Combinatorics, Probability and Computing (2020), **29**, pp. 163–189
doi:10.1017/S0963548319000294

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Edge-statistics on large graphs

Noga Alon[1,2,†], Dan Hefetz[3,‡], Michael Krivelevich[4,§] and Mykhaylo Tyomkyn[4,5,*,¶]

[1]Department of Mathematics, Princeton University, Princeton, New Jersey, USA, [2]School of Mathematics and Computer Science, Tel Aviv University, Tel Aviv, Israel, [3]Department of Computer Science, Ariel University, Ariel 40700, Israel, [4]School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997801, Israel and [5]Department of Mathematics, California Institute of Technology, Pasadena, CA 91125

*Corresponding author. Email: tyomkyn@caltech.edu

(Received 17 May 2018; revised 24 January 2019; first published online 14 November 2019)

Abstract

The inducibility of a graph H measures the maximum number of induced copies of H a large graph G can have. Generalizing this notion, we study how many induced subgraphs of fixed order k and size ℓ a large graph G on n vertices can have. Clearly, this number is $\binom{n}{k}$ for every n, k and $\ell \in \{0, \binom{k}{2}\}$. We conjecture that for every n, k and $0 < \ell < \binom{k}{2}$ this number is at most $(1/e + o_k(1))\binom{n}{k}$. If true, this would be tight for $\ell \in \{1, k-1\}$.

In support of our 'Edge-statistics Conjecture', we prove that the corresponding density is bounded away from 1 by an absolute constant. Furthermore, for various ranges of the values of ℓ we establish stronger bounds. In particular, we prove that for 'almost all' pairs (k, ℓ) only a polynomially small fraction of the k-subsets of $V(G)$ have exactly ℓ edges, and prove an upper bound of $(1/2 + o_k(1))\binom{n}{k}$ for $\ell = 1$.

Our proof methods involve probabilistic tools, such as anti-concentration results relying on fourth moment estimates and Brun's sieve, as well as graph-theoretic and combinatorial arguments such as Zykov's symmetrization, Sperner's theorem and various counting techniques.

2010 MSC Codes: Primary 05C35; Secondary 05D40

1. Introduction

Let k be a positive integer and let G be a finite graph of order at least k. Let $A = A_{G,k}$ be chosen uniformly at random from all subsets of $V(G)$ of size k and let $X_{G,k} = e(G[A])$. That is, $X_{G,k}$ is the random variable counting the number of edges of G with both endpoints in A. Naturally, the above quantities can also be interpreted as densities rather than probabilities, and we shall frequently switch between these two perspectives.

Given integers $n \geqslant k$ and $0 \leqslant \ell \leqslant \binom{k}{2}$, let $I(n, k, \ell) = \max\{\mathbb{P}(X_{G,k} = \ell) : |G| = n\}$, that is, $I(n, k, \ell)$ is the maximum density of induced subgraphs with k vertices and ℓ edges, taken over all graphs of order n. A standard averaging argument shows that $I(n, k, \ell)$ is a monotone decreasing function of n. Consequently, we define $\mathrm{ind}(k, \ell) := \lim_{n\to\infty} I(n, k, \ell)$ to be the *edge-inducibility*

[†]Research supported in part by NSF grant DMS-1855464, ISF grant 281/17, BSF grant 2018267 and the Simons Foundation.

[‡]The research of Dan Hefetz is supported by ISF grant 822/18.

[§]Partially supported by USA-Israel BSF grants 2014361 and 2018267, and by ISF grant 1261/17.

[¶]The research of Mykhaylo Tyomkyn is supported in part by ERC Starting Grant 633509.

© Cambridge University Press 2019

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

of k and ℓ. While this quantity is trivially 1 for $\ell \in \{0, \binom{k}{2}\}$ (simply take G to be a large empty or complete graph, respectively), it is natural to ask how large $\mathrm{ind}(k,\ell)$ can be for $0 < \ell < \binom{k}{2}$.

This question is closely related to the problem of determining the inducibilities of fixed graphs, a concept which was introduced in 1975 by Pippenger and Golumbic [16]. For a graph H, let $D_H(G)$ denote the number of induced subgraphs of G that are isomorphic to H, and let $I_H(n) = \max\{D_H(G) : |G| = n\}$. Again, the sequence $\{I_H(n)/\binom{n}{|H|}\}_{n=|H|}^{\infty}$ is monotone decreasing and thus converges to a limit $\mathrm{ind}(H)$, the *inducibility* of H. Recently there has been a surge of interest in this area (see *e.g.* [4], [10], [17] and [11]).

Observe that both types of inducibility are invariant under taking complements, *i.e.* $\mathrm{ind}(k,\ell) = \mathrm{ind}\,(k, \binom{k}{2} - \ell)$ and $\mathrm{ind}(H) = \mathrm{ind}(\overline{H})$. Note also that $\mathrm{ind}(H) \leqslant \mathrm{ind}(|H|, e(H))$. Moreover, if $|H| = k$ and $e(H) \in \{1, \binom{k}{2} - 1\}$, then $\mathrm{ind}(H) = \mathrm{ind}(k, e(H))$, as H is the unique (up to isomorphism) graph with k vertices and $e(H)$ edges.

Consider a random graph $G \sim G(n,p)$, where $p = \binom{k}{2}^{-1}$. A straightforward calculation shows that the expected value of the number of k-subsets of $V(G)$ which span precisely one edge is about $1/e$. This implies that $\mathrm{ind}(k,1) \geqslant 1/e + o_k(1)$ (as the $o_k(1)$ notation suggests, we will often think of k as an asymptotic quantity and, in particular, we will assume k to be sufficiently large wherever needed). In fact, we will see later that there are many constructions which achieve $1/e + o_k(1)$ as a lower bound for $\mathrm{ind}(k,1)$. Another construction, achieving the same asymptotic value for $\ell = k-1$, is the complete bipartite graph with the smaller part of size n/k, so that $\mathrm{ind}(k,k-1) \geqslant \mathrm{ind}(K_{1,k-1}) \geqslant 1/e + o_k(1)$. In fact, it is known [6] that $\mathrm{ind}(K_{1,k-1}) = 1/e + o_k(1)$. Note that the $o_k(1)$ term is necessary. For example, counting cherries in $K_{n/2,n/2}$ shows that $\mathrm{ind}(3,2) = \mathrm{ind}(K_{1,2}) \geqslant 3/4$ (in fact, it follows from Goodman's theorem that $\mathrm{ind}(3,1) = \mathrm{ind}(3,2) = 3/4$). Motivated by the aforementioned constructions (as well as some additional data), we conjecture that the lower bound of $1/e$ is asymptotically tight.

Conjecture 1.1. (Edge-statistics Conjecture). *For every $\varepsilon > 0$ there exists $k_0 = k_0(\varepsilon)$ such that, for all integers $k > k_0$ and $0 < \ell < \binom{k}{2}$, we have* $\mathrm{ind}(k,\ell) \leqslant 1/e + \varepsilon$.

For graph-inducibilities we make an analogous conjecture, which would be implied by the Edge-statistics Conjecture.

Conjecture 1.2. (Large Inducibility Conjecture).

$$\limsup\{\mathrm{ind}(H) : H \notin \{K_{|H|}, \overline{K}_{|H|}\}\} = 1/e.$$

Our first theorem in this paper constitutes a first step towards proving Conjecture 1.1. It asserts that $\mathrm{ind}(k,\ell)$ is bounded away from 1 by an absolute constant for every k and $0 < \ell < \binom{k}{2}$.

Theorem 1.3. *There exists an $\varepsilon > 0$ such that, for all positive integers k and ℓ which satisfy $0 < \ell < \binom{k}{2}$, we have*

$$\mathrm{ind}(k,\ell) < 1 - \varepsilon.$$

For clarity of presentation, we do not give explicit bounds on ε and refer to Section 6 for a discussion.

Note that it is not hard to prove that, for every positive integer k, we have $\mathrm{ind}(k,\ell) = 1$ if and only if $\ell \in \{0, \binom{k}{2}\}$. Indeed, if $0 < \ell < \binom{k}{2}$, then $\mathrm{ind}(k,\ell) < 1 - 4^{-k^2}$ is an easy consequence of Ramsey's theorem and the aforementioned monotonicity of $I(n,k,\ell)$. With a bit more effort, this bound can be improved to $1 - k^{-2}$. On the other hand, we do not see a simple argument that would upper-bound $\mathrm{ind}(k,\ell)$ away from 1 by an absolute constant as in Theorem 1.3. Note also that the related problem of *minimizing* graph-inducibilities has been extensively studied. In particular, Pippenger and Golumbic [16] showed that the inducibility of any k-vertex graph is at least

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

Table 1. Bounds on $\mathrm{ind}(k, \ell)$

$\ell = \ell(k)$	1	const.	$[\omega(1), O(k)]$	$[\omega(k), o(k^2)]$	$[\Omega(k^2), k^2/4]$
$\mathrm{ind}(k, \ell) \leqslant$	1/2	3/4	$1 - \varepsilon$	1/2	$O(k^{-0.1})$

$(1 + o_k(1))k!/k^k$. It follows that $\mathrm{ind}(H) > 0$ for every graph H and thus $\mathrm{ind}(k, \ell) > 0$ for every k and ℓ. We refer the reader to Section 6 for further discussion.

For various ranges of values of ℓ (viewed as a function of k) we establish much better upper bounds than the one stated in Theorem 1.3. First, for every ℓ satisfying $\min\{\ell, \binom{k}{2} - \ell\} = \omega(k)$, we prove an upper bound of 1/2.

Proposition 1.4. *For every $\varepsilon > 0$ there exist $C(\varepsilon) > 0$ and $k_0(\varepsilon) > 0$ such that the following holds. Let k and ℓ be integers satisfying $k \geqslant k_0$ and $Ck \leqslant \ell \leqslant \binom{k}{2} - Ck$. Then*

$$\mathrm{ind}(k, \ell) \leqslant \frac{1}{2} + \varepsilon.$$

Next, we prove Conjecture 1.1 'almost everywhere'. In fact, we prove a much stronger statement, namely that for every ℓ satisfying $\min\{\ell, \binom{k}{2} - \ell\} = \Omega(k^2)$ the quantity $\mathrm{ind}(k, \ell)$ is actually polynomially small in k – the right asymptotic behaviour, as can be seen by considering the random graph $G(n, \ell/\binom{k}{2})$, which gives $\mathrm{ind}(k, \ell) = \Omega(k^{-1})$.

Theorem 1.5. *For every positive integer k and ℓ such that $\min\{\ell, k^2/2 - \ell\} = \Omega(k^2)$, we have $\mathrm{ind}(k, \ell) = O(k^{-0.1})$.*

Lastly, we consider the case when ℓ is fixed (*i.e.* does not depend on k). Here we prove an upper bound of 3/4. In the interesting subcase $\ell = 1$, which corresponds to the inducibility of the one-edge graph (equivalently, of K_k^-, the complete graph with one edge removed), we prove an even better bound of 1/2.

Theorem 1.6. *For every fixed positive integer ℓ we have*

$$\mathrm{ind}(k, \ell) \leqslant \frac{3}{4} + o_k(1).$$

Moreover, for $\ell = 1$ we have

$$\mathrm{ind}(k, 1) \leqslant \frac{1}{2} + o_k(1).$$

Our results are summarized in Table 1. For various ranges of $\ell \leqslant k^2/4$, it states the best known upper bound on $\mathrm{ind}(k, \ell)$. Note that for $\ell \geqslant k^2/4$ the table can be extended symmetrically.

1.1 Notation

Throughout this paper, log stands for the natural logarithm, unless explicitly stated otherwise. For positive integers $n \geqslant k$ we let $(n)_k$ denote the *falling factorial* $\prod_{i=0}^{k-1}(n - i)$. The *symmetric difference* of two sets A and B, denoted by $A \triangle B$, is $(A \setminus B) \cup (B \setminus A)$.

Our graph-theoretic notation is standard and follows that of [5]. In particular, we use the following. For a graph G, let $V(G)$ and $E(G)$ denote its sets of vertices and edges respectively, and let $|G| = |V(G)|$ and $e(G) = |E(G)|$. The *complement* of G, denoted by $\overline{G}$, is the graph with vertex set $V(G)$ and edge set $\binom{V(G)}{2} \setminus E(G)$. For a set $S \subseteq V(G)$, let $G[S]$ denote the graph with vertex set S and edge set $\{uv \in E(G) : u, v \in S\}$. For disjoint sets $S, T \subseteq V(G)$, let $G[S, T]$ denote the bipartite graph with parts S and T and edge-set $\{uv \in E(G) : u \in S, v \in T\}$. For a set $S \subseteq V(G)$

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

and a vertex $v \in V(G)$, let $N_G(v, S) = \{u \in S : uv \in E(G)\}$ denote the *neighbourhood* of v in S and let $d_G(v, S) = |N_G(v, S)|$ denote the *degree* of v into S. We abbreviate $N_G(v, V(G))$ under $N_G(v)$ and $d_G(v, V(G))$ under $d_G(v)$; we refer to the former as the *neighbourhood* of v in G and to the latter as the *degree* of v in G. The *maximum degree* of a graph G is $\Delta(G) = \max\{d_G(u) : u \in V(G)\}$. Often, when there is no risk of confusion, we omit the subscript G from the notation above.

