Received: 6 May 2021 | Accepted: 2 February 2022 | Published on: 16 April 2022

DOI: 10.1002/rsa.21090

RESEARCH ARTICLE

WILEY

Sharp threshold for the Erdős–Ko–Rado theorem

József Balogh[1,2] | **Robert A. Krueger**[1] | **Haoran Luo**[1]

[1]Department of Mathematics, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

[2]Moscow Institute of Physics and Technology, Moscow, Russia

Correspondence

Robert A. Krueger, Department of Mathematics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.

Email: rak5@illinois.edu

Abstract

For positive integers n and k with $n \geq 2k+1$, the Kneser graph $K(n,k)$ is the graph with vertex set consisting of all k-sets of $\{1, \dots, n\}$, where two k-sets are adjacent exactly when they are disjoint. The independent sets of $K(n,k)$ are k-uniform intersecting families, and hence the maximum size independent sets are given by the Erdős–Ko–Rado Theorem. Let $K_p(n,k)$ be a random spanning subgraph of $K(n,k)$ where each edge is included independently with probability p. Bollobás, Narayanan, and Raigorodskii asked for what p does $K_p(n,k)$ have the same independence number as $K(n,k)$ with high probability. For $n = 2k+1$, we prove a hitting time result, which gives a sharp threshold for this problem at $p = 3/4$. Additionally, completing work of Das and Tran and work of Devlin and Kahn, we determine a sharp threshold function for all $n > 2k+1$.

KEYWORDS

hitting time, intersecting families, Kneser graph, transference

1 | INTRODUCTION

For a positive integer n, we denote by $[n]$ the set $\{1, 2, \dots, n\}$. For a set A, we denote by $\binom{A}{k}$ $\left(\text{resp. } \binom{A}{\leq k}\right)$ the collection of all subsets of A of size (resp. at most) k.

For positive integers n and k, the *Kneser graph* $K(n,k)$ is the graph on $\binom{[n]}{k}$, where two sets are adjacent exactly when they are disjoint. This graph has no edges unless $n \geq 2k$, $K(2k,k)$ is just a matching, and $K(n,1)$ is a complete graph, so we assume from now on that $n \geq 2k+1$ and $k \geq 2$. The independent sets of $K(n,k)$ are k-uniform intersecting families, important objects of research in

extremal combinatorics. For $x \in [n]$, we call $\mathcal{K}_x = \left\{ S \in \binom{[n]}{k} : x \in S \right\}$ the *star* based at x. Note that $\mathcal{K}_x$ is an independent set of $K(n,k)$ of size $\binom{n-1}{k-1}$. The classical Erdős–Ko–Rado theorem [11] states that $\alpha(K(n,k)) = \binom{n-1}{k-1}$ when $n \geq 2k$, where $\alpha(G)$ is the maximum size of an independent set in G. An extension by Hilton and Milner [17] implies that every maximum independent set of the Kneser graph is a star, when $n \geq 2k+1$.

Let $K_p(n,k)$ be a random spanning subgraph of $K(n,k)$, where each edge is included independently with probability $p = p(n,k)$. If $E = E(n,k)$ is an event defined on the probability space associated with $K_p(n,k)$, we say E occurs *with high probability* (*w.h.p.*) if the probability of E tends to 1 as $n \to \infty$. In general, we consider k to depend on n, so we need only take $n \to \infty$. Bollobás, Narayanan, and Raigorodskii [8] proposed a random variant of the Erdős–Ko–Rado theorem, following a recent trend of so-called transference results.

Question 1.1 ([8])**.** For what p does $\alpha(K_p(n,k)) = \alpha(K(n,k))$ with high probability?

Informally, this is asking whether the Erdős–Ko–Rado theorem still holds if we "forget" that a randomly selected portion of disjoint pairs are in fact disjoint. Since the probability that $\alpha(K_p(n,k)) = \binom{n-1}{k-1}$ increases with p, to answer Question 1.1, one wants to determine the "smallest" p for which $\alpha(K_p(n,k)) = \binom{n-1}{k-1}$ w.h.p.

An obvious necessary condition to $\alpha(K_p(n,k)) = \binom{n-1}{k-1}$ is the maximality of stars as independent sets. A *superstar based at* x is a set family of the form $\{A\} \cup \mathcal{K}_x$, where $A \notin \mathcal{K}_x$. Thus $\alpha(K_p(n,k)) > \binom{n-1}{k-1}$ if there exists a superstar spanning an independent set. The expected number of independent superstars in $K_p(n,k)$ is

$$n \binom{n-1}{k} (1-p)^{\binom{n-k-1}{k-1}},$$

which changes from $o(1)$ to $\omega(1)$ when $p \approx p_0$, where

$$p = p_0 := \begin{cases} 3/4, & \text{if } n = 2k+1, \\ \log\left(n \binom{n-1}{k}\right) / \binom{n-k-1}{k-1}, & \text{if } n > 2k+1, \end{cases} \tag{1}$$

and log denotes the natural logarithm. The only reason that $n = 2k+1$ seems different is because we lose the approximation $1 - p_0 \approx e^{-p_0}$ in that case. A straightforward second moment calculation (see [8], [9], or Observation 4.1) gives that for every constant $\varepsilon > 0$, for every $p \leq (1-\varepsilon)p_0$, $K_p(n,k)$ has an independent superstar w.h.p., and thus $\alpha(K_p(n,k)) > \binom{n-1}{k-1}$ w.h.p.

When they posed Question 1.1, Bollobás, Narayanan, and Raigorodskii [8] effectively answered it for $k = o(n^{1/3})$. They showed that for all constant $\varepsilon > 0$, for all $p \geq (1+\varepsilon)p_0$, $\alpha(K_p(n,k)) = \binom{n-1}{k-1}$ w.h.p., showing that p_0 is a "sharp threshold function" for this problem when $k = o(n^{1/3})$. Subsequently, Balogh, Bollobás, and Narayanan [2] made progress on Question 1.1 for all $k \leq (\frac{1}{2} - \gamma)n$, for every constant $\gamma > 0$. Das and Tran [9] formulated a removal lemma for $K(n,k)$, showed that p_0 is a sharp threshold function for $k \leq n/C$, for some constant C, and showed that p_0 is a "coarse" threshold function for $k \leq (\frac{1}{2} - \gamma)n$, for every constant $\gamma > 0$, meaning that there exists some constant $c = c(\gamma) > 1$ such that for all $p \geq cp_0$, $\alpha(K_p(n,k)) = \binom{n-1}{k-1}$. Concurrently, Devlin and Kahn [10] showed that p_0

is a coarse threshold function for all $n \geq 2k+1$, that is, they removed the dependence of c on γ. In particular, they made the first progress on Question 1.1 for $n = 2k+1$, showing that for some $\varepsilon > 0$, $\alpha(K_{(1-\varepsilon)}(2k+1,k)) = \binom{2k}{k-1}$ w.h.p. What remains is to show that p_0 is a sharp threshold for $k \geq n/C$, which is what we do in the present work.

In fact, we can prove something stronger, as most previous papers have done: above a threshold, not only are stars maximum independent sets, but they are unique. We say that a spanning subgraph of $K(n,k)$ is *EKR* if every maximum independent set is a star. Clearly, if $K_p(n,k)$ is EKR w.h.p., then $\alpha(K_p(n,k)) = \binom{n-1}{k-1}$ w.h.p. Thus, to resolve Question 1.1, it suffices to show that for $p \geq (1+\varepsilon)p_0$, $K_p(n,k)$ is EKR w.h.p.

Theorem 1.2. *Let $\varepsilon > 0$ be a constant. If $n \geq 2k+1$ and $p \geq (1+\varepsilon)p_0$, then $K_p(n,k)$ is EKR w.h.p.*

Perhaps Theorem 1.2 does not completely resolve Question 1.1 because of the ε. Bollobás, Narayanan, and Raigorodskii [8] raise the natural question of determining the "width" of the sharp threshold for Question 1.1: roughly, how small, in terms of n and k, may we take ε in Theorem 1.2? Das and Tran [9] proved Theorem 1.2 for all $\varepsilon \gg \frac{1}{k}$ and $k \leq n/C$. Such tight control of ε and the definition of p_0 indicate that in some sense the independence of superstars is the most difficult obstacle to overcome for $\alpha(K_p(n,k))$ to equal $\binom{n-1}{k-1}$. Das and Tran [9] suggested a stronger, hitting time version of Question 1.1 to capture this notion.

To give the statement, we need to consider the *random subgraph process*. Consider a graph G, and let $e_1, \ldots, e_m$ be the edges of G. Given a permutation $\pi : E(G) \to [m]$, we define $G_i(\pi)$ to be the spanning subgraph of G with edges e_k satisfying $\pi(e_k) \leq i$. Note that $G_i(\pi)$ has exactly i edges. If we select π uniformly at random, we can think of the sequence $G_1(\pi), G_2(\pi), \ldots, G_m(\pi)$ as adding one edge of G picked uniformly at random to $V(G)$ until there are no more edges to be added. For a monotone property P which G satisfies, we define the *hitting time* τ_P to be the minimum i such that G_i satisfies property P. Note that τ_P is a random variable depending on the permutation π, and G_{τ_P} is a random subgraph of G satisfying property P, but G_{τ_P} is not necessarily chosen uniformly from all such subgraphs.

For $n = 2k+1$, we prove that as soon as stars are maximal independent sets, stars are maximum independent sets w.h.p., where the "with high probability" is in the context of the underlying probability space of the random subgraph process. To be precise, let

$$\tau_{\text{super}} = \min\{i : K_i(n,k) \text{ has no independent superstars}\},$$
$$\tau_\alpha = \min\left\{i : \alpha(K_i(n,k)) = \binom{n-1}{k-1}\right\}.$$

Theorem 1.3. *With high probability, $\tau_\alpha = \tau_{\text{super}}$ when $n = 2k+1$.*

This theorem reduces Question 1.1 to: for what p is it true that $K_p(n,k)$ has no independent superstars? We defined p_0 to answer this question, so with an easy first moment calculation (see Observation 4.2), Theorem 1.3 gives the sharp threshold for $\alpha(K_p(2k+1,k)) = \binom{2k}{k-1}$.

Analogously, we can prove a hitting time theorem for the EKR property, the property that stars are the unique maximum independent sets. A *near-star based at x* is a set family of the form $\{A\} \cup (\mathcal{K}_x \backslash \{B\})$, where $A \notin \mathcal{K}_x$ and $B \in \mathcal{K}_x$. The obvious necessary condition to $K_p(n,k)$ being EKR is that there are no independent near-stars. Let

$$\tau_{\text{near}} = \min\{i : K_i(n,k) \text{ has no independent near-stars}\},$$
$$\tau_{\text{EKR}} = \min\{i : K_i(n,k) \text{ is EKR}\}.$$

Theorem 1.4. *W.h.p,* $\tau_{\mathrm{EKR}} = \tau_{\mathrm{near}}$ *when* $n = 2k + 1$.

In fact, p_0 is a sharp threshold function for both lacking independent superstars and lacking independent near-stars, so Theorem 1.4 proves Theorem 1.2 for $n = 2k + 1$. But since $\tau_{\mathrm{super}} < \tau_{\mathrm{near}}$ deterministically, neither Theorem 1.3 nor Theorem 1.4 implies the other.

A full resolution of Question 1.1 would be the proof of hitting time theorems for all n and k. We believe this ought to be true.

Conjecture 1.5. *W.h.p., for all* $n > 2k + 1$, $\tau_\alpha = \tau_{\mathrm{super}}$ *and* $\tau_{\mathrm{near}} = \tau_{\mathrm{EKR}}$.

Our methods for Theorems 1.4 and 1.3 extend easily to when $n - 2k = o(n)$, and a careful analysis of [8, 9] proves Conjecture 1.5 when $k = o(n)$. To name a particular case of Conjecture 1.5 which we do not know how to solve, consider $n = 4k$. See Section 4.5 for some technical remarks on Conjecture 1.5.

Having fully addressed Question 1.1, we now propose an extension of Theorems 1.3 and 1.4, and we mention other modern extensions of the Erdős–Ko–Rado theorem. With our techniques, one should be able to obtain finer information about the independent sets of $K_p(n, k)$ for arbitrary p, possibly answering the following generalization of Question 1.1: what is the size of the largest independent set of $K_p(n, k)$ not contained in a star? Conjecture 1.5 determines when the size of the largest independent set not contained in a star is at most $\binom{n-1}{k-1} - \ell + 1$, for $\ell \in \{1, 2\}$. The Hilton–Milner theorem [17] states that the largest independent set of $K(n, k)$ not contained in a star has size $\binom{n-1}{k-1} - \binom{n-k-1}{k-1} + 1$. Thus we can formulate a generalization of Conjecture 1.5 as follows. Define

$$\tau_\ell = \min\{i : \text{ for every } x \in [n] \text{ and } A \notin \mathcal{K}_x, d_{K_i(n,k)}(A, \mathcal{K}_x) \geq \ell\},$$

where $d_G(v, S)$ is the number of neighbors in G of v in $S \subseteq V(G)$. Note that $\tau_1 = \tau_{\mathrm{super}}$, $\tau_2 = \tau_{\mathrm{near}}$, and for $\ell = \binom{n-k-1}{k-1}$, $\tau_\ell = |E(K(n, k))|$ always. In this way, the $\tau_\ell / |E(K(n, k))|$ fill the space between p_0 and 1.

Question 1.6. Does the largest independent set of $K_{\tau_\ell}(n, k)$ not contained in a star have size $\binom{n-1}{k-1} - \ell + 1$ with high probability?

