On “stability” in the Erdős-Ko-Rado Theorem

Pat Devlin, Jeff Kahn

Abstract

Denote by $K_p(n, k)$ the random subgraph of the usual Kneser graph $K(n, k)$ in which edges appear independently, each with probability $p$. Answering a question of Bollobás, Narayanan, and Raigorodskii, we show that there is a fixed $p < 1$ such that a.s. (i.e., with probability tending to 1 as $k \to \infty$) the maximum independent sets of $K_p(2k+1, k)$ are precisely the sets $\{A \in V(K(2k+1, k)) : x \in A\} (x \in [2k+1])$.

We also complete the determination of the order of magnitude of the “threshold” for the above property for general $k$ and $n \geq 2k+2$. This is new for $k \sim n/2$, while for smaller $k$ it is a recent result of Das and Tran.

1 Introduction

The broad context of this paper is an effort, which has been one of the most interesting and successful combinatorial trends of the last couple decades, to understand how far some of the subject’s classical results remain true in a random setting. Since several nice accounts of these developments are available, we will not attempt a review (see, for example, the survey [13] or [6, 4] for discussions closer to present concerns) and mainly confine ourselves to the problem at hand.

Recall that, for integers $0 < k < n/2$, the Kneser graph, $K(n, k)$ has vertices the $k$-subsets of $[n] := \{1, 2, \ldots, n\}$, with two vertices adjacent if and only if they are disjoint sets. In what follows we set $K = \binom{[n]}{k}$ (the vertex set of $K(n, k)$). A star is one of the sets $K_x := \{A : x \in A\} (x \in [n])$. We also set $M = \binom{[n]}{k-1}$ (the size of a star) and write $C$ for the collection of $M$-subsets of $K$ that are not stars.

AMS 2010 subject classification: 05C35, 05D40, 05C80, 05C65, 05C69

Key words and phrases: Erdős-Ko-Rado Theorem, Kneser graph, random subgraph, threshold

* Supported by NSF grant DMS1201337.
In Kneser-graph terms the classical Erdős-Ko-Rado Theorem \cite{8} says that for $k < n/2$, the independence number of $K(n, k)$ is $M$ and, moreover, the only independent sets of this size are the stars.

Say a spanning subgraph $H$ of $K(n, k)$ has the \textit{EKR property} or \textit{is EKR} if each of its largest independent sets is a star. We are interested in this property for $H = K_p(n, k)$, the random subgraph of $K(n, k)$ in which edges appear independently, each with probability $p$. In particular we are interested in a question suggested and first studied by Bollobás, Narayanan, and Raigorodskii \cite{6}, viz.

\textbf{Question 1.} For what $p = p(n, k)$ is $K_p(n, k)$ likely to be EKR?

Formally, we would like to estimate the “threshold,” $p_c = p_c(n, k)$, which we define to be the unique $p$ satisfying

$$\Pr(K_p(n, k) \text{ is EKR}) = 1/2 \tag{1}$$

(which does turn out to be a threshold in the original Erdős-Rényi sense). Ideally (or nearly so) one hopes to identify some $p_0$, necessarily close to $p_c$, such that for fixed $\varepsilon > 0$, $K_p(n, k)$ is a.s. EKR if $p > (1 + \varepsilon)p_0$ and a.s. not EKR if $p < (1 - \varepsilon)p_0$. (As usual a property holds \textit{almost surely} (a.s.) if its probability tends to 1 as the relevant parameter—here $n$—tends to infinity).

Successively stronger results (some of this “ideal” type, some less precise) have been achieved by the aforementioned Bollobás \textit{et al.} \cite{6} and then by Balogh, Bollobás, and Narayanan \cite{4} and Das and Tran \cite{7}. Here we briefly discuss only \cite{7}, which subsumes the others.