The rest of this paper is organized as follows. In Section 2 we establish a number of facts and lemmas which will be useful later on when we upper-bound $\mathrm{ind}(k, \ell)$; we then prove Proposition 1.4. In Sections 3 and 4 we prove Theorems 1.3 and 1.5 respectively. Moving on to the fixed ℓ regime, in Section 5 we establish some additional tools and prove Theorem 1.6. In Section 6 we conclude the paper with several remarks and open problems.

2. Preliminaries

Recall that $A = A_{G,k}$ is the set chosen uniformly at random from all subsets of $V(G)$ of size k and $X_{G,k} = e(G[A])$. To simplify notation we abbreviate $X_{G,k}$ to X whenever there is no risk of confusion. Our first lemma provides a useful global–local criterion for handling edge-inducibilities.

Lemma 2.1. *Let k and ℓ be positive integers satisfying $0 < \ell < \binom{k}{2}$ and let $a = \mathrm{ind}(k, \ell)$. Let n be a sufficiently large integer and let G be a graph on n vertices which attains $I(n, k, \ell)$. Then, for every vertex $v \in V(G)$, we have $\mathbb{P}(X = \ell \mid v \in A) = a + o_n(1)$.*

Proof. The main idea of the proof is the same as in the proof of Lemma 2.4 from [10]. Double-counting yields

$$\begin{aligned} a + o_n(1) &= \mathbb{P}(X = \ell) \\ &= \frac{1}{k} \cdot \sum_{v \in V(G)} \mathbb{P}(X = \ell, v \in A) \\ &= \frac{1}{k} \cdot \sum_{v \in V(G)} \mathbb{P}(X = \ell \mid v \in A) \cdot \mathbb{P}(v \in A) \\ &= \frac{1}{k} \cdot \frac{k}{n} \cdot \sum_{v \in V(G)} \mathbb{P}(X = \ell \mid v \in A) \\ &= \frac{1}{n} \cdot \sum_{v \in V(G)} \mathbb{P}(X = \ell \mid v \in A). \end{aligned}$$

Let v^+ and v^- be the vertices with the largest and the smallest value of $\mathbb{P}(X = \ell \mid v \in A)$, respectively. Let the graph G' be obtained from G by Zykov's symmetrization [18], that is, remove v^- and add a twin copy of v^+ instead (say, the two copies of v^+ are not connected by an edge in G'). Then

$$\begin{aligned} \mathbb{P}(X = \ell) &\geqslant \mathbb{P}(X_{G',k} = \ell) \\ &\geqslant \mathbb{P}(X = \ell) - \mathbb{P}(X = \ell, v^- \in A) + \mathbb{P}(X = \ell, v^+ \in A) \\ &\quad - \mathbb{P}(X = \ell, v^- \in A, v^+ \in A) \\ &= \mathbb{P}(X = \ell) - \frac{k}{n} \cdot \mathbb{P}(X = \ell \mid v^- \in A) + \frac{k}{n} \cdot \mathbb{P}(X = \ell \mid v^+ \in A) \\ &\quad - \frac{k(k-1)}{n(n-1)} \cdot \mathbb{P}(X = \ell \mid v^+, v^- \in A) \\ &= \mathbb{P}(X = \ell) + \frac{k}{n} \cdot (\mathbb{P}(X = \ell \mid v^+ \in A) - \mathbb{P}(X = \ell \mid v^- \in A)) - O(n^{-2}), \end{aligned}$$

where the first inequality follows from our assumption that G maximizes $\mathbb{P}(X_{G,k} = \ell)$.

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

Therefore

$$\mathbb{P}(X = \ell \mid v^+ \in A) - \mathbb{P}(X = \ell \mid v^- \in A) = O(n^{-1}).$$

Hence,

$$\mathbb{P}(X = \ell \mid v^+ \in A) \leqslant a + O(n^{-1}) \quad \text{and} \quad \mathbb{P}(X = \ell \mid v^- \in A) \geqslant a - O(n^{-1}).$$

Since $a > 0$ (as noted in the Introduction), this concludes the proof of the lemma. □

For two vertices $v, w \in V(G)$ we now consider a subset of $V(G) \setminus \{v, w\}$ of size $k-1$, chosen uniformly at random among all such subsets. Our next lemma shows that, assuming that $\mathrm{ind}(k, \ell)$ is large, with a significant probability v and w will have the same degree into this set.

Lemma 2.2. *Let k and ℓ be positive integers satisfying $0 < \ell < \binom{k}{2}$, let $a = \mathrm{ind}(k, \ell)$, and suppose that $a > 1/2$. Let n be a sufficiently large integer and let G be a graph on n vertices which attains $I(n, k, \ell)$. Then, for any two vertices $v, w \in V(G)$, we have*

$$\mathbb{P}(e_G(v, A_{G\setminus\{v,w\},k-1}) = e_G(w, A_{G\setminus\{v,w\},k-1})) > 2a - 1 - o_n(1).$$

Proof. By Lemma 2.1 we have

$$\mathbb{P}(X = \ell \mid v \in A, w \notin A) = \mathbb{P}(X = \ell \mid v \in A) + o(1) = a + o_n(1) \tag{2.1}$$

and, symmetrically,

$$\mathbb{P}(X = \ell \mid w \in A, v \notin A) = a + o_n(1). \tag{2.2}$$

Setting $G'' = G \setminus \{v, w\}$ and $B = A_{G\setminus\{v,w\},k-1}$, identities (2.1) and (2.2) imply that

$$\mathbb{P}(X_{G'',k-1} + e_G(B, v) = \ell) = a + o_n(1) \quad \text{and} \quad \mathbb{P}(X_{G'',k-1} + e_G(B, w) = \ell) = a + o_n(1).$$

Let L_v (L_w, respectively) denote the event $X_{G'',k-1} + e_G(B, v) = \ell$ ($X_{G'',k-1} + e_G(B, w) = \ell$, respectively). Then

$$\mathbb{P}(e_G(B, v) = e_G(B, w)) \geqslant \mathbb{P}(L_v \cap L_w) = \mathbb{P}(L_v) + \mathbb{P}(L_w) - \mathbb{P}(L_v \cup L_w) > 2a - 1 - o_n(1).$$ □

Lemma 2.3. *For every $1/2 < a < 1$ there exists $C = C(a) > 0$ for which the following holds. Suppose that k and ℓ are positive integers with $0 < \ell < \binom{k}{2}$ such that $a = \mathrm{ind}(k, \ell)$. Suppose that n is sufficiently large and that G is a graph on n vertices which attains $I(n, k, \ell)$. Then, for any two vertices $v, w \in V(G)$, we have*

$$|N_G(v) \triangle N_G(w)| < Cn/k. \tag{2.3}$$

Moreover, as $a \to 1$, inequality (2.3) holds with $C \to 0$.

Note that the second part of the statement of Lemma 2.3 (*i.e.* the one referring to $a \to 1$) is only needed to prove Theorem 1.3 by contradiction, and is otherwise vacuous, as it contradicts Theorem 1.3.

Proof. We can assume k to be larger than any given absolute constant, since, in the first statement, the small values of k can be accommodated for by adjusting C, and the second statement implicitly assumes $k \to \infty$, since $a = \mathrm{ind}(k, \ell) < 1$ for any given $0 < \ell < \binom{k}{2}$, and takes only finitely many values when k is bounded.

Fix two arbitrary vertices $v, w \in V(G)$ and let

$$P = N_G(v) \setminus (w \cup N_G(w)), \ R = N_G(w) \setminus (v \cup N_G(v)) \quad \text{and} \quad Q = V(G) \setminus (\{v, w\} \cup P \cup R),$$

Let $B = A_{G\setminus\{v,w\},k-1}$. It then follows by Lemma 2.2 that

$$\mathbb{P}(|B \cap P| = |B \cap R|) > 2a - 1 - o_n(1). \tag{2.4}$$

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

Suppose that $|P \cup R| \geqslant cn/k$ for some absolute constant $c > 0$. Then

$$\mathbb{P}(B \cap (P \cup R) \neq \emptyset) \geqslant 1 - (1 + o(1))(1 - c/k)^{k-1} = 1 - e^{-c} + o_k(1) = \Omega_k(1). \tag{2.5}$$

In particular, if $|P \cup R| = \omega(n/k)$, then $\mathbb{P}(B \cap (P \cup R) \neq \emptyset) = 1 - o_k(1)$. Therefore, using (2.4) we obtain

$$\begin{aligned}\mathbb{P}(|B \cap P| = |B \cap R| \mid B \cap (P \cup R) \neq \emptyset) &= \frac{\mathbb{P}(|B \cap P| = |B \cap R| \wedge B \cap (P \cup R) \neq \emptyset)}{\mathbb{P}(B \cap (P \cup R) \neq \emptyset)} \\ &= \frac{2a - 1 - o(1)}{1 - o_k(1)} \\ &= 2a - 1 - o_k(1). \end{aligned} \tag{2.6}$$

Similarly, if $a = 1 - o_k(1)$ and $|P \cup R| = \Omega(n/k)$, then (2.4) implies that $\mathbb{P}(|B \cap P| = |B \cap R|) = 1 - o_k(1)$. Therefore

$$\begin{aligned}\mathbb{P}(|B \cap P| = |B \cap R| \mid B \cap (P \cup R) \neq \emptyset) &= \frac{\mathbb{P}(|B \cap P| = |B \cap R| \wedge B \cap (P \cup R) \neq \emptyset)}{\mathbb{P}(B \cap (P \cup R) \neq \emptyset)} \\ &= \frac{\mathbb{P}(B \cap (P \cup R) \neq \emptyset) - o_k(1)}{\mathbb{P}(B \cap (P \cup R) \neq \emptyset)} \\ &\overset{(2.5)}{=} 1 - o_k(1). \end{aligned} \tag{2.7}$$

Note that the above argument does not make any use of the graph structure of G. Indeed, the situation at hand can be viewed as an urn model, in which we have a large urn filled with $|P|$ pink balls and $|R|$ red balls, and we draw a fixed but otherwise arbitrary number $1 \leqslant s \leqslant k - 1$ of balls from the urn uniformly at random, without replacement. We would like to upper-bound the probability of drawing equally many pink balls and red balls. To this end, we first prove the following auxiliary claim.

Claim 2.4. *For every integer* $1 \leqslant t \leqslant \lfloor (k-1)/2 \rfloor$ *we have*

$$\mathbb{P}(|B \cap P| = |B \cap R| = t) \leqslant \frac{3}{4} \cdot \mathbb{P}(|B \cap (P \cup R)| = 2t).$$

Moreover, for every $\varepsilon > 0$ there exists $t_0 = t_0(\varepsilon)$ such that for every $t \geqslant t_0$ we have

$$\mathbb{P}(|B \cap P| = |B \cap R| = t) \leqslant \varepsilon \cdot \mathbb{P}(|B \cap (P \cup R)| = 2t).$$

Proof. Fix some $1 \leqslant t \leqslant \lfloor (k-1)/2 \rfloor$. By the log-concavity of the binomial coefficients we have

$$\binom{|P|}{t}\binom{|R|}{t} \leqslant \binom{(|P| + |R|)/2}{t}^2.$$

Moreover, using the fact that t is much smaller than $|P| + |R|$ (as $t < k$ and $|P| + |R| \geqslant cn/k$ by assumption, and thus can be assumed to be sufficiently large), straightforward calculations show that

$$\binom{(|P| + |R|)/2}{t}^2 \leqslant \frac{3}{4}\binom{|P| + |R|}{2t}.$$

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

We conclude that

$$\begin{aligned}\mathbb{P}(|B\cap P|=|B\cap R|=t) &= \left(\binom{|P|}{t}\binom{|R|}{t}\binom{n-2-|P|-|R|}{k-1-2t}\right)\Big/\binom{n-2}{k-1}\\ &\leqslant \frac{3}{4}\cdot\left(\binom{|P|+|R|}{2t}\binom{n-2-|P|-|R|}{k-1-2t}\right)\Big/\binom{n-2}{k-1}\\ &= \frac{3}{4}\cdot\mathbb{P}(|B\cap(P\cup R)|=2t).\end{aligned}$$

The second statement can be proved analogously; we omit the details. □

Coming back to the proof of Lemma 2.3, using Claim 2.4 with $a=1-o_k(1)$ we have

$$\begin{aligned}&\mathbb{P}(|B\cap P|=|B\cap R|>0)\\ &= \sum_{t=1}^{\lfloor(k-1)/2\rfloor}\mathbb{P}(|B\cap P|=|B\cap R|=t)\\ &\leqslant \frac{3}{4}\cdot\sum_{t=1}^{\lfloor(k-1)/2\rfloor}\mathbb{P}(|B\cap(P\cup R)|=2t)\\ &\leqslant \frac{3}{4}\cdot\mathbb{P}(|B\cap(P\cup R)|>0).\end{aligned}$$

Therefore

$$\mathbb{P}(|B\cap P|=|B\cap R|\mid B\cap(P\cup R)\neq\emptyset)\leqslant\frac{3}{4},$$

contrary to (2.7). Similarly, for any $a>1/2$, if $|P\cup R|=\omega(n/k)$, we obtain that for some $t=t(a)$ we have

$$\mathbb{P}(|B\cap P|=|B\cap R|\mid|B\cap(P\cup R)|>t)<\frac{1}{2}\cdot(2a-1),$$

contrary to (2.6). This concludes the proof of Lemma 2.3. □

Under closer inspection, Lemma 2.3 has the following immediate consequence.

