THE EDGE-STATISTICS CONJECTURE FOR $\ell \ll k^{6/5}$

ANDERS MARTINSSON, FRANK MOUSSET, ANDREAS NOEVER, AND MILOŠ TRUJIĆ

Abstract. Let $k$ and $\ell$ be positive integers. We prove that if $1 \leq \ell \leq o_k(k^{6/5})$, then in every large enough graph $G$, the fraction of $k$-vertex subsets that induce exactly $\ell$ edges is at most $1/e + o_k(1)$. Together with a recent result of Kwan, Sudakov, and Tran, this settles a conjecture of Alon, Hefetz, Krivelevich, and Tyomkyn.

1. Introduction

Given a graph $G$ and some $k \in \mathbb{N}$, let us write $X_{G,k}$ for the random variable corresponding to the number of edges induced by a subset $A \subseteq V(G)$ chosen uniformly at random among all subsets of size $k$. Define $I(n,k,\ell) := \max\{P[X_{G,k} = \ell] : v(G) = n\}$, the maximum probability of $X_{G,k} = \ell$ among all $n$-vertex graphs $G$. A standard averaging argument shows that the function $I(n,k,\ell)$ is decreasing in $n$, which implies that the limit

$$\text{ind}(k,\ell) := \lim_{n \to \infty} I(n,k,\ell)$$

exists. Observe that $\text{ind}(k,\ell) = \text{ind}(k,\binom{k}{2} - \ell)$. By considering the empty/complete graphs on $n$ vertices, it is moreover easy to see that $\text{ind}(k,0) = \text{ind}(k,\binom{k}{2}) = 1$, for all $k$. However, once we exclude the cases $\ell \in \{0, \binom{k}{2}\}$, it is sensible to suspect that $\text{ind}(k,\ell)$ becomes much smaller. For example, the quantitative version of Ramsey’s theorem implies that if $G$ is sufficiently large, then there is a positive probability that $A$ is either a clique or an independent set, which shows that $\text{ind}(k,\ell) < 1$ for all $\ell \notin \{0, \binom{k}{2}\}$.

The function $\text{ind}(k,\ell)$ was introduced by Alon, Hefetz, Krivelevich, and Tyomkyn [1], motivated by a connection to the notion of graph inducibility introduced earlier by Pippinger and Golumbic [8] (and which has recently become a rather popular topic, see for example [2, 4, 5, 6, 9]). In [1], Alon, Hefetz, Krivelevich, and Tyomkyn advanced three conjectures concerning the asymptotics of the function $\text{ind}(k,\ell)$ as $k \to \infty$.

Conjecture 1.1. For all $k, \ell \in \mathbb{N}$ with $0 < \ell < \binom{k}{2}$, we have $\text{ind}(k,\ell) \leq 1/e + o_k(1)$.

Conjecture 1.2. For all $k, \ell \in \mathbb{N}$ with $\min\{\ell, \binom{k}{2} - \ell\} = \omega_k(k)$, we have $\text{ind}(k,\ell) = o_k(1)$.

Conjecture 1.3. For all $k, \ell \in \mathbb{N}$ with $\min\{\ell, \binom{k}{2} - \ell\} = \Omega_k(k^2)$, we have $\text{ind}(k,\ell) = O_k(k^{-1/2})$.

Here, the subscript $k$ indicates that the asymptotic notation is understood as $k \to \infty$; for example, $o_k(1)$ denotes a function of $k$ tending to zero as $k \to \infty$. Several partial results on all three conjectures are given in [1].

Note that Conjecture 1.2 implies Conjecture 1.1 in the range where $\ell = \omega_k(k)$. Very recently, Kwan, Sudakov, and Tran [7] gave a proof of Conjecture 1.2 and showed that Conjecture 1.3 holds up to a polylogarithmic factor in $k$. The purpose of this paper is to give a proof of Conjecture 1.1 for all $1 \leq \ell \leq o_k(k^{6/5})$. Together with the result of [7], this result thus implies Conjecture 1.1 for all $\ell$.