A natural guess is that the value of $p_c$ is driven by the need to avoid independence of any $F \in \mathcal{C}$ that, for some $x$, satisfies $|\mathcal{F} \setminus K_x| = 1$. This turns out to suggest that $p_c = p_c(n, k)$ should be asymptotic to

$$p_0 = p_0(n, k) := \begin{cases} 
\frac{(n-k-1)}{k-1} \log(n(n-1)) & \text{if } n \geq 2k + 2, \\
3/4 & \text{if } n = 2k + 1
\end{cases}$$

(where, here and throughout, log is ln); namely, \cite{7} shows (strictly speaking only for $n \geq 2k + 2$) that for $p < (1 - \varepsilon)p_0$ (with $\varepsilon > 0$ fixed), $K_p(n, k)$ a.s. does contain independent $F$’s as above (implying $p_c > (1 - o(1)p_0)$, while it is easy to see that for $p > (1 + \varepsilon)p_0$ it a.s. does not. (Note $n = 2k + 1$ is not really special here: the form of $p_0$ changes because we lose the approximation of $1 - p$ by $e^{-p}$.)

In fact, Das and Tran show that, for some specified constant $C$, $p_c$ is indeed asymptotic to $p_0$ if $k < n/(3C)$, and is less than $Cnp_0/(n - 2k)$.
whenever $n \geq 2k + 2$, whence $p_c = O(p_0)$ if $k < (1/2 - \Omega(1))n$ (where the first implied constant depends on the second). Of course the estimate becomes less satisfactory as $k/n \to 1/2$, and in particular gives nothing for $n \in \{2k + 2, 2k + 1\}$. On the other hand, both [6] and [4] suggest that $n = 2k + 1$ is the most interesting case of the problem and ask whether one can at least show that $K_p(2k + 1, k)$ is a.s. EKR for some $p$ bounded away from 1. Here we prove such a result and also show that $p_c = O(p_0)$ remains true for general $n$ and $k$.

**Theorem 2.** There is a fixed $p < 1$ such that (for every $k$) $K_p(2k + 1, k)$ is a.s. EKR. There is a fixed $C$ such that for every $n$ and $k$, $K_p(n, k)$ is a.s. EKR for $p > Cp_0(n, k)$.

Again, one expects that $p_c \sim p_0$ in all cases and in particular that, as suggested in [4], $p_c(2k + 1, k) \to 3/4$; but we do not come close to these asymptotics and make no attempt to squeeze the best possible $\varepsilon$ and/or $C$ from our arguments.

It may be worth (briefly) comparing the present question with a similar one, introduced earlier by Balogh, Bohman and Mubayi [3], in which one considers a random induced subgraph of $K(n, k)$. Thus, one specifies only $\mathcal{H} = K_p$ (the random set in which each $A \in K$ is present with probability $p$, independent of other choices) and asks when the subgraph induced by $\mathcal{H}$ has the EKR property, now meaning that each largest independent set (that is, intersecting subfamily of $\mathcal{H}$) is a star $\mathcal{H}_x = \{A \in \mathcal{H} : x \in A\}$ for some $x$.

For $n = 2k + 1$ the situation here is similar to the one above: EKR should hold (a.s.) for any fixed $p > 3/4$, but even proving this for $p > 1 - \varepsilon$ with a fixed $\varepsilon > 0$—a problem suggested in [3]—does not seem easy; such a proof was given in [10] (using methods unrelated to those employed here).

But the resemblance may be superficial, and in fact the induced problem seems considerably subtler than the one considered here (as should probably be expected, e.g. since (i) the size of a largest star is itself a moving target and (ii) the most likely violators of EKR are not always families that are close to stars). See [9] for a guess as to what ought to be true here and [5, 9] for what’s known at this time.

The rest of the paper is devoted to the proof of Theorem 2. A single argument will suffice for both assertions, though, as noted below, not all of what we do is needed for $n = 2k + 1$. 
2 Proof

Notation. From now on we take $n = 2k + c$ and write $V$ for $[n]$ (so $K = \binom{V}{k}$).
For $H \subseteq K$, we let $H_x = \{A \in H : x \in A\}$, $H_H = H \setminus H_x$ ($x \in V$) and
$\Delta_H = \max\{|H_x| : x \in V\}$. As usual $|H_x|$ is the degree of $x$ in $H$. We use $M$ and $C$ as above and set $N = \binom{n-1}{k}$. For $F \in C$ we set $a_F = M - \Delta_F$ and
$e(F) = |\{\{A, B\} : A, B \in F, A \cap B = \emptyset\}|$ (the number of Kneser edges in $F$).