Lemma 2.5. *For every $1/2<a<1$ there exists $C=C(a)>0$ for which the following holds. Suppose that k and ℓ are positive integers with $0<\ell<\binom{k}{2}$ such that $a=\mathrm{ind}(k,\ell)$. Suppose that n is sufficiently large and that G is a graph on n vertices which attains $I(n,k,\ell)$. Suppose that $e(G)\leqslant\binom{n}{2}/2$. Then $\Delta(G)<Cn/k$. Moreover, as $a\to1$, the above holds with $C\to0$.*

Proof. We prove the first statement; the second can be proved analogously. It follows from Lemma 2.3 that $|d(v)-d(w)|=O(n/k)$ holds for any $v,w\in V(G)$. Let v be a vertex of minimum degree in G. Put $U:=N_G(v)$ and $W:=N_{\overline{G}}(v)$. Observe that $|U|\leqslant|W|$ since we assumed that $d(v)=\delta(G)$ and $e(G)\leqslant\binom{n}{2}/2$. Suppose for a contradiction that $|U|=\omega(n/k)$. We double-count the edges of the bipartite graph $G[U,W]$. Applying Lemma 2.3 to v and u for every $u\in U$, we derive that $d_{G[U,W]}(u)=O(n/k)$. In particular,

$$e(G[U,W])=O\left(|U|\cdot\frac{n}{k}\right)=O\left(|W|\cdot\frac{n}{k}\right).$$

On the other hand, applying Lemma 2.3 to v and w for every $w\in W$ yields

$$d_{G[U,W]}(w)\geqslant|U|-O\left(\frac{n}{k}\right)\geqslant\omega\left(\frac{n}{k}\right)-O\left(\frac{n}{k}\right)=\omega\left(\frac{n}{k}\right).$$

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

Therefore

$$|W| \cdot \omega\left(\frac{n}{k}\right) \leqslant e(G[U,W]) = O\left(|W| \cdot \frac{n}{k}\right),$$

which is clearly a contradiction, and thus $\delta(G) = O(n/k)$. Since, moreover, $|d(v) - d(w)| = O(n/k)$ holds for any $v, w \in V(G)$ by Lemma 2.3, we conclude that $\Delta(G) = O(n/k)$, as claimed. □

From now on, we denote $e(G)$ by m. For every $e \in E(G)$, let X_e be the indicator random variable for the event $e \in E(G[A])$, that is, $X_e = 1$ if both endpoints of e are in A and $X_e = 0$ otherwise. Observe that $X = \sum_{e \in E(G)} X_e$. Putting $\mu = \mathbb{E}(X)$, we have

$$\mu = \sum_{e \in E(G)} \mathbb{E}(X_e) = m \cdot \frac{k(k-1)}{n(n-1)} = m \cdot \frac{(k)_2}{n^2} \cdot (1 + O(1/n)). \tag{2.8}$$

This has the following consequence.

Lemma 2.6. *Let k and ℓ be positive integers satisfying $0 < \ell < \binom{k}{2}$ and let $a = \mathrm{ind}(k,\ell)$. Let n be a sufficiently large integer and let G be a graph with n vertices and m edges which attains $I(n,k,\ell)$. Then*

$$m \geqslant (1 - o_k(1))a\ell\frac{n^2}{k^2}.$$

In particular, if $a = 1 - o_k(1)$, then

$$m \geqslant (1 - o_k(1))\frac{n^2}{k^2}.$$

Proof. Suppose for a contradiction that $m < (1-\varepsilon)a\ell(n^2/k^2)$ for some constant $\varepsilon > 0$. It then follows by (2.8) that $\mu < (1 - \varepsilon/2)a\ell$. On the other hand, since $X \geqslant 0$ and by the choice of G we have $\mu \geqslant \ell \cdot \mathbb{P}(X = \ell) = (1 - o(1))a\ell$, a contradiction. □

Combining the above facts, we can immediately prove Proposition 1.4.

Proof of Proposition 1.4. Suppose for a contradiction that there exist $\varepsilon > 0$ and an integer $\ell = \ell(k)$ such that $Ck \leqslant \ell \leqslant \binom{k}{2} - Ck$ for some large $C > 0$, and $\mathrm{ind}(k,\ell) = a > 1/2 + \varepsilon$. Let G be a graph attaining $I(n,k,\ell)$, where n is sufficiently large. By symmetry we may assume that $e(G) \leqslant \binom{n}{2}/2$. Then, on the one hand, by Lemma 2.5 we have $\Delta(G) = O(n/k)$ entailing $e(G) = O(n^2/k)$. On the other hand, Lemma 2.6 implies that

$$e(G) \geqslant (1/2 - o(1))\ell \cdot \frac{n^2}{k^2} \geqslant \frac{C}{3} \cdot \frac{n^2}{k},$$

which is a contradiction for large enough C. We conclude that $a \leqslant 1/2 + o(1)$, as claimed. □

3. Proof of Theorem 1.3

Our goal in this section is to prove Theorem 1.3. Since, as noted in the Introduction, $\mathrm{ind}(k,\ell)$ is never identically 1 (assuming $0 < \ell < \binom{k}{2}$), it suffices to prove the theorem for large values of k. Hence, suppose now, for a contradiction, that the assertion of Theorem 1.3 is false for arbitrarily large values of k. That is, for every $\varepsilon > 0$ there exist arbitrarily large values of k such that, for some $0 < \ell < \binom{k}{2}$, there will be arbitrarily large values of n and graphs G on n vertices for which $\mathbb{P}(X_{G,k} = \ell) > 1 - \varepsilon$. We may assume that G maximizes $\mathbb{P}(X_{G,k} = \ell)$ over all n-vertex graphs and, by symmetry, that $e(G) \leqslant \binom{n}{2}/2$.

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

We would like to calculate the variance of $X_{G,k}$. Using Lemma 2.5 and our assumption that k is sufficiently large, we obtain

$$\begin{aligned}\mathbb{E}(X^2) &= \mathbb{E}\left(\left(\sum_{e\in E(G)} X_e\right)^2\right)\\ &= \mu + \sum_{\substack{(e,f)\in E(G)^2\\ e\cap f=\emptyset}} \mathbb{E}(X_eX_f) + \sum_{\substack{(e,f)\in E(G)^2\\ |e\cap f|=1}} \mathbb{E}(X_eX_f)\\ &= \mu + m^2\cdot\frac{(k)_4}{n^4}(1+O_k(1/n)) + S\cdot\frac{(k)_3}{n^3}(1+O_k(1/n)),\end{aligned}$$

where $S=\sum_{v\in V(G)} d(v)^2$. Therefore,

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}(X^2)-\mu^2\\ &= \left(m\cdot\frac{(k)_2}{n^2} - m^2\cdot\frac{(k)_2\cdot(k)_2}{n^4} + m^2\cdot\frac{(k)_4}{n^4} + S\cdot\frac{(k)_3}{n^3}\right)(1+O_k(1/n))\\ &= \left(m\cdot\frac{(k)_2}{n^2} + S\cdot\frac{(k)_3}{n^3} + m^2\cdot\frac{k(k-1)}{n^4}\cdot(k^2-5k+6-k^2+k)\right)(1+O_k(1/n))\\ &= \left(m\cdot\frac{k^2}{n^2} + S\cdot\frac{k^3}{n^3} - m^2\cdot\frac{4k^3}{n^4}\right)(1+o(1)).\end{aligned}$$

Since, by Lemma 2.5, we have $m=o(n^2/k)$ and consequently $4m^2k^3/n^4 = o(mk^2/n^2)$, we can write

$$\mathrm{Var}(X) = (1+o(1))\left(m\cdot\frac{k^2}{n^2} + S\cdot\frac{k^3}{n^3}\right). \tag{3.1}$$

We shall need the following anti-concentration inequality from [2], involving the fourth moment of a random variable.

Lemma 3.1. (Lemma 3.2(i) of [2]). *Let Y be a real random variable and suppose that its first, second and fourth moments satisfy $\mathbb{E}(Y)=0$, $\mathbb{E}(Y^2)=\sigma^2>0$ and $\mathbb{E}(Y^4)\leqslant b\sigma^4$ for some constant $b>0$. Then*

$$\mathbb{P}(Y>0)\geqslant\frac{1}{2^{4/3}b} \quad \textit{and} \quad \mathbb{P}(Y<0)\geqslant\frac{1}{2^{4/3}b}.$$

The aim in the rest of this section is to prove the following lemma, which directly implies Theorem 1.3.

Lemma 3.2. *There exists a constant $b>0$ such that*

$$\mathbb{E}((X_{G,k}-\mu)^4)\leqslant b\cdot\mathrm{Var}\,(X_{G,k})^2. \tag{3.2}$$

Having established this, we conclude the proof of Theorem 1.3 as follows. By Lemma 3.2 the random variable $Y:=X-\mu$ satisfies the assumptions of Lemma 3.1. Since the event $\{X=\ell\}$ is disjoint either from the event $\{Y>0\}$ or from the event $\{Y<0\}$, using Lemma 3.1 we obtain

$$\mathbb{P}(X=\ell)\leqslant 1-\frac{1}{2^{4/3}b}<1-\varepsilon,$$

contrary to our assumption that $\mathbb{P}(X=\ell)>1-\varepsilon$.

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

Proof of Lemma 3.2. First let us expand the right-hand side of (3.2):

$$\mathrm{Var}(X)^2 \overset{(3.1)}{=} (1+o(1))\left(m\cdot\frac{k^2}{n^2}+S\cdot\frac{k^3}{n^3}\right)^2 = \Theta\left(m^2\cdot\frac{k^4}{n^4}+mS\cdot\frac{k^5}{n^5}+S^2\cdot\frac{k^6}{n^6}\right). \tag{3.3}$$

Now, putting $p := (k)_2/(n)_2$, the left-hand side of (3.2) can be written as

$$\mathbb{E}((X-\mu)^4) = \mathbb{E}\left(\left(\sum_{e\in E(G)}(X_e-p)\right)^4\right) = \sum_{(e,f,g,h)\in E(G)^4}\mathrm{Cov}(e,f,g,h), \tag{3.4}$$

where, by abuse of notation,

$$\mathrm{Cov}(e,f,g,h) := \mathbb{E}((X_e-p)(X_f-p)(X_g-p)(X_h-p)). \tag{3.5}$$

The following simple technical claim will play a crucial role in our proof of Lemma 3.2.