Date: 12th August 2021.

Research supported by ISF grants 1028/16 and 1147/14, and ERC Starting Grant 633509 (FM), and by grant no. 200021 169242 of the Swiss National Science Foundation (MT).

Part of this work has been completed at a workshop of the research group of Angelika Steger in Buchboden in July 2018.
Theorem 1.4. For every \( \ell = \ell(k) \in \mathbb{N} \) such that \( 1 \leq \ell \leq o(k^{6/5}) \), we have
\[
\text{ind}(k, \ell) \leq 1/e + o_k(1).
\]

Even more recently, and independently of our own work, Fox and Sauermann [3] also gave a proof of Conjecture 1.1. The proof given here has the advantage that it is considerably shorter than the one in [3]. However, [3] contains some stronger bounds in certain ranges of \( \ell \) (e.g., it is shown that in fact \( \text{ind}(k, \ell) = o_k(1) \) when \( \omega_k(1) \leq \ell \leq o(k) \)), as well as results for the analogous problem in hypergraphs and other related results.

As noted in [1], the upper bound \( 1/e + o_k(1) \) in Theorem 1.4 is tight for example when \( \ell = 1 \), as can be seen by considering a random graph \( G_{n,p} \) where \( p = 1/\binom{k}{2} \). Similarly, the upper bound is tight for \( \ell = k - 1 \), as evidenced by the complete bipartite graph with parts of size \( n/k \) and \( (k-1)n/k \). It would be interesting to know whether the bound given by Theorem 1.4 is tight for some values of \( \ell \) besides 1 and \( k - 1 \).

2. A short proof for the case \( \ell = o_k(k) \)

Before presenting the full proof of Theorem 1.4, we give a short and self-contained proof for the case when \( \ell = o_k(k) \).

Proposition 2.1. For every \( \ell = \ell(k) \in \mathbb{N} \) such that \( 1 \leq \ell \leq o_k(k) \), we have
\[
\text{ind}(k, \ell) \leq 1/e + o_k(1).
\]

Proof. Choose \( k \) and \( \ell \) as in the statement and assume that \( n = n(k) \) is sufficiently large. Let \( G = (V, E) \) be a graph on \( n \) vertices and let \( v = (v_1, v_2, \ldots) \) be an infinite sequence of vertices chosen uniformly at random from \( V^\mathbb{N} \). We inductively colour the vertices in \( v \) with two colours, \textit{black} and \textit{green}, as follows:

1. \( v_1 \) is black;
2. \( v_i \) is green if and only if the graph induced by \( v_i \) and the \textit{black} vertices \( v_j \) with \( j < i \) contains at least \( \ell \) edges; otherwise, \( v_i \) is black.

Set \( L = L(v) := \min \{ i \geq 1 : \text{there are } k - 1 \text{ black vertices among } v_1, \ldots, v_i \} \) and \( L := \infty \) if there are fewer than \( k - 1 \) black vertices in \( v \). We then define \( Y_{G,k} = Y_{G,k}(v) \) as the random variable corresponding to the number of green vertices in the set \( \{ v_i : 1 \leq i < L \} \).

We first show that
\[
\mathbb{P}[X_{G,k} = \ell] \leq \mathbb{P}[Y_{G,k} = 1] + o_k(1). \tag{1}
\]

This can be seen as follows. Let \( \tilde{X}_{G,k} = e(\{ v_1, \ldots, v_k \}) \) and let \( A \) be the event that \( v_1, \ldots, v_k \) are all distinct. If \( n \) is sufficiently large given \( k \) (i.e., \( n = \omega(k^2) \)), then \( \mathbb{P}[A] = 1 - o_k(1) \). Thus
\[
\mathbb{P}[X_{G,k} = \ell] = \mathbb{P}[\tilde{X}_{G,k} = \ell \mid A] \leq \mathbb{P}[\tilde{X}_{G,k} = \ell / \mathbb{P}[A] \leq \mathbb{P}[\tilde{X}_{G,k} = \ell] + o_k(1). \tag{2}
\]