We hesitate to call this a conjecture, because while presumably our techniques would work for small ℓ, the situation for large ℓ is less clear.

Our proof method does not require too much structural information about $K(n, k)$; in fact, all that is needed is some regularity in the degrees, that the maximum independent sets are "spread out" in some sense, and some edge-isoperimetry, which is some upper bound on the number of edges a set of vertices can contain. The Kneser graph is a special class of so-called distance graphs, where similar problems have been studied [7], so it is likely our methods may be applied there as well. Arguably the most studied distance graph is the hypercube. There are many results concerning various graph parameters of the (edge-)random subgraph of the hypercube; see [21] for a comprehensive, but slightly outdated, survey. After writing this paper, we learned that some of those results use a method similar to our Lemma 3.1. In particular, Kostochka [20] gives a result similar to our Lemma 3.1 but in greater generality.

A more difficult, but natural, random variant of the Erdős–Ko–Rado theorem was proposed by Balogh, Bohman, and Mubayi [1]: instead of taking a random spanning subgraph of $K(n, k)$, we take a random induced subgraph of $K(n, k)$, including each vertex with probability p. For what p is a maximum

independent set of this random induced subgraph contained in a star? This is more naturally stated in the hypergraph version of the Erdős–Ko–Rado theorem, where we retain hyperedges with probability p. Balogh, Bohman, and Mubayi [1] proved a number of results, mostly for $k \leq n^{1/2-\epsilon}$. This is a particularly difficult problem because the property of interest is not monotone; for small p and large p the property holds, but sometimes for some p in between, the property does not hold. Hamm and Kahn improved the results for $n^{1/3} \ll k \leq \frac{1}{2}\sqrt{n \log n}$ [15] and $n = 2k + 1$ [16]. Gauy, Hàn and Oliveira [14] extended [1] and gave the asymptotic size of largest intersecting family for all k and almost all p. Balogh, Das, Delcourt, Liu, and Sharifzadeh [3] gave additional results up to $k \leq n/4$, but generally not as tight as [15]. Hamm and Kahn [16] optimistically conjecture that, just like in the problem addressed in this paper, the maximum independent sets are contained in stars (roughly) when the stars induce maximal independent sets.

Balogh, Das, Delcourt, Liu, and Sharifzadeh [3] also study the enumeration variant of the Erdős–Ko–Rado theorem, determining the order of magnitude of the log of the number of independent sets in $K(n, k)$. Balogh, Das, Liu, Sharifzadeh, and Tran [4] and independently Kupavskii and Frankl [12] strengthened this to say that most independent sets of $K(n, k)$ are contained in stars for $n \geq 2k + c\sqrt{k \log k}$, where c is a large constant in [4] and $c = 2$ in [12]. Balogh, Garcia, Li, and Wagner [5] push this $c\sqrt{k \log k}$ down to $100 \log k$. They conjecture it is true down to $n \geq 2k + 2$, since a simple computation shows that the families which show the tightness of the Hilton–Milner theorem [17] outnumber the trivial families for $n = 2k + 1$.

The rest of the paper is organized as follows. In Section 2 we collect results from prior works that we use in our proofs. In Section 3 we prove a lemma on counting sets of vertices of $K(n, k)$, in a similar spirit to the container method [6, 26]. We prove Theorems 1.3 and 1.4 in Section 4 almost simultaneously, definitively answering the $n = 2k + 1$ case of Question 1.1. Section 4.5 contains some technical remarks concerning Conjecture 1.5. In Section 5, we prove Theorem 1.2 for $2k+1 < n \leq Ck$, where C is a large constant, the other cases being proven by Theorem 1.4 or [8, 9].

1.1 | Standard estimates and notation

All logarithms are base e. We make extensive use of standard asymptotic notation to simplify our calculations. We say $f(n) = O(g(n))$ (resp. $f(n) = \Omega(g(n))$) if there exists $C > 0$ such that $f(n) \leq Cg(n)$ (resp. $f(n) \geq Cg(n)$) for all n sufficiently large. If $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$, then we say $f(n) = \Theta(g(n))$. We say $f(n) = o(g(n))$ (resp. $f(n) = \omega(g(n))$) if $f(n)/g(n) \to 0$ (resp. ∞) as $n \to \infty$. Since there is only one variable, n, tending to infinity, with the other variables being clearly dependent or clearly independent of n, we find the asymptotic notation unambiguous. Still, we use the notation judiciously. If desired, one could eliminate the use of asymptotic notation from this paper entirely, being explicit throughout with constant or logarithmic factors.

We often make use of the standard bound $(\frac{r}{s})^s \leq \binom{r}{s} \leq (\frac{er}{s})^s$ in the weaker form

$$\binom{r}{s}, \binom{r}{\leq s} = \exp\left(\Theta\left(s \log \frac{r}{s}\right)\right),$$

for $s \leq .99r$. We also use the following version of the Chernoff bound. We use $\mathbb{P}(E)$ to denote the probability of the event E and $\mathbb{E}[X]$ to denote the expected value of a random variable X.

Lemma 1.7 (Chernoff Bound)**.** *If X is binomially distributed with mean μ, then for $0 \leq \alpha \leq 1$ and $1 < \beta$, we have the lower tail bound*

$$\mathbb{P}(X \leq \alpha\mu) \leq \exp\left(-(1-\alpha+\alpha\log\alpha)\mu\right) \leq \exp\left(-(1-2\sqrt{\alpha})\mu\right), \tag{2}$$

and the upper tail bound

$$\mathbb{P}(X \geq \beta\mu) \leq \exp\left(-(1-\beta+\beta\log\beta)\mu\right) \leq \exp\left(-\mu\beta\log(\beta/e)\right). \tag{3}$$

2 | PRELIMINARIES

We make the dependence of k on n explicit wherever possible, and we assume that n is sufficiently large for our calculations to go through. In Section 4, we consider only $n = 2k+1$, and in Section 5 we consider only $2k+1 < n \leq Ck$, where C is a sufficiently large constant, since Theorem 1.2 was already proven for $n > Ck$ in [8, 9]. Despite this, we try to make statements not assuming any relationship between k and n so that our intermediate results are as useful as possible.

Let $\mathcal{T} = \{\mathcal{F} \subseteq V(K(n,k)) : |\mathcal{F}| = \binom{n-1}{k-1}, \mathcal{F} \text{ is not a star}\}$. By the union bound, the probability that $K_p(n,k)$ is not EKR is at most

$$\sum_{\mathcal{F}\in\mathcal{T}} (1-p)^{e(\mathcal{F})}. \tag{4}$$

We are done if we can show that (4) is $o(1)$ for the desired p. Unfortunately, $\mathcal{T}$ appears to be too large for this strategy to give a sharp threshold. Our strategy is to use the lower bounds on $e(\mathcal{F})$ from [9, 10], refine $\mathcal{T}$, and give improved bounds on the size of the refinement to make this strategy successful.

First, we introduce a framework which appeared in [10], although similar ideas appear in prior papers. Consider $\mathcal{F} \subseteq V(K(n,k))$ of size $\binom{n-1}{k-1}$. Let $x_{\mathcal{F}}$ be the smallest $x \in [n]$ minimizing $|\mathcal{F}\backslash\mathcal{K}_x|$. For ease of notation here, we write x for $x_{\mathcal{F}}$. Let $\mathcal{A}_{\mathcal{F}} = \mathcal{F}\backslash\mathcal{K}_x$, and $a_{\mathcal{F}} = |\mathcal{A}_{\mathcal{F}}|$. This $a_{\mathcal{F}}$ measures the "distance" from $\mathcal{F}$ to the nearest star.[1] Let $\mathcal{B}_{\mathcal{F}} = \mathcal{K}_x\backslash\mathcal{F}$, so that $|\mathcal{B}_{\mathcal{F}}| = a_{\mathcal{F}}$ since $|\mathcal{F}| = |\mathcal{K}_x|$. Note that $\mathcal{A}_{\mathcal{F}} \subseteq \binom{[n]\backslash\{x\}}{k}$ and $\mathcal{B}_{\mathcal{F}} \subseteq \mathcal{K}_x$; by $\overline{\mathcal{A}_{\mathcal{F}}}$ and $\overline{\mathcal{B}_{\mathcal{F}}}$ we mean $\overline{\mathcal{A}_{\mathcal{F}}} = \binom{[n]\backslash\{x\}}{k}\backslash\mathcal{A}_{\mathcal{F}}$ and $\overline{\mathcal{B}_{\mathcal{F}}} = \mathcal{K}_x\backslash\mathcal{B}_{\mathcal{F}} = \mathcal{K}_x \cap \mathcal{F}$. Since $\overline{\mathcal{B}_{\mathcal{F}}}$ is an intersecting set system, $E(\mathcal{F})$, the set of edges spanned by $\mathcal{F}$, is partitioned into $E(\mathcal{A}_{\mathcal{F}})$ and $E(\mathcal{A}_{\mathcal{F}}, \overline{\mathcal{B}_{\mathcal{F}}})$, the set of edges with one endpoint in $\mathcal{A}_{\mathcal{F}}$ and the other in $\overline{\mathcal{B}_{\mathcal{F}}}$. We mostly focus on the bipartite structure between $\mathcal{A}_{\mathcal{F}}$ and $\mathcal{B}_{\mathcal{F}}$,[2] but whenever we discuss neighborhoods, denoted by $N(\mathcal{A})$ for $\mathcal{A} \subseteq V(K(n,k))$, and edges, with $e(\mathcal{A}) = |E(\mathcal{A})|$ and $e(\mathcal{A},\mathcal{A}') = |E(\mathcal{A},\mathcal{A}')|$, the context is always $K(n,k)$ unless explicitly stated otherwise. To summarize,

$$\mathcal{F} = \mathcal{A}_{\mathcal{F}} \cup (\mathcal{K}_{x_{\mathcal{F}}}\backslash\mathcal{B}_{\mathcal{F}}),$$

where $x_{\mathcal{F}}$ is chosen to minimize $|\mathcal{A}_{\mathcal{F}}| = |\mathcal{B}_{\mathcal{F}}|$. Note that if $a_{\mathcal{F}} = 0$, then $\mathcal{F}$ is a star, and if $a_{\mathcal{F}} = 1$, then $\mathcal{F}$ is a near-star. We drop the $\mathcal{F}$ subscript when it is clear from context.

We begin with three helpful observations about the above framework. First, by a simple averaging argument, we observe that for every $\mathcal{F} \in \mathcal{T}$, $a_{\mathcal{F}}$ is not too large.

Observation 2.1. For every $\mathcal{F} \in \mathcal{T}$, $a_{\mathcal{F}} \leq \frac{n-k}{n}\binom{n-1}{k-1}$.

[1] In [23], $a_{\mathcal{F}}$ is called the *diversity* of $\mathcal{F}$. The earliest extensions of the Erdős–Ko–Rado theorem, for example [17], gave the maximum size of independent $\mathcal{F} \subseteq V(K(n,k))$ satisfying restrictions on $a_{\mathcal{F}}$.

[2] It can be helpful to see the hypercube in disguise in these definitions. If one takes the complements of every set of $\mathcal{A}_{\mathcal{F}}$ in $[n]\backslash\{x\}$, and one also removes the element x from every set of $\mathcal{B}_{\mathcal{F}}$, then the disjointness relation between $\mathcal{A}_{\mathcal{F}}$ and $\mathcal{B}_{\mathcal{F}}$ becomes the subset-superset relation with ground set $[n]\backslash\{x\}$.

Proof. Note that $\sum_{x\in[n]} |\mathcal{F}\cap\mathcal{K}_x| = k|\mathcal{F}|$ for any $\mathcal{F}$. Thus for some x we have

$$|\mathcal{F}\backslash\mathcal{K}_x| = \binom{n-1}{k-1} - |\mathcal{F}\cap\mathcal{K}_x| \le \frac{n-k}{n}\binom{n-1}{k-1}.$$

■

Second, unless $a_{\mathcal{F}}$ is very large, $\mathcal{F}$ is not close to $\mathcal{K}_y$ for any $y \neq x_{\mathcal{F}}$.

Observation 2.2. For every $\mathcal{F}\in\mathcal{T}$ and $y\in[n]$ with $y\neq x_{\mathcal{F}}$, we have

$$|\mathcal{F}\backslash\mathcal{K}_y| \ge \binom{n-2}{k-1} - |\mathcal{F}\backslash\mathcal{K}_x|.$$

Proof. Observe that $|\mathcal{F}\backslash\mathcal{K}_y| \ge |(\mathcal{K}_x\cap\mathcal{F})\backslash\mathcal{K}_y| \ge |\mathcal{K}_x\backslash\mathcal{K}_y| - |\mathcal{K}_x\backslash\mathcal{F}| = \binom{n-2}{k-1} - |\mathcal{F}\backslash\mathcal{K}_x|$. ■

Third, we relate $e(\overline{\mathcal{A}}_{\mathcal{F}}, \mathcal{B}_{\mathcal{F}})$ to $e(\mathcal{A}_{\mathcal{F}}, \overline{\mathcal{B}}_{\mathcal{F}})$, the latter of which is more directly related to $e(\mathcal{F})$.

Observation 2.3. For every $\mathcal{F}\in\mathcal{T}$, $e(\overline{\mathcal{A}}_{\mathcal{F}}, \mathcal{B}_{\mathcal{F}}) = \binom{n-k-1}{k} a_{\mathcal{F}} + e(\mathcal{A}_{\mathcal{F}}, \overline{\mathcal{B}}_{\mathcal{F}})$.