Claim 3.3. *For every* 4*-tuple* $(e,f,g,h)\in E(G)^4$ *we have*

$$\mathrm{Cov}(e,f,g,h) = O(\mathbb{E}(X_eX_fX_gX_h)).$$

Proof. Fix some 4-tuple $(e,f,g,h)\in E(G)^4$. Let B be some subset of $\{e,f,g,h\}$. Let H_B denote the graph spanned by the edges in B and let $t = |V(H_B)|$. Let H'_B be a graph obtained from H_B by deleting one of its edges (together with any of its endpoints if its degree in H is 1), adding two new vertices, and connecting them by an edge. Let t' denote the number of vertices of H'_B; clearly $t\leqslant t'\leqslant k$. Since $k\ll n$, a straightforward calculation shows that

$$\mathbb{E}\left(p\cdot\prod_{a\in E(H_B)\cap E(H'_B)}X_a\right) = \mathbb{E}\left(\prod_{a\in E(H'_B)}X_a\right) = O\left(\frac{k^{t'}}{n^{t'}}\right) = O\left(\frac{k^t}{n^t}\right) = O\left(\mathbb{E}\left(\prod_{a\in E(H_B)}X_a\right)\right).$$

Therefore, for every $B\subseteq\{e,f,g,h\}$ we have

$$p^{4-|B|}\cdot\mathbb{E}\left(\prod_{a\in B}X_a\right) = O(\mathbb{E}(X_eX_fX_gX_h)).$$

We conclude that

$$\mathrm{Cov}(e,f,g,h) = O(\mathbb{E}(X_eX_fX_gX_h)),$$

as claimed. □

We now distinguish between several cases, depending on how the four edges e,f,g,h are arranged. Suppose first that e,f,g,h are pairwise distinct. Let H be the four-edge graph spanned by e,f,g and h. Note that H is a subgraph of G, but *not* necessarily an induced one. The options for H are:

(i) $H = 4K_2$,
(ii) $H = 2K_2 + K_{1,2}$,
(iii) $H = 2K_{1,2}$,
(iv) $H = K_2 + K_{1,3}$,
(v) $H = K_2 + P_4$,
(vi) $H = K_2 + K_3$,
(vii) $H = K_{1,4}$,
(viii) $H = K^+_{1,3}$, a 3-star with one edge subdivided into two,
(ix) $H = P_5$,
(x) $H = C_4$,
(xi) $H = K^+_3$, a triangle with one pendant edge.

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

There are also several 'degenerate' cases, in which two of the edges e, f, g, h coincide; we will deal with these cases at the end of the proof.

We claim that in each of the above listed cases, after some cancellation, the respective contributions can be bounded from above using the terms appearing on the right-hand side of (3.3). As a helpful piece of notation, we denote the number of unlabelled (not necessarily induced) copies of H in G by $N(H)$.

Case (i). $H = 4K_2$. Given a fixed 4-tuple (e, f, g, h) forming H, it follows by (3.5) that

$$\begin{aligned}\mathrm{Cov}(e,f,g,h) &= \mathbb{E}(X_eX_fX_gX_h) - 4\mathbb{E}(X_eX_fX_g)p + 6\mathbb{E}(X_eX_f)p^2 - 4\mathbb{E}(X_e)p^3 + p^4\\ &= \frac{1}{n^8}\big[(k)_8 - 4(k)_6(k)_2 + 6(k)_4((k)_2)^2 - 3((k)_2)^4\big](1+O_k(1/n))\\ &= \frac{1}{n^8}\big[k^8 - 28k^7 - 4k^8 + 64k^7 + 6k^8 - 48k^7 - 3k^8 + 12k^7 + O(k^6)\big]\\ &= O\left(\frac{k^6}{n^8}\right).\end{aligned}$$

Taking the sum over all such tuples (e, f, g, h), we obtain an overall contribution of at most

$$O\left(\frac{m^4k^6}{n^8}\right) = O\left(m^2 \cdot \frac{k^4}{n^4}\right) = O(\mathrm{Var}(X)^2),$$

where the penultimate equality holds since $m = O(n^2/k)$ follows from Lemma 2.5 and the last equality holds by (3.3).

Case (ii). $H = 2K_2 + K_{1,2}$. Let e and f denote the two edges that share a vertex. It follows by (3.5) and a straightforward calculation that

$$\begin{aligned}\mathrm{Cov}(e,f,g,h) &= \big[\mathbb{E}(X_eX_fX_gX_h) - \mathbb{E}(X_eX_fX_g)p - \mathbb{E}(X_eX_fX_h)p + \mathbb{E}(X_eX_f)p^2\big](1+O_k(1/n))\\ &= \frac{1}{n^7}\big[(k)_7 - 2(k)_5(k)_2 + (k)_3((k)_2)^2\big](1+O_k(1/n))\\ &= \frac{1}{n^7}\big[k^7 - 21k^6 - 2k^7 + 22k^6 + k^7 - 5k^6 + O(k^5)\big]\\ &= \frac{1}{n^7}(-4k^6 + O(k^5)) < 0.\end{aligned}$$

Thus, we can ignore such tuples (e, f, g, h) as their contribution is negative.

Case (iii). $H = 2K_{1,2}$. A straightforward calculation shows that

$$\mathrm{Cov}(e,f,g,h) = (1+o(1))(\mathbb{E}(X_eX_fX_gX_h)) = O(k^6/n^6).$$

Moreover, the total number of such 4-tuples (e, f, g, h) is at most $4! \cdot S^2$. Hence,

$$\sum_{(e,f,g,h)\cong H} \mathrm{Cov}(e,f,g,h) = O\left(S^2 \cdot \frac{k^6}{n^6}\right) = O(\mathrm{Var}(X)^2),$$

where the last equality holds by (3.3).

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

Case (iv). $H = K_2 + K_{1,3}$. It follows from Claim 3.3 that $\mathrm{Cov}(e,f,g,h) = O(k^6/n^6)$. Note that a copy of $K_{1,3}$ in G can be viewed as a copy of P_3 with an additional edge attached to its middle vertex. This implies that $N(K_{1,3}) = O(N(P_3)\cdot\Delta(G)) = O(Sn/k)$, where the last equality holds by Lemma 2.5. Therefore,

$$\sum_{(e,f,g,h)\cong H} \mathrm{Cov}(e,f,g,h) = O(mS\cdot k^5/n^5) = O(\mathrm{Var}(X)^2),$$

where the last equality holds by (3.3).

Case (v). $H = K_2 + P_4$. This case is analogous to Case (iv) with P_4 in place of $K_{1,3}$. Since a copy of P_4 in G can be viewed as a copy of P_3 with another vertex connected by an edge to one of its end vertices, we have $N(P_4) = O(N(P_3)\cdot\Delta(G)) = O(Sn/k)$, where the last equality holds by Lemma 2.5. We conclude that

$$\sum_{(e,f,g,h)\cong H} \mathrm{Cov}(e,f,g,h) = O(m\cdot N(P_4)\cdot k^6/n^6) = O(mS\cdot k^5/n^5) = O(\mathrm{Var}(X)^2),$$

where the last equality holds by (3.3).

Case (vi). $H = K_2 + K_3$. This is similar to Cases (iv) and (v). It follows from Claim 3.3 that

$$\mathrm{Cov}(e,f,g,h) = O(k^5/n^5).$$

Hence,

$$\sum_{(e,f,g,h)\cong H} \mathrm{Cov}(e,f,g,h) = O\left(m\cdot N(K_3)\cdot\frac{k^5}{n^5}\right) = O(mS\cdot k^5/n^5) = O(\mathrm{Var}(X)^2),$$

where the second equality holds since $N(K_3) = O(S)$ and the last equality holds by (3.3).

Case (vii). $H = K_{1,4}$. It follows from Claim 3.3 that

$$\sum_{(e,f,g,h)\cong H} \mathrm{Cov}(e,f,g,h) = O\left(N(K_{1,4})\cdot\frac{k^5}{n^5}\right). \tag{3.6}$$

Note that

$$N(K_{1,4}) = O\left(\sum_{v\in V(G)} d(v)^4\right) = O\left(\Delta(G)^2\cdot\sum_{v\in V(G)} d(v)^2\right) = O\left(\frac{n^2}{k^2}\cdot S\right) = O(mS),$$

where the penultimate equality holds by Lemma 2.5 and the last equality holds by Lemma 2.6. We conclude that

$$\sum_{(e,f,g,h)\cong H} \mathrm{Cov}(e,f,g,h) = O\left(mS\cdot\frac{k^5}{n^5}\right) = O(\mathrm{Var}(X)^2),$$

where the last equality holds by (3.3).

Case (viii). $H = K_{1,3}^+$. This is very similar to Case (vii) with $K_{1,3}^+$ in place of $K_{1,4}$. Since a copy of $K_{1,3}^+$ in G can be viewed as a copy of P_3 with two leaves attached to one of its end vertices, it follows that

$$N(K_{1,3}^+) = O(S\cdot\Delta(G)^2) = O(mS).$$

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

Therefore

$$\sum_{(e,f,g,h)\cong H} \mathrm{Cov}(e,f,g,h) = O\left(N(K_{1,3}^+)\cdot\frac{k^5}{n^5}\right) = O\left(mS\cdot\frac{k^5}{n^5}\right) = O(\mathrm{Var}(X)^2),$$

where the last equality holds by (3.3).

Case (ix). $H = P_5$. This is again very similar to Case (vii) with P_5 in place of $K_{1,4}$. Since a copy of P_5 in G can be viewed as a copy of P_3 with a leaf attached to each of its end vertices, it follows that

$$N(P_5) = O(S\cdot\Delta(G)^2) = O(mS).$$

Therefore

$$\sum_{(e,f,g,h)\cong H} \mathrm{Cov}(e,f,g,h) = O\left(N(P_5)\cdot\frac{k^5}{n^5}\right) = O\left(mS\cdot\frac{k^5}{n^5}\right) = O(\mathrm{Var}(X)^2),$$

where the last equality holds by (3.3).

Case (x). $H = C_4$. It follows from Claim 3.3 that

$$\sum_{(e,f,g,h)\cong H} \mathrm{Cov}(e,f,g,h) = O\left(N(C_4)\cdot\frac{k^4}{n^4}\right).$$

It is evident that $N(C_4) = O(N(P_4))$. Since, moreover, we can view P_4 as P_3 with an additional edge appended to one of its leaves, it follows that

$$N(P_4) = O(N(P_3)\cdot\Delta(G)) = O(S\cdot\Delta(G)) = O\left(S\cdot\frac{n}{k}\right),$$

where the last equality holds by Lemma 2.5. Therefore

$$\sum_{(e,f,g,h)\cong H} \mathrm{Cov}(e,f,g,h) = O\left(N(C_4)\cdot\frac{k^4}{n^4}\right) = O\left(S\cdot\frac{k^3}{n^3}\right) = O\left(mS\cdot\frac{k^5}{n^5}\right) = O(\mathrm{Var}(X)^2),$$

where the penultimate equality holds by Lemma 2.6 and the last equality holds by (3.3).

Case (xi). $H = K_3^+$. Using the calculations made in Case (x) we obtain

$$N(K_3^+) = O(N(P_4)) = O\left(S\cdot\frac{n}{k}\right).$$

It thus follows by Claim 3.3 that

$$\sum_{(e,f,g,h)\cong H} \mathrm{Cov}(e,f,g,h) = O\left(N(K_3^+)\cdot\frac{k^4}{n^4}\right) = O\left(S\cdot\frac{k^3}{n^3}\right) = O\left(mS\cdot\frac{k^5}{n^5}\right) = O(\mathrm{Var}(X)^2),$$

where the last equality holds by (3.3).

The degenerate cases

We now consider the cases where some edge of $\{e,f,g,h\}$ appears more than once; in those cases, the corresponding graph H has at most three edges:

(xii) $H = 3K_2$,
(xiii) $H = K_2 + K_{1,2}$,

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

(xiv) $H = K_{1,3}$,
(xv) $H = P_4$,
(xvi) $H = K_3$,
(xvii) $H = K_{1,2}$,
(xviii) $H = 2K_2$,
(xix) $H = K_2$.

Case (xii). $H = 3K_2$. Without loss of generality, suppose that e, f, g are independent and that $h = e$. Since, clearly, $X_e^2 = X_e$, we obtain

$$\begin{aligned}\mathrm{Cov}(e,f,g,h) &= \mathbb{E}((X_e-p)^2(X_f-p)(X_g-p))\\ &= (\mathbb{E}(X_eX_fX_g) - p\mathbb{E}(X_eX_f) - p\mathbb{E}(X_eX_g) + p^2\mathbb{E}(X_e))(1+O_k(1/n))\\ &= \frac{1}{n^6}((k)_6 - 2(k)_2(k)_4 + ((k)_2)^3)(1+O_k(1/n))\\ &= \frac{-4k^5+O(k^4)}{n^6}\\ &< 0.\end{aligned}$$

Thus, the total contribution of all such tuples is negative.

Case (xiii). $H = K_2 + K_{1,2}$. It follows from Claim 3.3 that

$$\sum_{(e,f,g,h)\cong H} \mathrm{Cov}(e,f,g,h) = O\left(mN(K_{1,2})\cdot\frac{k^5}{n^5}\right) = O\left(mS\cdot\frac{k^5}{n^5}\right) = O(\mathrm{Var}(X)^2),$$

where the last equality holds by (3.3).