Next, since \( \ell \) edges can span at most \( 2\ell \) vertices, it follows by symmetry that
\[
\mathbb{P}[\tilde{X}_{G,k} = \ell] \geq \mathbb{P}[\tilde{X}_{G,k-1} = \ell] \geq \mathbb{P}[\tilde{X}_{G,k} = \ell] \cdot \frac{k - 2\ell}{k} \geq \mathbb{P}[\tilde{X}_{G,k} = \ell] - o_k(1), \tag{3}
\]

where the last inequality uses \( \ell = o(k) \) (in fact, this is the only place where we use this assumption). Finally, and crucially, observe that \( \tilde{X}_{G,k} = \tilde{X}_{G,k-1} = \ell \) implies \( Y_{G,k} = 1 \): if \( \tilde{X}_{G,k} = \tilde{X}_{G,k-1} = \ell \), then at least one green vertex must appear before the \((k - 1)\)st black vertex, and if there is more than one such green vertex, then \( \tilde{X}_{G,k} > \ell \). From this, together with (2) and (3), it follows that \( \mathbb{P}[X_{G,k} = \ell] \leq \mathbb{P}[Y_{G,k} = 1] + o_k(1) \), as claimed. Therefore, it suffices to show that \( \mathbb{P}[Y_{G,k} = 1] \leq 1/e \).
Let $u = (u_1, \ldots, u_{k-1})$ be a sequence of $k - 1$ (not necessarily distinct) vertices of $G$. Let $U(u)$ be the event that $u_1, \ldots, u_{k-1}$ are the first $k - 1$ black vertices in $v$. Now observe that if $\mathbb{P}[U(u)]$ is nonzero, then the conditional distribution of $Y_{G,k}$ given $U(u)$ is given by the sum

$$\text{Geom}(p_1) + \text{Geom}(p_2) + \cdots + \text{Geom}(p_{k-2})$$

of independent geometric distributions with parameters

$$p_i := \frac{1}{n}\{v \in V : e(\{u_1, \ldots, u_i, v\}) \geq \ell\}.$$  

Indeed, suppose that we have chosen the vertices $v_1, v_2, \ldots, v_t = u_i$ up to $u_i$. From then on, each vertex that we choose from the set $\{v \in V : e(\{u_1, \ldots, u_i, v\}) \geq \ell\}$ is green, while the first vertex that we choose outside of this set is the next black vertex $u_{i+1}$. It follows that

$$\mathbb{P}[Y_{G,k} = 1 \mid U(u)] = \sum_{i=1}^{k-2} p_i \prod_{j=1}^{k-2} (1 - p_j) \leq \sum_{i=1}^{k-2} p_i \cdot e^{-\sum_{j=1}^{k-2} p_j} \leq 1/e,$$

using that $f(x) = xe^{-x}$ is maximised for $x = 1$. Since this is true for every relevant choice of $u$, we also have $\mathbb{P}[Y_{G,k} = 1] \leq 1/e$ unconditionally. The proposition then follows using (1). \qed

3. Proof of Theorem 1.4

We use the following simple facts about hypergeometric random variables.