Proof. The degree of vertices in $\binom{[n]\backslash\{x\}}{k}$ to $\mathcal{K}_x$ is $\binom{n-k-1}{k-1}$, while the degree of vertices in $\mathcal{K}_x$ to $\binom{[n]\backslash\{x\}}{k}$ is $\binom{n-k}{k}$. Thus

$$\begin{aligned} e(\overline{\mathcal{A}}_{\mathcal{F}}, \mathcal{B}_{\mathcal{F}}) &= \binom{n-k}{k} a_{\mathcal{F}} - e(\mathcal{A}_{\mathcal{F}}, \mathcal{B}_{\mathcal{F}}) = \binom{n-k}{k} a_{\mathcal{F}} - \binom{n-k-1}{k-1} a_{\mathcal{F}} + e(\mathcal{A}_{\mathcal{F}}, \overline{\mathcal{B}}_{\mathcal{F}}) \\ &= \binom{n-k-1}{k} a_{\mathcal{F}} + e(\mathcal{A}_{\mathcal{F}}, \overline{\mathcal{B}}_{\mathcal{F}}). \end{aligned}$$

■

Let

$$\mathcal{T}_x(a) = \{\mathcal{F}\in\mathcal{T} : x_{\mathcal{F}} = x,\ a_{\mathcal{F}} = a\}.$$

We define subsets $\mathcal{T}^i$ of $\mathcal{T}$ for $i\in[5]$ as is convenient. Whenever $\mathcal{T}^i$ is defined, we define $\mathcal{T}^i_x(a)$ to be $\mathcal{T}^i\cap\mathcal{T}_x(a)$. For reference, we list the definitions of all the $\mathcal{T}^i$ here, along with the subsection where they first appear; we introduce them formally when we need them, so some notation is not yet defined:

$$\mathcal{T}^1 = \left\{\mathcal{F}\in\mathcal{T} : e(\mathcal{F}) \le \frac{5}{p_0} a_{\mathcal{F}} \log \frac{\binom{n-1}{k}}{a_{\mathcal{F}}}\right\}, \quad \text{(Section 2.1)}$$

$$\mathcal{T}^2 = \left\{\mathcal{F}\in\mathcal{T}^1 : \mathcal{B}_{\mathcal{F}} \subseteq N(\mathcal{A}_{\mathcal{F}})\right\}, \quad \text{(Section 2.2)}$$

$$\mathcal{T}^3 = \left\{\mathcal{F}\in\mathcal{T}^2 : \mathcal{F} \text{ is 2-linked}\right\}, \quad \text{(Section 4.2)}$$

$$\mathcal{T}^4 = \left\{\mathcal{F}\in\mathcal{T}^1 : \left|A^{<1/\sqrt{k}}_{\mathcal{F}}\right| \le \frac{a_{\mathcal{F}}}{\log\log k}\right\}, \quad \text{(Section 5.2)}$$

$$\mathcal{T}^5 = \left\{\mathcal{F}\in\mathcal{T}^4 : e(\mathcal{A}_{\mathcal{F}}) \le \frac{1}{2} e(\mathcal{F})\right\}. \quad \text{(Section 5.4)}$$

As an example of how these definitions are used, in Lemma 2.8, we easily show that no $\mathcal{F} \in \mathcal{T}\backslash\mathcal{T}^1$ is independent in $K_p(n,k)$ for $p \geq p_0$ w.h.p., because by definition, the $\mathcal{F} \in \mathcal{T}\backslash\mathcal{T}^1$ have many edges. In other cases, like Proposition 2.12, we show that if some $\mathcal{F} \in \mathcal{T}^1$ is independent in $K_p(n,k)$, then some $\mathcal{F} \in \mathcal{T}^2$ is independent in $K_p(n,k)$, which reduces the problem of showing that no $\mathcal{F} \in \mathcal{T}^1$ is independent to showing that no $\mathcal{F} \in \mathcal{T}^2$ is independent.

2.1 | Bounding $e(\mathcal{F})$

Since $\mathcal{F}$ is determined by $x = x_{\mathcal{F}}$, $\mathcal{A}_{\mathcal{F}} \subseteq \binom{[n]\backslash\{x\}}{k}$, and $\mathcal{B}_{\mathcal{F}} \subseteq \mathcal{K}_x$, we trivially have

$$|\mathcal{T}_x(a)| \leq \binom{\binom{n-1}{k}}{a}\binom{\binom{n-1}{k-1}}{a} \leq \binom{\binom{n-1}{k}}{a}^2. \tag{5}$$

To obtain a threshold for Question 1.1, Devlin and Kahn [10] combined (5) with the following very nontrivial lower bound on $e(\mathcal{F})$:

Theorem 2.4 ([10]). *There exists $c > 0$ such that for all $n \leq 2k + k/6$ and $\mathcal{F} \in \mathcal{T}$,*

$$e(\mathcal{F}) > c\frac{1}{k}\binom{n-k-1}{k-1} a_{\mathcal{F}} \log \frac{\binom{n-1}{k}}{a_{\mathcal{F}}}. \tag{6}$$

As we are interested in all n and k, we need a lower bound on $e(\mathcal{F})$ for not just $n \leq 2k + k/6$. We can remove the assumption on n in Theorem 2.4 with a similar bound due to Das and Tran [9]. It is implicit in their paper, so for completeness, we derive it from their removal lemma for $\mathcal{T}$.

Theorem 2.5 ([9]). *There is an absolute constant $D > 1$ such that if $n \geq 2k + 1$ and $\mathcal{F} \in \mathcal{T}$ with $e(\mathcal{F}) < \beta\binom{n-1}{k-1}\binom{n-k-1}{k-1}$, where $\beta \leq \frac{n-2k}{(20D)^2 n}$, then there is an $x \in [n]$ such that*

$$|\mathcal{F}\backslash\mathcal{K}_x| \leq D\beta\frac{n}{n-2k}\binom{n-1}{k-1}.$$

Theorem 2.6 ([9]). *There exists $c > 0$ such that for all $n \geq 2k + 1$ and $\mathcal{F} \in \mathcal{T}$,*

$$e(\mathcal{F}) \geq c\frac{n-2k}{n}\binom{n-k-1}{k-1} a_{\mathcal{F}}. \tag{7}$$

Proof. Let $\mathcal{F} \in \mathcal{T}_x(a)$. Let D be the constant given in Theorem 2.5, and let

$$\beta = \frac{a}{(20D)^2 \frac{n}{n-2k}\binom{n-1}{k-1}} \leq \frac{n-2k}{(20D)^2 n},$$

which follows from $a = |\mathcal{F}\backslash\mathcal{K}_x| \leq |\mathcal{F}| = \binom{n-1}{k-1}$. Since $x_{\mathcal{F}} = x$, we have that for all stars $\mathcal{K}_y$,

$$|\mathcal{F}\backslash\mathcal{K}_y| \geq |\mathcal{F}\backslash\mathcal{K}_x| = a > \frac{a}{400D} = D\beta\frac{n}{n-2k}\binom{n-1}{k-1}.$$

Applying the contrapositive of Theorem 2.5 to $\mathcal{F}$ with the above β, we have

$$e(\mathcal{F}) \geq \beta \binom{n-1}{k-1}\binom{n-k-1}{k-1} = \frac{1}{(20D)^2} \cdot \frac{n-2k}{n}\binom{n-k-1}{k-1} a,$$

which proves the theorem with $c = 1/(20D)^2$. ∎

For convenience, we combine Theorems 2.4 and 2.6 so that they work for all n and k.

Theorem 2.7. *There exists an absolute constant $\theta > 0$ such that for all $n \geq 2k+1$ and $\mathcal{F} \in \mathcal{T}$,*

$$e(\mathcal{F}) \geq \theta \frac{1}{p_0} a_{\mathcal{F}} \log \frac{\binom{n-1}{k}}{a_{\mathcal{F}}}, \tag{8}$$

and

$$e(\mathcal{F}) \geq \theta \frac{1}{p_0} a_{\mathcal{F}} \frac{n-2k}{n} \log \binom{n-1}{k}. \tag{9}$$

Proof. Note that, from (1), we have

$$p_0 \geq \frac{k \log \frac{n-1}{k}}{\binom{n-k-1}{k-1}} \geq \frac{k \log 2}{\binom{n-k-1}{k-1}},$$

so we can replace (6) with $e(\mathcal{F}) \geq c\frac{1}{p_0} a_{\mathcal{F}} \log \frac{\binom{n-1}{k}}{a_{\mathcal{F}}}$, albeit with a different constant c. For $2k + k/6 < n$, we have

$$\frac{1}{p_0} a \log \frac{\binom{n-1}{k}}{a} \leq \binom{n-k-1}{k-1} a \frac{\log \binom{n-1}{k}}{\log\left(n\binom{n-1}{k}\right)} \leq \binom{n-k-1}{k-1} a \cdot 13 \frac{n-2k}{n},$$

so by Theorem 2.6, $e(\mathcal{F}) \geq c\frac{1}{p_0} a_{\mathcal{F}} \log \frac{\binom{n-1}{k}}{a_{\mathcal{F}}}$ holds for all n and k, again with a different constant c, yielding (8). Equation (9) follows directly from Theorem 2.6 and (1). ∎

From here on, we fix θ sufficiently small given by Theorem 2.7. Using this, we can repeat the argument of Devlin and Kahn [10], which is just upper bounding (4), to show that we do not need to consider those $\mathcal{F}$ with $e(\mathcal{F})$ large. Let

$$\mathcal{T}^1 = \left\{ \mathcal{F} \in \mathcal{T} : e(\mathcal{F}) \leq \frac{5}{p_0} a_{\mathcal{F}} \log \frac{\binom{n-1}{k}}{a_{\mathcal{F}}} \right\}.$$

Lemma 2.8. *Assume $k = \omega(1)$ and let $p \geq \frac{49}{50} p_0$. Then no $\mathcal{F} \in \mathcal{T} \setminus \mathcal{T}^1$ is independent in $K_p(n,k)$ w.h.p.*

Proof. Let $x \in [n]$. By Observation 2.1, for every $\mathcal{F} \in \mathcal{T}$, we have $\binom{n-1}{k}/a_{\mathcal{F}} \geq \frac{n}{k} > 2$. Hence by (5), we have

$$|\mathcal{T}_x(a)| \le \left(\binom{n-1}{k}\right)^2 \le \exp\left(2a\log\frac{e\binom{n-1}{k}}{a}\right) \le \exp\left(\left(2+\frac{2}{\log 2}\right)a\log\frac{\binom{n-1}{k}}{a}\right).$$

Using $(1-p)^{e(\mathcal{F})} \le \exp(-pe(\mathcal{F}))$ and the union bound, the probability that some $\mathcal{F} \in \mathcal{T}_x(a)\backslash\mathcal{T}_x^1(a)$ is independent is at most

$$\exp\left(\left(2+\frac{2}{\log 2}\right)a\log\frac{\binom{n-1}{k}}{a} - \frac{49}{10}a\log\frac{\binom{n-1}{k}}{a}\right) \le \exp\left(-\frac{1}{70}a\log\frac{\binom{n-1}{k}}{a}\right).$$

Taking the union bound over all a, the probability that $\mathcal{F} \in \mathcal{T}\backslash\mathcal{T}^1$ with $x_{\mathcal{F}} = x$ is independent is at most

$$\begin{aligned}
&\sum_{a=1}^{n}\exp\left(-\frac{1}{70}a\log\frac{\binom{n-1}{k}}{a}\right) + \sum_{a=n+1}^{\binom{n-1}{k}}\exp\left(-\frac{1}{70}a\log\frac{\binom{n-1}{k}}{a}\right)\\
&\quad\le n\exp\left(-\frac{1}{70}\log\binom{n-1}{k}\right) + \binom{n-1}{k}\exp\left(-\frac{1}{70}n\log\frac{\binom{n-1}{k}}{n}\right)\\
&\quad\le \exp(-\omega(\log n)) + \exp(-\omega(n)) = o(1/n),
\end{aligned}$$

since $\log n = o(\log\binom{n-1}{k})$ for $k = \omega(1)$. A union bound over all x completes the proof. ∎

In the previous proof, we took a union bound over certain $\mathcal{F}$ with $a_{\mathcal{F}} = a$ and $x_{\mathcal{F}} = x$, and then we took a union bound over all a and x. Our proofs often take this form, and since the latter union bound can often be executed as above, we omit these details going forward.

Note that if $\mathcal{F} \in \mathcal{T}^1$, we get the following upper bound on $a_{\mathcal{F}}$, which will be particularly useful when $n - 2k$ is large.

Lemma 2.9. *For every $\mathcal{F} \in \mathcal{T}^1$, we have*

$$\frac{\theta}{5}\cdot\frac{n-2k}{n}\log\binom{n-1}{k} \le \log\frac{\binom{n-1}{k}}{a_{\mathcal{F}}}.$$

Proof. For $\mathcal{F} \in \mathcal{T}^1$, by (9), we have

$$\theta\frac{1}{p_0}a_{\mathcal{F}}\frac{n-2k}{n}\log\binom{n-1}{k} \le e(\mathcal{F}) \le \frac{5}{p_0}a_{\mathcal{F}}\log\frac{\binom{n-1}{k}}{a_{\mathcal{F}}}.$$

∎

2.2 | Neighborhood assumption

Let $\mathcal{T}^2 = \left\{\mathcal{F} \in \mathcal{T}^1 : \mathcal{B}_{\mathcal{F}} \subseteq N(\mathcal{A}_{\mathcal{F}})\right\}$. Our first reduction is that for most $\mathcal{F} \in \mathcal{T}^1$, there exists $\mathcal{F}' \in \mathcal{T}^2$ such that $E(\mathcal{F}') \subseteq E(\mathcal{F})$. This is helpful because if we know that $\mathcal{F}'$ is not independent, then we can automatically conclude that $\mathcal{F}$ is not independent. To make this reduction, we need a lemma concerning

the vertex boundary of $\mathcal{F} \in \mathcal{T}$ in $K(n,k)$, which will also be useful for the next reduction. We derive this from the Kruskal–Katona theorem—in fact, we only need a weaker version due to Lovász. For real z, we define $\binom{z}{k} = \frac{z(z-1)\cdots(z-k+1)}{k!}$.