In view of [7] we may assume

$$k > 6c.$$  \hfill (2)

We also assume henceforth that

$$p > \begin{cases} 
1 - \varepsilon & \text{if } c = 1, \\
Cp_0(n, k) & \text{if } c \geq 2
\end{cases} \hfill (3)$$

for suitable fixed $C$ and $\varepsilon > 0$ (namely, ones that support our arguments) and want to show that then

$$\Pr(\text{some } F \in C \text{ is independent in } K_p(n, k)) = o(1).$$

Perhaps surprisingly, this is given by a straight union bound; that is, there are lower bounds on the sizes of the various $e(F)$’s that imply (with $F$ running over $C$)

$$\sum (1 - p)^{e(F)} = o(1). \hfill (4)$$

This contrasts with (e.g.) [10], where a naive union bound gives nothing.

The rest of our discussion is devoted to the proof of (4), and we assume from now on that $F \in C$. Notice that we always have

$$a_F/N \leq k/n \hfill (5)$$

(since the trivial $\Delta_F \geq k|F|/n = kM/n$ gives $a_F \leq (1 - k/n)M = kN/n$).

The next assertion is the main point.

Lemma 3. There is a fixed $\vartheta > 0$ such that for any $F \in C$,

$$e(F) > \vartheta k^{-1} \binom{n-k-1}{k-1} a_F \log(N/a_F). \hfill (6)$$
We first observe that this easily gives \((4)\). Noting that (for any \(a\)) the number of \(F\)'s with \(a_F = a\) is at most \((M_a)(N_a)\) (choose a maximum degree vertex \(x\) of \(F\) and then the \(a\)-sets \(K_x \setminus F \subseteq K_x\) and \(F \subseteq K_x\)), we find that, with \(\vartheta\) as in Lemma 3, the sum in \((4)\) is then less than

\[
\sum_{0 < a \leq kN/n} \left\{ \binom{M_a}{a} \binom{N_a}{a} \exp\left[ -\vartheta k^{-1}(n-k-1) a \log(N/a) \right] : 0 < a \leq kN/n \right\},
\]

\((7)\)

where \(\xi = \log(1/\varepsilon)\) if \(n = 2k + 1\) and otherwise \(\xi = p\). We may bound the summand using \((M_a)(N_a) < \exp[2 \vartheta \log(eN/a)]\) and

\[
\xi \vartheta k^{-1}(n-k-1) \geq \left\{ \begin{array}{ll}
\vartheta \log(1/\varepsilon) & \text{if } n = 2k + 1, \\
C \vartheta k^{-1} \log(n(n-1)) & \text{otherwise,}
\end{array} \right.
\]

and the expression in \((7)\) is then easily seen to be small if (say) \(\varepsilon < e^{-5/\vartheta}\) or \(C > 4/\vartheta\) (for \(n = 2k + 1\) and \(n \geq 2k + 2\) respectively).

The proof of Lemma 3 divides into three regimes, depending on \(a_F\). The first of these—\(a_F\) not too small—is handled as in [7], from which we recall only what we need (see their Theorem 1.2):

**Theorem 4.** There is a fixed \(K\) such that for any \(2 \leq k < n/2\): if \(F \in C\) satisfies \(a_F > K \zeta(n) M\) with \(\zeta \leq \frac{c}{(10K)^2 n}\), then \(e(F) > \zeta M(n-k-1)\).

It will be convenient to assume (as we may) that \(K \geq 1\). Theorem 4 gives \((6)\) for any \(F\) satisfying \(a_F > M/(100K)\) (with \(\vartheta\) something like \(.01K^{-2}\)), so we assume from now on that this is not the case.

For smaller values we need to say a little about graphs belonging to the “Johnson scheme” (e.g. [12]). For positive integers \(k \leq m\) we use \(J_i(m, k)\) for the graph on \(V_{m,k} := \left[ \begin{array}{c} m \\ k \end{array} \right] \) with \(A, B\) adjacent \((A \sim_i B)\) iff \(|A \Delta B| = 2i\). Here we take \(m = n - 1\) and will be interested in \(i \in \{1, c\}\). Uniform measure on \(V_{m,k}\) will be denoted \(\mu_k\).

