.1 with $\sigma_1 \le \theta\sigma$. Let $\mathcal{F}_i \subset \binom{[n]}{k}$ for $i \in [\sigma_1]$ and $\mathcal{F}_i \subset \binom{[n]}{k-1}$ for $i \in [\sigma_1+1, \sigma]$. Suppose $\mathcal{F}_i$ is $(C_1\sigma_1, \varepsilon\sigma_1 k/n)$-uncapturable for $i \in [\sigma_1]$, that $\mu(\mathcal{F}_i) \ge 1 - \theta$ for $i \in [\sigma_1+1, \sigma_2]$, and $\mu(\mathcal{F}_i) \ge \beta := e^{-k/C_1} + C_1 sk/n$ for $i \in [\sigma_2+1, \sigma]$. Then $\mathcal{F}_1, \dots, \mathcal{F}_\sigma$ cross contain $H_1^+, \dots, H_\sigma^+$.*

Next we deduce Theorem 1.9 from Lemma 5.2.

Proof of Theorem 1.9. Let $G \in \mathcal{G}(r,s,\Delta)$ with $\sigma(G) = \sigma$ and $C \gg C_1 \gg \theta^{-1} \gg \delta^{-1} \gg \varepsilon^{-1} \gg r\Delta$. Suppose $\mathcal{F} \subset \binom{[n]}{k}$ with $C \le k \le n/Cs$ is G^+-free with $|\mathcal{F}| > |\mathcal{S}_{n,k,\sigma-1}| - \delta\binom{n-1}{k-1}$. We need to find $J \in \binom{[n]}{\sigma-1}$ with $|\mathcal{F}_J^\emptyset| \le \varepsilon\binom{n-1}{k-1}$.

As in the proof of Theorem 1.8 we let $J = \{i \in [n] : \mu(\mathcal{F}_i^i) \ge \beta\}$, where $\beta := e^{-k/C_1} + C_1 sk/n$. We recall that $|J| \le \sigma - 1$ and $\mathcal{F}_J^\emptyset$ is $(a, \mu(\mathcal{F}_J^\emptyset)/2)$-uncapturable with $a = \mu(\mathcal{F}_J^\emptyset)n/4k\beta$. Replacing '$\varepsilon$' in that proof by $.1\theta^2$ we obtain $|\mathcal{F}_J^\emptyset| \le .1\theta^2|\mathcal{S}_{n,k,\sigma-1}| \le .2\theta^2(\sigma-1)\binom{n-1}{k-1}$. We may assume $\sigma \ge 2\theta^{-1}$, otherwise $|\mathcal{F}_J^\emptyset| \le \theta\binom{n-1}{k-1}$. As $|\mathcal{F}_J^\emptyset| \ge |\mathcal{F}| - |\mathcal{S}_{n,k,J}| \ge (.9(\sigma - 1 - |J|) - \delta)\binom{n-1}{k-1}$ we deduce $|J| > (1 - .3\theta^2)(\sigma - 1)$, so $1 \le \sigma_1 := \sigma - |J| \le 1 + .3\theta^2\sigma \le \theta\sigma$.

Now we let $S, S_1, H_1, \dots, H_\sigma$ be as in Setup 5.1, where we can greedily choose $S_1 \subset S$ with $|S_1| = \sigma_1$ such that $\{G^x_x : x \in S_1\}$ are vertex disjoint, as any partial choice of S_1 forbids at most $\sigma_1(\Delta r)^2 < \sigma$ vertices of S. We write $J = \{j_{\sigma_1+1}, \dots, j_\sigma\}$, let $\mathcal{F}_i = \mathcal{F}_J^\emptyset$ for $i \in [\sigma_1]$ and $\mathcal{F}_i = \mathcal{F}_J^{j_i}$ for $i \in [\sigma_1+1, \sigma]$, where we can assume $|\mathcal{F}_{\sigma_1+1}| \ge \dots \ge |\mathcal{F}_\sigma|$. We note that $\mu(\mathcal{F}_{\sigma_2}) > 1 - \theta$, as otherwise we would have the contradiction $|\mathcal{F}| < |\mathcal{F}_J^\emptyset| + \big(\sigma_2 - \sigma_1 + (\sigma - \sigma_2)(1-\theta)\big)\binom{n-1}{k-1} < \big((1 + .2\theta^2)\sigma - \sigma_1 - \theta(\sigma - \sigma_2)\big)\binom{n-1}{k-1} < |\mathcal{S}_{n,k,\sigma-1}| - \delta\binom{n-1}{k-1}$.

Now we must have $\mu(\mathcal{F}_J^\emptyset) \le \varepsilon\sigma_1 k/n$; otherwise $\mathcal{F}_J^\emptyset$ is $(C_1\sigma_1, \varepsilon\sigma_1 k/2n)$-uncapturable, so $\mathcal{F}_1, \dots, \mathcal{F}_\sigma$ cross contain $H_1^+, \dots, H_\sigma^+$ by Lemma 5.2, contradicting $\mathcal{F}$ being G^+-free. As $|\mathcal{F}_J^\emptyset| \ge |\mathcal{F}| - |\mathcal{S}_{n,k,J}| \ge (.9(\sigma_1 - 1) - \delta)\binom{n-1}{k-1}$ we deduce $.9(\sigma_1 - 1) - \delta \le \varepsilon\sigma_1$, so $\sigma_1 = 1$ and $\mu(\mathcal{F}_J^\emptyset) \le \varepsilon k/n$. □

The remainder of the section will be devoted to the proof of Lemma 5.2. Similarly to the proofs of our previous embedding results (Lemmas 2.9 and 2.10), the strategy will be to find shadow embeddings and then lifting embeddings. However, there are further technical challenges to overcome in the current setting, particularly when the uniformity k of our families is small, when we need to 'pause' the shadow embedding after embedding $H'_i = H_i$ for $i \in [\sigma_1]$, then lift this part of the embedding, then complete the shadow embedding, and finally lift the remainder of the embedding. The shadow embedding lemma will be presented in the next subsection. The third subsection contains further results on upgrading uncapturability to globalness, which we call 'enhanced upgrading', as they obtain globalness parameters that are significantly stronger than one might expect, and this will be a crucial technical ingredient of the proof. In the fourth subsection we establish an improved lifting result that allows for a much weaker uncapturability assumption than that in Lemma 3.1. We conclude with the proof of Lemma 5.2 in the final subsection.

5.2 Shadow embeddings

Here we extend the argument used in Lemma 4.9 to prove the following lemma that will be applied to show that the fat shadows of $\mathcal{F}_1, \dots, \mathcal{F}_\sigma$ as in Lemma 5.2 cross contain $H_1, \dots, H_\sigma$. Whereas before we were embedding into nearly complete hypergraphs, now many of our hypergraphs will be quite sparse, which makes the embedding more challenging: the idea is to replace the naive greedy arguments by Theorem 4.6, here making key use of our observation that we can assume G is r-partite.

Lemma 5.3. *Let $C \gg \eta^{-1} \gg K \gg r\Delta$ and $0 < \theta < \eta$. Let $G, H_1, \dots, H_\sigma$ be as in Setup 5.1 and $\mathcal{G}_1, \dots, \mathcal{G}_\sigma \subset \binom{[n]}{r}$ with $n > C\sigma$. Suppose $\mu(\mathcal{G}_i) \ge 1 - \eta$ for $i \in [\sigma_2 + 1, \sigma]$, $\mu(\mathcal{G}_i) \ge 1 - \theta$ for $i \in [\sigma_1 + 1, \sigma_2]$ and $\mu(\mathcal{G}_i) \ge \theta^{1/2r} + n^{-1/K} + r\Delta\sigma_1/n$ for $i \in [\sigma_1]$. Let $c = 1 - \theta^{1/r}$. Then $\partial_c\mathcal{G}_1, \dots, \partial_c\mathcal{G}_{\sigma_2}$ cross contain $H'_1, \dots, H'_{\sigma_2}$ and $\mathcal{G}_1, \dots, \mathcal{G}_\sigma$ cross contain $H_1, \dots, H_\sigma$.*

Proof. For each $i \in [\sigma_1 + 1, \sigma_2]$ we define $\mathcal{G}_i^r, \dots, \mathcal{G}_i^0$ recursively by $\mathcal{G}_i^r = \mathcal{G}_i$ and $\mathcal{G}_i^{j-1} = \partial^{j-1}_{1-\theta^{1/r}}\mathcal{G}_i^j$ for $j \in [r]$. Clearly each $\mathcal{G}_i^j \subset \partial_{c_j}\mathcal{G}_i$ where $c_j = 1 - (r - j)\theta^{1/r}$.

We claim that each $\mu(\mathcal{G}_i^j) \ge 1 - \theta^{j/r}$. To see this, we argue by induction on $r - j$. For $r - j = 0$ we have $\mu(\mathcal{G}_i^r) \ge 1 - \theta$ by assumption. For the induction step, consider any $j \in [r]$ and uniformly random $A \subset B \subset [n]$ with $|A| = j - 1$ and $|B| = j$. Given any $A \notin \mathcal{G}_i^{j-1}$ we have $\mathbb{P}(B \notin \mathcal{G}_i^j) \ge \theta^{1/r}$, so $1 - \mu(\mathcal{G}_i^j) \ge \theta^{1/r}(1 - \mu(\mathcal{G}_i^{j-1}))$. The claim follows.

Next we will construct a cross embedding ϕ of $H'_1, \dots, H'_{\sigma_2}$ in $\partial_c\mathcal{G}_1, \dots, \partial_c\mathcal{G}_{\sigma_2}$. We recall that $H'_i = H_i$ for $i \in [\sigma_1]$ and all H'_i are defined on V_1, which is the disjoint union of $V(H_1), \dots, V(H_{\sigma_1})$. We proceed in σ_1 steps, defining ϕ on $V(H_t)$ at step t. When ϕ has been defined on $U_t := \bigcup_{i \le t} V(H_i)$, we say ϕ is *t-good* if $\phi(e \cap U_t) \in \mathcal{G}_i^{|e \cap U_t|}$ for each $i \in [\sigma_2]$ and $e \in \mathcal{G}_i$ with $e \cap U_t \neq \emptyset$.

We note that if ϕ is t-good then $\phi(H_i) \subset \mathcal{G}_i^r = \mathcal{G}_i = \partial_c\mathcal{G}_i$ for all $i \in [t]$ and if ϕ is σ_1-good then $\phi(H_i) \subset \partial_c\mathcal{G}_i$ for all $i \in [\sigma_2]$. As ϕ defined on $U_0 = \emptyset$ is trivially 0-good, it remains to show for any $t \in [\sigma_1]$ that we can extend any $(t-1)$-good ϕ to a t-good embedding.

For clarity of exposition, we start by showing the case $t = 1$. Obtain $\mathcal{H}_1$ from $\mathcal{G}_1$ by removing any edge e such that $f \notin \mathcal{G}_i^{|f|}$ for some $\emptyset \neq f \subset e$ and $i \in [\sigma_2]$ with $V(H_i) \cap V(H_1) \neq \emptyset$. There are at most $r\Delta^2$ such i, so by a union bound and the above claim we have $\mu(\mathcal{H}_1) \ge \mu(\mathcal{G}_1) - r\Delta^2 2^r \theta^{1/r} > n^{-1/K}$. We can assume that G is r-partite, so by Theorem 4.6 we can find an embedding ϕ'_1 of $N_1 := \{e \in G : e \cap V(H_1) \neq \emptyset\}$ in $\mathcal{H}_1$. Now $\phi = \phi' \mid_{V(H_1)}$ is 1-good.

Now we consider general $t \in [\sigma_1]$. Obtain $\mathcal{H}_t$ from $(\mathcal{G}_t)^{\emptyset}_{\phi(U_{t-1})}$ by removing any edge e such that $f \notin \mathcal{G}_i^{|f|}$ for some $\emptyset \neq f \setminus \phi(A') \subset e$ where $A \in H_i$ with $V(H_i) \cap V(H_t) \neq \emptyset$ and $A' = A \cap U_{t-1}$. For any such non-empty A', as ϕ is $(t-1)$-good we have $\phi(A') \in \mathcal{G}_i^{|A'|}$, so $\mu((\mathcal{G}_i^j)^{A'}_{A'}) \ge 1 - (j - |A'|)\theta^{1/r}$ for any $|A'| \le j \le r$. Thus a union bound gives $\mu(\mathcal{H}_t) \ge \mu(\mathcal{G}_t) - |U_{t-1}|k/n - r\Delta^2 2^r r\theta^{1/r} > n^{-1/K}$. Now as in the case $t = 1$ we obtain a t-good extension by embedding $N_t := \{e \in G : e \cap V(H_t) \neq \emptyset\}$ in $\mathcal{H}_t$ and restricting to $V(H_t)$.