Theorem 2.10 (Kruskal [22], Katona [19], Lovász (problem 13.31(b) of [24])). *Let $\mathcal{F}$ be a nonempty family of k-sets, and let $z \geq k$ be a real number such that $|\mathcal{F}| = \binom{z}{k}$. Then for every $\ell \leq k$,*

$$|\{S : |S| = \ell, \exists F \in \mathcal{F} \text{such that } S \subseteq F\}| \geq \binom{z}{\ell}.$$

Lemma 2.11. *Let $x \in [n]$. For every nonempty $\mathcal{A} \subseteq \binom{[n]\setminus\{x\}}{k}$ with $|\mathcal{A}| < \binom{n-2}{k-1}$, we have $|N(\mathcal{A}) \cap \mathcal{K}_x| > |\mathcal{A}|$. Moreover, for every $\mathcal{F} \in \mathcal{T}_x(a)$, we have $|N(\mathcal{A}_\mathcal{F}) \cap \mathcal{K}_x| > a$.*

Proof. Consider $\mathcal{A}^\star = \{([n]\setminus\{x\})\setminus A : A \in \mathcal{A}\} \subseteq \binom{[n]\setminus\{x\}}{n-k-1}$. Then $|\mathcal{A}^\star| = |\mathcal{A}|$ and

$$\left|\left\{S \in \binom{[n]\setminus\{x\}}{k-1} : \exists A \in \mathcal{A}^\star \text{ such that } S \subseteq A\right\}\right| = |N(\mathcal{A}) \cap \mathcal{K}_x|. \tag{10}$$

Since $|\mathcal{A}| < \binom{n-2}{n-k-1}$, there exists a real number $z < n-2$ such that $|\mathcal{A}| = \binom{z}{n-k-1}$ and $n-k-1 \leq z$. Applying Theorem 2.10 to $\mathcal{A}^\star$ and using (10), we have

$$|N(\mathcal{A}) \cap \mathcal{K}_x| \geq \binom{z}{k-1} > \binom{z}{n-k-1} = |\mathcal{A}|, \tag{11}$$

where the second inequality follows because $\binom{n-2}{k-1} = \binom{n-2}{n-k-1}$, $n-k-1 > k-1$, and $z < n-2$.

To prove the second part of the lemma, note that by Observation 2.1, for any $\mathcal{F} \in \mathcal{T}$,

$$|\mathcal{A}_\mathcal{F}| \leq \frac{n-k}{n}\binom{n-1}{k-1} = \frac{n-1}{n}\binom{n-2}{k-1},$$

and we are done by applying the first part of the lemma. ∎

Now we can give the reduction from $\mathcal{T}^1$ to $\mathcal{T}^2$, which we employ in Sections 4.5 and 5.6.

Proposition 2.12. *For every $\mathcal{F} \in \mathcal{T}_x^1(a)$ with $a < \frac{1}{3}\binom{n-2}{k-1}$, there exists $\mathcal{F}' \in \mathcal{T}_x^2(a)$ satisfying $\mathcal{A}_{\mathcal{F}'} = \mathcal{A}_\mathcal{F}$ and $E(\mathcal{F}') \subseteq E(\mathcal{F})$.*

Proof. Let $\mathcal{F} \in \mathcal{T}_x(a)$. Since $a \leq |N(\mathcal{A}_\mathcal{F}) \cap \mathcal{K}_x|$ by Lemma 2.11, there exists $\mathcal{B}'$ of size a such that $N(\mathcal{A}_\mathcal{F}) \cap \mathcal{B}_\mathcal{F} \subseteq \mathcal{B}' \subseteq N(\mathcal{A}_\mathcal{F}) \cap \mathcal{K}_x$. Let $\mathcal{F}' = \mathcal{A}_\mathcal{F} \cup (\mathcal{K}_x\setminus\mathcal{B}')$, so

$$E(\mathcal{F}') = E(\mathcal{A}_\mathcal{F}) \cup E(\mathcal{A}_\mathcal{F}, \overline{\mathcal{B}'}) \subseteq E(\mathcal{A}_\mathcal{F}) \cup E(\mathcal{A}_\mathcal{F}, \overline{N(\mathcal{A}_\mathcal{F}) \cap \mathcal{B}_\mathcal{F}}) = E(\mathcal{F}).$$

By Observation 2.2, for any $y \in [n]$ with $y \neq x$,

$$|\mathcal{F}'\setminus\mathcal{K}_y| \geq |\mathcal{F}\setminus\mathcal{K}_y| - |\mathcal{B}'| \geq \binom{n-2}{k-1} - 2a > a = |\mathcal{F}'\setminus\mathcal{K}_x|.$$

Thus $x_{\mathcal{F}'} = x$ and so $\mathcal{F}' \in \mathcal{T}_x^2(a)$ and $\mathcal{A}_{\mathcal{F}'} = \mathcal{A}_\mathcal{F}$. ∎

3 | MAIN LEMMA

One might hope that, with the extra condition $\mathcal{T}^1$ imposes over $\mathcal{T}$, we have much stronger upper bounds on $|\mathcal{T}_x^1(a)|$ than (5). This is the content of the following technical lemma, which is one of our main innovations.

Note that $\binom{n-k-1}{k-1}$ is the degree of vertices in $\binom{[n]\setminus\{x\}}{k}$ to $\mathcal{K}_x$, and note that $\binom{n-k}{k} = \frac{n-k}{k}\binom{n-k-1}{k-1}$ is the degree of vertices in $\mathcal{K}_x$ to $\binom{[n]\setminus\{x\}}{k}$. For each $\mathcal{F}$ and $\delta = \delta(n,k)$, define $\mathcal{A}_{\mathcal{F}}^{\geq\delta} = \{A \in \mathcal{A}_{\mathcal{F}} : d(A,\mathcal{B}_{\mathcal{F}}) \geq \delta\binom{n-k-1}{k-1}\}$, where $d(A,\mathcal{B}_{\mathcal{F}}) = |N(A)\cap\mathcal{B}_{\mathcal{F}}|$. Let $\mathcal{A}_{\mathcal{F}}^{<\delta} = \mathcal{A}_{\mathcal{F}}\setminus\mathcal{A}_{\mathcal{F}}^{\geq\delta}$. We use the structure between $\mathcal{A}_{\mathcal{F}}$ and $\mathcal{B}_{\mathcal{F}}$ for $\mathcal{F} \in \mathcal{T}^1$ to give upper bounds for the number of $\mathcal{A}_{\mathcal{F}}^{\geq\delta}$ across all $\mathcal{F} \in \mathcal{T}_x^1(a)$.

Lemma 3.1. *If* $\delta = \omega\left(\frac{n\log\frac{n}{n-2k}}{k\log\binom{n-1}{k}}\right)$ *and* $\delta \leq 1/2$, *then the number of choices for* $\mathcal{A}_{\mathcal{F}}^{\geq\delta}$ *across all* $\mathcal{F} \in \mathcal{T}_x^1(a)$ *is at most* $\exp\left(o\left(a\log\frac{\binom{n-1}{k}}{a}\right)\right)$.

Proof. Let $D = \binom{n-k-1}{k-1}$, and choose p_1 so that $p_1 = \omega(1/\delta D)$ and $\frac{p_1}{p_0}\frac{n}{k}\log\frac{n}{n-2k} = o(1)$. This is possible because the assumption on δ implies that

$$\frac{1}{p_0}\frac{n}{k}\log\frac{n}{n-2k} = o(\delta D).$$

To each $\mathcal{A}_{\mathcal{F}}^{\geq\delta}$ we give a certificate $(\mathcal{Y},\mathcal{A}_1,\mathcal{A}_2,\mathcal{A}_3)$ with the following properties:

1. $\mathcal{Y} \subseteq \mathcal{K}_x$, $|\mathcal{Y}| \leq 3ap_1$,
2. $\mathcal{A}_1 \subseteq N(\mathcal{Y})$, $|\mathcal{A}_1| \leq \frac{p_1}{p_0}\cdot\frac{30n}{\theta k}\cdot a\log\frac{\binom{n-1}{k}}{a}$, $|N(\mathcal{Y})\setminus\mathcal{A}_1| \leq a$,
3. $\mathcal{A}_2 \subseteq N(\mathcal{Y})\setminus\mathcal{A}_1$, $|\mathcal{A}_2| \leq \frac{10}{\log\binom{n-1}{k}}a\log\frac{\binom{n-1}{k}}{a}$,
4. $\mathcal{A}_3 \subseteq \binom{[n]\setminus\{x\}}{k}$, $|\mathcal{A}_3| = o(a)$,
5. $\mathcal{A}_{\mathcal{F}}^{\geq\delta} = (N(\mathcal{Y})\setminus(\mathcal{A}_1\cup\mathcal{A}_2))\cup\mathcal{A}_3$.

By the last property, we can reconstruct $\mathcal{A}_{\mathcal{F}}^{\geq\delta}$ from its certificate. We postpone counting the number of choices of possible certificates until after we prove the existence of such certificates.

Fix $\mathcal{F} \in \mathcal{T}_x^1(a)$, and let $\mathcal{A}^{\geq\delta} = \mathcal{A}_{\mathcal{F}}^{\geq\delta}$ and $\mathcal{A}^{<\delta} = \mathcal{A}_{\mathcal{F}}^{<\delta}$. We prove that there is a particular choice of $\mathcal{Y}$ for which the elements of the certificate are $\mathcal{Y}$, $\mathcal{A}_1 := N(\mathcal{Y})\setminus\mathcal{A}$, $\mathcal{A}_2 := (N(\mathcal{Y})\setminus\mathcal{A}_1)\cap\mathcal{A}^{<\delta}$, and $\mathcal{A}_3 := \mathcal{A}^{\geq\delta}\setminus N(\mathcal{Y})$. Select a random subset $\mathcal{Y}$ of $\mathcal{B}$, including each vertex independently with probability p_1. Then the expected size of $\mathcal{Y}$ is

$$\mathbb{E}|\mathcal{Y}| = ap_1.$$

The probability that $A \in \mathcal{A}^{\geq\delta}$ is not in the neighborhood of $\mathcal{Y}$ is at most $(1-p_1)^{\delta D}$, so

$$\mathbb{E}|\mathcal{A}^{\geq\delta}\setminus N(\mathcal{Y})| \leq a(1-p_1)^{\delta D} \leq ae^{-p_1\delta D} = ae^{-\omega(1)} = o(a).$$

Note that by Observation 2.3, Theorem 2.7, and the definition of $\mathcal{T}^1$,

$$\mathbb{E}|N(\mathcal{Y})\setminus\mathcal{A}| \leq \mathbb{E}\sum_{v\in\mathcal{Y}} d(v,\overline{\mathcal{A}}) = \mathbb{E}\sum_{v\in\mathcal{B}} 1_{v\in\mathcal{Y}}d(v,\overline{\mathcal{A}}) = p_1 e(\overline{\mathcal{A}},\mathcal{B})$$

$$= p_1\left(\binom{n-k-1}{k}a + e(\mathcal{A},\overline{\mathcal{B}})\right) = p_1\left(\frac{\log\left(n\binom{n-1}{k}\right)}{k}\cdot\frac{1}{p_0}a(n-2k) + e(\mathcal{A},\overline{\mathcal{B}})\right)$$

$$\le p_1\left(\frac{n}{\theta k}\cdot\frac{\log\left(n\binom{n-1}{k}\right)}{\log\binom{n-1}{k}} + 1\right)e(\mathcal{F}) \le \frac{p_1}{p_0}\cdot\frac{10n}{\theta k}\cdot a\log\frac{\binom{n-1}{k}}{a}.$$

By Markov's inequality, there exists a choice of $\mathcal{Y}$ such that

$$|\mathcal{Y}| \le 3ap_1,$$
$$|\mathcal{A}^{\ge\delta}\backslash N(\mathcal{Y})| \le o(a),$$
$$|N(\mathcal{Y})\backslash\mathcal{A}| \le \frac{p_1}{p_0}\cdot\frac{30n}{\theta k}\cdot a\log\frac{\binom{n-1}{k}}{a}.$$

Since $N(\mathcal{Y})\backslash\mathcal{A}_1 \subseteq A$ by definition, all that remains to be shown is the upper bound on $|\mathcal{A}_2|$. Observe that

$$(1-\delta)D|\mathcal{A}^{<\delta}| \le e(\mathcal{A},\overline{\mathcal{B}}) \le e(\mathcal{F}) \le \frac{5}{p_0}a\log\frac{\binom{n-1}{k}}{a},$$

so since $\delta \le 1/2$,

$$|\mathcal{A}_2| \le |\mathcal{A}^{<\delta}| \le \frac{10}{\log\binom{n-1}{k}}a\log\frac{\binom{n-1}{k}}{a}.$$