We use \(\beta_i(A)\) for the size of the edge boundary of \(A \subseteq V_{m,k}\) in \(J_i(m, k)\); that is,

\[
\beta_i(A) = |\{\{A, A'\} : A \in A, A' \in (V_{m,k} \setminus A), A \sim_i A'\}|.
\]

The following lower bounds on \(\beta_c\) and \(\beta_1\) will suffice for our purposes.
For $\beta_c$ we use a standard version of the eigenvalue-expansion connection due to Alon and Milman [2] (see e.g. [1, Theorem 9.2.1]), which (here) says that for any $A \subseteq V_{m,k}$,

$$\beta_c(A) \geq \lambda|A|(1 - \mu_k(A)),$$

with $\lambda$ the smallest positive eigenvalue of the Laplacian of $J_c(m,k)$ (the matrix $DI_N - A$, where $D = \binom{k}{c}(\binom{m-k}{c}$ and $A$ are the degree and adjacency matrix of $J_c(m,k)$). We assert that (assuming (2))

$$\lambda = \frac{m}{k} \binom{k}{c} \left( \frac{m - k - 1}{c - 1} \right).$$

(9)

Proof. (This ought to be known, but we couldn’t find a reference.) The eigenvalues of $A$ are (again, see e.g. [12])

$$\lambda_j := \sum_{i=0}^c (-1)^i S_i^j, \quad j = 0, \ldots, k,$$

where $S_i^j := \binom{i}{j} \binom{k-j}{c-i} \binom{m-k-j}{c-i}$. In particular, $\lambda_0 = \binom{k}{c}(\binom{m-k}{c}$,

$$\lambda_1 = \binom{k-1}{c} \binom{m-k-1}{c-1} - \binom{k-1}{c} \binom{m-k-1}{c-1} = \binom{k-1}{c} \binom{m-k-1}{c} \frac{km-k^2-cm}{km-k^2-cm+cm}$$

(10)

and $\lambda_0 - \lambda_1 = \lambda$ (the value in (9)), so we just need to show that $\lambda_j \leq \lambda_1$ for $j \geq 2$. In fact it is enough to show that

$$S_i^j \leq \lambda_1 \text{ whenever } j \geq 2,$$

(11)

since log-concavity of the sequences $\left( S_i^j \right)_i$ implies log-concavity of $(S_i^j)_i$ and thus $\lambda_j \leq \max_i S_i^j$.

Routine manipulations (using the expression for $\lambda_1$ in (10)) give

$$S_i^j / \lambda_1 = \frac{km-k^2}{km-k^2-cm} \binom{k-1}{j-i} \binom{m-k}{j-i} \binom{i}{j-i} \frac{1}{\binom{i}{j}} \leq \frac{1}{\binom{i}{j}} \frac{km-k^2}{km-k^2-cm}.$$

For $0 < i < j$, the r.h.s. is less than 1 since $km-k^2 < 2(km-k^2-cm)$, as follows easily from (2). On the other hand, it is easy to see (using (2)) that each of $S_0^j$ and $S_2^j$ is less than $\lambda_1$, which gives (11) for $i \in \{0, j\}$, since $S_0^j$ and $S_2^j$ are decreasing in $j$. 

\[ \square \]
For $\beta_1$ we use an instance of a result of Lee and Yau [11] (estimating the log-Sobolev constant for $J_1(m,k)$): there is a fixed $\gamma > 0$ such that, for any $k$ as in (2) and $A \subseteq \binom{[m]}{k}$,

$$\beta_1(A) > \gamma m |A| \log(1/\mu_k(A)). \quad (12)$$

Proof of Lemma 3. As already noted, Theorem 4 gives Lemma 3 when $a_F > M/(100D)$, so we assume this is not the case.

We assume (w.l.o.g.) that $x = n$ is a maximum degree vertex of $F$ and set $A = F_x$ and $B = \{V \setminus T : T \in K_x \setminus F\}$ (so $|A| = |B| = a_F$).

As above we take $m = n - 1$. The rest of our discussion takes place in the universe $V \setminus \{x\} = [m]$. We use $\Gamma_l$ for $\binom{[m]}{l}$—thus $A \subseteq \Gamma_k$ (our earlier $V_{m,k}$) and $B \subseteq \Gamma_{k+c}$—and set $\bar{A} = \Gamma_k \setminus A$ and $\bar{B} = \Gamma_{k+c} \setminus B$.