Thus we have constructed a cross embedding ϕ of $H'_1, \dots, H'_{\sigma_2}$ in $\partial_c\mathcal{G}_1, \dots, \partial_c\mathcal{G}_{\sigma_2}$. To complete the proof we extend ϕ to a cross embedding $H_1, \dots, H_\sigma$ in $\mathcal{G}_1, \dots, \mathcal{G}_\sigma$, which requires $\phi(e \setminus V_1) \in (\mathcal{G}_i)^{e \cap V_1}_{e \cap V_1}$ for all $e \in H_i$, $i \in [\sigma_1 + 1, \sigma]$; this is possible by Lemma 4.9. □

5.3 Enhanced upgrading

This subsection provides further results on upgrading uncapturability to globalness with enhanced parameters that will be crucial in later proofs. We start by showing that every family has a restriction that is global or large.

Lemma 5.4. *Let $b, r \in \mathbb{N}$, $\alpha > 1$ and $\mathcal{F} \subset \binom{[n]}{k}$ with $k \ge br$. Then there is $B \subset [n]$ with $|B| \le br$ such that if $\mu(\mathcal{F}^B_B) < \alpha^b \mu(\mathcal{F})$ then $\mathcal{F}^B_B$ is $(r, \alpha\mu(\mathcal{F}^B_B))$-global with $\mu(\mathcal{F}^B_B) \ge \alpha^{1_{B \neq \emptyset}}\mu(\mathcal{F})$.*

Proof. We consider $\mathcal{F}_0, \mathcal{F}_1, \dots$, where $\mathcal{F}_0 = \mathcal{F}$, and if $i < b$ and $\mathcal{F}_i$ is not $(r, \alpha\mu(\mathcal{F}_i))$-global then we let $\mathcal{F}_{i+1} = (\mathcal{F}_i)^{B_i}_{B_i}$ so that $|B_i| \le r$ and $\mu(\mathcal{F}_{i+1}) > \alpha\mu(\mathcal{F}_i)$. When this sequence terminates at some $\mathcal{F}_t$ we let $B = \bigcup_{i \le t} B_i$. Clearly $\mathcal{F}^B_B = \mathcal{F}_t$ has the required properties. □

By iterating the previous result we obtain the following upgrading lemma.

Lemma 5.5. *Suppose $b,r,m \in \mathbb{N}$ and for each $i \in [m]$ that $\alpha_i > 1$ and $\mathcal{F}_i \subset \binom{[n]}{k_i}$ with $rb \le k_i \le n/2rm\alpha_i$ is (rbm, β_i)-uncapturable with $\alpha_i^b \beta_i > 2rmk_i/n$. Then there are disjoint $B_1, \dots, B_m$ with each $|B_i| \le rb$ such that, setting $\mathcal{G}_i = (\mathcal{F}_i)_B^{B_i}$ where $B = \bigcup_i B_i$, if $\mu(\mathcal{G}_i) < \alpha_i^b \beta_i/2$ then $\mathcal{G}_i$ is $(r, 4\alpha_i \mu(\mathcal{G}_i))$-global with $\mu(\mathcal{G}_i) > \alpha_i^{1_{B_i \neq \emptyset}} \beta_i/2$.*

Proof. We will choose $B_1, \dots, B_m$ sequentially and define $\mathcal{F}_i^0, \dots, \mathcal{F}_i^m$ for $i \in [m]$ by $\mathcal{F}_i^0 = \mathcal{F}_i$, $\mathcal{F}_j^i = (\mathcal{F}_j^{i-1})_{B_i}^{\emptyset}$ for $j \neq i$ and $\mathcal{F}_i^i = (\mathcal{F}_i^{i-1})_{B_i}^{B_i}$. At step i, we have $\mu(\mathcal{F}_i^{i-1}) \ge \beta_i$ by uncapturability of $\mathcal{F}_i$, so by Lemma 5.4 we can choose B_i with $|B_i| \le rb$ such that if $\mu(\mathcal{F}_i^i) < \alpha_i^b \mu(\mathcal{F}_i^{i-1})$ then $\mathcal{F}_i^i$ is $(r, \alpha\mu(\mathcal{F}_i^i))$-global with $\mu(\mathcal{F}_i^i) \ge \alpha_i^{1_{B_i \neq \emptyset}} \beta_i$. After step m, for any $i \in [m]$ we have $\mathcal{G}_i^m = \mathcal{G}_i = (\mathcal{F}_i)_B^{B_i}$. If $\mu(\mathcal{F}_i^i) \ge \alpha_i^b \mu(\mathcal{F}_i^{i-1})$ then $\mu(\mathcal{G}_i) \ge \alpha_i^b \beta_i - rmk_i/n \ge \alpha_i^b \beta_i/2$. Otherwise, $\mathcal{F}_i^i$ is $(r, \alpha_i \mu(\mathcal{F}_i^i))$-global with $\mu(\mathcal{F}_i^i) \ge \alpha_i^{1_{B_i \neq \emptyset}} \mu(\mathcal{F})$, and $(n/2k_i\alpha_i, \mu(\mathcal{F}_i^i)/2)$-uncapturable by Lemma 2.4, so $\mu(\mathcal{G}_i) > \mu(\mathcal{F}_i^i)/2 \ge \alpha_i^{1_{B_i \neq \emptyset}} \beta_i/2$, and $\mathcal{G}_i$ is $(r, 4\alpha_i \mu(\mathcal{G}_i))$-global by Lemma 2.2. □

For our final upgrading lemma we apply the previous one twice: the idea is that the globalness from the first application provides the second application with much better uncapturability.

Lemma 5.6. *Suppose $b,r,m \in \mathbb{N}$ and for each $i \in [m]$ that $\mathcal{F}_i \subset \binom{[n]}{k_i}$ with $rb \le k_i \le n/2rmb^2$ is $(2m, \beta_i)$-uncapturable with $\beta_i > 8rmk_i/bn$. Then there are disjoint $B_1, \dots, B_m$ with each $|B_i| \le rb+2$ such that, setting $\mathcal{G}_i = (\mathcal{F}_i)_B^{B_i}$ where $B = \bigcup_i B_i$, if $\mu(\mathcal{G}_i) < 2^b \beta_i/8$ then $\mathcal{G}_i$ is $(r, 8\mu(\mathcal{G}_i))$-global with $\mu(\mathcal{G}_i) > 2^{1_{B_i \neq \emptyset}} \beta_i/8$.*

Proof. We start by applying Lemma 5.5 with $(b, 1, 2)$ in place of (α_i, r, b). This gives disjoint $S_1, \dots, S_m$ with each $|S_i| \le 2$ such that, setting $\mathcal{H}_i = (\mathcal{F}_i)_S^{S_i}$ where $S = \bigcup_i S_i$, if $\mu(\mathcal{H}_i) < b^2 \beta_i/2$ then $\mathcal{H}_i$ is $(1, 4b\mu(\mathcal{H}_i))$-global with $\mu(\mathcal{H}_i) > \beta_i/2$.

We claim that each $\mathcal{H}_i$ is $(rbm, \beta_i/4)$-uncapturable. Indeed, this holds by a union bound if $\mu(\mathcal{H}_i) \ge b^2\beta_i/2$, as then $\mu((\mathcal{H}_i)_B^{\emptyset}) \ge \mu(\mathcal{H}_i) - |J|k_i/n \ge \beta_i/4$ whenever $|J| \le rbm$, as $\beta_i \ge 8rmk_i/bn$. On the other hand, if $\mathcal{H}_i$ is $(1, 4b\mu(\mathcal{H}_i))$-global with $\mu(\mathcal{H}_i) > \beta_i/2$ then $\mathcal{H}_i$ is $(n/2bk_i, \mu(\mathcal{H}_i)/2)$-uncapturable by Lemma 2.4, so $(rbm, \beta_i/4)$-uncapturable, as $k_i \le n/2rmb^2$.

Now we can apply Lemma 5.5 again to $\mathcal{H}_1, \dots, \mathcal{H}_m$ with $(2, r, b)$ in place of (α_i, r, b). This gives disjoint $S'_1, \dots, S'_m$ with each $|S'_i| \le rb$ such that, setting $\mathcal{G}_i = (\mathcal{H}_i)_{S'}^{S'_i}$ where $S' = \bigcup_i S'_i$, if $\mu(\mathcal{G}_i) < 2^b \beta_i/8$ then $\mathcal{G}_i$ is $(r, 8\mu(\mathcal{G}_i))$-global with $\mu(\mathcal{G}_i) > 2^{1_{S'_i \neq \emptyset}} \beta_i/8$. Thus $B_i = S_i \cup S'_i$ for $i \in [m]$ are as required. □

5.4 Refined capturability for matchings

Here we prove the following sharper version of Lemma 3.1, obtaining cross matchings under a much weaker uncapturability condition.

Lemma 5.7. *Let $C \gg K \gg d \ge 1$ and $\mathcal{F}_i \subset \binom{[n]}{k_i}$ with $k \le k_i \le Kk$ for $i \in [s]$, where $2d \le k \le n/Cs$. Suppose $\mathcal{F}_i$ is $(2dm, (2mk_i/n)^d)$-uncapturable for $i \in [m]$ and $\mu(\mathcal{F}_i) > 12(s + Km \log \frac{n}{mk})k_i/n$ for $i > m$. Then $\mathcal{F}_1, \dots, \mathcal{F}_s$ cross contain a matching.*

Proof. We start by upgrading uncapturability to globalness. We apply Lemma 5.5 with $r = 1$, $b = 2d$, $\alpha_i = \sqrt{n/mk_i}$, $\beta_i = (mk_i/n)^d$ noting that each $rb \le k_i \le n/2rm\alpha_i$ and $\alpha_i^b \beta_i = 2^d > 2rmk_i/n$, obtaining $B = \bigcup_i B_i$ with each $|B_i| \le 2d$ such that each $\mathcal{G}_i = (\mathcal{F}_i)_B^{B_i}$ is $(r, 4\alpha_i\mu(\mathcal{G}_i))$-global with $\mu(\mathcal{G}_i) > (2mk_i/n)^d/2$. We note by Lemma 2.4 that $\mathcal{G}_i$ is $(n/8\alpha_i k_i, (2mk_i/n)^d/4)$-uncapturable. Now we pass to the biased setting: we let $p_i = k_i/n$ and note that $\mathcal{H}_i = \mathcal{G}_i^{\uparrow}$ is $(n/8\alpha_i k_i, (2mk_i/n)^d/16)$-uncapturable by Lemma 3.2.

Now we will apply Theorem 3.5.2 to choose $S_1, \dots, S_m$ with each $|S_i| < K \log \frac{n}{mk}$ and define $\mathcal{H}_i^0, \dots, \mathcal{H}_i^m$ for $i \in [s]$ by $\mathcal{H}_i^0 = \mathcal{H}_i$, $\mathcal{H}_j^i = (\mathcal{H}_j^{i-1})_{S_i}^{\emptyset}$ for $j \neq i$ and $\mathcal{H}_i^i = (\mathcal{H}_i^{i-1})_{S_i}^{S_i}$. At step i, we have $\mu(\mathcal{H}_i^{i-1}) \geq (2mk_i/n)^d/16$ by uncapturability of $\mathcal{H}_i$, as $\sum_{j<i} |S_j| < Km \log \frac{n}{mk}$ and $n/8\alpha_i k_i \geq \frac{1}{8}\sqrt{nm/Kk}$, using $n/mk \geq C \gg K$.

Applying Theorem 3.5.2 with $\eta < 1/2d$ and $\sqrt{K}$ in place of K we obtain $S_i \subset [n]$ with $|S_i| \leq \sqrt{K} \log \mu(\mathcal{H}_i^{i-1})^{-1} < K \log \frac{n}{mk}$ and $\mu_{Kp_i}(\mathcal{H}_i^i) \geq \mu^\eta > \sqrt{mp_i}$, so $\mu_{Kp_i}(\mathcal{H}_i^m) \geq \sqrt{mp_i} - |S|Kp_i > 3m(Kp_i)$. For $i > m$, by Lemma 3.2 and a union bound we have $\mu_{p_i}(\mathcal{H}_i^m) > \mu(\mathcal{F}_i)/4 - |S|p_i > 3sp_i$. Thus by Lemma 3.8 there is a cross matching in $\mathcal{H}_1^m, \dots, \mathcal{H}_s^m$, and so in $\mathcal{F}_1, \dots, \mathcal{F}_s$. □

5.5 Lifted embeddings

We conclude this section by proving Lemma 5.2 which completes the proof of Theorem 1.9. As mentioned earlier, the proof becomes more complicated as the uniformity k of our family decreases. When it is quite large we can bound the fat shadow using Fairness, but otherwise we must rely on the weaker estimates from Lemma 4.4, so there are additional technical challenges, resolved by enhanced upgrading and in one case pausing the shadow embedding for a preliminary lifting step.