We count the number of possible certificates by counting each part of the certificate sequentially. Since $p_1 = o(1)$, $|\mathcal{Y}| = o(a)$, so the number of choices for $\mathcal{Y}$ is at most

$$\binom{\binom{n-1}{k}}{o(a)} \le \exp\left(o\left(a\log\frac{\binom{n-1}{k}}{a}\right)\right).$$

Note that

$$|N(\mathcal{Y})| \le \binom{n-k}{k}|\mathcal{Y}| \le 3ap_1\binom{n-k}{k},$$

so given $\mathcal{Y}$, the number of choices for $\mathcal{A}_1$ is at most, using Lemma 2.9,

$$\binom{3ap_1\binom{n-k}{k}}{\le \frac{p_1}{p_0}\cdot\frac{30n}{\theta k}\cdot a\log\frac{\binom{n-1}{k}}{a}} \le \exp\left(O\left(\frac{p_1}{p_0}\cdot\frac{n}{k}\cdot a\log\frac{\binom{n-1}{k}}{a}\log\frac{p_0\binom{n-k}{k}}{\frac{30n}{\theta k}\log\frac{\binom{n-1}{k}}{a}}\right)\right)$$
$$\le \exp\left(O\left(\frac{p_1}{p_0}\cdot\frac{n}{k}\cdot a\log\frac{\binom{n-1}{k}}{a}\log\frac{6(n-k)}{n-2k}\right)\right) = \exp\left(O\left(\frac{p_1}{p_0}\cdot\frac{n}{k}\log\frac{n}{n-2k}\right)a\log\frac{\binom{n-1}{k}}{a}\right)$$
$$= \exp\left(o\left(a\log\frac{\binom{n-1}{k}}{a}\right)\right).$$

Given $\mathcal{Y}$ and $\mathcal{A}_1$, the number of possible choices for $\mathcal{A}_2$ is at most, using Observation 2.1,

$$\left(\begin{array}{c} a \\ \leq \frac{10}{\log\binom{n-1}{k}} a \log \frac{\binom{n-1}{k}}{a} \end{array}\right) \leq \exp\left(o\left(a \log \frac{\binom{n-1}{k}}{a}\right)\right).$$

Finally, since $|\mathcal{A}_3| = o(a)$, the number of choices for $\mathcal{A}_3$ is at most

$$\left(\begin{array}{c} \binom{n-1}{k} \\ o(a) \end{array}\right) \leq \exp\left(o\left(a \log \frac{\binom{n-1}{k}}{a}\right)\right).$$

■

Let $\mathcal{B}_{\mathcal{F}}^{\geq\delta} = \left\{B \in \mathcal{B}_{\mathcal{F}} : d(B, \mathcal{A}_{\mathcal{F}}) \geq \delta\binom{n-k}{k}\right\}$ and $\mathcal{B}_{\mathcal{F}}^{<\delta} = \mathcal{B}_{\mathcal{F}} \backslash \mathcal{B}_{\mathcal{F}}^{\geq\delta}$. There is a nearly identical lemma for $\mathcal{B}_{\mathcal{F}}^{\geq\delta}$ as well.

Lemma 3.2. *If* $\delta = \omega\left(\frac{k}{n}\frac{\log\log\binom{n-1}{k}}{\log\binom{n-1}{k}}\right)$ *and* $\delta \leq 1/2$, *then the number of* $\mathcal{B}_{\mathcal{F}}^{\geq\delta}$ *across all* $\mathcal{F} \in \mathcal{T}_x^1(a)$ *is at most* $\exp\left(o\left(a \log \frac{\binom{n-1}{k}}{a}\right)\right)$.

The proof is essentially the same as the proof of the lemma for $\mathcal{A}_{\mathcal{F}}^{\geq\delta}$, except that we use $D = \binom{n-k}{k}$, $\frac{p_1}{p_0}\log\log\binom{n-1}{k} = o(1)$, and for $\mathcal{X} \subseteq \mathcal{A}$ randomly chosen with probability p_1, we have

$$\mathbb{E}|N(X)\backslash\mathcal{B}| \leq p_1 e(\mathcal{A}, \overline{\mathcal{B}}) \leq 5\frac{p_1}{p_0} a \log \frac{\binom{n-1}{k}}{a}.$$

3.1 | Large diversity when $n - 2k$ is small

The following proposition demonstrates our most basic way to show that the $\mathcal{F} \in \mathcal{T}^1$ are not independent in $K_p(n,k)$ w.h.p.: we use Lemmas 3.1 and 3.2 to bound the number of choices for $\mathcal{A}_{\mathcal{F}}^{\geq\delta}$ and $\mathcal{B}_{\mathcal{F}}^{\geq\delta}$, while we bound the number of $\mathcal{A}_{\mathcal{F}}^{<\delta}$ and $\mathcal{B}_{\mathcal{F}}^{<\delta}$ by controlling the size of those sets with $e(\mathcal{F})$. This trick works when $n - 2k$ is small and $a_{\mathcal{F}}$ is large.

Proposition 3.3. *For any constant* $\varepsilon > 0$, *when* $n - 2k = o(n)$, *no* $\mathcal{F} \in \mathcal{T}^1$ *with* $\log\frac{\binom{n-1}{k}}{a_{\mathcal{F}}} = o(n)$ *is independent in* $K_p(n,k)$ *w.h.p., where* $p = (1-\varepsilon)p_0$.

Proof. By Lemmas 3.1 and 3.2, since

$$\frac{n\log\frac{n}{n-2k}}{k\log\binom{n-1}{k}} = o(1) \quad \text{and} \quad \frac{k}{n}\frac{\log\log\binom{n-1}{k}}{\log\binom{n-1}{k}} = o(1),$$

the number of distinct $\mathcal{A}_{\mathcal{F}}^{\geq 1/2}$ and $\mathcal{B}_{\mathcal{F}}^{\geq 1/2}$ across all $\mathcal{F} \in \mathcal{T}_x^1(a)$ is at most

$$\exp\left(o\left(a\log\frac{\binom{n-1}{k}}{a}\right)\right).$$

Note that since $\log \frac{\binom{n-1}{k}}{a} = o(n)$ and $\mathcal{F} \in \mathcal{T}^1$,

$$\frac{1}{2}\binom{n-k-1}{k-1}|\mathcal{A}^{<1/2}| \le e(\mathcal{A}, \overline{\mathcal{B}}) \le e(\mathcal{F}) \le \frac{5}{p_0} a \log \frac{\binom{n-1}{k}}{a} = o\left(\frac{1}{p_0} an\right),$$

and since $\log \binom{n-1}{k} = \Theta(n)$, we have $|\mathcal{A}^{<1/2}| = o(a)$. Thus the number of such $\mathcal{A}^{<1/2}$ is at most

$$\binom{\binom{n-1}{k}}{o(a)} \le \exp\left(o\left(a \log \frac{\binom{n-1}{k}}{a}\right)\right).$$

We perform a similar calculation for $\mathcal{B}^{<1/2}$. Using Observation 2.3, (8), and $n - 2k = o(n)$,

$$\frac{1}{2}\binom{n-k}{k}|\mathcal{B}^{<1/2}| \le e(\overline{\mathcal{A}}, \mathcal{B}) = \binom{n-k-1}{k} a + e(\mathcal{A}, \overline{\mathcal{B}}) = \binom{n-k-1}{k} a + o\left(\frac{an}{p_0}\right),$$

so $|\mathcal{B}^{<1/2}| = o(a)$, and we can proceed as before.

Taking the union bound and using (8), we have that the probability some $\mathcal{F} \in \mathcal{T}_x^1(a)$ is independent is at most

$$\exp\left(o\left(a \log \frac{\binom{n-1}{k}}{a}\right)\right) \cdot (1-p)^{\theta \frac{1}{p_0} a \log \frac{\binom{n-1}{k}}{a}} \le \exp\left(-\frac{\theta}{2} a \log \frac{\binom{n-1}{k}}{a}\right).$$

We take a union bound over all possible choices for a and x to finish the proof. ∎

4 | HITTING TIME FOR $n = 2k+1$

For the entirety of this section, let $n = 2k + 1$, fix a constant $\varepsilon > 0$ sufficiently small, and let p and p' be such that

$$p = 1 - 4^{-(1-\varepsilon)} < \frac{3}{4} < 1 - 4^{-(1+\varepsilon)} = p'.$$

We prove Theorems 1.3 and 1.4 in Section 4.5 as quick corollaries of a characterization of independent sets in $K_{\tau_{\text{super}}}(n,k)$ that holds with high probability, which we prove in Section 4.4. To get this characterization, we first show in Section 4.1 that, in a rigorous sense, $K_p(n,k)$ is a subgraph of $K_{\tau_{\text{super}}}(n,k)$ w.h.p. In Section 4.2, we describe a method of efficiently counting the number of possible $\mathcal{A}_{\mathcal{F}}$ using components in an auxiliary graph. We apply this method in Section 4.3 to show that no "connected" $\mathcal{F}$ is independent in $K_p(n,k)$ w.h.p., and we remove the connectedness condition in Section 4.4.

4.1 | Approximation by the binomial model

Let G be a graph. Recall the definition of a random subgraph process of G: take a permutation of $E(G)$ uniformly at random and consider the initial segments of this permutation as a sequence of random spanning subgraphs of G. It will be convenient for us to generalize this model to the *continuous time random subgraph process*, see section 1.1 of [18], which also generalizes the $K_p(n,k)$ model, that is,

the model where we include each edge independently with a fixed probability. Assign to each edge e of G a uniformly randomly chosen real number $f(e)$ between 0 and 1. The underlying probability space is now the set of functions $f : E(G) \to [0,1]$ with the uniform distribution. Let $G_p(f)$ be the spanning subgraph of G with $E(G_p(f)) = \{e \in E(G) : f(e) \le p\}$. For deterministic p, that is, p which do not depend on f, G_p is simply the random subgraph of G chosen by including each edge with probability p. However, we can let p be a random variable, depending on f and thus the structure of the random subgraph. This relates this model to the random subgraph process: with probability 1, f is injective, so as we slowly increase p from 0 to 1, we randomly add edges one by one. We can analogously define the hitting time $q_P(f)$ for a monotone property P in the G_p model to be the smallest $p \in [0,1]$ such that $G_p(f)$ satisfies property P. If τ_P is the hitting time for P in the random subgraph process, then $G_{q_P} = G_{\tau_P}$ as distributions. Define

$$q_1(f) = q_1(n,k,f) = \min\{p \in [0,1] : K_p(n,k,f) \text{ has no independent superstars}\},$$

$$q_2(f) = q_2(n,k,f) = \min\{p \in [0,1] : K_p(n,k,f) \text{ has no independent near-stars}\}.$$

To prove Theorem 1.3, it suffices to show that $\alpha(K_{q_1(f)}(n,k,f)) = \binom{n-1}{k-1}$ w.h.p., where the "with high probability" is taken with respect to the uniform distribution of $f : E(K(n,k)) \to [0,1]$. Likewise, to prove Theorem 1.4, it suffices to show that $K_{q_2(f)}(n,k,f)$ is EKR w.h.p. To this end, we first show that p_0 is a sharp threshold for both containing independent superstars and containing independent near-stars, thus giving the bounds $(1-\varepsilon)p_0 \le q_1(f) < q_2(f) \le (1+\varepsilon)p_0$ w.h.p., for every constant[3] $\varepsilon > 0$. Proofs of the lower bound appear in [8] and [9]: a second moment calculation works for all $n > 2k+1$ in [8], while the simple calculation from [9] works for a smaller range including $n = 2k+1$ just as effectively. We repeat the argument of [9] here for $n = 2k+1$ for completeness.

Observation 4.1. There exists an independent superstar in $K_p(2k+1,k)$ w.h.p., where $p = 1-4^{-(1-\varepsilon)}$. That is, $p < q_1(f)$ w.h.p.

Proof. Since the superstars based at 1 are edge-disjoint, the probability that $K_p(n,k)$ contains no independent superstar based at 1 is

$$\left(1-(1-p)^{\binom{n-k-1}{k-1}}\right)^{\binom{n-1}{k}} = \left(1-4^{-(1-\varepsilon)k}\right)^{\binom{2k}{k}}$$
$$\le \exp\left(-4^{-(1-\varepsilon)k}\binom{2k}{k}\right) \le \exp\left(-\frac{4^{\varepsilon k}}{2\sqrt{k}}\right) \to 0.$$

■

The following gives a proof of Theorem 1.2 for $n = 2k+1$ from Theorem 1.4.

Observation 4.2. W.h.p., $K_{p'}(2k+1,k)$ has no independent near-stars, where $p' = 1-4^{-(1+\varepsilon)}$. That is, $q_2(f) < p'$ w.h.p.

Proof. We call a near-star $\{A\} \cup (\mathcal{K}_x \backslash \{B\})$, where $A \notin \mathcal{K}_x$ and $B \in \mathcal{K}_x$, *maximal* if A and B are disjoint. If a nonmaximal near-star is independent, then it is contained in an independent superstar.

[3] In fact, ε can be taken much smaller than a constant, but for $n = 2k+1$ constant ε is sufficient.

Therefore if there are no independent superstars and no independent maximal near-stars, then there are no independent near-stars. Note that $\binom{n-k-1}{k-1} = k$ is the degree of vertices of $\binom{[n]\setminus\{x\}}{k}$ to $\mathcal{K}_x$. The probability $A \notin \mathcal{K}_x$ forms an independent superstar or maximal near-star based at x is

$$(1-p')^k + kp'(1-p')^{k-1} \le 4k4^{-(1+\varepsilon)k}.$$

Thus the expected number of independent superstars and maximal near-stars is at most

$$n\binom{n-1}{k} \cdot 4k4^{-(1+\varepsilon)k} \le k^2 4^{-\varepsilon k} \to 0,$$

so by Markov's inequality $K_{p'}(n,k)$ has no independent near-stars w.h.p. ∎

Going forward, the dependence of the hitting times $q_1(f)$ and $q_2(f)$ and random graphs $K_p(n,k,f)$ on f will be implicit.