Case (xiv). $H = K_{1,3}$. It follows from Claim 3.3 that

$$\begin{aligned}&\sum_{(e,f,g,h)\cong H} \mathrm{Cov}(e,f,g,h)\\ &= O\left(N(K_{1,3})\cdot\frac{k^4}{n^4}\right) = O\left(\Delta(G)\cdot N(K_{1,2})\cdot\frac{k^4}{n^4}\right) = O\left(S\cdot\frac{k^3}{n^3}\right) = O\left(mS\cdot\frac{k^5}{n^5}\right) = O(\mathrm{Var}(X)^2),\end{aligned}$$

where the third equality holds by Lemma 2.5, the fourth equality holds by Lemma 2.6, and the last equality holds by (3.3).

Case (xv). $H = P_4$. It follows from Claim 3.3 that

$$\sum_{(e,f,g,h)\cong H} \mathrm{Cov}(e,f,g,h) = O\left(N(P_4)\cdot\frac{k^4}{n^4}\right) = O\left(S\cdot\frac{k^3}{n^3}\right) = O\left(mS\cdot\frac{k^5}{n^5}\right) = O(\mathrm{Var}(X)^2),$$

where the second equality was proved in Case (x) and the last equality holds by (3.3).

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

Case (xvi). $H = K_3$. It follows from Claim 3.3 that

$$\sum_{(e,f,g,h)\cong H} \mathrm{Cov}(e,f,g,h) = O\left(N(K_3)\cdot\frac{k^3}{n^3}\right) = O\left(S\cdot\frac{k^3}{n^3}\right) = O\left(mS\cdot\frac{k^5}{n^5}\right) = O(\mathrm{Var}(X)^2),$$

where the last equality holds by (3.3).

Case (xvii). $H = K_{1,2}$. It follows from Claim 3.3 that

$$\sum_{(e,f,g,h)\cong H} \mathrm{Cov}(e,f,g,h) = O\left(N(K_{1,2})\cdot\frac{k^3}{n^3}\right) = O\left(S\cdot\frac{k^3}{n^3}\right) = O\left(mS\cdot\frac{k^5}{n^5}\right) = O(\mathrm{Var}(X)^2),$$

where the last equality holds by (3.3).

Case (xviii). $H = 2K_2$. It follows from Claim 3.3 that

$$\sum_{(e,f,g,h)\cong H} \mathrm{Cov}(e,f,g,h) = O\left(N(2K_2)\cdot\frac{k^4}{n^4}\right) = O\left(m^2\cdot\frac{k^4}{n^4}\right) = O(\mathrm{Var}(X)^2),$$

where the last equality holds by (3.3).

Case (xix). $H = K_2$. It follows from Claim 3.3 that

$$\sum_{(e,f,g,h)\cong H} \mathrm{Cov}(e,f,g,h) = O\left(m\cdot\frac{k^2}{n^2}\right) = O\left(m^2\cdot\frac{k^4}{n^4}\right) = O(\mathrm{Var}(X)^2),$$

where the last equality holds by (3.3).

To conclude, we have considered every possible case and in each one we have shown that its respective contribution to $\mathbb{E}((X-\mu)^4)$ is of order of magnitude $O(\mathrm{Var}(X)^2)$. Since the number of cases is constant, this completes the proof of Lemma 3.2 and thus also of Theorem 1.3. □

4. Very small inducibility for almost all ℓ

Our aim in this section is to prove Theorem 1.5. Let k and ℓ be positive integers for which $\min\{k, \binom{k}{2} - \ell\} = \Omega(k^2)$. Since $\mathrm{ind}(k,\ell)$ is defined as $\lim_{n\to\infty} I(n,k,\ell)$ and the latter sequence is monotone decreasing, it suffices to show that $I(2k,k,\ell) = O(k^{-0.1})$. Suppose then that we have a $2k$-vertex graph $G = (V,E)$, in which we are selecting a k-vertex set $A \subseteq V$ uniformly at random. As in the proofs in previous sections, we may assume that G maximizes $\mathbb{P}(X_{G,k} = \ell)$ amongst all $2k$-vertex graphs.

Claim 4.1. $\min\{e(G), e(\overline{G})\} = \Omega(k^2)$.

Proof. Since $|G| = 2k$ and G contains an induced subgraph of order k and size $\ell = \Omega(k^2)$ (as, clearly, $I(2k,k,\ell) > 0$), we have that $e(G) = \Omega(k^2)$. The statement for $\overline{G}$ follows analogously. □

A pair of distinct vertices $\{u,v\} \subseteq V$ will be called *distinguished* if $|N_G(u) \triangle N_G(v)| = \Theta(k)$. Let $D = D(G)$ be the set of all distinguished pairs in G.

Claim 4.2. $|D| = \Theta(k^2)$.

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

Proof. It is evident that $|D| = O(k^2)$. To see that $|D| = \Omega(k^2)$, note that

$$\sum_{\{u,v\}\in\binom{V}{2}} |N(u)\triangle N(v)| = \sum_{\{u,v\}\in\binom{V}{2}} (d(u)+d(v)-2|N(u)\cap N(v)|)$$

$$= (|V|-1)\sum_{v\in V} d(v) - 2\sum_{v\in V}\binom{d(v)}{2}$$

$$= 2\left[(2k-1)e(G) - \sum_{v\in V}\binom{d(v)}{2}\right].$$

By a theorem of Ahlswede and Katona [1], for a graph H of given order and size the quantity $\sum_v \binom{d(v)}{2}$, which corresponds to the number of cherries $K_{1,2}$, is maximized when either H or its complement is a union of a clique and an independent set. In either case it follows by Claim 4.1 and a straightforward calculation that

$$\sum_{\{u,v\}\in\binom{V}{2}} |N(u)\triangle N(v)| \geqslant \sum_{\{u,v\}\in\binom{V(H)}{2}} |N_H(u)\triangle N_H(v)| = \Omega(k^3).$$

Since the maximum degree of G is $O(k)$, we conclude that $D(G) = \Omega(k^2)$, as claimed. □

Claim 4.3. *For any graph F with $2k$ vertices and $\Omega(k^2)$ edges, and any integer t satisfying $k - \sqrt{k} \leqslant t \leqslant k + \sqrt{k}$, we have*

$$\mathbb{P}\left(X_{F,t} \leqslant \frac{e(F)}{5}\right) = o(k^{-1}).$$

Proof. Let $B \subseteq V(F)$ be a random set of size t and let $\overline{B} = V(F)\setminus B$ be its complement. Fix an arbitrary vertex $v \in V$. It follows by standard tail estimates for the hypergeometric distribution that

$$\mathbb{P}(|d(v,B) - d(v,\overline{B})| > k^{0.9}) = o(k^{-2})$$

(in fact, the right-hand side is exponentially small in a positive constant power of k; see for example [12, Theorem 2.10]). Hence, by the union bound, with probability $1 - o(k^{-1})$, it holds that $|d(v,B) - d(v,\overline{B})| \leqslant k^{0.9}$ for every $v \in V(F)$. Summing over all vertices in B, we obtain

$$|2e(F[B]) - e(F[B,\overline{B}])| \leqslant \sum_{v\in B} |d(v,B) - d(v,\overline{B})| \leqslant 2k^{1.9} = o(e(F)).$$

Similarly, summing over all vertices in $\overline{B}$, we obtain

$$|2e(F[\overline{B}]) - e(F[B,\overline{B}])| = o(e(F)).$$

Hence, with probability $1 - o(k^{-1})$ we have

$$e(F[B]) = e(F[\overline{B}]) + o(e(F)) = \frac{1}{2}\cdot e(F[B,\overline{B}]) + o(e(F)) = (1/4 + o(1))e(F).$$

In particular, $X_{F,t} = e(F[B]) \geqslant e(F)/5$ holds with probability $1 - o(k^{-1})$. □

Now, let us sample the set A in two steps as follows. We first sample a set $A_0 \subseteq V$ of size $k - k^{0.2}$ uniformly at random, then we sample a set $A_1 \subseteq V\setminus A_0$ of size $k^{0.2}$ in a manner which will be specified later, and finally we set $A = A_0 \cup A_1$. Note that A_1 will be sampled in a way which will ensure that the resulting set A will indeed be chosen uniformly amongst all subsets of $V(G)$ of size k.

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

Given A_0, a distinguished pair $\{u, v\}$ is said to be *bad* if $|d(u, A_0) - d(v, A_0)| \leqslant k^{0.4}$ and *good* otherwise. Our next claim shows that, with sufficiently high probability, most distinguished pairs are good.

Claim 4.4. *With probability* $1 - O(k^{-0.1})$ *there are at most* $|D|/6$ *bad pairs.*

Proof. Fix some distinguished pair $\{u, v\}$ and let $X_{uv} = 1$ if $\{u, v\}$ is bad and $X_{uv} = 0$ otherwise. Let $Y_{uv} = |A_0 \cap (N(u) \triangle N(v))|$ and let $s = |N(u) \triangle N(v)|$; recall that $\{u, v\}$ is a distinguished pair and thus $s = \Theta(k)$. Then

$$\begin{aligned}\mathbb{P}(X_{uv} = 1) &= \sum_{t=0}^{s} \mathbb{P}(X_{uv} = 1 | Y_{uv} = t) \cdot \mathbb{P}(Y_{uv} = t) \\ &= \sum_{t=\varepsilon k}^{0.9s} \mathbb{P}(X_{uv} = 1 | Y_{uv} = t) \cdot \mathbb{P}(Y_{uv} = t) + O(k^{-1}), \end{aligned} \tag{4.1}$$

where the first equality holds by the law of total probability and the second equality holds for a sufficiently small constant $\varepsilon > 0$ since, by standard tail estimates for the hypergeometric distribution, we have that $\mathbb{P}(Y_{uv} \leqslant \varepsilon k) = O(k^{-1})$ and $\mathbb{P}(Y_{uv} \geqslant 0.9s) = O(k^{-1})$.

Assume without loss of generality that $|N(u) \setminus N(v)| \geqslant |N(v) \setminus N(u)|$; observe that by the definition of a distinguished pair it thus follows that $r := |N(u) \setminus N(v)| = \Theta(k)$. Let $Z_{uv} = |A_0 \cap (N(u) \setminus N(v))|$. Observe that for any $\varepsilon k \leqslant t \leqslant s$ we have

$$\mathbb{P}(X_{uv} = 1 | Y_{uv} = t) \leqslant \sum_{i=(t-k^{0.4})/2}^{(t+k^{0.4})/2} \mathbb{P}(Z_{uv} = i | Y_{uv} = t). \tag{4.2}$$

Note that $Z_{uv} | Y_{uv} = t$ is a hypergeometric random variable with parameters s, t and r. Since, moreover, $\mathbb{E}(Z_{uv}) = rt/s = \Theta(k)$, a straightforward calculation shows that for every $(t - k^{0.4})/2 \leqslant i \leqslant (t + k^{0.4})/2$ we have

$$\mathbb{P}(Z_{uv} = i | Y_{uv} = t) = O(k^{-1/2}). \tag{4.3}$$

Combining (4.1), (4.2) and (4.3), we obtain

$$\mathbb{P}(X_{uv} = 1) = O(k^{-0.1}). \tag{4.4}$$

Now let $X = \sum_{\{u,v\} \in D} X_{uv}$ be the random variable which counts the number of bad pairs in D. It follows by (4.4) that

$$\mathbb{E}(X) = \sum_{\{u,v\} \in D} \mathbb{E}(X_{uv}) = O(k^{1.9}).$$

Applying Markov's inequality to X, we conclude that $\mathbb{P}(X \geqslant |D|/6) = O(k^{-0.1})$. □

Claim 4.5. *With probability* $1 - O(k^{-0.1})$*, the set* A_0 *will have the following property: there exist disjoint sets* $X, Y \subseteq V \setminus A_0$ *such that* $|X| = |Y| = \Theta(k)$ *and* $d(x, A_0) - d(y, A_0) \geqslant k^{0.4}$ *holds for every* $x \in X$ *and* $y \in Y$*.*

Proof. Recall that $|D| = \Theta(k^2)$ holds by Claim 4.2. By considering an auxiliary graph F with $V(F) = V$ and $E(F) = \{uv : \{u, v\} \in D\}$ to which we apply Claim 4.3, we infer that, with probability $1 - o(k^{-1})$, at least $|D|/5$ of the distinguished pairs are disjoint from A_0. On the other hand it follows by Claim 4.4 that, with probability $1 - O(k^{-0.1})$, at most $|D|/6$ of the distinguished pairs are bad. We conclude that, with probability $1 - O(k^{-0.1})$, there is a set $D' \subseteq D \cap \binom{V \setminus A_0}{2}$ of size ck^2, for a constant $c > 0$, such that $|d(u, A_0) - d(v, A_0)| \geqslant k^{0.4}$ for every $\{u, v\} \in D'$.