For $S \subseteq \Gamma_k$ and $T \subseteq \Gamma_{k+c}$, set

$$\Lambda(S, T) = |\{(A, B) \in S \times T : A \subseteq B\}|.$$

Notice that

$$e(F) = \Lambda(A, \bar{B}) + e(A) \geq \Lambda(A, \bar{B}). \quad (13)$$

We next observe that lower bounds on the $\beta$'s imply lower bounds on the quantities $\Lambda(A, \bar{B})$:

**Proposition 5.** For any $F \in C$,

$$\Lambda(A, \bar{B}) \geq \max \left\{ \left(2^{k-1} \beta_c(A), \frac{1}{2^{c-1}} \binom{k+c-2}{c-1} \beta_1(A) \right) - \left(\binom{k+c-1}{c-1}\right)|A|/2 \right\}. \quad (14)$$

Of course in view of (13) this gives the same lower bound on $e(F)$.

**Proof.** The combination of

$$\Lambda(A, \bar{B}) + \Lambda(A, \bar{A}) = \Lambda(A, \Gamma_{k+c}) = \binom{k+c-1}{c} |A|$$

and

$$\Lambda(A, B) + \Lambda(\bar{A}, B) = \Lambda(\bar{B}, B) = \binom{k+c}{c} |B| = \binom{k+c}{c} |A|$$

gives

$$\Lambda(\bar{A}, B) = \Lambda(A, \bar{B}) + \binom{k+c-1}{c-1} |A|. \quad (15)$$

For the second bound in (14) we work in the ("Johnson") graph $J_1(m,k)$. Write $\Phi$ for the number of triples $(A, B, A') \in A \times \Gamma_{k+c} \times A$ with $A \sim A'$.
and \( A \cup A' \subseteq B \). Since each relevant pair \((A, A')\) admits exactly \( \binom{k+c-2}{c-1} \) choices of \( B \), we have
\[
\Phi = \beta_1(A) \binom{k+c-2}{c-1}.
\]
(16)

On the other hand, for each of the above triples, either \((A, B)\) is one of the pairs counted by \( \Lambda(A, \bar{B}) \) or \((A', B)\) is one of the pairs counted by \( \Lambda(\bar{A}, B) \) (and not both). In the first case the number of choices of \( A' \) is at most the number of neighbors of \( A \) contained in \( B \), namely \( ck \), and similarly in the second case. This with (15) gives
\[
\Phi \leq (\Lambda(A, \bar{B}) + \Lambda(\bar{A}, B)) ck = ck(2\Lambda(A, \bar{B}) + \binom{k+c-1}{c-1}|A|),
\]
and then combining with (16) yields the stated bound.

The argument for the first bound is similar and we just indicate the changes. We work in \( J_c(m, k) \) and consider triples as above but with \( A \sim c \bar{A} \) (so \( B = A \cup A' \)). The number of triples, which is now just \( \beta c(A) \), is bounded above by
\[
(\binom{k}{c} + \binom{k}{c}) \binom{k+c-1}{c-1} = (\binom{k}{c} + \binom{k+c-1}{c-1}|A|)
\]
(\( \binom{k}{c} \) being the number of neighbors—now in \( J_c(m, k) \)—of \( A \) contained in \( B \) when \( A \in \Gamma_k, B \in \Gamma_{k+c} \) and \( A \subseteq B \)), and the desired bound follows.

Finally, combining (14) with (13) and our earlier bounds on the \( \beta \)'s (see (8)-(12)) yields (with \( \gamma \) as in (12))
\[
e(F) \geq \frac{|A|}{2} \left( \binom{k+c-2}{c-1} \max\{ (1 - \frac{m}{k} \mu_k(A)), \frac{m \log \frac{1}{\mu_k(A)}}{ck} - \frac{k+c-1}{ck} \} \right).
\]
(17)