4.2 | Connected components

Similar to how we reduced from $\mathcal{T}^1$ to $\mathcal{T}^2$ in Section 2.2, we reduce to "connected components" to deal with $\mathcal{F} \in \mathcal{T}^2$ with $a_{\mathcal{F}}$ small. Our goal is to produce a subcollection $\mathcal{T}^3$ of $\mathcal{T}^2$ such that: for $\mathcal{F} \in \mathcal{T}^2$ with $a_{\mathcal{F}}$ small, there exists $\mathcal{F}' \in \mathcal{T}^3$ such that $E(\mathcal{F}') \subseteq E(\mathcal{F})$, and the number of $\mathcal{A}_{\mathcal{F}}$ across all $\mathcal{F} \in \mathcal{T}^3$ is small. This reduction is helpful because it often reduces the number of $\mathcal{F}$ we take a union bound over.

We define $\mathcal{T}^3$ by a connectedness condition in an auxiliary graph. Let $J_x(n,k)$ be the auxiliary graph on $\binom{[n]\setminus\{x\}}{k}$ defined by A is adjacent to A' if and only if $N(A) \cap N(A') \cap \mathcal{K}_x \neq \emptyset$ in $K(n,k)$, or equivalently, $|A \cup A'| \le n-k$.[4] If A is adjacent to A' in $J_x(n,k)$, A and A' are said to be *2-linked with respect to x*. For $\mathcal{A}' \subseteq \mathcal{A} \subseteq \binom{[n]\setminus\{x\}}{k}$, we say $\mathcal{A}'$ is a *2-linked component of* $\mathcal{A}$ *with respect to x* if $\mathcal{A}'$ is a connected component of the subgraph of $J_x(n,k)$ induced by $\mathcal{A}$. We say $\mathcal{F} \subseteq \binom{[n]}{x}$ is *2-linked with respect to x* if the subgraph of $J(n,k)$ induced by $\mathcal{F}\setminus\mathcal{K}_x$ is connected.

Let $\mathcal{T}^3 = \{\mathcal{F} \in \mathcal{T}^2 : \mathcal{F}$ is 2-linked with respect to $x_{\mathcal{F}}\}$. The extra condition that $\mathcal{T}^3$ imposes over $\mathcal{T}^2$ cuts down on the number of choices for $\mathcal{A}_{\mathcal{F}}$. To see this, we need the following lemma about counting induced connected subgraphs.

Lemma 4.3 ([13], lemma 2.1)**.** *Let G be a graph on n vertices of maximum degree at most d. The number of $A \subseteq V(G)$ with $|A| = a$ such that $G[A]$ is connected is at most $n(ed)^a$.*

Since $J_x(2k+1,k)$ is k^2-regular, we immediately get the following lemma.

Lemma 4.4. *The number of $\mathcal{A}_{\mathcal{F}}$ with $\mathcal{F} \in \mathcal{T}_x^3(a)$ is at most $\binom{2k}{k}(ek^2)^a$ for $n = 2k+1$.*

4.3 | Small diversity and 2-linked $\mathcal{F}$

Proposition 3.3 already handles $\mathcal{F}$ with $a_{\mathcal{F}}$ large; the following proposition deals with when $a_{\mathcal{F}}$ is small. Here we bound the number of $\mathcal{F}$ by counting 2-linked $\mathcal{F}$ with $\mathcal{B}_{\mathcal{F}} \subseteq N(\mathcal{A}_{\mathcal{F}})$. When $a_{\mathcal{F}}$ is very small,

[4]The Johnson "scheme" is a collection of graphs $J(n,k,c)$, defined on $\binom{[n]}{k}$ with U and V adjacent if and only if $|U\setminus V| = c$ (see [25]). The Kneser graph is $J(n,k,k)$, and $J(n,k,1)$ and $J(n,k,n-2k)$ were used to establish the main result of [10]. The union of $J_x(n,k)$ over all x is the union of $J(n,k,c)$ over all $c \le n-2k$.

we use a trivial bound on $e(\mathcal{F})$ instead of Theorem 2.7. We extend this proposition to all $\mathcal{F}$ in $\mathcal{T}^2$ in Section 4.4.

Proposition 4.5. *No $\mathcal{F} \in \mathcal{T}^3$ with $\log \frac{\binom{n-1}{k}}{a_{\mathcal{F}}} = \omega(\log k)$ is independent in $K_p(n,k)$ except if $a_{\mathcal{F}} = 1$ w.h.p., where $p = 1 - 4^{-(1-\varepsilon)}$ and $n = 2k+1$.*

Proof. By Lemma 4.4, the number of 2-linked $\mathcal{A}_{\mathcal{F}}$ with $\mathcal{F} \in \mathcal{T}_x^3(a)$ is at most

$$\binom{2k}{k} \exp\left(O(a \log k)\right) = \binom{2k}{k} \exp\left(o\left(a \log \frac{\binom{n-1}{k}}{a}\right)\right),$$

since $\log \frac{\binom{n-1}{k}}{a} = \omega(\log k)$. The number of choices for $\mathcal{B} \subseteq N(\mathcal{A})$, given $\mathcal{A}$, is at most

$$\binom{a\binom{n-k-1}{k-1}}{a} = \binom{ak}{a} \le \exp\left(O\left(a \log k\right)\right) = \exp\left(o\left(a \log \frac{\binom{n-1}{k}}{a}\right)\right).$$

Note that $e(\mathcal{F}) \ge a(\binom{n-k-1}{k-1} - a) \ge \frac{7}{8}ak$ for $a \le k/8$, so the probability some $\mathcal{F} \in \mathcal{T}_x^3(a)$ with $a \le k/8$ is independent is at most

$$\binom{2k}{k} \exp\left(o\left(a \log \frac{\binom{n-1}{k}}{a}\right)\right)(1-p)^{\frac{7}{8}ak} \le 4^k \exp\left(o\left(a \log \frac{\binom{n-1}{k}}{a}\right)\right) 4^{-(1-\varepsilon)\frac{7}{8}ak} \le 4^{-ak/2},$$

using $a \ge 2$. For $a \ge k/8$, we use (8), so the probability that some $\mathcal{F}$ is independent is at most

$$\exp\left(\log\binom{2k}{k} + o\left(a \log \frac{\binom{n-1}{k}}{a}\right) - \theta a \log \frac{\binom{n-1}{k}}{a}\right) = \exp\left(-\frac{\theta}{2} a \log \frac{\binom{n-1}{k}}{a}\right).$$

In both cases, we can take a union bound over all such a and x to finish the proof. ∎

4.4 | Reduction to 2-linked components

Now we extend Proposition 4.5 from $\mathcal{T}^3$ to $\mathcal{T}^2$ by reducing $\mathcal{F} \in \mathcal{T}^2$ to its components in $\mathcal{T}^3$. Combined with Proposition 3.3, this gives a characterization of the independent sets of $K_{q_1}(n,k)$, of which Theorems 1.3 and 1.4 are a quick corollary.

Proposition 4.6. *Let $H = K_{q_1}(n,k)$. W.h.p., the only $\mathcal{F} \in \mathcal{T}$ which are independent in H have the following form: $\mathcal{F} = \{A_1, \dots, A_m\} \cup (\mathcal{K}_x \backslash \{B_1, \dots, B_m\})$ where $E_H(\mathcal{F}, \mathcal{K}_x) = \{A_1B_1, \dots, A_mB_m\}$ and $m < \frac{1}{4}\binom{n-2}{k-1}$.*

Proof. Let $H = K_{q_1}(n,k)$. By Observation 4.1, $q_1 > p$ w.h.p., so assume this is the case. Furthermore, assume that H satisfies the conclusions of Lemma 2.8 and Propositions 3.3 and 4.5. Let $\mathcal{F} \in \mathcal{T}$ be

independent in H, so by Lemma 2.8, $\mathcal{F} \in \mathcal{T}_x^1(a)$ for some x. By Proposition 3.3, $a < \frac{1}{3}\binom{n-2}{k-1}$, so let $\mathcal{F}' \in \mathcal{T}_x^2(a)$ be a set family given by applying Proposition 2.12 to $\mathcal{F}$.

Let $\mathcal{A}_1, \dots, \mathcal{A}_m$ be the connected components of $\mathcal{A}_{\mathcal{F}}$ in $J(n,k)$, and let $\mathcal{B}_i = \mathcal{B}_{\mathcal{F}'} \cap N(\mathcal{A}_i)$ for $i \in [m]$. Let $i \in [m]$ be such that $|\mathcal{A}_i| \geq |\mathcal{B}_i|$. We claim that, with the assumptions on H, $|\mathcal{A}_i| = |\mathcal{B}_i| = 1$. Since there are no edges in $J_x(n,k)$ between distinct $\mathcal{A}_i$, the $N(\mathcal{A}_i) \cap \mathcal{K}_x$ and thus the $\mathcal{B}_i$ are pairwise disjoint. Because $\mathcal{F}' \in \mathcal{T}^2$, we have

$$\sum_{i=1}^{m} |\mathcal{B}_i| = |\mathcal{B}_{\mathcal{F}'}| = |\mathcal{A}_{\mathcal{F}'}| = \sum_{i=1}^{m} |\mathcal{A}_i|. \tag{12}$$

With our claim that $|\mathcal{A}_i| \geq |\mathcal{B}_i|$ implies $|\mathcal{A}_i| = |\mathcal{B}_i| = 1$, this gives that $|\mathcal{A}_i| = |\mathcal{B}_i| = 1$ for all $i \in [m]$, Let $\mathcal{A}_i = \{A_i\}$ and $\mathcal{B}_i = \{B_i\}$. Since the $N(\mathcal{A}_i)$ are disjoint, we have that $\mathcal{F}' = \{A_1, \dots, A_m\} \cup (\mathcal{K}_x \backslash \{B_1, \dots, B_m\})$ where $E_H(\mathcal{F}', \mathcal{K}_x) = \{A_1B_1, \dots, A_mB_m\}$. By construction, $\mathcal{A}_{\mathcal{F}} = \mathcal{A}_{\mathcal{F}'}$, so unless $\mathcal{B}_{\mathcal{F}} = \mathcal{B}_{\mathcal{F}'}$, $\mathcal{F}$ will contain some edge A_iB_i. Thus $\mathcal{F} = \mathcal{F}'$, and $\mathcal{F}$ has the desired structure.

All that remains is to prove the claim. Since $|\mathcal{A}_i| \leq a < \binom{n-2}{k-1}$, by Lemma 2.11, $|N(\mathcal{A}_i) \cap \mathcal{K}_x| \geq |\mathcal{A}_i|$, so there exists $\mathcal{B}'$ of size $|\mathcal{A}_i|$ such that $N(\mathcal{A}_i) \cap \mathcal{B}_{\mathcal{F}'} \subseteq \mathcal{B}' \subseteq N(\mathcal{A}_i) \cap \mathcal{K}_x$. Let $\mathcal{F}'' = \mathcal{A}_i \cup (\mathcal{K}_x \backslash \mathcal{B}')$, and observe that

$$E(\mathcal{F}'') = E(\mathcal{A}_i) \cup E(\mathcal{A}_i, \overline{\mathcal{B}'}) \subseteq E(\mathcal{A}_{\mathcal{F}}) \cup E(\mathcal{A}_i, \overline{N(\mathcal{A}_i) \cap \mathcal{B}_{\mathcal{F}'}}) \subseteq E(\mathcal{F}').$$

By Observation 2.2, for any $y \in [n]$ with $y \neq x$,

$$|\mathcal{F}'' \backslash \mathcal{K}_y| \geq |\mathcal{F}' \backslash \mathcal{K}_y| - |\mathcal{B}'| - |\mathcal{A}_{\mathcal{F}}| \geq \binom{n-2}{k-1} - 3a > a \geq |\mathcal{A}_i| = |\mathcal{F}'' \backslash \mathcal{K}_x|,$$

which follows since $a < \frac{1}{4}\binom{n-2}{k-1}$ by Proposition 3.3. Thus $x_{\mathcal{F}''} = x$ and so $\mathcal{A}_{\mathcal{F}''} = \mathcal{A}_i$, $\mathcal{B}_{\mathcal{F}''} = \mathcal{B}'$, and $\mathcal{F}'' \in \mathcal{T}_x^3(|\mathcal{A}_i|)$. Since H satisfies the conclusion of Proposition 4.5, we must have $a_{\mathcal{F}''} = 1$, so $|\mathcal{A}_i| = 1$. If $|\mathcal{B}_i| = 0$, then $\mathcal{A}_i \cup \mathcal{K}_x$ is an independent superstar in H, which is forbidden. Thus $|\mathcal{A}_i| = |\mathcal{B}_i| = 1$. ∎

4.5 | Proof of Theorems 1.3 and 1.4

Compiling the results of this section, we first prove the simpler Theorem 1.4 and then Theorem 1.3.