**Lemma 3.1.** Let $X$ be hypergeometric random variable counting the number of successes obtained when sampling $m$ elements without replacement from a population of size $N$ containing $Np$ successes. Assume $m^2/N \to 0$ and $m \to \infty$. If $mp \to \lambda < \infty$, then

$$\max_i \left| \mathbb{P}[X = i] - \frac{\lambda^i e^{-\lambda}}{i!} \right| \to 0,$$

where the maximum is taken over all nonnegative integers. On the other hand, if $mp(1-p) \to \infty$, then

$$\max_i \mathbb{P}[X = i] \to 0.$$  

**Proof.** Let $Y$ be defined in the same way as $X$ except that the $m$ elements are sampled with replacement. Let $A$ be the event that all $m$ elements are distinct in this experiment. Then we have $\mathbb{P}[X = i] = \mathbb{P}[Y = i \mid A]$. The assumption $m^2/N \to 0$ implies that $\mathbb{P}[A] \to 1$ and hence

$$\max_i \left| \mathbb{P}[X = i] - \mathbb{P}[Y = i] \right| \to 0.$$  

Note that $\mathbb{P}[Y = i] = \frac{\lambda^i p^i (1-p)^{m-i}}{\binom{m}{i}}$. Then the first assertion follows from the usual Poisson approximation to the binomial distribution. Similarly, the second assertion follows from the de Moivre–Laplace theorem. \qed

Let $k$ and $\ell$ be such that $1 \leq \ell \leq o_k(k^{6/5})$ and assume that $G$ is graph with $n$ vertices, where we assume that $n = n(k)$ is sufficiently large to support our arguments. We always interpret asymptotic statements as $k \to \infty$, and thus omit the subscript $k$ in the asymptotic notation from now on. We say that an event holds with high probability (w.h.p. for short) if the probability that it holds approaches 1 as $k \to \infty$.

For two events $\mathcal{E} = \mathcal{E}(k)$ and $\mathcal{F} = \mathcal{F}(k)$ (which can thus also depend on $\ell$, $G$, and $n$), we say that $\mathcal{E}$ is essentially contained in $\mathcal{F}$, and write $\mathcal{E} \ll \mathcal{F}$, if $\mathbb{P}(\mathcal{E} \setminus \mathcal{F}) = o(1)$.

As in the introduction, let $A$ denote a uniformly random subset of $V(G)$ of size $k$ and set $X_{G,k} = e(A)$. Throughout the proof, we let $\mathcal{E}$ denote the event that $X_{G,k} = \ell$.

Observe that it is enough to show that $\mathcal{E}$ is essentially contained in an event of probability $1/e + o(1)$. To define this event, let first $(w_k)_{k \geq 1}$ be a sequence of positive real numbers that goes to infinity at a sufficiently slow rate and, for every integer $d \geq 0$, define the event

$$\mathcal{D}_d := \{\text{all but at most } w_k \sqrt{\ell} \text{ vertices in } A \text{ have degree } d \text{ in } G[A]\}.$$
In particular, we choose $w_k$ such that $w_k \sqrt{t} = o(k)$. Our main goal is to show that there exists some deterministic value $d = d(G, k, \ell)$ such that $E \subseteq E \cap D_d$. This is sufficient by the following claim.

**Claim 3.2.** **For every** $d \geq 0$, **we have**

$$P[E \cap D_d] \leq 1/e + o(1).$$

**Proof.** Assume first that $d \geq 1$. Let $v$ be a vertex chosen uniformly at random among the vertices in $A$. Since $D_d$ implies that all but $o(k)$ vertices of $A$ have degree $d$ in $G[A]$, we have $P[e(v, A) = d \mid D_d] = 1 - o(1)$ and thus

$$P[E \cap D_d] \leq P[D_d] \leq (1 + o(1))P[e(v, A) = d] \leq P[e(v, A) = d] + o(1).$$

Note also that we have $P[E \cap D_d] = 0$ unless $kd \leq 3\ell$ and so using $\ell = o(k^{6/5})$ we can assume $d = o(k^{1/5})$. Note next that we can generate the pair $(v, A)$ by choosing first a uniformly random vertex $v$ in $V(G)$ and then choosing the remaining $k - 1$ vertices of $A$ uniformly among the $(k - 1)$-element subsets of $V(G) \setminus \{v\}$. In particular, if we fix the choice of $v$, then $e(v, A)$ follows a hypergeometric distribution with sample size $k - 1$ and a population of size $n - 1$ comprising $d_G(v)$ successes. If $d_G(v) \geq n/2$ then it follows from $d = o(k)$ and Markov’s inequality that $e(v, A) > d$ with probability $1 - o(1)$. On the other hand, if $d_G(v) \leq n/2$, then Lemma 3.1 implies that we either have $P[e(v, A) = d] = o(1)$ or