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

Let $(u_1, \dots, u_{k+k^{0.2}})$ be an ordering of the vertices of $V \setminus A_0$ by non-increasing order of degrees into A_0, that is, $d(u_i, A_0) \geqslant d(u_j, A_0)$ for every $1 \leqslant i < j \leqslant k + k^{0.2}$. Let $q = ck/3$, let $X = \{u_1, \dots, u_q\}$, and let $Y = \{u_{k+k^{0.2}-q+1}, \dots, u_{k+k^{0.2}}\}$. Observe that $X \cap Y = \emptyset$ and $|X| = |Y| = q = \Theta(k)$. It thus remains to prove that $d(x, A_0) - d(y, A_0) \geqslant k^{0.4}$ holds for every $x \in X$ and $y \in Y$. Suppose for a contradiction that there exist $x \in X$ and $y \in Y$ such that $d(x, A_0) - d(y, A_0) < k^{0.4}$. Since, by definition, $d(x, A_0) \geqslant d(u, A_0) \geqslant d(y, A_0)$ for every $u \in V \setminus (A_0 \cup X \cup Y)$, it follows that $\{u, v\} \cap (X \cup Y) \neq \emptyset$ for every $\{u, v\} \in D'$. Therefore

$$\begin{aligned} ck^2 = |D'| &\leqslant \binom{|X \cup Y|}{2} + |X \cup Y| \cdot |V \setminus (A_0 \cup X \cup Y)| \\ &\leqslant 2q(q + k + k^{0.2} - 2q) \\ &\leqslant 2ck/3 \cdot k \\ &< ck^2, \end{aligned}$$

which is clearly a contradiction. □

Let A_0 be a set chosen randomly as described above. Let X and Y be disjoint subsets of $V \setminus A_0$ such that $|X| = |Y| = ck$ for some $c > 0$, and $d(x, A_0) - d(y, A_0) \geqslant k^{0.4}$ holds for every $x \in X$ and $y \in Y$; such sets exist with probability $1 - O(k^{-0.1})$ by Claim 4.5. Now, we choose a random set $A_1 \subseteq V \setminus A_0$ of size $k^{0.2}$ as follows. First, we choose $k^{0.2}$ pairwise disjoint vertex pairs from $V \setminus A_0$ uniformly at random. Then, from each such pair, we choose uniformly at random exactly one element to be in A_1, all choices being mutually independent.

Claim 4.6. *Let A_1 be chosen randomly as described above and let $A = A_0 \cup A_1$. Then, for any integer ℓ, we have $\mathbb{P}(e_G(A) = \ell) = O(k^{-0.1})$.*

Proof. Let M denote the set of all $k^{0.2}$ randomly chosen pairs. Let M_{XY} denote the set of chosen pairs which have one element in X and one in Y, and let $m = |M_{XY}|$. We claim that $m = \Omega(k^{0.2})$ with probability $1 - e^{-\Omega(k^{0.2})}$. Indeed, consider choosing the pairs which constitute M one by one. Suppose that we have already chosen j pairs for some $0 \leqslant j < k^{0.2}$, and now choose the $(j+1)$th pair. Let M_j denote the set of all vertices in the union of these j pairs; clearly $|M_j| = O(k^{0.2})$. Let $T_{j+1} = 1$ if the $(j+1)$th pair has one element in X and one in Y, and $T_{j+1} = 0$ otherwise. Then

$$\mathbb{P}(T_{j+1} = 1) = \frac{|X \setminus M_j| \cdot |Y \setminus M_j|}{\binom{2k}{2}} \geqslant \frac{c^2}{3}.$$

Note that $m = \sum_{j=1}^{k^{0.2}} T_j$. Let $Z \sim \mathrm{Bin}(k^{0.2}, c^2/3)$ and observe that, by the above calculation, m stochastically dominates Z. Hence, using standard bounds on the tail of the binomial distribution, we conclude that

$$\mathbb{P}(m < c^2 k^{0.2}/6) \leqslant \mathbb{P}(Z < c^2 k^{0.2}/6) \leqslant \mathbb{P}(Z < \mathbb{E}(Z)/2) < e^{-c^2 k^{0.2}/24}.$$

Let $\{x_1, y_1\}, \{x_2, y_2\}, \dots, \{x_m, y_m\}$ be the elements of M_{XY}, where

$$\{x_1, \dots, x_m\} \subseteq X \quad \text{and} \quad \{y_1, \dots, y_m\} \subseteq Y.$$

Fix any choice of one element from every pair in $M \setminus M_{XY}$. With any choice of one element from every pair in M_{XY}, we associate a binary vector $\bar{z} = (z_1, \dots, z_m)$ in a natural way, namely, for every $1 \leqslant i \leqslant m$, $z_i = 1$ if we chose x_i to be in A_1 and $z_i = 0$ if we chose y_i. For each such vector $\bar{z}$, we denote the resulting random k-subset of $V(G)$ by $A_{\bar{z}}$. For two such vectors $\bar{z}$ and $\bar{w}$, we say that $\bar{z} > \bar{w}$ if $z_i \geqslant w_i$ for every $1 \leqslant i \leqslant m$ and $\bar{z} \neq \bar{w}$. We claim that $e_G(A_{\bar{z}}) > e_G(A_{\bar{w}})$ whenever $\bar{z} > \bar{w}$. Indeed, let $1 \leqslant i \leqslant m$ be an index for which $z_i = 1$ and $w_i = 0$. Then

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

$$e_G(A_{\bar{z}}) - e_G(A_{\bar{w}}) \geqslant d(x_i, A_0) - d(y_i, A_0) - e_G(A_{\bar{w}} \setminus A_0) \geqslant k^{0.4} - \binom{k^{0.2}}{2} > 0.$$

It follows that, for any integer ℓ, the elements of $\{\bar{z} \in \{0,1\}^m : e_G(A_{\bar{z}}) = \ell\}$ form an anti-chain. Hence, by Sperner's theorem we conclude that

$$\mathbb{P}(e_G(A) = \ell) \leqslant \binom{m}{\lfloor m/2 \rfloor} \Big/ 2^m = O(1/\sqrt{m}) = O(k^{-0.1}). \qquad \square$$

Claim 4.6 implies that $I(2k, k, \ell) = O(k^{-0.1})$. Thus indeed $\mathrm{ind}(k, \ell) = O(k^{-0.1})$ and the proof of Theorem 1.5 is complete. Note that the proof of Claim 4.6 resembles (and was inspired by) the Littlewood–Offord problem and its solution by Erdős [7].

5. When ℓ is fixed: Proof of Theorem 1.6

Our aim in this section is to show that, under certain natural conditions, $X_{G,k}$ exhibits a Poisson-like behaviour. The following proposition makes this precise.

Proposition 5.1. *Suppose that $1 \leqslant \ell \ll k \ll n$ are integers and that G is a graph with n vertices and m edges. Suppose that*

$$\lim_{k\to\infty} m \cdot \frac{k^2}{n^2} = \mu$$

for some constant $\mu = \mu(\ell) > 0$ and that $\Delta := \Delta(G) = o_k(n/k)$. Then

$$\mathbb{P}(X_{G,k} = \ell) = (1 + o_k(1))e^{-\mu} \cdot \frac{\mu^\ell}{\ell!}.$$

Proof. Throughout this proof the $o(\cdot)$-notation will solely refer to $o_k(\cdot)$; hence we omit the subscript k for readability. By Brun's sieve (see *e.g.* Theorem 8.3.1 in [3]), in order to prove that

$$\mathbb{P}(X_{G,k} = \ell) = (1 + o(1))e^{-\mu} \cdot \frac{\mu^\ell}{\ell!},$$

it suffices to show that

$$\mathbb{E}\left[\binom{X}{r}\right] = (1 + o(1)) \cdot \frac{\mu^r}{r!}$$

holds for every fixed integer $1 \leqslant r \leqslant r(\ell)$.

Let $e_1, \ldots, e_m$ be an arbitrary ordering of the edges of G. For every $1 \leqslant j \leqslant m$, let $X_j = 1$ if both endpoints of e_j are in the random k-set A, and $X_j = 0$ otherwise. Observe that $X_{G,k} = \sum_{j=1}^m X_j$. In particular,

$$\mathbb{E}(X_{G,k}) = \sum_{j=1}^m \mathbb{E}(X_j) = m \cdot \binom{n-2}{k-2} \Big/ \binom{n}{k} = m \cdot \frac{k(k-1)}{n(n-1)}.$$

It thus follows by the definition of μ that

$$\mathbb{E}\left[\binom{X}{1}\right] = \mathbb{E}(X) = (1 + o(1))\mu.$$

Now, fix some positive integer r. Let $1 \leqslant j_1 < \cdots < j_r \leqslant m$ be arbitrary indices. Assume first that $e_{j_1}, \ldots, e_{j_r}$ form a matching in G. Then

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

$$\mathbb{P}(X_{j_1} = 1 \wedge \cdots \wedge X_{j_r} = 1) = \binom{n-2r}{k-2r} \Big/ \binom{n}{k} = (1+o(1))\left(\frac{k}{n}\right)^{2r}.$$

Moreover, the number of ways to choose indices $1 \leqslant j_1 < \cdots < j_r \leqslant m$ for which $e_{j_1}, \ldots, e_{j_r}$ form a matching in G is $(1+o(1))\binom{m}{r}$ (trivially, it is at most $\binom{m}{r}$). Indeed, by our assumption on Δ, this number is at least

$$\frac{1}{r!}\prod_{t=0}^{r-1}(m-2t\Delta) \geqslant \frac{1}{r!}\prod_{t=0}^{r-1}(m-2tmk/n) = \frac{m^r}{r!}\cdot(1+o(1))e^{-2k\binom{r}{2}/n} = (1+o(1))\binom{m}{r},$$

where the last equality holds since, by assumption, n is sufficiently large with respect to k and r.

Next, assume that $e_{j_1}, \ldots, e_{j_r}$ do not form a matching in G. Let H be the graph whose edges are $e_{j_1}, \ldots, e_{j_r}$ and whose vertices are the endpoints of $e_{j_1}, \ldots, e_{j_r}$. Let $C_1, \ldots, C_t$ denote the connected components of H, and let $c_i = |C_i|$ for every $1 \leqslant i \leqslant t$. Then

$$\mathbb{P}(X_{j_1} = 1 \wedge \cdots \wedge X_{j_r} = 1) = \binom{n-\sum_{i=1}^t c_i}{k-\sum_{i=1}^t c_i} \Big/ \binom{n}{k} = (1+o(1))\left(\frac{k}{n}\right)^{\sum_{i=1}^t c_i}.$$

Assume without loss of generality that $c_1 \geqslant \cdots \geqslant c_t$. Since H is not a matching, we must have $c_1 \geqslant 3$. For all integers $a \geqslant 2$ and $b \geqslant a-1$, the number of ways to choose b edges of G that form a connected component on a vertices, is at most

$$m(a-1)!\Delta^{a-2}\binom{a}{2}^{b-a+1} = \begin{cases} O(m) & \text{if } a = 2, \\ m \cdot o[(n/k)^{a-2}] & \text{if } a > 2. \end{cases}$$

Indeed, we begin by choosing an arbitrary edge of G (in m ways), then, one by one, we choose $a-2$ additional edges to form a tree on a vertices (in at most $2\Delta \cdot 3\Delta \cdot \ldots \cdot (a-1)\Delta = (a-1)!\Delta^{a-2}$ ways), and then we choose the remaining $b-a+1$ edges such that their endpoints are among the a vertices of the tree.