(Replacing \( \beta_c(A) \) in (14) by the lower bound provided by (8) and (9) gives
\[
e(F) \geq (2\binom{k}{c})^{-1} \frac{m}{k} \binom{k+c-2}{c-1}|A|(1 - \mu_k(A)) - \binom{k+c-1}{c-1} \frac{|A|}{2}
\]
\[
= \left[ \frac{|A|}{2} \left( \frac{m}{k} \binom{m-1}{c-1}(1 - \mu_k(A)) - \binom{k+c-1}{c-1} \right) \right]
\]
\[
= \left[ \frac{|A|}{2} \binom{k+c-2}{c-1} \left[ 1 - \frac{m}{k} \mu_k(A) \right] \right],
\]
and replacing \( \beta_1 \) by the lower bound from (12) yields
\[
e(F) \geq \frac{1}{2ck} \binom{k+c-2}{c-1} \gamma m|A| \log \frac{1}{\mu_k(A)} - \binom{k+c-1}{c-1} \frac{|A|}{2},
\]
which is easily seen to be equal to the second bound in (17).
It only remains to observe that this does what we want, namely that (for suitable $\vartheta$) the expression in (17) is at least as large as the bound in (6), or, equivalently, that the max in (17) is at least

$$\frac{2(k+c-1)}{ck}\vartheta \log(\frac{N/a_F}{c/k}) < \frac{4}{c}\vartheta \log(1/\mu_k(A));$$

we assert that this is true provided $\vartheta < \frac{\gamma}{5}$. If $\log(1/\mu_k(A)) \leq \frac{c}{\gamma}$, then the r.h.s. of (18) is less than the first term in the max (which is essentially 1 since $\mu_k(A) = |A|/N < |A|/M$, which we are assuming is less than .01$D^{-1}$; see following Theorem 4).

If, on the other hand, $\log(1/\mu_k(A)) > \frac{c}{\gamma}$, then the second term in the max is at least

$$\frac{2}{c} \frac{2k+c-1}{k} \log \frac{1}{\mu_k(A)} - \frac{2}{c} \frac{k+c-1}{k} \log \frac{1}{\mu_k(A)} = \frac{\gamma}{c} \log \frac{1}{\mu_k(A)},$$

which is again greater than the r.h.s. of (18).

For $n = 2k+1$ we could avoid the machinery used above for intermediate values of $|A|$ (namely (8), (9) and the first bound in (14)) by choosing $\zeta$ in Theorem 4 to handle $\log(1/\mu_k(A)) \leq \frac{c}{\gamma}$ (and adjusting $\vartheta$ accordingly).

References

[1] N. Alon and J.H. Spencer, The Probabilistic Method, Wiley, New York, 2008.

[2] N. Alon and V.D. Milman, Eigenvalues, expanders and superconcentrators, pp. 320-322 in Proc. 25th FOCS, IEEE, New York, 1984.

[3] J. Balogh, T. Bohman and D. Mubayi, Erdős-Ko-Rado in random hypergraphs. Combin. Probab. Comput. 18 (2009), 629-646.

[4] J. Balogh, B. Bollobás and B. Narayanan, Transference for the Erdős-Ko-Rado Theorem, submitted.

[5] J. Balogh, S. Das, M. Delcourt, H. Liu, and M. Sharifzadeh, The typical structure of intersecting families of discrete structures, arXiv:1408.2559 [math.CO].

[6] B. Bollobás, B. Narayanan and A. Raigorodskii, On the stability of the Erdős-Ko-Rado theorem, arXiv:1408.1288 [math.CO]
[7] S. Das and T. Tran, A simple removal lemma for large nearly-intersecting families. [arXiv:1412.7885] [math.CO]

[8] P. Erdős, C. Ko and R. Rado, Intersection theorems for systems of finite sets, *Quart. J. Math. Oxford Ser. (2)* 12 (1961), 313-320.

[9] A. Hamm and J. Kahn, On Erdős-Ko-Rado for random hypergraphs I, submitted. [arXiv:1412.5085]

[10] A. Hamm and J. Kahn, On Erdős-Ko-Rado for random hypergraphs II, submitted. [arXiv:1406.5793]

[11] T.-Y. Lee and H.-T. Yau, Logarithmic Sobolev inequality for some models of random walks, *Ann. Probab.* 26 (1998), 1855-1873.

[12] F.J. MacWilliams and N.J.A. Sloane, *The Theory of Error-correcting Codes*, vol. 2, North-Holland, Amsterdam, 1977.

[13] V. Rödl and M. Schacht, Extremal results in random graphs, Erdős Centennial, vol. 25 series (2013) *Bolyai Soc. Math. Stud.*, 535-583.

Department of Mathematics
Rutgers University
Piscataway NJ 08854
prd41@math.rutgers.edu
jkahn@math.rutgers.edu