$$P[e(v, A) = d] = \frac{\lambda^d e^{-\lambda}}{d^d} + o(1)$$

for some $\lambda \geq 0$. Optimising the value of $\lambda$, we see that

$$\frac{\lambda^d e^{-\lambda}}{d^d} \leq \frac{d^d e^{-d}}{d^d} \leq 1/e,$$

where the last inequality uses $d \geq 1$.

Suppose next that $d = 0$. In this case we can proceed as in the proof of Proposition 2.1, if, instead of (1), we can show that

$$P[E \cap D_0] \leq P[Y_{G,k} = 1] + o(1). \quad (4)$$

Assume the process is the same as in Proposition 2.1 and that $Y_{G,k}$, $\tilde{X}_{G,k}$, and $A$ are defined in the same way. Then (4) can be seen as follows. Let $\tilde{D}_0$ be the event that all but at most $w_k \sqrt{t} = o(k)$ of the vertices $v_1, \ldots, v_k$ are isolated in $G[\{v_1, \ldots, v_k\}]$. We have

$$P[E \cap D_0] = P[\tilde{D}_0] \text{ and } \tilde{X}_{G,k} = \ell \mid A \leq P[\tilde{D}_0] \text{ and } \tilde{X}_{G,k} = \ell \mid P[A] \leq P[\tilde{D}_0] \text{ and } \tilde{X}_{G,k} = \ell + o(1).$$

Since each permutation of $v_1, \ldots, v_k$ is equally likely, we further have

$$P[\tilde{X}_{G,k} = \tilde{X}_{G,k-1} = \ell] \geq P[\tilde{D}_0] \text{ and } \tilde{X}_{G,k} = \ell - \frac{w_k \sqrt{t}}{k},$$

where the error term in the right hand side is $o(1)$ provided $w_k$ increases slowly enough. As $\tilde{X}_{G,k} = \tilde{X}_{G,k-1} = \ell$ implies $Y_{G,k} = 1$ deterministically, the proof of (4) is complete. \hfill \Box

It remains to show that there is some $d = d(G, k, \ell)$ such that $E \subseteq D_d$. We do this over a series of claims. First, let us define the event

$$D_* = \bigcup_{d \geq 0} D_d = \{\text{all but at most } w_k \sqrt{t} \text{ vertices in } A \text{ have the same degree in } G[A]\}.$$

The first claim we need is the following:

**Claim 3.3.** **We have** $E \subseteq D_*$. 

The somewhat technical proof of Claim 3.3 is deferred to the end of the paper. With this claim at hand, we continue with the proof of the theorem. We partition the vertices of $G$ into two sets:

- the heavy vertices $V_{\text{heavy}} := \{v \in V(G) : \deg_G(v) \geq n\ell^{1/3}/k\}$;
- the light vertices $V_{\text{light}} := \{v \in V(G) : \deg_G(v) < n\ell^{1/3}/k\}$.

We first show that we can assume that there are not too many heavy vertices.

**Claim 3.4.** Assume that $\ell = \omega(1)$ and that $G$ contains more than $5\ell^{2/3} n/k$ heavy vertices. Then $\Pr[\mathcal{E}] = o(1)$.