Proof of Lemma 5.2. Let $C \gg C_1 \gg \theta^{-1} \gg \varepsilon^{-1} \gg r\Delta$ and $C < k < n/Cs$. Let $G, H_1, \dots, H_\sigma$ be as in Setup 5.1 with $\sigma_1 \leq \theta\sigma$. Let $\mathcal{F}_i \subset \binom{[n]}{k}$ for $i \in [\sigma_1]$ and $\mathcal{F}_i \subset \binom{[n]}{k-1}$ for $i \in [\sigma_1 + 1, \sigma]$. Suppose $\mathcal{F}_i$ is $(C_1\sigma_1, \varepsilon\sigma_1 k/n)$-uncapturable for $i \in [\sigma_1]$, that $\mu(\mathcal{F}_i) \geq 1 - \theta$ for $i \in [\sigma_1 + 1, \sigma_2]$, and $\mu(\mathcal{F}_i) \geq \beta := e^{-k/C_1} + C_1 sk/n$ for $i \in [\sigma_2 + 1, \sigma]$. We need to show that $\mathcal{F}_1, \dots, \mathcal{F}_\sigma$ cross contain $H_1^+, \dots, H_\sigma^+$.

We consider cases according to the size of k. We start with the case $k \geq \sqrt{C_1} \log \frac{n}{\sigma_1}$, for which we will use enhanced upgrading. We apply Lemma 5.6 to $\mathcal{F}_1, \dots, \mathcal{F}_{\sigma_1}$ with $m = \sigma_1$, $b = C_1 + \log_2 \frac{s}{m}$, each $\beta_i = \varepsilon mk/n$ and $2r$ in place of r, noting that $2rb \leq k \leq n/2rmb^2$ and $\beta_i > 8rmk/bn$. This gives disjoint $B_1, \dots, B_m$ with each $|B_i| \leq 2rb + 2$ such that, setting $\mathcal{G}_i = (\mathcal{F}_i)_B^{B_i}$ where $B = \bigcup_i B_i$, if $\mu(\mathcal{G}_i) < 2^b \varepsilon mk/8n$ then $\mathcal{G}_i$ is $(2r, 8\mu(\mathcal{G}_i))$-global with $\mu(\mathcal{G}_i) > \varepsilon mk/8n > m/n \geq e^{-k/\sqrt{C_1}}$. For $i \in [\sigma_1 + 1, \sigma]$, writing $\mathcal{G}_i = (\mathcal{F}_i)_B^\emptyset$, we have $\mu(\mathcal{G}_i) \geq \mu(\mathcal{F}_i) - |B|k/n \geq e^{-k/C_1} + C_1 sk/2n$.

By Fairness (Proposition 4.3), with $\sqrt{C_1}$ in place of C, writing $c_i = (1 - \varepsilon)\mu(\mathcal{G}_i)$ for $i \in [\sigma]$ we have $\mu(\partial_{c_i}^{r'} \mathcal{G}_i) \geq 1 - \varepsilon$ for $r' \in \{r - 1, r\}$, so $\partial_{c_1}\mathcal{G}_1, \dots, \partial_{c_\sigma}\mathcal{G}_\sigma$ cross contain a copy $\phi(H_1), \dots, \phi(H_\sigma)$ of $H_1, \dots, H_\sigma$ by Lemma 4.9. We write $V' = \operatorname{Im} \phi$ and consider $\mathcal{H}_1, \dots, \mathcal{H}_s$ corresponding to the edges $A_1, \dots, A_s$ of $H_1, \dots, H_\sigma$, where for each edge A_j of H_i with $i \in [\sigma]$ we let $\mathcal{H}_j = (\mathcal{G}_i)_{V'}^{\phi(A_j)}$. To complete the proof of this case it suffices to show that $\mathcal{H}_1, \dots, \mathcal{H}_s$ cross contain a matching.

To do so, we verify the conditions of Lemma 3.1. Consider any $A_j \in H_i$. If $i > \sigma_1$ or $i \in [\sigma_1]$ with $\mu(\mathcal{G}_i) \geq 2^b \varepsilon mk/8n > C_1^2 sk/n$ then $\mu(\mathcal{H}_j) \geq c_i - |V'|k/n > C_1 sk/3n$. Now consider $i \in [\sigma_1]$ such that $\mathcal{G}_i$ is $(2r, 8\mu(\mathcal{G}_i))$-global with $\mu(\mathcal{G}_i) > \varepsilon mk/8n$. Then $\mathcal{H}_j$ and $\mathcal{H}'_j = (\mathcal{G}_i)_{\phi(A_j)}^{\phi(A_j)}$ are $(r, 16\mu(\mathcal{G}_i))$-global by Lemma 2.2. As $\mu(\mathcal{H}'_j) > c_i = (1 - \varepsilon)\mu(\mathcal{G}_i)$, by Lemma 2.4 $\mathcal{H}'_j$ is $(n/40k, \mu(\mathcal{H}'_j)/2)$-uncapturable, so $\mu(\mathcal{H}_j) \geq \mu(\mathcal{H}'_j)/2 > \varepsilon mk/20n$, and $\mathcal{H}_j$ is $(n/80k, \mu(\mathcal{H}_j)/2)$-uncapturable again by Lemma 2.4. Thus the required conditions hold.

Henceforth we can assume $k < \sqrt{C_1} \log \frac{n}{\sigma_1}$. In this case we upgrade uncapturability to globalness using Lemma 2.6 to obtain disjoint $S_1, \dots, S_{\sigma_1}$ with each $|S_i| \leq 2r$ such that, setting $\mathcal{G}_i = (\mathcal{F}_i)_S^{S_i}$ where $S = \bigcup_i S_i$, whenever $\mu(\mathcal{G}_i) < \beta$ we have $S_i = \emptyset$ and $\mathcal{G}_i$ is $(2r, 2\beta)$-global with $\mu(\mathcal{G}_i) > \varepsilon\sigma_1 k/n$. For $i > \sigma_1$ we set $\mathcal{G}_i = (\mathcal{F}_i)_S^\emptyset$ and note that $\mu(\mathcal{G}_i) \geq \mu(\mathcal{F}_i) - |S|k/n > \beta/2$. As before, for any $i \notin [\sigma_1 + 1, \sigma_2]$ with $\mu(\mathcal{G}_i) > \beta/2$ Fairness gives $\mu(\partial_{c_i}^{r'} \mathcal{G}_i) \geq 1 - \varepsilon$ for $r' \in \{r - 1, r\}$, where $c_i = (1 - \varepsilon)\mu(\mathcal{G}_i)$. For $i \in [\sigma_1 + 1, \sigma_2]$ we have the better bound $\mu(\partial_{c_i}^{r'} \mathcal{G}_i) \geq 1 - \sqrt{\theta}$ where $c_i = 1 - \sqrt{\theta}$ from Lemma 4.2. For $i \in I := \{i : \mu(\mathcal{G}_i) < \beta/2\}$ we note that $\mathcal{G}_i$ is G^+-free, as $S_i = \emptyset$, so we can bound the fat shadow by Lemma 4.4: we take $\ell = k$, use $(2\varepsilon)^{-1} \gg r\Delta$ in place of C, and write $c_i = \mu(\mathcal{G}_i)/2\binom{k}{r} \geq \mu(\mathcal{G}_i)/2k^r$, to obtain

$$\mu(\partial_{c_i}^r \mathcal{G}_i) \geq \left((\mu(\mathcal{G}_i)/2 - (s/n)^{2k\varepsilon})n/sk^2\right)^{r/(k-1)} \geq z := (\sigma_1/sk^2)^{2r/k} - (s/n)^{r\varepsilon}.$$

Next we consider the case that $k \geq 2C_1 \log \frac{s}{\sigma_1}$. Then $z \geq 1 - \varepsilon$, so $\partial_{c_1}\mathcal{G}_1, \dots, \partial_{c_\sigma}\mathcal{G}_\sigma$ cross contain a copy $\phi(H_1), \dots, \phi(H_\sigma)$ of $H_1, \dots, H_\sigma$ by Lemma 4.9. With notation as in the previous case, it remains to show that $\mathcal{H}_1, \dots, \mathcal{H}_s$ cross contain a matching. To do so, we verify the conditions of Lemma 5.7, taking $m = |I|$, $d = 2$ and $K = \varepsilon^{-1}$. Consider any $A_j \in H_i$. If $i \notin I$ then $\mu(\mathcal{H}_j) \ge \beta/3 - |\operatorname{Im}\phi|k/n > 12(s + \varepsilon^{-1}|I|\log\frac{n}{k|I|})k/n$, as $|I|/n \le \sigma_1/n < e^{-k/\sqrt{C_1}}$, so $|I|k/n \cdot \log\frac{n}{k|I|} < k^2 e^{-k/\sqrt{C_1}} < \beta^2$. Now suppose $i \in I$, so that $\mathcal{G}_i$ is $(2r, 2\beta)$-global with $\mu(\mathcal{G}_i) > \varepsilon\sigma_1 k/n$. Then $\mathcal{H}_j$ and $\mathcal{H}'_j = (\mathcal{G}_i)^{\phi(A_j)}_{\phi(A_j)}$ are $(r, 4\beta)$-global by Lemma 2.2. As $\mu(\mathcal{H}'_j) > c_i \ge \mu(\mathcal{G}_i)/2k^r$, by Lemma 2.4 $\mathcal{H}'_j$ is $(a, \mu(\mathcal{H}'_j)/2)$-uncapturable, where $a = \mu(\mathcal{G}_i)n/8k\beta > \varepsilon\sigma_1/8\beta > rs \ge |\operatorname{Im}\phi|$ as $\sigma_1/s \ge e^{-k/2C_1} \ge \sqrt{\beta}$, since $ks/n < k\Delta\sigma_1/n < \Delta k e^{-k/\sqrt{C_1}}$. Hence $\mu(\mathcal{H}_j) \ge \mu(\mathcal{H}'_j)/2 > \mu(\mathcal{G}_i)/4k^r > 2(2|I|k/n)^2$, and $\mathcal{H}_j$ is $(4|I|, \mu(\mathcal{H}_j)/2)$-uncapturable again by Lemma 2.4. Thus the required conditions hold.

It remains to consider the case $k < 2C_1 \log\frac{s}{\sigma_1}$. We start by applying 5.3 to $(\partial^r_{c_i}\mathcal{G}_i : i \in [\sigma_2])$ with $\theta_0 = \sqrt{\sigma_1/\sigma} \le \sqrt{\theta}$ in place of θ, recalling for $i \in [\sigma_1 + 1, \sigma_2]$ that $\mu(\partial^r_{c_i}\mathcal{G}_i) \ge 1 - \sqrt{\theta} \ge 1 - \theta_0$ and $\mu(\partial^r_{c_i}\mathcal{G}_i) \ge 1 - \varepsilon$ for $i \in [\sigma_1] \setminus I$, and noting for $i \in I$ that $\mu(\partial^r_{c_i}\mathcal{G}_i) \ge \theta_0^{1/2r} + n^{-\varepsilon} + r\Delta\sigma_1/n$. This gives a cross embedding ϕ of $H'_1, \dots, H'_{\sigma_2}$ in $(\partial_{cc_i}\mathcal{G}_i : i \in [\sigma_2])$, where $c = 1 - \theta_0^{1/r}$.

Next we extend $(\phi(H'_i) : i \in [\sigma_1]) = (\phi(H_i) : i \in [\sigma_1])$ to a cross embedding $(\phi(H^+_i) : i \in [\sigma_1])$ in $(\mathcal{G}_i : i \in [\sigma_1])$, by finding a cross matching in $(\mathcal{H}_j : j \in [s_1])$ corresponding to the edges $A_1, \dots, A_{s_1}$ of $H_1, \dots, H_{\sigma_1}$, where for each edge A_j of H_i with $i \in [\sigma_1]$ we let $\mathcal{H}_j = (\mathcal{G}_i)^{\phi(A_j)}_{\operatorname{Im}\phi}$. This is possible by Lemma 5.7, which applies similarly to the previous case, where for uncapturability of $\mathcal{H}'_j$ we note that now $|\operatorname{Im}\phi| \le rs_1 \le r\Delta\sigma_1$.