Proof of Theorem 1.4. Let $H = K_{q_2}(n,k)$. By Proposition 4.6, since $q_1 < q_2$, w.h.p. the only set families in $\mathcal{T}$ which are independent in H are $\{A_1, \dots, A_m\} \cup (\mathcal{K}_x \backslash \{B_1, \dots, B_m\})$ where $E_H(\mathcal{F}, \mathcal{K}_x) = \{A_1B_1, \dots, A_mB_m\}$. But $\{A_1\} \cup (\mathcal{K}_x \backslash \{B_1\})$ is a near-star, which is not independent in H. Thus no $\mathcal{F} \in \mathcal{T}$ is independent in H w.h.p., as desired. ∎

Proof of Theorem 1.3. Let $H = K_{q_1}(n,k)$. Consider a set family $\mathcal{F} \subseteq V(H)$ of size $\binom{n-1}{k-1} + 1$ which is independent in H. Let x minimize $|\mathcal{K}_x \backslash \mathcal{F}|$, and let $B \in \mathcal{K}_x \cap \mathcal{F}$. By Proposition 4.6, w.h.p. $\mathcal{F} \backslash \{B\} = \{A_1, \dots, A_m\} \cup (\mathcal{K}_y \backslash \{B_1, \dots, B_m\})$, for some $y \in [n]$, where $E_H(\mathcal{F} \backslash \{B\}, \mathcal{K}_y) = \{A_1B_1, \dots, A_mB_m\}$. This means that $|\mathcal{K}_y \backslash \mathcal{F}| < \frac{1}{4}\binom{n-2}{k-1}$, so by Observation 2.2, $y = x$. Thus $B = B_i$ for some i, so $\mathcal{F}$ contains the edge A_iB_i, a contradiction. ∎

Remark. It is straightforward to generalize the proofs in this section to all k with $n - 2k = o(n)$. The main difference is that the degrees in the auxiliary graph $J_x(n,k)$ increase, but not so dramatically

when $n-2k=o(n)$. However, for say $n=4k$, the graph $J_x(n,k)$ becomes complete, and so Lemma 4.3 no longer provides an efficient way to count the number of $\mathcal{A}_{\mathcal{F}}$.

One may also obtain hitting time results for $k=o(n)$ via a different approach, namely a careful modification of the argument in [9], which is a slight variation on the argument in [8]. For $\mathcal{F}$ with small $a_{\mathcal{F}}$ one counts maximal independent sets in $K(n,k)$, while $\mathcal{F}$ with large $a_{\mathcal{F}}$ are already handled by Lemma 2.8 and (9), which imply

$$\log \frac{\binom{n-1}{k}}{a} > \frac{\theta}{5}\cdot\frac{n-2k}{n}\log\binom{n-1}{k} \sim k\log\frac{n}{k}.$$

When k is very small, further methods from [8] must be used. One must also be careful with the choice of ε, as a constant ε does not work with these techniques. Unfortunately, these techniques break down when $k=\Omega(n)$ and in that case only handle $\mathcal{F}$ with $a_{\mathcal{F}}=o(n)$.

5 | SHARP THRESHOLD FOR $n \le Ck$

As Theorem 1.2 is proven for $n=2k+1$ in Section 4, and Das and Tran [9] proved Theorem 1.2 for $k \le n/C$ for some constant C, we assume for this section that $2k+1<n<Ck$, fixing this constant C sufficiently large. Additionally fix a constant $\varepsilon>0$ sufficiently small, and let $p=(1+\varepsilon)p_0$.

In Section 5.1, we prove an important assumption about $K_{(1+\varepsilon)p_0}(n,k)$ that we are unable to make about the hitting time version $K_{\tau_{\text{near}}}(n,k)$. Given this, in Section 5.2 we show that for every $\mathcal{F}$ in consideration, $|\mathcal{A}_{\mathcal{F}}^{<\delta}|$ is small, and so, with Lemma 3.1, the number of choices for $\mathcal{A}_{\mathcal{F}}$ is small. When $n-2k$ is small, this allows us to quickly finish in Section 5.3. Otherwise, this allows us to eliminate those $\mathcal{F}$ with $e(\mathcal{A}_{\mathcal{F}})$ large from consideration in Section 5.4. Finally, in Section 5.5 we deal with counting $\mathcal{B}^{<\delta}$, and we wrap up the proof of Theorem 1.2 in Section 5.6.

While our proof likely works for smaller ε, giving tighter bounds on the "width of the window" for this threshold, we do not optimize our choice of ε here, mostly because we think the stronger hitting time version of Conjecture 1.5 ought to be true.

5.1 | Minimum degree assumption

First we show that not only are near-stars not independent in $K_p(n,k)$ w.h.p., but every superstar contains at least δk edges in $K_p(n,k)$ for some constant $\delta>0$.

Lemma 5.1. *Let $H=K_p(n,k)$, where $p=(1+\varepsilon)p_0$. There exists a constant $\delta=\delta(\varepsilon)>0$ such that, w.h.p., for every $x\in[n]$ and $A\in\binom{[n]\setminus\{x\}}{k}$, we have $d_H(A,\mathcal{K}_x)\ge\delta k$.*

Proof. Let $x\in[n]$ and $S\in\binom{[n]\setminus\{x\}}{k}$. Let

$$\delta=\frac{\varepsilon^2(1+\varepsilon)\log\left(n\binom{n-1}{k}\right)}{25k},$$

and note that δ is constant, since $\log\left(n\binom{n-1}{k}\right)=\Theta(k)$. Since $d_H(A,\mathcal{K}_x)$ is binomially distributed with mean

$$p\binom{n-k-1}{k-1}=(1+\varepsilon)\log\left(n\binom{n-1}{k}\right)=25\delta k/\varepsilon^2,$$

by Lemma 1.7, the probability that $d_H(A, \mathcal{K}_x)$ is less than δk is at most

$$\exp\left(-\left(1-2\sqrt{\varepsilon^2/25}\right)(1+\varepsilon)\log\left(n\binom{n-1}{k}\right)\right) \le \left(n\binom{n-1}{k}\right)^{-1-\varepsilon/2}.$$

Taking the union bound over all $n\binom{n-1}{k}$ choices for x and A, the lemma follows. ∎

We fix the δ given by Lemma 5.1 for the remainder of this section. The specific value of δ is not important; all that matters is that δ remains constant as $n \to \infty$.

Remark. Lemma 5.1 already implies that $\mathcal{F}$ with $a_{\mathcal{F}} < \delta k$ are not independent in $K_p(n,k)$ w.h.p. This is because for such $\mathcal{F}$, $\mathcal{B}_{\mathcal{F}}$ is too small to absorb all the edges from a single vertex of $\mathcal{A}_{\mathcal{F}}$.

5.2 | Few vertices of low degree in $\mathcal{A}_{\mathcal{F}}$

Vertices of $\mathcal{A}_{\mathcal{F}}$ with low degree to $\mathcal{B}_{\mathcal{F}}$ in $K(n,k)$ will likely have relatively low degree to $\mathcal{B}_{\mathcal{F}}$ in $K_p(n,k)$. But if $K_p(n,k)$ satisfies the minimum degree assumption, then there must be edges between $\mathcal{A}_{\mathcal{F}}$ and $\overline{\mathcal{B}_{\mathcal{F}}}$, so $\mathcal{F}$ is not independent. We formalize this with the following lemma, which upper bounds the number of vertices of $\mathcal{A}_{\mathcal{F}}$ with low degree to $\mathcal{B}_{\mathcal{F}}$ and with Lemma 3.1 upper bounds the total number of choices for $\mathcal{A}_{\mathcal{F}}$. Let

$$\mathcal{T}^4 = \left\{\mathcal{F} \in \mathcal{T}^1 : \left|\mathcal{A}_{\mathcal{F}}^{<1/\sqrt{k}}\right| \le \frac{a_{\mathcal{F}}}{\log\log k}\right\}.$$

Proposition 5.2. *No $\mathcal{F} \in \mathcal{T}^1 \setminus \mathcal{T}^4$ is independent in $K_p(n,k)$ w.h.p., where $p = (1+\varepsilon)p_0$.*

Proof. Let $H = K_p(n,k)$, and consider $\mathcal{F} \in \mathcal{T}_x^1(a) \setminus \mathcal{T}_x^4(a)$. For $A \in \mathcal{A}_{\mathcal{F}}^{<1/\sqrt{k}}$, $d_H(A, \mathcal{B}_{\mathcal{F}})$ is binomially distributed with mean $pd_{K(n,k)}(A, \mathcal{B}_{\mathcal{F}}) \le (1+\varepsilon)\frac{1}{\sqrt{k}}\log\binom{n-1}{k}$. By Lemma 1.7, the probability that this degree is at least δk is at most

$$\exp\left(-\delta k \log \frac{\delta k}{e(1+\varepsilon)\frac{1}{\sqrt{k}}\log\binom{n-1}{k}}\right) \le \binom{n-1}{k}^{-3\log\log k}.$$

Since the edges incident between different $A \in \mathcal{A}_{\mathcal{F}}$ and $\mathcal{B}_{\mathcal{F}}$ are distinct, the probability that every $A \in \mathcal{A}_{\mathcal{F}}^{<1/\sqrt{k}}$ satisfies $d_H(A, \mathcal{B}_{\mathcal{F}}) \ge \delta k$ is at most $\binom{n-1}{k}^{-3a}$. If for some $A \in \mathcal{A}_{\mathcal{F}}$, $d_H(A, \mathcal{B}_{\mathcal{F}}) < \delta k$, but $d_H(A, \mathcal{K}_x) \ge \delta k$, then $\mathcal{F}$ is not independent. By (5), applying the union bound over all $\mathcal{F}$ yields that the probability that some $\mathcal{F} \in \mathcal{T}_x^1(a) \setminus \mathcal{T}_x^4(a)$ is independent, and H satisfies the minimum degree assumption, is at most

$$\binom{\binom{n-1}{k}}{a}\binom{\binom{n-1}{k-1}}{a}\binom{n-1}{k}^{-3a} \le \binom{n-1}{k}^{-a}.$$

Taking a union bound over all a and x and applying Lemma 5.1 finishes the proof. ∎

Using the definition of $\mathcal{T}^4$, we can conclude the following.

Lemma 5.3. *The number of $\mathcal{A}_{\mathcal{F}}$ across all $\mathcal{F} \in \mathcal{T}_x^4(a)$ is at most* $\exp\left(o\left(a\log\frac{\binom{n-1}{k}}{a}\right)\right)$.

Proof. By Lemma 3.1, since

$$\frac{1}{\sqrt{k}} = \omega\left(\frac{n\log\frac{n}{n-2k}}{k\log\binom{n-1}{k}}\right),$$

the number of $\mathcal{A}_{\mathcal{F}}^{\geq 1/\sqrt{k}}$ across all $\mathcal{F} \in \mathcal{T}_x^1(a)$ is at most

$$\exp\left(o\left(a\log\frac{\binom{n-1}{k}}{a}\right)\right).$$

Since $\left|\mathcal{A}_{\mathcal{F}}^{<1/\sqrt{k}}\right| = o(a)$ for all $\mathcal{F} \in \mathcal{T}_x^4(a)$, the number of choices for $\mathcal{A}_{\mathcal{F}}^{<1/\sqrt{k}}$ is at most

$$\binom{\binom{n-1}{k}}{o(a)} = \exp\left(o\left(a\log\frac{\binom{n-1}{k}}{a}\right)\right).$$

Multiplying these counts finishes the proof. ∎

5.3 | When $n - 2k$ is small

When $n - 2k = o(n)$, we can essentially repeat the proof of Proposition 4.5, although the calculations are easier this time around.

Proposition 5.4. *When $n - 2k = o(n)$, no $\mathcal{F} \in \mathcal{T}^2 \cap \mathcal{T}^4$ with $\log\frac{\binom{n-1}{k}}{a} = \omega\left(\log\binom{n-k-1}{n-2k}\right)$ is independent in $K_p(n,k)$, where $p = (1+\varepsilon)p_0$*

Proof. By Lemma 5.3, the number of $\mathcal{A}_{\mathcal{F}}$ across all $\mathcal{F} \in \mathcal{T}_x^4(a)$ is at most

$$\exp\left(o\left(a\log\frac{\binom{n-1}{k}}{a}\right)\right).$$

Recall that for $\mathcal{F} \in \mathcal{T}^2$, $\mathcal{B}_{\mathcal{F}} \subseteq N(\mathcal{A}_{\mathcal{F}})$. Given $\mathcal{A}$, the number of choices for $\mathcal{B} \subseteq N(\mathcal{A})$ is at most

$$\binom{a\binom{n-k-1}{k-1}}{a} = \binom{a\binom{n-k-1}{n-2k}}{a} \leq \exp\left(a\log\binom{n-k-1}{n-2k}\right) \leq \exp\left(o\left(a\log\frac{\binom{n-1}{k}}{a}\right)\right).$$

Using (8), the probability that some $\mathcal{F} \in \mathcal{T}_x^4(a)$ is independent is at most

$$\exp\left(o\left(a\log\frac{\binom{n-1}{k}}{a}\right) - (1+\varepsilon)\theta a\log\frac{\binom{n-1}{k}}{a}\right) \leq \exp\left(-\frac{\theta}{2}a\log\frac{\binom{n-1}{k}}{a}\right).$$

Taking a union bound over all a and x finishes the proof. ∎

5.4 | Few disjoint pairs in $\mathcal{A}$

In the next subsection, we need lower bounds on $e(\mathcal{A}, \overline{\mathcal{B}}) = e(\mathcal{F}) - e(\mathcal{A})$. Unfortunately, lower bounds for just $e(\mathcal{F})$ are given in Theorem 2.7, so we supply an upper bound on $e(\mathcal{A})$ here. Let

$$\mathcal{T}^5 = \left\{ \mathcal{F} \in \mathcal{T}^4 : e(\mathcal{A}_\mathcal{F}) \le \frac{1}{2} e(\mathcal{F}) \right\}.$$