Hence, the total number of ways to choose r edges of G that form a graph H consisting of connected components of orders $c_1 \geqslant \cdots \geqslant c_t \geqslant 2$, where $c_1 \geqslant 3$, is at most

$$\prod_{i=1}^t m \cdot o\left[\left(\frac{n}{k}\right)^{c_i-2}\right] = m^t \cdot o\left[\left(\frac{n}{k}\right)^{\sum_{i=1}^t c_i - 2t}\right].$$

For every $1 \leqslant t \leqslant r$, let L_t denote the set of all integer vectors $(c_1, \ldots, c_t)$ such that $c_1 \geqslant \cdots \geqslant c_t \geqslant 2$, $c_1 \geqslant 3$, and $\sum_{i=1}^t c_i < 2r$. Combining everything together we conclude that

$$\begin{aligned}
\mathbb{E}\left[\binom{X}{r}\right] &= (1+o(1))\binom{m}{r}\cdot(1+o(1))\left(\frac{k}{n}\right)^{2r} \\
&\quad + \sum_{t=1}^{r}\sum_{(c_1,\ldots,c_t)\in L_t} m^t \cdot o\left[\left(\frac{n}{k}\right)^{\sum_{i=1}^t c_i - 2t}\right]\cdot(1+o(1))\left(\frac{k}{n}\right)^{\sum_{i=1}^t c_i} \\
&= (1+o(1))\left(\frac{1}{r!}\left(\frac{mk^2}{n^2}\right)^r + \sum_{t=1}^{r}\sum_{(c_1,\ldots,c_t)\in L_t} o\left[\left(\frac{mk^2}{n^2}\right)^t\right]\right) \\
&= (1+o(1))\left[\frac{\mu^r}{r!} + o(\mu^r)\right] \\
&= (1+o(1))\cdot\frac{\mu^r}{r!},
\end{aligned}$$

where the penultimate equality holds since r, t and $\mu > 0$ are fixed. □

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

The first consequence of Proposition 5.1 is that it provides a whole plethora of constructions demonstrating that $\mathrm{ind}(k,1) \geqslant 1/e + o_k(1)$. Indeed, let n be a sufficiently large integer and let G be any graph with n vertices, $(1+o_k(1))(n^2/k^2)$ edges, and maximum degree $o_k(n/k)$. Then, G satisfies the assumptions of Proposition 5.1 with $\mu = 1$, implying that $\mathrm{ind}(k,1) \geqslant 1/e + o_k(1)$. This rich family of constructions includes, in particular, the random graphs $G(n, \binom{k}{2}^{-1})$ (with high probability) which were mentioned in the Introduction, and the pairwise disjoint union of $\binom{k}{2}$ cliques, on $n\binom{k}{2}^{-1}$ vertices each.

Moreover, by combining Proposition 5.1 with some of our previous arguments as well as some new ones, we can prove Theorem 1.6.

5.1 The $1/2$*-bound for* $\ell = 1$

Let k be a sufficiently large integer, let $a = \mathrm{ind}(k,1)$, and let G be a graph on n vertices, attaining the maximum density of induced one-edged graphs, that is, $\mathbb{P}(X_{G,k} = 1) = a + o_n(1)$.

Our aim is to show that either $a \leqslant 1/2 + o_k(1)$, or that the conditions of Proposition 5.1 are satisfied. In the latter case we will be done, as applying Proposition 5.1 would imply $a \leqslant 1/e + o_k(1)$.

Claim 5.2. $\Delta(G) \leqslant 10n/k$.

Proof. Let v be a vertex of maximum degree in G and suppose for a contradiction that $d(v) > 10n/k$. It follows by Lemma 2.1 and the discussion succeeding the proof of Proposition 5.1 that

$$\mathbb{P}(X_{G,k} = 1 \mid v \in A) = a + o_n(1) \geqslant 1/e + o_k(1).$$

However, a straightforward calculation which uses the Poisson approximation shows that

$$\begin{aligned}\mathbb{P}(X_{G,k} > 1 \mid v \in A) &\geqslant \mathbb{P}(|A \cap N_G(v)| > 1 \mid v \in A)\\ &> 1 - e^{-10} - 10e^{-10} + o_k(1)\\ &> 1 - \mathbb{P}(X_{G,k} = 1 \mid v \in A),\end{aligned}$$

which is an obvious contradiction. □

Claim 5.3. *If* $\Delta(G) = o_k(n/k)$*, then* $a \leqslant 1/e + o_k(1)$.

Proof. Recall that $e(G) = \Omega(n^2/k^2)$ holds by Lemma 2.6. Assume first that $e(G) \leqslant 20n^2/k^2$. By compactness we may assume that

$$m = (1 + o_k(1))\mu \frac{n^2}{k^2}$$

for some constant $0 < \mu \leqslant 20$, whereby the conditions of Proposition 5.1 are satisfied. Applying it yields

$$a + o_n(1) = \mathbb{P}(X_{G,k} = 1) = (1 + o_k(1))\mu e^{-\mu} \leqslant 1/e + o_k(1).$$

Assume then that $e(G) > 20n^2/k^2$. Let G' be an arbitrary subgraph of G with $20n^2/k^2$ edges. Since $\Delta'(G) \leqslant \Delta(G) = o_k(n/k)$, the graph G' satisfies the conditions of Proposition 5.1 with $\mu = 20$. Therefore

$$a + o_n(1) = \mathbb{P}(X_{G,k} = 1) \leqslant \mathbb{P}(X_{G',k} \leqslant 1) = 21e^{-20} + o_k(1) \leqslant 1/e + o_k(1).$$ □

By Claims 5.2 and 5.3 and by compactness, we may assume that $\Delta(G) = cn/k$ for some constant $c > 0$.

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

Let v be a vertex of maximum degree in G, let $Q = N_G(v)$, and let $R = V(G) \setminus (Q \cup \{v\})$. By the Poisson approximation of the hypergeometric distribution, with probabilities of approximately e^{-c} and ce^{-c}, respectively, A will contain exactly 0 or 1 vertices from Q. Therefore, invoking Lemma 2.1, we obtain

$$\begin{aligned} a + o_n(1) &= \mathbb{P}(X_{G,k} = 1 | v \in A) \\ &= \mathbb{P}(X_{G,k} = 1 | v \in A \text{ and } |A \cap Q| = 0) \cdot \mathbb{P}(|A \cap Q| = 0 | v \in A) \\ &\quad + \mathbb{P}(X_{G,k} = 1 | v \in A \text{ and } |A \cap Q| = 1) \cdot \mathbb{P}(|A \cap Q| = 1 | v \in A) \\ &\leqslant (1 + o_k(1)) \cdot (\mathbb{P}(X_{G[R],k-1} = 1) \cdot e^{-c} + \mathbb{P}(X_{G[R],k-2} = 0) \cdot ce^{-c}). \end{aligned} \tag{5.1}$$

Claim 5.4. *For any graph H and any positive integers k and t we have*

$$\mathbb{P}(X_{H,k-1} = t) \geqslant \frac{k - 2t}{k} \cdot \mathbb{P}(X_{H,k} = t).$$

Proof. Sample uniformly at random a vertex set $A_{k-1} \subseteq V(H)$ of size $k-1$, using the following two steps. First sample k vertices uniformly at random and without replacement, obtaining a set $A_k \subseteq V(H)$. Then, choose a vertex $u \in A_k$ uniformly at random and put $A_{k-1} = A_k \setminus \{u\}$. Observe that if in the first step we sampled a t-edge graph, then the probability of 'destroying' it in the second step is at most $2t/k$. Therefore

$$\mathbb{P}(X_{H,k-1} = t) \geqslant \mathbb{P}(e(G[A_{k-1}]) = t | e(G[A_k]) = t) \cdot \mathbb{P}(e(G[A_k]) = t) \geqslant \frac{k - 2t}{k} \cdot \mathbb{P}(X_{H,k} = t). \qquad \square$$

Claim 5.5. *For any graph H and any positive integers k and t we have*

$$\mathbb{P}(X_{H,k-1} = t) \leqslant \mathbb{P}(X_{H,k} = t) + \frac{2t + 2}{k}.$$

Proof. Consider the same two-step sampling as in the proof of Claim 5.4. Note that if, in the first step, the set A_k contains more than t edges, then the probability of obtaining a t-edge graph in the second step is at most $(2t + 2)/k$ (this is because, for every $s \geqslant 1$, a graph with $t + s$ edges contains at most $2t + 2$ vertices of degree s). Thus,

$$\begin{aligned} \mathbb{P}(X_{H,k-1} = t) &= \sum_{i=0}^{\binom{k}{2}} \mathbb{P}(X_{H,k} = i) \cdot \mathbb{P}(X_{H,k-1} = t | e(G[A_k]) = i) \\ &\leqslant \mathbb{P}(X_{H,k} = t) + \sum_{i=t+1}^{\binom{k}{2}} \mathbb{P}(X_{H,k} = i) \cdot \mathbb{P}(X_{H,k-1} = t | e(G[A_k]) = i) \\ &\leqslant \mathbb{P}(X_{H,k} = t) + \frac{2t + 2}{k} \sum_{i=t+1}^{\binom{k}{2}} \mathbb{P}(X_{H,k} = i) \\ &\leqslant \mathbb{P}(X_{H,k} = t) + \frac{2t + 2}{k}. \end{aligned}$$

$\square$

Let b denote $\mathbb{P}(X_{G[R],k-2} = 1)$. Then, applying Claim 5.4 to (5.1), we obtain

$$\begin{aligned} a &\leqslant (1 + o_k(1)) \cdot (\mathbb{P}(X_{G[R],k-1} = 1) \cdot e^{-c} + \mathbb{P}(X_{G[R],k-2} = 0) \cdot ce^{-c}) \\ &\leqslant (1 + o_k(1)) \cdot (\mathbb{P}(X_{G[R],k-2} = 1) \cdot e^{-c} + \mathbb{P}(X_{G[R],k-2} = 0) \cdot ce^{-c}) \\ &\leqslant (1 + o_k(1)) \cdot (be^{-c} + (1 - b)ce^{-c}). \end{aligned}$$

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

Observe that if $c \geqslant 1$, then

$$\begin{aligned} a &\leqslant (1+o_k(1)) \cdot (be^{-c} + (1-b)ce^{-c}) \\ &\leqslant (1+o_k(1)) \cdot (bce^{-c} + (1-b)ce^{-c}) \\ &\leqslant 1/e + o_k(1). \end{aligned}$$

Assume then that $c \leqslant 1$. Observe that, in this case, $be^{-c} + (1-b)ce^{-c}$ is an increasing function in b. Moreover,

$$b \leqslant (1+o_k(1))\mathbb{P}(X_{G[R],k} = 1) \leqslant (1+o_k(1))a,$$

where the first inequality holds by Claim 5.5 (applied twice), and the second inequality holds since $a = \mathrm{ind}(k, 1)$ and R is large. Hence,

$$(1+o_k(1))a \leqslant be^{-c} + (1-b)ce^{-c} \leqslant ae^{-c} + (1-a)ce^{-c}.$$

This implies that

$$a \leqslant (1+o_k(1))\frac{ce^{-c}}{1-e^{-c}+ce^{-c}} = (1+o_k(1))\frac{c}{e^c-1+c} \leqslant 1/2 + o_k(1).$$

5.2 The 3/4-bound for a fixed ℓ

We use an argument similar to that used in the case $\ell = 1$, but slightly more technical. Let ℓ be a positive integer and let $k \gg \ell$. Let $a = \mathrm{ind}(k, \ell)$, and note that a is bounded away from 0 by a constant $\varepsilon(\ell)$. This can be seen, for instance, by considering an appropriate random graph. Our aim is to prove that $a \leqslant 3/4 + o_k(1)$.

Let G be a graph on n vertices, attaining the maximum density of induced ℓ-edged graphs, that is, $\mathbb{P}(X_{G,k} = \ell) = a + o_n(1)$. Using arguments analogous to those used to prove Claims 5.2 and 5.3, we may assume that $\Delta(G) = cn/k$ for some constant $c > 0$. Let v be a vertex of maximum degree in G, let $Q = N_G(v)$, and let $R = V(G) \setminus (Q \cup \{v\})$. By the Poisson approximation of the hypergeometric distribution, for each fixed t, with probability of approximately $e^{-c}c^t/t!$, the set A will contain exactly t vertices from Q. It thus follows by Lemma 2.1 that

$$\begin{aligned} a + o_n(1) &= \mathbb{P}(X_{G,k} = \ell | v \in A) \\ &= \sum_{t=0}^{\ell} \mathbb{P}(X_{G,k} = \ell | v \in A \text{ and } |A \cap Q| = t) \cdot \mathbb{P}(|A \cap Q| = t | v \in A) \\ &\leqslant (1+o_k(1)) \cdot \left(\mathbb{P}(X_{G[R],k-1} = \ell) \cdot e^{-c} + \sum_{t=1}^{\ell} \mathbb{P}(X_{G[R],k-1-t} \leqslant \ell - t) \cdot \frac{c^t}{t!} e^{-c} \right). \end{aligned} \tag{5.2}$$

Let b denote $\mathbb{P}(X_{G[R],k-\ell-1} = \ell)$. Then, applying Claim 5.4 to (5.2) a constant (depending only on ℓ) number of times, we obtain

$$\begin{aligned} a &\leqslant (1+o_k(1)) \cdot \left(\mathbb{P}(X_{G[R],k-1} = \ell) \cdot e^{-c} + \sum_{t=1}^{\ell} \mathbb{P}(X_{G[R],k-1-t} \leqslant \ell - t) \cdot \frac{c^t}{t!} e^{-c} \right) \\ &\leqslant (1+o_k(1)) \cdot \left(\mathbb{P}(X_{G[R],k-\ell-1} = \ell) \cdot e^{-c} + \sum_{t=1}^{\ell} \mathbb{P}(X_{G[R],k-\ell-1} \leqslant \ell - t) \cdot \frac{c^t}{t!} e^{-c} \right) \\ &\leqslant (1+o_k(1)) \cdot \left(be^{-c} + (1-b) \sum_{t=1}^{\ell} \frac{c^t}{t!} e^{-c} \right). \end{aligned}$$