Finally, we extend to a cross embedding $(\phi(H^+_i) : i \in [\sigma])$ in $(\mathcal{G}_i : i \in [\sigma])$ by finding a cross copy of $(A_j \setminus V_1 : s_1 < j \le s)$ in $(\mathcal{H}_j : s_1 < j \le s)$, where for each edge A_j of H_i with $\sigma_1 < i \le \sigma$ we let $\mathcal{H}_j = (\mathcal{G}_i)^{\phi(A_j \cap V_1)}_{\operatorname{Im}\phi}$. This is possible by Lemma 2.9, as each $\mu(\mathcal{H}_j) \ge \mu(\mathcal{G}_i) - \Delta\sigma_1 k^2/n > \beta/4$, using $k < 2C_1 \log\frac{s}{\sigma_1}$ and $\sigma_1 \le \theta\sigma$. $\square$

6 The Huang–Loh–Sudakov Conjecture

Here we prove Theorem 1.2, which establishes the Huang–Loh–Sudakov Conjecture. In the first subsection we prove a strong stability version that has independent interest. We then deduce the exact result in the second subsection.

6.1 A strong stability result

Here we prove the following strong approximate version of the Huang–Loh–Sudakov conjecture, which will be refined to obtain the exact result in the following subsection.

Theorem 6.1. *Let $0 < C^{-1} \ll \varepsilon$ and $\mathcal{F}_i \subset \binom{[n]}{k_i}$ with $C \le k_i \le n/Cs$ for all $i \in [s]$. If $\mathcal{F}_1, \dots, \mathcal{F}_s$ are cross free of a matching and each $|\mathcal{F}_i| \ge |\mathcal{S}_{n,k_i,s-1}| - (1-\varepsilon)\binom{n-1}{k_i-1}$ then there is $J \in \binom{[n]}{s-1}$ so that $|\mathcal{F}_i \setminus \mathcal{S}_{n,k_i,J}| \le \varepsilon\binom{n-1}{k_i-1}$ for all $i \in [s]$.*

The idea of the proof will be to consider $A = \{a_1, \dots, a_\ell\} \subset [n]$ maximal such that there are distinct $b_1, \dots, b_\ell$ so that all $(\mathcal{F}_{b_i})^{a_i}_{a_i}$ are large. This motivates the setting of the following lemma.

Lemma 6.2. *Let $0 < C^{-1} \ll \beta \ll \varepsilon \le 1$ and $m, \ell, n, s, k_1, \dots, k_s \in \mathbb{N}$ with $\ell \le m \le s$ and each $k_i \le n/Cs$. Suppose $\mathcal{F}_i \subset \binom{[n]}{k_i}$ and $J_i := \{j \in [n] : \mu((\mathcal{F}_i)^j_j) \ge \beta\}$ for each $i \in [s]$ are such that*

(a) there are distinct $a_1, \dots, a_\ell \in [n]$ with $a_i \in J_i$ for $i \in [\ell]$;

(b) $\mu((\mathcal{F}_i)^{\emptyset}_{J_i}) \ge \varepsilon(m - |J_i|)k_i/n$ and $J_i \subset A := \{a_1, \dots, a_\ell\}$ for each $i \in [\ell+1, m]$;

(c) $\mu(\mathcal{F}_i) \ge Ck_i s/n$ for all $i \in [m+1, s]$.

Then $\mathcal{F}_1, \dots, \mathcal{F}_s$ cross contain a matching.

Proof. It suffices to check the conditions of Lemma 3.1 for $\mathcal{G}_1, \dots, \mathcal{G}_s$ defined by $\mathcal{G}_i = (\mathcal{F}_i)^{a_i}_A$ for $i \in [\ell]$ and $\mathcal{G}_i = (\mathcal{F}_i)^{\emptyset}_A$ otherwise. We do so with $m - \ell$ in place of m and $(\mathcal{G}_i : \ell < i \le m)$ in place of $\mathcal{F}_1, \dots, \mathcal{F}_m$. For $i \in [s] \setminus [m]$ we have $\mu(\mathcal{G}_i) \ge \mu(\mathcal{F}_i) - |A|k_i/n \ge Ck_i s/2n$. Similarly, for $i \in [\ell]$ we have $\mu(\mathcal{G}_i) \ge \mu((\mathcal{F}_i)^{a_i}_{a_i}) - |A|k_i/n \ge \beta/2 \ge (\beta/2)(Ck_i s/n) \ge C^{1/2}k_i s/2n$. For $i \in [\ell+1, m]$ we note by definition of J_i that $\mathcal{G}_i$ is $(1, 2\beta)$-global with $\mu(\mathcal{G}_i) \ge \mu((\mathcal{F}_i)^{\emptyset}_{J_i}) - |A \setminus J_i|\beta k_i/n \ge \varepsilon(m-\ell)k_i/n$, so $(\varepsilon(m-\ell)/4\beta, \varepsilon(m-\ell)k_i/2n)$-uncapturable by Lemma 2.4. Thus the required conditions hold. □

We deduce our stability result as follows.

Proof of Theorem 6.1. Let $0 < C^{-1} \ll \beta \ll \varepsilon \le 1/2$ and $\mathcal{F}_i \subset \binom{[n]}{k_i}$ with $k_i \le n/Cs$ for all $i \in [s]$. Let $J_1, \dots, J_s$ be as in Lemma 6.2. Let $A = \{a_1, \dots, a_\ell\} \subset [n]$ be maximal such that there are distinct $b_1, \dots, b_\ell$ with $a_i \in J_{b_i}$ for all $i \in [\ell]$. Without loss of generality we may assume $b_i = i$ for all $i \in [\ell]$. By maximality, we have $J_i \subset \{a_1, \dots, a_\ell\}$ for all $i \in [\ell+1, s]$.

We may assume $\ell < s$, and that $\mu\big((\mathcal{F}_h)^{\emptyset}_{J_h}\big) < .1\varepsilon(s - |J_h|)k_h/n$ for some $h \in [\ell+1, s]$, otherwise Lemma 6.2 provides the required cross matching. Noting that $|\mathcal{S}_{n,k_h,s-1}| - (1-\varepsilon)\binom{n-1}{k_h-1} \le |\mathcal{F}_h| \le |\mathcal{S}_{n,k_h,J_h}| + .1\varepsilon(s - |J_h|)\binom{n-1}{k_h-1}$, we see that $|J_h| = s - 1 = \ell$, $h = s$ and $J_h = A$. Now for each $i \in [s-1]$, as $a_i \in A = J_h$ we can apply the same argument switching the roles of $\mathcal{F}_i$ and $\mathcal{F}_h$ to deduce $\mu\big((\mathcal{F}_i)^{\emptyset}_{J_i}\big) < .1\varepsilon k_h/n$ and $J_i = A$. The theorem follows. □

6.2 The exact result

To complete the proof of the Huang–Loh–Sudakov Conjecture we will upgrade the approximate result of the previous subsection to an exact result via the following bootstrapping lemma (stated in a more general form than needed here as we will also use it for our other Turán results).

Lemma 6.3. *Let $C \gg \beta^{-1} \gg d \ge 1$ and $\mathcal{F}_i \subset \binom{[n]}{k_i}$ for all $i \in [s]$ with $\sum_{i=1}^s k_i \le n/C$. Suppose $\mathcal{F}_1, \dots, \mathcal{F}_s$ are cross free of some hypergraph $G = \{e_1, \dots, e_s\}$ with $|e_i| = k_i$ for each $i \in [s]$ and $e_s \cap \bigcup_{i=1}^{s-1} e_i = \emptyset$. If $\sum_{i=1}^{s-1}(1 - \mu(\mathcal{F}_i)) \le \alpha \in (0, \beta)$ then $\mu(\mathcal{F}_s) \le (\alpha k_s/n)^d$.*

Proof. Let $k = n - n/C$ and $\mathcal{G}_s = \mathcal{F}_s^{\uparrow} \cap \binom{[n]}{k}$. Then $\mathcal{F}_1, \dots, \mathcal{F}_{s-1}, \mathcal{G}_s$ are cross free of G' obtained from G by enlarging e_s to e'_s of size k. Suppose for contradiction that $\mu(\mathcal{F}_s) > (\alpha k_s/n)^d$. Let $t \in [k_s]$ be minimal so that $|\mathcal{F}_s| = (\alpha k_s/n)^d\binom{n}{k_s} \ge \binom{n-t}{k_s-t}$. Then $(\alpha k_s/n)^d < (k_s/n)^{t-1}$, so if $t > 2d$ then $\alpha < (k_s/n)^{t/2d}$. By Kruskal-Katona $|\mathcal{G}_s| \ge \binom{n-t}{k-t}$, so $\mu(\mathcal{G}_s) \ge (1 - 2/C)^t > \sqrt{\alpha}$, as if $t \le 2d$ then $(1-2/C)^t > (1-2/C)^{2d} > \sqrt{\beta}$ or otherwise $\alpha^{2d/t} < k_s/n \le C^{-1} < (1-2/C)^{4d}$. Now we let $\phi : V(G') \to [n]$ be a uniformly random injection. Let E be the event that $\phi(e'_s) \notin \mathcal{G}_s$ or $\phi(e_i) \notin \mathcal{F}_i$ for some $i \in [s-1]$. Then $1 = \mathbb{P}(E) \le 1 - \mu(\mathcal{G}_s) + \sum_{i \in [s-1]}(1 - \mu(\mathcal{F}_i)) < 1 - \sqrt{\alpha} + \alpha$, contradiction. □

Theorem 1.2 will now follow by combining Theorem 6.1 and Lemma 6.3.

Proof of Theorem 1.2. Let $0 < 1/C \ll \varepsilon \ll 1$ and $\mathcal{F}_i \subset \binom{[n]}{k_i}$ with $|\mathcal{F}_i| \ge |\mathcal{S}_{n,k_i,s-1}|$ and $k_i \le \frac{n}{Cs}$ for all $i \in [s]$. Suppose $\mathcal{F}_1, \dots, \mathcal{F}_s$ have no cross matching. By Theorem 6.1 there is $J \in \binom{[n]}{s-1}$ such that $\mu\big((\mathcal{F}_i)^{\emptyset}_J\big) = \varepsilon_i k_i/|V|$ with $V = [n] \setminus J$ and $\varepsilon_i \le \varepsilon$ for all $i \in [s]$. We may assume that ε_s is maximal.

Next we claim that we can list the elements of J as $\mathbf{j} = (j_1, \dots, j_{s-1})$ so that

$$M_{\mathbf{j}} := \sum_{i \in [s-1]} \mu\big((\mathcal{F}_i)^{j_i}_J\big) \ge s - 1 - \varepsilon_s.$$

To see this, we note that $\mathbb{E}_{\mathbf{j}} M_{\mathbf{j}} = \mathbb{E}_{i \in [s-1]} \sum_{j \in J} \mu\big((\mathcal{F}_i)^j_J\big)$ when $\mathbf{j}$ is uniformly random. As each $(\mathcal{F}_i)^I_J \subset (\mathcal{S}_{n,k_i,s-1})^I_J$ whenever $\emptyset \ne I \subset J$ and $\mu(\mathcal{F}_i) \ge \mu(\mathcal{S}_{n,k_i,s-1})$, we have $0 \le \mu(\mathcal{F}_i) - \mu(\mathcal{S}_{n,k_i,s-1}) \le \mu((\mathcal{F}_i)^{\emptyset}_J) - k|V|^{-1}\sum_{j \in J}(1 - \mu((\mathcal{F}_i)^j_J))$, so $\sum_{j \in J}\mu((\mathcal{F}_i)^j_J)) \ge s - 1 - \varepsilon_s$. The claim follows.

Now let $\mathcal{H}_i = (\mathcal{F}_i)^{j_i}_J \subset \binom{V}{k-1}$ for all $i \in [s-1]$, and $\mathcal{H}_s = (\mathcal{F}_s)^{\emptyset}_J \subset \binom{V}{k-1}$. Then $\mathcal{H}_1, \dots, \mathcal{H}_s$ have no cross matching, $\sum_{i \in [s-1]}(1 - \mu(\mathcal{H}_i)) \le \varepsilon_s$ and $\mu(\mathcal{H}_s) = \varepsilon_s k_s/|V|$. Therefore $\varepsilon_s = 0$ by Lemma 6.3 with $d = 1$. By choice of ε_s we deduce $\varepsilon_i = 0$ for all $i \in [s]$. Thus $\mathcal{F}_i = \mathcal{S}_{n,k_i,J}$ for all $i \in [s]$. □

7 Critical graphs

In this section we prove Theorem 1.6, which gives exact Turán results for expanded critical r-graphs of bounded degree. In fact, we will prove the following strong stability version.

Theorem 7.1. *Let $G \in \mathcal{G}(r, \Delta, s)$ be critical and $C \gg \beta^{-1} \gg dr\Delta$.*
Suppose $\mathcal{F} \subset \binom{[n]}{k}$ with $C \le k \le n/Cs$ is G^+-free and $|\mathcal{F}| \ge |\mathcal{S}_{n,k,\sigma-1}| - \varepsilon\binom{n-1}{k-1}$ with $\varepsilon \in (0, \beta)$.
Then there is $J \in \binom{[n]}{\sigma-1}$ with $|\mathcal{F} \setminus \mathcal{S}_{n,k,J}| \le \varepsilon^d \binom{n-1}{k-1}$.
Furthermore, if $k \le \sqrt{n}$ and $|\mathcal{F}| \ge |\mathcal{S}_{n,k,J}| - \beta\binom{n-r}{k-r}$ then $\mathcal{F} \subset \mathcal{S}_{n,k,J}$.