Corollary 5.5. *No $\mathcal{F} \in \mathcal{T}^4 \backslash \mathcal{T}^5$ is independent in $K_p(n,k)$ w.h.p., where $p = (1+\varepsilon)p_0$.*

Proof. By Lemma 5.3, the number of $\mathcal{A}_\mathcal{F}$ across all $\mathcal{F} \in \mathcal{T}_x^4(a)$ is at most

$$\exp\left(o\left(a \log \frac{\binom{n-1}{k}}{a} \right) \right).$$

The probability that a particular $\mathcal{A}_\mathcal{F}$ is independent, where $\mathcal{F} \in \mathcal{T}^4 \backslash \mathcal{T}^5$, is $(1-p)^{e(\mathcal{A})} \le (1-p)^{\frac{1}{2}e(\mathcal{F})}$, so the probability that some $\mathcal{F} \in \mathcal{T}_x^4(a) \backslash \mathcal{T}_x^5(a)$ is independent is at most, using (8),

$$\exp\left(o\left(a \log \frac{\binom{n-1}{k}}{a} \right) - \frac{1}{2}\theta a \log \frac{\binom{n-1}{k}}{a} \right).$$

A union bound over all choices of a and x finishes the proof. ∎

5.5 | When $n - 2k$ is large

If the number of $\mathcal{B}_\mathcal{F}$ across all $\mathcal{F} \in \mathcal{T}_x^5(a)$ were small, then we could proceed as in Proposition 5.4. Lemma 3.2 guarantees few choices for $\mathcal{B}_\mathcal{F}^{\ge\eta}$ for reasonable η, so we could try to prove that there are few choices for $\mathcal{B}_\mathcal{F}^{<\eta}$. Unfortunately, the strategy we used for $\mathcal{A}_\mathcal{F}^{<\eta}$ will not work here, since although we can make a similar minimum degree assumption, there would be no analogue of Proposition 5.2, as $e_H(\mathcal{F}) = 0$ does not guarantee that $e_H(\overline{\mathcal{A}}, \mathcal{B}) = 0$.

In fact, we are unable to have good control of $|\mathcal{B}^{<\eta}|$ for reasonable η. Instead, by taking a union bound over all $\mathcal{A}_\mathcal{F}$ and $\mathcal{B}_\mathcal{F}^{\ge\eta}$, we will show that w.h.p. there are many edges between $\mathcal{A}_\mathcal{F}$ and $\overline{\mathcal{B}_\mathcal{F}^{\ge\eta}}$ for every $\mathcal{F}$ — too many to be absorbed by $\mathcal{B}_\mathcal{F}^{<\eta}$, so these $\mathcal{F}$ cannot be independent.

Proposition 5.6. *When $n - 2k = \omega(\log n)$ and $k = \omega(1)$, no $\mathcal{F} \in \mathcal{T}^5$ is independent in $K_p(n,k)$ w.h.p., where $p = (1+\varepsilon)p_0$.*

Proof. Let $H = K_p(n,k)$. Consider $\mathcal{F} \in \mathcal{T}_x^5(a)$. Define

$$\eta = \frac{\theta e^{-81/\theta}}{16} \cdot \frac{k}{n-k} \cdot \frac{\log \frac{\binom{n-1}{k}}{a}}{\log\left(n\binom{n-1}{k}\right)} \ge \Omega\left(\frac{k}{n-k} \cdot \frac{n-2k}{n} \right),$$

which follows from Lemma 2.9. By Lemma 5.3, the number of $\mathcal{A}_{\mathcal{F}}$ across all $\mathcal{F} \in \mathcal{T}_x^5(a)$ is at most

$$\exp\left(o\left(a\log\frac{\binom{n-1}{k}}{a}\right)\right).$$

By Lemma 3.2, the number of distinct $\mathcal{B}_{\mathcal{F}}^{\geq\eta}$ across all $\mathcal{F} \in \mathcal{T}^3$ is at most

$$\exp\left(o\left(a\log\frac{\binom{n-1}{k}}{a}\right)\right),$$

since

$$\eta = \Omega\left(\frac{k}{n}\cdot\frac{n-2k}{n}\right) = \omega\left(\frac{k}{n}\frac{\log\log\binom{n-1}{k}}{\log\binom{n-1}{k}}\right).$$

Thus the number of $(\mathcal{A}, \mathcal{B}^{\geq\eta})$ pairs is at most

$$\exp\left(o\left(a\log\frac{\binom{n-1}{k}}{a}\right)\right).$$

Observe that $e_H(\mathcal{A}, \overline{\mathcal{B}^{\geq\eta}})$ is binomially distributed with mean

$$pe_{K(n,k)}(\mathcal{A}, \overline{\mathcal{B}^{\geq\eta}}) \geq \frac{1}{2}(1+\varepsilon)p_0 e(\mathcal{F}) \geq \frac{\theta}{2}a\log\frac{\binom{n-1}{k}}{a}.$$

By Lemma 1.7, the probability that $e_H(\mathcal{A}, \overline{\mathcal{B}^{\geq\eta}}) \leq \frac{\theta}{16}a\log\frac{\binom{n-1}{k}}{a}$ is at most

$$\exp\left(-\Omega\left(a\log\frac{\binom{n-1}{k}}{a}\right)\right).$$

We may take the union bound over all such $\mathcal{F}$, a, and x to get that, w.h.p., $e_H(\mathcal{A}_{\mathcal{F}}, \overline{\mathcal{B}_{\mathcal{F}}^{\geq\eta}}) \geq \frac{\theta}{16}a\log\frac{\binom{n-1}{k}}{a}$ for all $\mathcal{F} \in \mathcal{T}^5$.

Similarly, observe that $e_H(\mathcal{A}, \mathcal{B}^{<\eta})$ is binomially distributed with mean

$$pe_{K(n,k)}(\mathcal{A}, \mathcal{B}^{<\eta}) \leq (1+\varepsilon)p_0|\mathcal{B}^{<\eta}|\eta\binom{n-k}{k} \leq (1+\varepsilon)\frac{\theta}{16}e^{-81/\theta}a\log\frac{\binom{n-1}{k}}{a},$$

where in the second inequality we simply used $|\mathcal{B}^{<\eta}| \leq a$. By Lemma 1.7, the probability that $e_H(\mathcal{A}, \mathcal{B}^{<\eta}) \geq \frac{\theta}{16}a\log\frac{\binom{n-1}{k}}{a}$ is at most

$$\exp\left(-5a\log\frac{\binom{n-1}{k}}{a}\right).$$

Using (5), we can take a union bound over all $\mathcal{F} \in \mathcal{T}_x^3(a)$, and then union bound over all a and x. Thus w.h.p., $e_H(\mathcal{A}, \mathcal{B}^{<\eta}) < \frac{\theta}{16} a \log \frac{\binom{n-1}{k}}{a}$ for all $\mathcal{F} \in \mathcal{T}^5$.

Thus w.h.p., for all $\mathcal{F} \in \mathcal{T}^5$,

$$e_H(\mathcal{A}_\mathcal{F}, \overline{\mathcal{B}}_\mathcal{F}) \geq e_H(\mathcal{A}_\mathcal{F}, \overline{\mathcal{B}_\mathcal{F}^{\geq\eta}}) - e_H(\mathcal{A}_\mathcal{F}, \mathcal{B}_\mathcal{F}^{<\eta}) > 0,$$

so no such $\mathcal{F}$ is independent in H. ∎

5.6 | Proof of Theorem 1.2

Proof of Theorem 1.2. By Lemma 2.8, the only $\mathcal{F} \in \mathcal{T}$ which are independent in $K_p(n,k)$ are in $\mathcal{T}^1$ w.h.p.

For $n - 2k \leq \sqrt{n}$, say, Proposition 3.3 gives that the only $\mathcal{F} \in \mathcal{T}^1$ which are independent in $K_p(n,k)$ satisfy $\log \frac{\binom{n-1}{k}}{a} = o(n)$. For each such $\mathcal{F}$, there exists $\mathcal{F}' \in \mathcal{T}^2$ with $E(\mathcal{F}'') \subseteq E(\mathcal{F})$ by Proposition 2.12. Since

$$\log \binom{n-k-1}{n-2k} = O\left((n-2k)\log \frac{n}{n-2k}\right) = o(n),$$

Propositions 5.2 and 5.4 show that $\mathcal{F}'$ is not independent w.h.p., completing the proof for $n - 2k \leq \sqrt{n}$.

For $n - 2k \geq \sqrt{n}$, Proposition 5.2 gives that the only $\mathcal{F} \in \mathcal{T}^1$ which are independent in $K_p(n,k)$ are in $\mathcal{T}^4$. The theorem then follows from Corollary 5.5 and Proposition 5.6. ∎

REFERENCES

1. J. Balogh, T. Bohman, and D. Mubayi, *Erdős-Ko-Rado in random hypergraphs*, Comb. Prob. Comput. **18** (2009), no. 5, 629–646.
2. J. Balogh, B. Bollobás, and B. P. Narayanan, *Transference for the Erdős-Ko-Rado theorem*, Forum Math. Sigma **3** (2015), e23.
3. J. Balogh, S. Das, M. Delcourt, H. Liu, and M. Sharifzadeh, *Intersecting families of discrete structures are typically trivial*, J. Comb. Theory Ser. A **132** (2015), 224–245.
4. J. Balogh, S. Das, H. Liu, M. Sharifzadeh, and T. Tran, *Structure and supersaturation for intersecting families*, Electron. J. Comb. **26** (2019), P2.34.
5. J. Balogh, R. I. Garcia, L. Li, and A. Z. Wagner, *Intersecting families of sets are typically trivial*, arXiv preprint arXiv:2104.03260. 2021.
6. J. Balogh, R. Morris, and W. Samotij, *Independent sets in hypergraphs*, J. Amer. Math. Soc. **28** (2015), no. 3, 669–709.
7. L. I. Bogolubsky, A. S. Gusev, M. M. Pyaderkin, and A. M. Raigorodskii, *Independence numbers and chromatic numbers of the random subgraphs of some distance graphs*, Sbornik Math. **206** (2015), no. 10, 1340–1374.
8. B. Bollobás, B. P. Narayanan, and A. M. Raigorodskii, *On the stability of the Erdős-Ko-Rado theorem*, J. Comb. Theory Ser. A **137** (2016), 64–78.
9. S. Das and T. Tran, *Removal and stability for Erdős-Ko-Rado*, SIAM J. Discrete Math. **30** (2016), no. 2, 1102–1114.
10. P. Devlin and J. Kahn, *On "stability" in the Erdős-Ko-Rado theorem*, SIAM J. Discrete Math. **30** (2016), no. 2, 1283–1289.
11. P. Erdős, C. Ko, and R. Rado, *Intersection theorems for systems of finite sets*, Q. J. Math. **12** (1961), no. 1, 313–320.
12. P. Frankl and A. Kupavskii, *Erdős-Ko-Rado theorem for $\{0, \pm 1\}$-vectors*, J. Comb. Theory Ser. A **155** (2018), 157–179.
13. D. Galvin and J. Kahn, *On phase transition in the hard-core model on $\mathbb{Z}^d$*, Comb. Prob. Comput. **13** (2004), no. 2, 137–164.

14. M. M. Gauy, H. Hàn, and I. C. Oliveira, *Erdős-Ko-Rado for random hypergraphs: asymptotics and stability*, Comb. Prob. Comput. **26** (2017), no. 3, 406–422.
15. A. Hamm and J. Kahn, *On Erdős-Ko-Rado for random hypergraphs I*, Comb. Prob. Comput. **28** (2019), no. 6, 881–916.
16. A. Hamm and J. Kahn, *On Erdős-Ko-Rado for random hypergraphs II*, Comb. Prob. Comput. **28** (2019), no. 1, 61–80.
17. A. J. W. Hilton and E. C. Milner, *Some intersection theorems for systems of finite sets*, Q. J. Math. **18** (1967), no. 1, 369–384.
18. S. Janson, T. Łuczak, and A. Ruciński, Random graphs, John Wiley & Sons, New York: 2000.
19. G. O. H. Katona, "A theorem of finite sets," Theory of graphs, Akadémia Kiadó, Budapest, 1968, pp. 187–207.
20. A. V. Kostochka, *On the number of connected subgraphs with small edge-boundary in regular graphs*, Random Struct. Algor **5** (1994), no. 1, 147–154.
21. A. V. Kostochka, A. A. Sapozhenko, and K. Weber, *On random cubical graphs*, Ann. Discrete Math. **51** (1992), 155–160.
22. J. B. Kruskal, *The number of simplices in a complex*, Math. Optim. Tech. **10** (1963), 251–278.
23. A. Kupavskii, *Diversity of uniform intersecting families*, Eur. J. Comb. **74** (2018), 39–47.
24. L. Lovász, Combinatorial problems and exercises, North-Holland, Amsterdam, 1993.
25. F. J. MacWilliams and N. J. A. Sloane, The theory of error-correcting codes, North-Holland, Amsterdam, 1977.
26. D. Saxton and A. Thomason, *Hypergraph containers*, Invent. Math. **201** (2015), no. 3, 925–992.

How to cite this article: J. Balogh, R. A. Krueger, and H. Luo, *Sharp threshold for the Erdős–Ko–Rado theorem*, Random Struct. Algorithms. **62** (2023), 3–28. https://doi.org/10.1002/rsa.21090