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

Assume first that $c < \log 2$. In this case we have

$$\sum_{t=1}^{\ell} \frac{c^t}{t!} < e^c - 1 < 2 - 1 = 1,$$

which implies that $be^{-c} + (1-b)\sum_{t=1}^{\ell} e^{-c}c^t/t!$ is an increasing function in b. Now observe that

$$b \leqslant (1 + o_k(1))\mathbb{P}(X_{G[R],k} = \ell) \leqslant (1 + o_k(1))a,$$

where the first inequality holds by Claim 5.5 (applied $\ell + 1$ times), and the second inequality holds since $a = \mathrm{ind}(k, \ell)$ and R is large. Hence,

$$(1 + o_k(1))a \leqslant be^{-c} + (1-b)\sum_{t=1}^{\ell} \frac{c^t}{t!}e^{-c} \leqslant ae^{-c} + (1-a)\sum_{t=1}^{\ell} \frac{c^t}{t!}e^{-c}.$$

This implies that

$$(1 + o_k(1))a \leqslant \frac{e^{-c}\sum_{t=1}^{\ell} c^t/t!}{1 - e^{-c} + e^{-c}\sum_{t=1}^{\ell} c^t/t!} = \frac{\sum_{t=1}^{\ell} c^t/t!}{e^c - 1 + \sum_{t=1}^{\ell} c^t/t!} < \frac{\sum_{t=1}^{\ell} c^t/t!}{2\sum_{t=1}^{\ell} c^t/t!} = \frac{1}{2} < \frac{3}{4}.$$

It thus remains to consider the case $c \geqslant \log 2$. Recall that $d_G(v) = \Delta(G) = cn/k$ and let w be a vertex of minimum degree in G.

Claim 5.6. $d_G(w) = \delta(G) = o_k(n/k)$.

Proof. Suppose that $\delta(G) \geqslant \alpha n/k$ for some constant $\alpha = \alpha(\ell) > 0$. Let G' be a subgraph of G obtained by selecting each edge of G independently at random with probability β/k, where $\beta = \beta(\ell, \alpha)$ is a sufficiently large constant. Since $n \gg k$, with high probability all degrees in G' will be between $(1 - o(1))\alpha\beta n/k^2$ and $(1 + o(1))c\beta n/k^2$, and by compactness we may assume that $e(G) = (1 + o_k(1))\mu n^2/k^2$ for some constant $\mu = \mu(\ell, \alpha, \beta) > \alpha\beta/3 > 0$. Hence G' satisfies the conditions of Proposition 5.1, implying that the distribution of $X_{G',k}$ is approximately Poisson with parameter μ. However, since β was chosen to be sufficiently large, this means that $A_{G',k}$, and *a fortiori* $A_{G,k}$ will induce more than ℓ edges with probability $1 - o_k(1)$. □

Now, let $B = A_{G\setminus\{v,w\},k-1}$ denote a subset of $V(G) \setminus \{v, w\}$ of size $k - 1$, chosen uniformly at random among all such subsets. Assuming that $a > 1/2$ (as otherwise we are done), we apply Lemma 2.2 to v and w to infer

$$\mathbb{P}(e_G(v, B) = e_G(w, B)) > 2a - 1 - o_n(1).$$

However, since $d_G(w) = o_k(n/k)$ holds by Claim 5.6 and $d_G(v) \geqslant \log 2(n/k)$ holds by assumption, a straightforward calculation shows that

$$\mathbb{P}(e_G(w, B) = 0) = 1 - o_k(1) \quad \text{and} \quad \mathbb{P}(e_G(v, B) = 0) \leqslant (1 + o_k(1))e^{-\log 2} = 1/2 + o_k(1).$$

It follows that $1/2 + o_k(1) > 2a - 1 - o_n(1)$, implying $a \leqslant 3/4 + o_k(1)$, as claimed.

6. Concluding remarks

Bounds on ε in Theorem 1.3

While, for clarity of presentation, we did not make an effort to calculate ε explicitly, it is not difficult to see that this constant is not too small; when k is sufficiently large, the value $\varepsilon = 1/100$ is certainly sufficient, and with some care, one can improve it to $\varepsilon = 1/10$. On the other hand, looking at some of our arguments, it is evident that in order to go below $1/2$ for every sufficiently large k and every $0 < \ell < \binom{k}{2}$, one would need new ideas.

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

Upper bounds for fixed ℓ

It is not hard to see that our argument in the proof of Theorem 1.6 in fact gives better bounds than $1/2 + o(1)$ for $\ell = 1$ and $3/4 + o(1)$ in the general case. Take, for example, the case $\ell = 1$. For $\Delta(G) = cn/k$ we obtain

$$a \leqslant (1 + o(1))\frac{c}{e^c - 1 + c},$$

where the right-hand side attains its maximum of $1/2$ when c is close to zero. On the other hand, when $\Delta(G) = o(n/k)$, Proposition 5.1 (Brun's sieve) implies that $a \leqslant 1/e + o(1)$. Interpolating between these two arguments shows that indeed

$$\mathrm{ind}(k, 1) < 1/2 - \Omega(1).$$

Strengthening Conjecture 1.1 in some ranges

Theorem 1.5 demonstrates that the assertion of Conjecture 1.1 holds, with room to spare, for almost all values of k and ℓ. In particular, $\mathrm{ind}(k, \ell)$ tends to 0 with k as long as ℓ and $\binom{k}{2} - \ell$ are quadratic in k. We believe that in the above statement 'quadratic' can be replaced with 'superlinear'.

Conjecture 6.1. *For all pairs (k, ℓ) satisfying $\min\{\ell, \binom{k}{2} - \ell\} = \omega(k)$, we have $\mathrm{ind}(k, \ell) = o(1)$.*

On the other hand, our construction for $\ell = k - 1$ can be straightforwardly extended to show that for any fixed integer C we have $\mathrm{ind}(k, C(k - C)) = \Omega(1)$.

Note also that the argument appearing in the proof of Theorem 1.5 can be applied in a wider range, namely, it can be used to prove that $\mathrm{ind}(k, \ell)$ is polynomially small in k whenever

$$\min\left\{\ell, \binom{k}{2} - \ell\right\} = \Omega(k^{2-\delta})$$

for some explicit constant $\delta > 0$.

It would be interesting to determine the 'correct' power of k in various ranges. For instance, when ℓ and $\binom{k}{2} - \ell$ are quadratic in k, Theorem 1.5 yields $\mathrm{ind}(k, \ell) = O(k^{-0.1})$, whereas, for some values of $\ell = \Theta(k^2)$, we have $\mathrm{ind}(k, \ell) = \Omega(k^{-1/2})$; take for instance $\ell = k^2/4$ and $G = K_{n/2,n/2}$. We believe that the latter bound is tight.

Conjecture 6.2. *For all pairs (k, ℓ) satisfying $\min\{\ell, \binom{k}{2} - \ell\} = \Omega(k^2)$, we have $\mathrm{ind}(k, \ell) = O(k^{-1/2})$.*

Minimum edge-inducibility

A natural counterpart to Conjecture 1.1 would be to determine the asymptotic value of

$$\eta(k) := \min\left\{\mathrm{ind}(k, \ell) : 0 \leqslant \ell \leqslant \binom{k}{2}\right\},$$

as k tends to ∞. Note that this question has been well studied in the setting of graph-inducibilities [10, 11, 16, 17]. In particular, it is known [16, 17] that

$$\min\{\mathrm{ind}(H) : |H| = k\} = (1 + o(1))\frac{k!}{k^k} = e^{-k+o(1)}.$$

This is in stark contrast to the lower bound of $\eta(k) = \Omega(1/k)$ for edge-inducibilities, achieved by the random graphs $G(n, \ell/\binom{k}{2})$.

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

3/4 *as a general upper bound*

As noted in the Introduction, $\mathrm{ind}(k,\ell) < 1$ for every positive integer k and every $0 < \ell < \binom{k}{2}$. It seems plausible that in fact $\mathrm{ind}(k,\ell) \leqslant 3/4$ for all such pairs (k,ℓ). Note that if true, this bound would be tight since, as noted in the Introduction, $\mathrm{ind}(3,1) = \mathrm{ind}(3,2) = 3/4$. Such a result would simultaneously improve Theorems 1.3 and 1.6.

Hypergraphs

The concepts of both graph- and edge-inducibility extend naturally to r-uniform hypergraphs. Since the lower bound constructions of $1/e$ for $\ell = 1$ and for the star $K_{1,k-1}$ extend to higher uniformities as well, it makes sense to ask if Conjectures 1.1 and 1.2 would also hold in this more general setting. Needless to say, we expect these questions to be difficult.

Further questions

The constant of $1/e$, being closely related to the Poisson distribution, makes an appearance in many combinatorial and probabilistic set-ups. In particular, we would like to draw the reader's attention to the interesting conjectures of Feige [8] and Rudich (see [13]).

Addendum

In the period when this paper was under review, our results were extended in several directions. In particular, The Edge-statistics Conjecture (Conjecture 1.1) was proved in the superlinear regime by Kwan, Sudakov and Tran [14]. In the sublinear regime the conjecture was proved by Fox and Sauermann [9] and, independently, by Martinsson, Mousset, Noever and Trujić [15]. Thus, the results of [14] combined with either [9] or [15] prove the assertion of Conjecture 1.1. Some questions which remain open can be found in these papers.

Acknowledgement

We thank the anonymous referee for carefully reading our paper and for providing helpful remarks.

References

[1] Ahlswede, R. and Katona, G. O. H. (1978) Graphs with maximal number of adjacent pairs of edges. *Acta Math. Acad. Sci. Hungar.* **32** 97–120.
[2] Alon, N., Gutin, G. and Krivelevich, M. (2004) Algorithms with large domination ratio. *J. Algorithms* **50** 118–131.
[3] Alon, N. and Spencer, J. H. (2016) *The Probabilistic Method*, fourth edition, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley.
[4] Balogh, J., Hu, P., Lidický B. and Pfender, F. (2016) Maximum density of induced 5-cycle is achieved by an iterated blow-up of 5-cycle. *Eur. J. Combin.* **52** 47–58.
[5] Bollobás, B. (1998) *Modern Graph Theory*, Springer.
[6] Brown, J. I. and Sidorenko, A. (1994) The inducibility of complete bipartite graphs. *J. Graph Theory* **18** 629–645.
[7] Erdős, P. (1945) On a lemma of Littlewood and Offord. *Bull. Amer. Math. Soc.* **51** 898–902.
[8] Feige, U. (2004) On sums of independent random variables with unbounded variance, and estimating the average degree in a graph. In *Proceedings of the Thirty-Sixth Annual ACM Symposium on the Theory of Computing*, ACM, pp. 594–603.
[9] Fox, J. and Sauermann, L. (2018) A completion of the proof of the Edge-statistics Conjecture. arXiv:1809.01352
[10] Hefetz, D. and Tyomkyn, M. (2018) On the inducibility of cycles. *J. Combin. Theory Ser. B* **133** 243–258.
[11] Král, D., Norin, S. and Volec, J. (2018) A bound on the inducibility of cycles. arXiv:1801.01556
[12] Janson, S., Łuczak, T. and Ruciński, A. (2000) *Random Graphs*, Wiley.

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press

[13] Kahn, J., Saks, M. and Smyth, C. (2011) The dual BKR inequality and Rudich's conjecture. *Combin. Probab. Comput* **20** 257–266.
[14] Kwan, M., Sudakov, B. and Tran, T. (2019) Anticoncentration for subgraph statistics. *J. Lond. Math. Soc.* **99** 757–777.
[15] Martinsson, A., Mousset, F., Noever, A. and Trujić, M. (2018) The edge-statistics conjecture for $\ell \ll k^{6/5}$. arXiv:1809.02576
[16] Pippenger, N. and Golumbic, M. C. (1975) Inducibility of graphs. *J. Combin. Theory Ser. B* **19** 189–203.
[17] Yuster, R. (2018) On the exact maximum induced density of almost all graphs and their inducibility. arXiv:1801.01047
[18] Zykov, A. A. (1949) On some properties of linear complexes. *Mat. Sbornik* N.S. **24** 163–188.

Cite this article: Alon N, Hefetz D, Krivelevich M and Tyomkyn M (2020). Edge-statistics on large graphs. *Combinatorics, Probability and Computing* **29**, 163–189. https://doi.org/10.1017/S0963548319000294

https://doi.org/10.1017/S0963548319000294 Published online by Cambridge University Press