In the first subsection we will describe the strategy of the proof and complete the proof, assuming a certain bootstrapping lemma that will be proved in the second subsection.

7.1 Strategy

Recall that an r-graph G is *critical* if it has an edge e such that $\sigma(G \setminus e) = \tau(G \setminus e) < \tau(G) = \sigma(G)$. Thus we can adopt the following set-up.

Setup 7.2. Let $G \in \mathcal{G}'(r, s, \Delta)$ be critical. Fix a crosscut S in $G^+(r+1)$ with $|S| = \sigma := \sigma(G)$ and $\{G^x_x : x \in S\} = \{H_i : i \in [\sigma]\}$ with $|H_\sigma| = 1$. Let $I = \{i \in [\sigma - 1] : V(H_i) \cap V(H_\sigma) \neq \emptyset\}$.

The following bootstrapping lemma will be proved in the next subsection. It shows that if we cannot find a cross embedding of $H_1^+, \dots, H_\sigma^+$ as in the above set up, if all but one of the families are nearly complete then the last must be very small.

Lemma 7.3. *Let $G, H_1, \dots, H_\sigma$ be as in Setup 7.2. Let $C \gg \beta^{-1} \gg dr\Delta$ and $\mathcal{F}_i \subset \binom{[n]}{k_i}$ with $k_i \in [k/2, k]$ for $i \in [\sigma]$, where $C \le k \le n/Cs$. Suppose $\mathcal{F}_\sigma$ is G^+-free, $\sum_{i=1}^{\sigma-1}(1 - \mu(\mathcal{F}_i)) \le \varepsilon \le \beta$, $\mu(\mathcal{F}_\sigma) \ge \varepsilon^d k/n$ and $1 - \mu(\mathcal{F}_i) \le \varepsilon_0 := 2\varepsilon/\sigma$ for all $i \in I$. Then $\mathcal{F}_1, \dots, \mathcal{F}_\sigma$ cross contain $H_1^+, \dots, H_\sigma^+$.*

We conclude this subsection by deducing Theorem 7.1 from Lemma 7.3.

Proof of Theorem 7.1. By Theorem 1.9 (refined junta approximation) there is $J \in \binom{[n]}{\sigma-1}$ such that $|\mathcal{F} \setminus \mathcal{S}_{n,k,J}| = \delta\binom{n-1}{k-1}$ with $\delta^{-1} \gg dr\Delta$. We write $J = \{j_1, \dots, j_{\sigma-1}\}$, $\mathcal{F}_i = \mathcal{F}_J^{j_i}$ for $i \in [\sigma - 1]$ and $\mathcal{F}_\sigma = \mathcal{F}_J^\emptyset$. Note that $\mathcal{F}_\sigma$ is G^+-free. We may assume $I = [|I|]$ and $|\mathcal{F}_1| \ge \dots \ge |\mathcal{F}_{\sigma-1}|$. Now

$$\begin{aligned} \mu(\mathcal{F}) &\le \mu(\mathcal{F}_J^\emptyset) + \mu(\mathcal{S}_{n,k,J}) - \tfrac{k-1}{n-|J|}\sum_{i=1}^{\sigma-1}(1 - \mu(\mathcal{F}_i)) \\ &\le \delta k/n + \mu(\mathcal{F}) + \varepsilon k/n - \tfrac{k}{2n}\sum_{i=1}^{\sigma-1}(1 - \mu(\mathcal{F}_i)), \end{aligned}$$

so $\sum_{i=1}^{\sigma-1}(1-\mu(\mathcal{F}_i)) \le 2(\varepsilon+\delta)$. Now for each $i \in I$ we have $1-\mu(\mathcal{F}_i) \le 4r\Delta(\varepsilon+\delta)/\sigma$ as if $\sigma \le 2|I| \le 2r\Delta$ this follows from $1 - \mu(\mathcal{F}_i) \le 2(\varepsilon + \delta)$, or otherwise from $1 - \mu(\mathcal{F}_i) \le \frac{2(\varepsilon+\delta)}{\sigma-|I|}$.

As $\mathcal{F}_1, \dots, \mathcal{F}_\sigma$ are cross free of $H_1^+, \dots, H_\sigma^+$ as in Setup 7.2, Lemma 7.3 with $(2r\Delta(\varepsilon + \delta), 2d)$ in place of (ε, d) gives $\delta k/n = \mu(\mathcal{F}_\sigma) < (2r\Delta(\varepsilon+\delta))^{2d}k/n$. As $\varepsilon^{-1}, \delta^{-1} \gg dr\Delta$ we have $((2r\Delta)(\varepsilon+\delta))^{2d} = (2r\Delta)^{2d}\sum_{i=0}^{2d}\binom{2d}{i}\varepsilon^i\delta^{2d-i} < (\varepsilon^d + \delta)/2$, so $\delta < \varepsilon^d$, i.e. $|\mathcal{F}_J^\emptyset| = |\mathcal{F}_\sigma| < \varepsilon^d\binom{n-1}{k-1}$.

Finally, let $k \le \sqrt{n}$ and suppose for contradiction that $|\mathcal{F}| \ge |\mathcal{S}_{n,k,J}| - \beta\binom{n-r}{k-r}$ but there is some $A \in \mathcal{F} \setminus \mathcal{S}_{n,k,J}$. By the previous statement with $d = 1$ and $\varepsilon = \beta\binom{n-r}{k-r}\binom{n-1}{k-1}^{-1}$ we have $|\mathcal{F}_J^\emptyset| \le \beta\binom{n-r}{k-r}$, so $|\mathcal{S}_{n,k,J} \setminus \mathcal{F}| \le 2\beta\binom{n-r}{k-r}$. We fix any $R \in \binom{A}{r}$ and a bijection $\phi : A_s \to R$, where $H_\sigma = \{A_s\}$ and define $\mathcal{G}_1, \dots, \mathcal{G}_{s-1}$ by $\mathcal{G}_j = (\mathcal{F}_i)_A^{\phi(A'_j)}$ whenever A_j is an edge of H_i with $A'_j = A_j \cap A_s$. For each $j \in [s-1]$, writing $r_j = |A'_j| + 1 \in [r]$, we have $\binom{n-k-r_j}{k-r_j} - |\mathcal{G}_j| \le |\mathcal{S}_{n,k,J} \setminus \mathcal{F}|$, so as $\binom{n-k-r}{k-r} \ge .1\binom{n}{k-r}$ for $k \le \sqrt{n}$ we have $1 - \mu(\mathcal{G}_j) \le 20\beta < 1/2$. However, now $\mathcal{G}_1, \dots, \mathcal{G}_{s-1}$ cross contain $A_1 \setminus A_s, \dots, A_{s-1} \setminus A_s$ by Lemma 2.9, so we have the required contradiction. □

7.2 Bootstrapping

Now we complete the proof of Theorem 7.1 by proving Lemma 7.3. The idea is to reduce to the case that the critical edge is disjoint from all other edges, so that we can apply Lemma 6.3.

Proof of Lemma 7.3. Let $G, H_1, \dots, H_\sigma$ be as in Setup 7.2. Let $C \gg \beta^{-1} \gg dr\Delta$ and $\mathcal{F}_i \subset \binom{[n]}{k_i}$ with $k_i \in [k/2, k]$ for $i \in [\sigma]$, where $C \le k \le n/Cs$. Suppose $\sum_{i=1}^{s-1}(1-\mu(\mathcal{F}_i)) \le \varepsilon \le \beta$, $\mu(\mathcal{F}_\sigma) \ge \varepsilon^d k/n$ and $1-\mu(\mathcal{F}_i) \le \varepsilon_0 := 2\varepsilon/\sigma$ for all $i \in I$.

We need to show that $\mathcal{F}_1, \dots, \mathcal{F}_\sigma$ cross contain $H_1^+, \dots, H_\sigma^+$. Write $G = \{A_1, \dots, A_s\}$ where $H_\sigma = \{A_s\}$ and $A = A_s \cap \bigcup_{i<s} A_i$. It suffices to find an injection $\phi : A \to [n]$ such that Lemma 6.3 provides a cross embedding of $e_1^+, \dots, e_s^+$ in $\mathcal{G}_1, \dots, \mathcal{G}_s$, where for each edge $A_j \in H_i$ we define $e_j = A_j \setminus A_s$ and $\mathcal{G}_j = (\mathcal{F}_i)_{\phi(A)}^{\phi(A \cap A_j)}$. We note that if $A \cap A_j = \emptyset$ then $1-\mu(\mathcal{G}_j) \le 2(1-\mu(\mathcal{F}_i))$ for any choice of ϕ. Also, for uniformly random ϕ we have $\mathbb{P}(\mu(\mathcal{G}_j) \ge 1-\sqrt{\varepsilon_0}) > 1-\sqrt{\varepsilon_0}$ whenever $i \in I$ by Lemma 4.2.

Next suppose $\mu(\mathcal{F}_\sigma) \ge e^{-k\beta}$. Then Fairness (Proposition 4.3) gives $\mathbb{P}(\mu(\mathcal{G}_s) \ge \mu(\mathcal{F}_\sigma)/2) > 1/2$. By a union bound we can fix ϕ with $\sum_{i=1}^{s-1}(1-\mu(\mathcal{G}_i)) \le 2\varepsilon + |I|\sqrt{\varepsilon_0} \le \alpha := 2\Delta\sqrt{\varepsilon}$ and $\mu(\mathcal{G}_s) \ge \mu(\mathcal{F}_\sigma)/2 \ge (\alpha k/n)^{3d}$. Then Lemma 6.3 applies as required.

It remains to consider the case $\mu(\mathcal{F}_\sigma) < e^{-k\beta}$. We will apply Lemma 4.4 to show that we can fix ϕ with $\sum_{i=1}^{s-1}(1-\mu(\mathcal{G}_i)) \le 2\varepsilon + |I|\sqrt{\varepsilon_0} \le \alpha := 2\Delta\sqrt{\varepsilon}$ as above and $\mu(\mathcal{G}_s) \ge c := \mu(\mathcal{F}_\sigma)/2k^r \ge e^{k\beta}\mu(\mathcal{F}_\sigma) \cdot \mu(\mathcal{F}_\sigma)/2k^r \ge \mu(\mathcal{F}_\sigma)^2 \ge (\alpha k/n)^{6d}$. Again this will suffice by Lemma 6.3. Lemma 4.4 with $\ell = k$ gives $\mathbb{P}(\mu(\mathcal{G}_s) \ge c) \ge (\mu(\mathcal{F}_\sigma)/2)^{r/k} \ge \varepsilon^{1/4} n^{-2r/k}$, so we are done unless $\varepsilon^{1/4} n^{-2r/k} < |I|\sqrt{\varepsilon_0}$, which implies $\sigma^2 n^{-8r/k} < (2\Delta)^4 \varepsilon$. As $\varepsilon \ll \Delta^{-1}$ this implies $k < n^\beta$, say. Furthermore, we can assume $\mathcal{F}_\sigma$ is $(2r, \mu(\mathcal{F}_\sigma)\beta n/sk)$-global, otherwise we can apply the above argument with some $(\mathcal{F}_\sigma)_R^R$ in place of $\mathcal{F}_\sigma$ to get $\mathbb{P}(\mu(\mathcal{G}_s) \ge c) \ge (\mu(\mathcal{F}_\sigma)\beta n/2sk)^{r/k} \ge \varepsilon^{1/4} s^{-2r/k} > |I|\sqrt{\varepsilon_0}$.

Now we claim that $\partial_c^r \mathcal{F}_\sigma$ is G-free. This will suffice to complete the proof, as then Lemma 4.4 gives the improved estimate $\mu(\partial_c^r \mathcal{F}_\sigma) \ge (\varepsilon^d/ks)^{2r/k} - (s/n)^\beta > |I|\sqrt{\varepsilon_0}$, using $s \le r\sigma < n^{8r/k}$. To see the claim, we suppose $\phi(G) \subset \partial_c^r \mathcal{F}_\sigma$ and will obtain a contradiction by finding a cross matching in $\mathcal{H}_1, \dots, \mathcal{H}_s$, where for each edge A_j of G we let $\mathcal{H}_j = (\mathcal{F}_\sigma)_{\operatorname{Im}\phi}^{\phi(A_j)}$. We verify the conditions of Lemma 5.7, with $(s, s, d, 2)$ in place of (s, m, d, K). As $\mathcal{F}_\sigma$ is $(2r, \mu(\mathcal{F}_\sigma)\beta n/sk)$-global, each $\mathcal{H}_j$ is $(r, 2\mu(\mathcal{F}_\sigma)\beta n/sk)$-global by Lemma 2.2. Also, $\mathcal{F}_\sigma$ is $(\beta^{-1}s, \mu(\mathcal{F}_\sigma)/2)$-uncapturable by Lemma 2.4, so each $\mu(\mathcal{H}_j) \ge \mu(\mathcal{F}_\sigma)/2 \ge \varepsilon^d k/2n$, and each $\mathcal{H}_j$ is $(s/2\beta, \varepsilon^d k/4n)$-uncapturable by Lemma 2.4. As $\sigma^2 n^{-8r/k} < (2\Delta)^4\varepsilon$ and $k < n^\beta$ we have $\varepsilon^d k/n > (3sk/n)^d$, and so the conditions of Lemma 5.7 hold. But this is a contradiction, as then $\mathcal{H}_1, \dots, \mathcal{H}_s$ cross contain a matching. Therefore $\partial_c^r \mathcal{F}_\sigma$ is G-free, as claimed. □

8 The Füredi–Jiang–Seiver Conjecture

In this section we prove the Füredi–Jiang–Seiver Conjecture on the Turán numbers of expanded paths. As previously mentioned, for paths of odd length the conjecture follows from our result on critical graphs (Theorem 1.6), so it remains to consider paths of even length. We will consider the more general setting of expansions of (normal) graphs (r-graphs with $r = 2$) satisfying the following generalised criticality property. Recall that we denote the crosscut and transversal numbers of an r-graph G by $\sigma(G)$ and $\tau(G)$, and that $\sigma(G) \ge \tau(G)$. Consider any G with $\tau(G) = \sigma(G)$. We say G is *a_1-degree-critical* if (i) $\sigma(G - x) < \sigma(G)$ for some x of degree $|G_x^x| \le a_1$, and (ii) $\tau(G - x) = \tau(G)$ for any x with $|G_x^x| < a_1$. We say G is *a_2-matching-critical* if (i) $\sigma(G \setminus M) < \sigma(G)$ for some matching M with $|M| \le a_2$, and (ii) $\tau(G \setminus M) = \tau(G)$ for any matching M with $|M| < a_2$. We say G is *(a_1, a_2)-critical* if it is both a_1-degree-critical and a_2-matching-critical.

We note that even paths and cycles are $(2, 2)$-critical, and that any G is critical (in the sense defined above) if and only if G is $(a_1, 1)$-critical, where a_1 is the minimum possible degree of any vertex belonging to any minimum size crosscut of G^+. The significance of the generalised definition is that it enables to show that the following natural construction is extremal for the Turán problem for

G^+. For any $T \subset [n]$ we write $\mathcal{G}_{n,k}(T) = \{A \in \binom{[n]}{k} : T \subset A\}$ for the family in $\binom{[n]}{k}$ generated by T. For $\mathcal{T} \subset \{0,1\}^n$ we write $\mathcal{G}_{n,k}(\mathcal{T}) = \bigcup_{T \in \mathcal{T}} \mathcal{G}_{n,k}(T)$. We let $\mathcal{F}_{n,k,G} = \mathcal{G}_{n,k}(\mathcal{T})$ where $\mathcal{T}$ is the disjoint union of $\sigma(G) - 1$ singletons and a graph $F_{a_1 a_2}$ with as many edges as possible subject to having no vertex of degree $\geq a_1$ or matching of size $\geq a_2$. Then $\mathcal{F}_{n,k,G}$ is G^+-free by definition of (a,b)-criticality. We will show that it is extremal. When G is a path of even length this will complete the proof of the Füredi–Jiang–Seiver Conjecture.

Theorem 8.1. *Let $G \in \mathcal{G}(2, \Delta, s)$ be (a_1, a_2)-critical, $C \gg a_2\Delta$ and $C \leq k \leq n/Cs$. Then $ex(n, G^+(k)) = |\mathcal{F}_{n,k,G}|$.*

Moreover, we will prove the following strong stability version.

Theorem 8.2. *Let $G \in \mathcal{G}(2, \Delta, s)$ be (a_1, a_2)-critical and $C \gg \beta^{-1} \gg a_2 d\Delta$.*

Suppose $\mathcal{F} \subset \binom{[n]}{k}$ with $C \leq k \leq n/Cs$ is G^+-free. If $|\mathcal{F}| \geq |\mathcal{S}_{n,k,\sigma-1}|$ then $|\mathcal{F} \setminus \mathcal{G}_{n,k}(\mathcal{T})| \leq \beta^{-1}\binom{n-3}{k-3}$ for some $\mathcal{T} = \{\{x\} : x \in J\} \cup F$ where $J \in \binom{[n]}{\sigma-1}$ and $F \subset \binom{[n]\setminus J}{2}$ with $|F| \leq |F_{a_1 a_2}|$.

Moreover, if $|\mathcal{F}| \geq |\mathcal{F}_{n,k,G}| - \varepsilon\binom{n-2}{k-2}$ with $\varepsilon \in (0, \beta)$ then $\mu(\mathcal{F} \setminus \mathcal{G}) \leq (\varepsilon k/n)^d$ for some copy $\mathcal{G}$ of $\mathcal{F}_{n,k,G}$, where if $k \leq \sqrt{n}$ then $\mathcal{F} \subset \mathcal{G}$.

Throughout this section we adopt the following set up.

Setup 8.3. Let $G \in \mathcal{G}'(2, s, \Delta)$ be (a_1, a_2)-critical with $\sigma(G) = \sigma$. Let $\mathcal{B} = \{B_i : i \in [a]\}$ be a r-graph matching with $r \in [2]$, and $\mathcal{B}' = \{B'_i : i \in [a]\} \subset G$, where if $r = 2$ then $a = a_2$ and each $B'_i = B_i$ or if $r = 1$ then $a = a_1$ and each $B'_i = B_i \cup \{x\}$ for some vertex x of degree a. Let $S = \{s_1, \ldots, s_{\sigma-1}\}$ be a crosscut in $(G \setminus \mathcal{B}')^+$ and let $H_i = G_{s_i}^{s_i}$ for $i \in [\sigma - 1]$. Let $I = \{i \in [\sigma - 1] : V(H_i) \cap V(\mathcal{B}) \neq \emptyset\}$.

We prove a bootstrapping lemma in the next subsection and then deduce Theorem 8.2 in the following subsection.

8.1 Bootstrapping

In this subsection we prove the following bootstrapping lemma, which is analogous to Lemma 7.3, except that rather than concluding that some family is small we conclude that some family is capturable.

Lemma 8.4. *With notation as in Setup 8.3, let $C \gg \beta^{-1} \gg ad\Delta$ and $C \leq k \leq n/Cs$. Let $\mathcal{F}_i \subset \binom{[n]}{k_i}$ with $k_i \in [k/2, k]$ for $i \in [\sigma - 1]$ and $\mathcal{F}'_i \subset \binom{[n]}{k'_i}$ with $k'_i \in [k/2, k]$ for $i \in [a]$ be such that $\mathcal{F}_1, \ldots, \mathcal{F}_{\sigma-1}, \mathcal{F}'_1, \ldots, \mathcal{F}'_a$ are cross free of $H_1^+, \ldots, H_{\sigma-1}^+, B_1^+, \ldots, B_a^+$. Suppose $\sum_{i=1}^{s-1}(1 - \mu(\mathcal{F}_i)) \leq \varepsilon \leq \beta$ and $1 - \mu(\mathcal{F}_i) \leq \varepsilon_0 := 2\varepsilon/\sigma$ for all $i \in I$. Then some $\mathcal{F}'_i$ is $(\beta^{-1}, \gamma_i + (k/n)^d)$-capturable, where $\gamma_i < \varepsilon^d$, and if $\mathcal{F}'_i$ is G^+-free then $\gamma_i < \varepsilon^d k/n$.*

The proof requires the following lemma which is analogous to Lemma 6.3.

Lemma 8.5. *Let $C \gg C' \gg ad$, $\mathcal{F}_i \subset \binom{[n]}{k_i}$ for $i \in [s]$ and $\mathcal{F}'_i \subset \binom{[n]}{k'_i}$ for $i \in [a]$ with $\sum_{i=1}^{s} k_i + \sum_{i=1}^{a} k'_i \leq n/C$. Suppose $(\mathcal{F}_1, \ldots, \mathcal{F}_s, \mathcal{F}'_1, \ldots, \mathcal{F}'_a)$ are cross free of $G = (e_1, \ldots, e_s, e'_1, \ldots, e'_a)$ with each $|e_i| = k_i$, $|e'_i| = k'_i$ and $e \cap e'_i = \emptyset$ for all $i \in [a]$ and $e'_i \neq e \in G$. If $\sum_{i=1}^{s}(1 - \mu(\mathcal{F}_i)) < 1/2$ then some $\mathcal{F}'_i$ is $(C', (k'_i/n)^d)$-capturable.*

Proof. Let $k = n/2a$ and for each $i \in [a]$ let $\mathcal{G}_i = (\mathcal{F}'_i)^{\uparrow} \cap \binom{[n]}{k}$. Then $(\mathcal{F}_1, \ldots, \mathcal{F}_s, \mathcal{G}_1, \ldots, \mathcal{G}_a)$ are cross free of G' obtained from G by enlarging each e'_i to e^*_i of size k. Suppose for contradiction that each $\mathcal{F}'_i$ is $(C', (k'_i/n)^d)$-uncapturable. Then an argument of Dinur and Friedgut, applying Russo's Lemma and Friedgut's junta theorem (see Lemma 2.7 in [4]), shows that each $\mu(\mathcal{G}_i) > 1 - 1/2a$. Consider a uniformly random injection $\phi : V(G') \to [n]$. Let E be the event that some $\phi(e_i) \notin \mathcal{F}_i$ or some $\phi(e^*_i) \notin \mathcal{G}_i$. Then $1 = \mathbb{P}(E) \leq \sum_{i \in [s]}(1 - \mu(\mathcal{F}_i)) + \sum_{i \in [a]}(1 - \mu(\mathcal{G}_i)) < 1/2 + 1/2$, contradiction. □

Proof of Lemma 8.4. With notation as in Setup 8.3, let $C \gg \beta^{-1} \gg b \gg d \gg a\Delta$ and $C \le k \le n/Cs$. Let $\mathcal{F}_i \subset \binom{[n]}{k_i}$ with $k_i \in [k/2, k]$ for $i \in [\sigma-1]$ and $\mathcal{F}'_i \subset \binom{[n]}{k'_i}$ with $k'_i \in [k/2, k]$ for $i \in [a]$ be such that $\mathcal{F}_1, \dots, \mathcal{F}_{\sigma-1}, \mathcal{F}'_1, \dots, \mathcal{F}'_a$ are cross free of $H_1^+, \dots, H_{\sigma-1}^+, B_1^+, \dots, B_a^+$. Suppose $\sum_{i=1}^{s-1}(1-\mu(\mathcal{F}_i)) \le \varepsilon \le \beta$ and $1-\mu(\mathcal{F}_i) \le \varepsilon_0 := 2\varepsilon/\sigma$ for all $i \in I$. Suppose for contradiction that each $\mathcal{F}'_i$ is $(\beta^{-1}, \gamma_i + (k/n)^d)$-uncapturable, where either $\gamma_i \ge \varepsilon^d$ or $\mathcal{F}'_i$ is G^+-free and $\gamma_i \ge \varepsilon^d k/n$.

We start by upgrading uncapturability to globalness. By Lemma 5.5 with $(b, 4, a)$ in place of (b, r, m) and each $\alpha_i = n/kb$, $\beta_i = \gamma_i + (k/n)^d$, noting that $8b \le k \le n/8a(n/bk)$, $4ba < \beta^{-1}$ and $(n/kb)^b (k/n)^d > n/k \gg 1$, there is a set S' partitioned into $S'_1, \dots, S'_a$ with each $|S'_i| \le 8b$ such that each $\mathcal{G}_i^0 := (\mathcal{F}'_i)_{S'}^{S'_i}$ is $(8, 4\mu(\mathcal{G}_i^0)n/kb)$-global with $\mu(\mathcal{G}_i^0) > \alpha_i^{1_{S'_i \ne \emptyset}} \beta_i/2$. We have $2\mu(\mathcal{G}_i^0) > \varepsilon^d + (k/n)^d$, unless $\mathcal{F}'_i$ is G^+-free and $S'_i = \emptyset$, in which case $\mathcal{G}_i^0$ is a restriction of $\mathcal{F}'_i$, so is also G^+-free, with $2\mu(\mathcal{G}_i^0) > \varepsilon^d k/n + (k/n)^d$.

Next we define $\mathcal{G}'_i := (\mathcal{F}'_i)_S^{S_i}$ with enhanced globalness, obtaining S partitioned into $S_1, \dots, S_a$ by letting $S_i = S'_i$ if $\mathcal{G}_i^0$ is $(4, \mu(\mathcal{G}_i^0)\beta n/sk)$-global, or otherwise letting $S_i = S'_i \cup R_i$ where $|R_i| \le 4$ and $\mathcal{G}_i^1 := (\mathcal{G}_i^0)_{R_i}^{R_i}$ has $\mu(\mathcal{G}_i^1) > \mu(\mathcal{G}_i^0)\beta n/sk$. We also define $\mathcal{G}_i = (\mathcal{F}_i)_S^{\emptyset}$ for $i \in [\sigma-1]$ and note that each $1 - \mu(\mathcal{G}_i) \le 2(1-\mu(\mathcal{F}_i))$.

By Lemma 2.2, each $\mathcal{G}_i^1$ or $\mathcal{G}'_i$ is $(4, 2\mu(\mathcal{G}_i^0)\beta n/sk)$-global if $R_i = \emptyset$ or $(4, 8\mu(\mathcal{G}_i^0)n/kb)$-global otherwise. By Lemma 2.4, each $\mathcal{G}_i^1$ is $(b/8, \mu(\mathcal{G}_i^1)/2)$-uncapturable, so $\mu(\mathcal{G}'_i) > \mu(\mathcal{G}_i^1)/2 \ge \mu(\mathcal{G}_i^0)/2$. Thus $2\beta^{-1}\mu(\mathcal{G}'_i) \ge \gamma'_i + (k/n)^d$, and

(i) $\mathcal{G}'_i$ is $(4, 8\mu(\mathcal{G}'_i)n/kb)$-global with $\gamma'_i \ge \varepsilon^d/s$, or

(ii) $\mathcal{G}'_i$ is G^+-free and $(4, 2\mu(\mathcal{G}'_i)\beta n/sk)$-global with $\gamma'_i \ge \varepsilon^d k/n$.

Indeed, if option (i) does not hold then $\mathcal{G}_i^0$ is G^+-free with $2\mu(\mathcal{G}_i^0) > \varepsilon^d k/n + (k/n)^d$, and also is $(4, \mu(\mathcal{G}_i^0)\beta n/sk)$-global, so $R_i = \emptyset$ and $\mathcal{G}'_i$ is a restriction of $\mathcal{G}_i^0$, so is also G^+-free.

We will show that $\mathcal{G}_1, \dots, \mathcal{G}_{\sigma-1}, \mathcal{G}'_1, \dots, \mathcal{G}'_a$ cross contain $H_1^+, \dots, H_{\sigma-1}^+, B_1^+, \dots, B_a^+$, thus obtaining the required contradiction. It suffices to find an injection $\phi : B \to [n]$, where $B = \bigcup_{i=1}^a B_i$, such that Lemma 8.5 provides a cross embedding of $e_1^+, \dots, e_s^+$ in $\mathcal{G}_1, \dots, \mathcal{G}_s$, where for each edge $A_j \in H_i$ we define $e_j = A_j \setminus B$ and $\mathcal{H}_j = (\mathcal{G}_i)_{\phi(B)}^{\phi(B \cap A_j)}$, or if $A_j = B_i$ we define $e_j = A_j \setminus B = \emptyset$ and $\mathcal{H}_j = (\mathcal{G}'_i)_{\phi(B)}^{\phi(B_i)}$.

We note that if $B \cap A_j = \emptyset$ then each $1 - \mu(\mathcal{H}_j) \le 2(1-\mu(\mathcal{G}_i))$ for any ϕ. We consider ϕ obtained by choosing independent uniformly random injections $\phi_i : B_i \to [n]$ for each $i \in [a]$. Then $\mathbb{P}(\phi \text{ is injective}) \ge 1 - 2a^2/n$ and $\mathbb{P}(\mu(\mathcal{H}_j) \ge 1 - \sqrt{\varepsilon_0}) > 1 - 2\sqrt{\varepsilon_0}$ whenever $A_j \in \bigcup_{i \in I} H_i$ by Lemma 4.2. We write E_i for the event that $\phi_i(B_i) \in \partial_{c_i}\mathcal{G}'_i$, where $c_i = b^{-.3}\mu(g'_i)$. It suffices to show that conditional on E_i each $\mathcal{H}'_i := (\mathcal{G}'_i)_{\phi(B)}^{\phi_i(B_i)}$ is $(\sqrt{b}, (k/n)^{2d})$-uncapturable, and that $\mathbb{P}(E_i) \ge \varepsilon_0^{1/3a}$.

For uncapturability, we recall that $\mathcal{G}'_i$ is $(4, 8\mu(\mathcal{G}'_i)n/kb)$-global with $2\beta^{-1}\mu(\mathcal{G}'_i) \ge (k/n)^d$. Thus $\mathcal{H}'_i$ and $\mathcal{H}''_i := (\mathcal{G}'_i)_{\phi_i(B_i)}^{\phi_i(B_i)}$ are $(2, 8\mu(\mathcal{G}'_i)n/kb)$-global by Lemma 2.2. Conditional on E_i we have $\mu(\mathcal{H}''_i) > c_i$, so $\mathcal{H}''_i$ is $(b^{.7}/16, \mu(\mathcal{H}''_i)/2)$-uncapturable by Lemma 2.4. Then $\mu(\mathcal{H}'_i) \ge \mu(\mathcal{H}''_i)/2 \ge b^{-.3}\mu(\mathcal{G}'_i)/4$, so $\mathcal{H}'_i$ is $(b^{.7}/32, \mu(\mathcal{H}'_i)/2)$-uncapturable by Lemma 2.4, and so $(\sqrt{b}, (k/n)^{2d})$-uncapturable.

It remains to show $\mathbb{P}(E_i) \ge \varepsilon_0^{1/3a}$. We may assume $\mu(\mathcal{G}'_i) < e^{-k\beta}$, otherwise this holds easily by Fairness (Proposition 4.3). As $2\beta^{-1}\mu(\mathcal{G}'_i) \ge (k/n)^d$ this gives $k < n^\beta$. By Lemma 4.4 with $\ell = b^{.1}$ we are done unless $\varepsilon_0^{1/3a} > \mathbb{P}(E_i) = \mu(\partial_{c_i}\mathcal{G}'_i) \ge (\mu(\mathcal{G}'_i)/2)^{2/\ell}$, which implies $\gamma'_i + (k/n)^d \le 2\beta^{-1}\mu(\mathcal{G}'_i) < (\varepsilon/s)^{b^{.05}}$. As $\gamma'_i < \varepsilon^d/s$ we have option (ii) above, so $\mathcal{G}'_i$ is G^+-free. As $\varepsilon^d k/n \le \gamma'_i < (\varepsilon/s)^{b^{.05}}$ we also have $s < \varepsilon n^{b^{-.05}}$.

Now we claim that $\partial_{c_i}^2\mathcal{G}'_i$ is G-free. This will suffice to complete the proof, as then Lemma 4.4 gives the improved estimate $\mu(\partial_{c_i}^2\mathcal{G}'_i) \ge (\varepsilon^d k/sb + k/n - (s/n)^2)^{b^{-.02}} > (\varepsilon/s)^{b^{-.01}}$. To see the claim, we suppose $\phi'(G) \subset \partial_{c_i}^2\mathcal{G}'_i$ and will obtain a contradiction by finding a cross matching in $\mathcal{A}_1, \dots, \mathcal{A}_s$, where for each edge A_j of G we let $\mathcal{A}_j = (\mathcal{G}'_i)_{\mathrm{Im}\,\phi'}^{\phi'(A_j)}$. We verify the conditions of Lemma 5.7, with $(s, s, d, 2)$ in place of (s, m, d, K). As $\mathcal{G}'_i$ is $(4, 2\mu(\mathcal{G}'_i)\beta n/sk)$-global, each $\mathcal{H}_j$ is $(2, 4\mu(\mathcal{G}'_i)\beta n/sk)$-global by Lemma 2.2. Also, $\mathcal{G}'_i$ is $(s/4\beta, \mu(\mathcal{G}'_i)/2)$-uncapturable by Lemma 2.4, so each $\mu(\mathcal{H}_j) \ge \mu(\mathcal{G}'_i)/2 \ge \beta\varepsilon^d k/4n$, and each $\mathcal{H}_j$ is $(s/8\beta, \beta\varepsilon^d k/8n)$-uncapturable by Lemma 2.4. As $s < \varepsilon n^{b^{-.05}}$ and $k < n^\beta$ we have $\beta\varepsilon^d k/8n > (3sk/n)^d$, so the required conditions hold. $\square$

8.2 Strong stability

We conclude with the proof of the main result of this section.

Proof of Theorem 8.2. Let $G \in \mathcal{G}(2, \Delta, s)$ be (a_1, a_2)-critical and $C \gg \beta^{-1} \gg b \gg d \gg a_2\Delta$. Suppose $\mathcal{F} \subset \binom{[n]}{k}$ with $C \le k \le n/Cs$ is G^+-free and $|\mathcal{F}| \ge |\mathcal{S}_{n,k,\sigma-1}|$.

By Theorem 1.9 (refined junta approximation) there is $J \in \binom{[n]}{\sigma-1}$ such that $|\mathcal{F} \setminus \mathcal{S}_{n,k,J}| = \delta\binom{n-1}{k-1}$ with $\delta^{-1} \gg bd\Delta$. We write $J = \{j_1, \dots, j_{\sigma-1}\}$, let $\mathcal{F}_i = \mathcal{F}_J^{j_i}$ for $i \in [\sigma-1]$, say with $|\mathcal{F}_1| \ge \dots \ge |\mathcal{F}_{\sigma-1}|$, and note that $\mathcal{F}_J^\emptyset$ is G^+-free. As in the proof of Theorem 7.1, we have $\sum_{i=1}^{\sigma-1}(1 - \mu(\mathcal{F}_i)) \le 2\delta$, so $1 - \mu(\mathcal{F}_i) \le 4r\Delta\delta/\sigma$ for any $i \le \min\{r\Delta, \sigma-1\}$.

As G is a_2-matching-critical, we can define $H_1^2, \dots, H_{\sigma-1}^2$, $B_1^2, \dots, B_{a_2}^2$ and I^2 as in Setup 8.3 with $r = 2$ and $a = a_2$, where we identify I^2 with $[|I^2|]$. Letting $\mathcal{F}'_i = \mathcal{F}_J^\emptyset$ for $i \in [a_2]$, we have $\mathcal{F}_1, \dots, \mathcal{F}_{\sigma-1}, \mathcal{F}'_1, \dots, \mathcal{F}'_{a_2}$ cross free of $(H_1^2)^+, \dots, (H_{\sigma-1}^2)^+, (B_1^2)^+, \dots, (B_{a_2}^2)^+$, so $\mathcal{F}_J^\emptyset$ is $(b, (2\delta)^d k/n + (k/n)^d)$-capturable by Lemma 8.4. We fix $J' \in \binom{[n]\setminus J}{b}$ so that $\mu(\mathcal{F}_{J\cup J'}^\emptyset) < (2\delta)^d k/n + (k/n)^d$.

As G is a_1-degree-critical, we can define $H_1^1, \dots, H_{\sigma-1}^1$, $B_1^1, \dots, B_a^1$ and I^1 as in Setup 8.3 with $r = 1$ and $a = a_1$, where we identify I^1 with $[|I^1|]$. For each $x \in J'$, letting $\mathcal{F}'_i = \mathcal{F}_{J\cup\{x\}}^x$ for $i \in [a_1]$, we have $\mathcal{F}_1, \dots, \mathcal{F}_{\sigma-1}, \mathcal{F}'_1, \dots, \mathcal{F}'_{a_1}$ cross free of $(H_1^1)^+, \dots, (H_{\sigma-1}^1)^+, (B_1^1)^+, \dots, (B_{a_2}^1)^+$, so $\mathcal{F}_{J\cup\{x\}}^x$ is $(b, (2\delta)^d k/n + (k/n)^d)$-capturable by Lemma 8.4. We fix $J_x \in \binom{[n]\setminus(J\cup\{x\})}{b}$ so that $\mu((\mathcal{F}_{J\cup\{x\}}^x)_{J_x}^\emptyset) < (2\delta)^d k/n + (k/n)^d$.

Let $F = \{T \in \binom{[n]\setminus J}{2} : \mu(\mathcal{F}_{T\cup J}^T) > bk/n\}$. Then $F \subset F' := \{xy : x \in J', y \in J_x\}$ and $|F'| \le b^2$. Writing $\mathcal{T} = \{\{x\} : x \in J\} \cup F$, we have $|\mathcal{F} \setminus \mathcal{G}_{n,k}(\mathcal{T})| \le |\mathcal{F}_{J\cup J'}^\emptyset| + \sum_{x\in J'} |\mathcal{F}_{J\cup\{x\}\cup J_x}^x| + \sum_{T\in F'} |\mathcal{F}_{T\cup J}^T|$, so $\mu(\mathcal{F} \setminus \mathcal{G}_{n,k}(\mathcal{T})) \le ((2\delta)^d k/n + (k/n)^d)(1 + bk/n) + (bk/n)^3$. Writing $\mathcal{G} := \mathcal{G}_{n,k}(\mathcal{T})$, as $|\mathcal{F} \setminus \mathcal{G}| \ge |\mathcal{F} \setminus \mathcal{S}_{n,k,J}| - |\mathcal{G}_{n,k}(F)|$ we also have $\mu(\mathcal{F} \setminus \mathcal{G}) \ge \delta k/n - (bk/n)^2$. We deduce $\delta k/n \le (2\delta)^d k/n + 2(bk/n)^2$, so $\delta \le 3bk/n$, giving $|\mathcal{F} \setminus \mathcal{G}| \le 2b^3\binom{n-3}{k-3}$.

To complete the proof of the first statement of the theorem, it remains to show $|F| \le |F_{a_1a_2}|$. To see this, note that otherwise F contains some $F_0 = (T_i : i \in [a_r])$, where $r = 2$ and F_0 is a matching or $r = 1$ and F_0 is a star. Writing $\mathcal{F}'_i = \mathcal{F}_{J\cup T_i}^{T_i}$, we have $\mathcal{F}_1, \dots, \mathcal{F}_{\sigma-1}, \mathcal{F}'_1, \dots, \mathcal{F}'_{a_2}$ cross free of $(H_1^r)^+, \dots, (H_{\sigma-1}^r)^+, (B_1^r)^+, \dots, (B_{a_r}^r)^+$, so some $\mathcal{F}'_i$ is $(b/2, (k/n)^d)$-capturable by Lemma 8.4. However, $\mu(\mathcal{F}'_i) > bk/n$ as $T_i \in \mathcal{F}$, so we have a contradiction.

Now suppose $|\mathcal{F}| \ge |\mathcal{F}_{n,k,G}| - \varepsilon\binom{n-2}{k-2}$ with $\varepsilon \in (0, \beta)$. We have

$$\mu(\mathcal{F}) \le \mu(\mathcal{F} \setminus \mathcal{G}) + \mu(\mathcal{G}) - \tfrac{k}{2n}\sum_{i=1}^{\sigma-1}(1 - \mu(\mathcal{F}_i)) - \tfrac{k^2}{2n^2}\sum_{T\in F}(1 - \mu(\mathcal{F}_{J\cup T}^T)),$$

where $\mu(\mathcal{F} \setminus \mathcal{G}) \le 2(bk/n)^3$ and $\mu(\mathcal{G}) \le \mu(\mathcal{F}_{n.k.G}) - (|F_{a_1a_2}| - |F|)k^2/2n^2 \le \mu(\mathcal{F}) + (|F_{a_1a_2}| - |F| + 2\varepsilon)k^2/2n^2$. Thus $|F| = |F_{a_1a_2}|$, so $\mathcal{G} := \mathcal{G}_{n,k}(\mathcal{T})$ is a copy of $\mathcal{F}_{n,k,G}$, and

$$\sum_{i=1}^{\sigma-1}(1 - \mu(\mathcal{F}_i)) + \sum_{T\in F}(1 - \mu(\mathcal{F}_{J\cup T}^T)) \le 3\varepsilon.$$

Next we suppose for contradiction that $\mu(\mathcal{F}\setminus\mathcal{G}) > (\varepsilon k/n)^d$. We fix some $T \in \binom{[n]\setminus J}{2} \setminus F$ with $\mu(\mathcal{F}_{J\cup T}^T) > (\varepsilon k/n)^{d+2}$. By maximality of $F_{a_1a_2}$ we can fix a matching $T_1, \dots, T_{a_2}$ in F with $T_{a_2} = T$. Writing $\mathcal{F}'_i = \mathcal{F}_{J\cup T_i}^{T_i}$, we have $\mathcal{F}_1, \dots, \mathcal{F}_{\sigma-1}, \mathcal{F}'_1, \dots, \mathcal{F}'_{a_2}$ cross free of $(H_1^2)^+, \dots, (H_{\sigma-1}^2)^+, (B_1^2)^+, \dots, (B_{a_2}^2)^+$. Thus Lemma 6.3 gives the required contradiction, so $\mu(\mathcal{F} \setminus \mathcal{G}) \le (\varepsilon k/n)^d$, as required.

Finally, let $k \le \sqrt{n}$ and suppose for contradiction that there is some $A \in \mathcal{F} \setminus \mathcal{G}$. From the previous statement we have $|\mathcal{G} \setminus \mathcal{F}| \le 2\beta\binom{n-2}{k-2}$. We fix any $T \in \binom{[n]\setminus J}{2}$ with $T \subset A$, a matching

$T_1, \dots, T_{a_2}$ in F with $T_{a_2} = T$, and a bijection $\phi : B^2_{a_2} \to T$. Writing $A'_j = A_j \cap A_s$ for each edge A_j of G, where $A_s = B^2_{a_2}$, we define $\mathcal{G}_1, \dots, \mathcal{G}_{s-1}$ by $\mathcal{G}_j = (\mathcal{F}_i)_A^{\phi(A'_j)}$ if $A_j \in H_i$ with $i \in [\sigma - 1]$ or $\mathcal{G}_j = (\mathcal{F}_J^{\emptyset})_A^{\phi(A'_j)}$ if $A_j = B^2_i$ with $i \in [a_2 - 1]$. For each $j \in [s-1]$, writing $r_j = |A'_j| + 1 \in [2]$, we have $\binom{n-k-r_j}{k-r_j} - |\mathcal{G}_j| \le |\mathcal{G} \setminus \mathcal{F}|$, so as $\binom{n-k-2}{k-2} \ge .1\binom{n}{k-2}$ for $k \le \sqrt{n}$ we have $1 - \mu(\mathcal{G}_j) \le 20\beta < 1/2$. However, now $\mathcal{G}_1, \dots, \mathcal{G}_{s-1}$ cross contain $A_1 \setminus A_s, \dots, A_{s-1} \setminus A_s$ by Lemma 2.9, so we have the required contradiction. $\square$

Acknowledgment

We would like to thank Yuval Filmus, Ehud Friedgut, Gil Kalai, Nathan Keller, Guy Kindler, and Muli Safra for various helpful comments and suggestions.

References

[1] Béla Bollobás, David E. Daykin, and Paul Erdős. Sets of independent edges of a hypergraph. *Quart. J. Math. Oxford Ser.(2)*, 27(105):25–32, 1976.

[2] Béla Bollobás and Andrew G Thomason. Threshold functions. *Combinatorica*, 7(1):35–38, 1987.

[3] David Conlon, Jacob Fox, and Benny Sudakov. Ramsey numbers of sparse hypergraphs. *Random Structures & Algorithms*, 35(1):1–14, 2009.

[4] Irit Dinur and Ehud Friedgut. Intersecting families are essentially contained in juntas. *Combinatorics, Probability and Computing*, 18(1-2):107–122, 2009.

[5] Irit Dinur and Samuel Safra. On the hardness of approximating minimum vertex cover. *Annals of Mathematics*, pages 439–485, 2005.

[6] David Ellis, Yuval Filmus, and Ehud Friedgut. Triangle-intersecting families of graphs. *Journal of the European Mathematical Society*, 14:841–885, 2012.

[7] David Ellis, Ehud Friedgut, and Haran Pilpel. Intersecting families of permutations. *Journal of the American Mathematical Society*, 24(3):649–682, 2011.

[8] David Ellis, Nathan Keller, and Noam Lifshitz. Stability versions of Erdős–Ko–Rado type theorems, via isoperimetry. *Journal of the European Mathematical Society*, 21:3857–3902, 2019.

[9] David Ellis and Bhargav Narayanan. On symmetric 3-wise intersecting families. *Proceedings of the American Mathematical Society*, 145(7):2843–2847, 2017.

[10] P. Erdős and T. Gallai. On maximal paths and circuits of graphs. *Acta Math. Acad. Sci. Hungar.*, 10:337–356, 1959.

[11] Paul Erdős. A problem on independent r-tuples. *Ann. Univ. Sci. Budapest*, 8:93–95, 1965.

[12] Paul Erdős, Chao Ko, and Richard Rado. Intersection theorems for systems of finite sets. *The Quarterly Journal of Mathematics*, 12(1):313–320, 1961.

[13] Peter Frankl. Improved bounds for Erdős matching conjecture. *Journal of Combinatorial Theory, Series A*, 120(5):1068–1072, 2013.

[14] Peter Frankl and Zoltán Füredi. Exact solution of some Turán-type problems. *Journal of Combinatorial Theory, Series A*, 45(2):226–262, 1987.

[15] Peter Frankl and Andrey Kupavskii. Simple juntas for shifted families. *Discrete Analysis*, 14, 2020.

[16] Peter Frankl and Andrey Kupavskii. The Erdős matching conjecture and concentration inequalities. *Journal of Combinatorial Theory, Series B*, 157:366–400, 2022.

[17] Peter Frankl, Vojtech Rödl, and Andrzej Ruciński. On the maximum number of edges in a triple system not containing a disjoint family of a given size. *Combinatorics, Probability and Computing*, 21(1-2):141–148, 2012.

[18] Peter Frankl and Norihide Tokushige. Weighted multiply intersecting families. *Studia Scientiarum Mathematicarum Hungarica*, 40(3):287–291, 2003.

[19] Ehud Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–35, 1998.

[20] Zoltán Füredi. Linear trees in uniform hypergraphs. *European Journal of Combinatorics*, 35:264–272, 2014.

[21] Zoltán Füredi and Tao Jiang. Hypergraph Turán numbers of linear cycles. *Journal of Combinatorial Theory, Series A*, 123(1):252–270, 2014.

[22] Zoltán Füredi, Tao Jiang, and Robert Seiver. Exact solution of the hypergraph Turán problem for k-uniform linear paths. *Combinatorica*, 34(3):299–322, 2014.

[23] Spencer Greenberg and Mehryar Mohri. Tight lower bound on the probability of a binomial exceeding its expectation. *Statistics & Probability Letters*, 86:91 – 98, 2014.

[24] Hao Huang, Po-Shen Loh, and Benny Sudakov. The size of a hypergraph and its matching number. *Combinatorics, Probability and Computing*, 21(03):442–450, 2012.

[25] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs*. Wiley-Interscience, 2000.

[26] Gyula Katona. A theorem of finite sets. In *Classic Papers in Combinatorics*, pages 381–401. Springer, 2009.

[27] Peter Keevash. Hypergraph Turán problems. *Surveys in combinatorics*, 392:83–140, 2011.

[28] Peter Keevash. Hypergraph matchings and designs. *Proceedings of the 2018 ICM*, 2018.

[29] Peter Keevash, Noam Lifshitz, Eoin Long, and Dor Minzer. Hypercontractivity for global functions and sharp thresholds. *Journal of the American Mathematical Society*, 37:245–279, 2024.

[30] Nathan Keller and Noam Lifshitz. The junta method for hypergraphs and Chvátal's simplex conjecture. *Advances in Mathematics*, 392:107991, 2021.

[31] Alexandr Kostochka, Dhruv Mubayi, and Jacques Verstraëte. Turán problems and shadows I: paths and cycles. *Journal of Combinatorial Theory, Series A*, 129:57–79, 2015.

[32] Alexandr Kostochka, Dhruv Mubayi, and Jacques Verstraëte. Turán problems and shadows II: trees. *Journal of Combinatorial Theory, Series B*, 122:457–478, 2017.

[33] Joseph B Kruskal. The number of simplices in a complex. *Mathematical optimization techniques*, page 251, 1963.

[34] László Lovász. *Combinatorial Problems and Exercises*. North-Holland, Amsterdam, 1993.

[35] László Lovász and M.D. Plummer. *Matching theory*. AMS Chelsea Publishing, 2009.

[36] Hongliang Lu, Yan Wang, and Xingxing Yu. A better bound on the size of rainbow matchings. *Journal of Combinatorial Theory, Series A*, 195:105700, 2023.

[37] Tomasz Łuczak and Katarzyna Mieczkowska. On Erdős' extremal problem on matchings in hypergraphs. *Journal of Combinatorial Theory, Series A*, 124:178–194, 2014.

[38] Dhruv Mubayi and Jacques Verstraëte. A survey of Turán problems for expansions. In *Recent Trends in Combinatorics*, pages 117–143. Springer, 2016.

[39] Alexander Schrijver. *Combitorial Optimization: Polyhedra and Efficiancy*. Springer-Verlag Berlin Heidelberg, 2003.