**Proof.** We generate $A$ by first choosing a random set $A_1$ of size $k/2$ and then choosing another random set $A_2 \subseteq V(G) \setminus A_1$ of size $k/2$. We have

$$E[|A_1 \cap V_{\text{heavy}}|] = \frac{|V_{\text{heavy}}| k}{2n} \geq \frac{5\ell^{2/3}}{2} = \omega(1).$$

In particular, the Chernoff bounds for the hypergeometric distribution imply that w.h.p. $A_1$ contains at least $2.49\ell^{2/3}$ heavy vertices. Expose the set $A_1$ and assume that this is the case. Then every (fixed) vertex $v \in A_1 \cap V_{\text{heavy}}$ satisfies $E[e(v, A_2)] \geq (1 - o(1))\ell^{1/3}/2$. Hence, again by the Chernoff bounds and a union bound over an arbitrary set $S \subseteq A_1 \cap V_{\text{heavy}}$ of size $2.49\ell^{2/3}$, we get

$$\Pr[\exists v \in S : e(v, A_2) < 0.49\ell^{1/3}] = o(1).$$

In particular, the union $A = A_1 \cup A_2$ w.h.p. contains at least

$$2.49\ell^{2/3} \cdot 0.49\ell^{1/3} > \ell$$

edges of $G$, implying $\Pr[\mathcal{E}] = o(1)$. \hfill $\square$

**Claim 3.5.** Let $Z := \sum_{v \in A} e(v, A)$. Assume that $\ell = \omega(1)$. Then either

$$\Pr[\mathcal{E}] = o(1)$$

or

$$\text{Var}[X_{G,k} - Z] \leq 30\ell^{5/3}. $$

**Proof.** Let $H := e(A \cap V_{\text{heavy}})$ and $L := e(A \cap V_{\text{light}})$ and observe that $X_{G,k} - Z = H - L$. Using the elementary inequality $(a - b)^2 \leq 2a^2 + 2b^2$, we have

$$\text{Var}[X_{G,k} - Z] = \text{Var}[H - L] \leq 2\text{Var}[H] + 2\text{Var}[L].$$

For any edge $e \in G$, let $X_e$ denote the indicator random variable for the event that both endpoints of $e$ are contained in $A$. We have

$$\text{Var}[H] = \sum_{e \in G} \sum_{f \in G} \text{Cov}[X_e, X_f]$$

and

$$\text{Var}[L] = \sum_{e \in G} \sum_{f \in G} \text{Cov}[X_e, X_f].$$

For each of these sums, an elementary calculation shows that $\text{Cov}[X_e, X_f] \leq 0$ if $e$ and $f$ do not have a common endpoint. On the other hand, if $e$ and $f$ intersect in exactly one endpoint, we have $\text{Cov}[X_e, X_f] \leq \E[X_e] \cdot \E[X_f] \cdot (k/n)$. Lastly, we have $\text{Cov}[X_e, X_e] = \text{Var}[X_e] \leq \E[X_e]$. Let $\mu_1 := \E[H]$ and $\mu_2 := \E[L]$. Since we may assume $|V_{\text{heavy}}| \leq 5\ell^{2/3} n/k$ (as otherwise Claim 3.4 implies $\Pr[\mathcal{E}] = o(1)$), we then obtain

$$\text{Var}[H] \leq e(\text{heavy}) \cdot \E[X_e] + 2 \cdot e(\text{heavy}) \cdot \frac{5\ell^{2/3} n}{k} \cdot \E[X_e] \cdot \frac{k}{n} \leq (1 + o(1))\mu_1 \cdot 10\ell^{2/3}.$$
Similarly, using the fact that every light vertex has degree at most \( n\ell^{1/3}/k \), we get
\[
\text{Var}[L] \leq e(V_{\text{light}}) \cdot E[X_e] + 2 \cdot e(V_{\text{light}}) \cdot \frac{n\ell^{1/3}}{k} \cdot \frac{k}{n} = (1 + o(1))\mu_2 \cdot 2\ell^{1/3}.
\]

If either of \( \mu_1 \) or \( \mu_2 \) is greater than, say, 1.01\( \ell \), then by Chebyshev’s inequality, the corresponding random variable \( H \) or \( L \) is concentrated around its expectation, which (since \( H, L \leq X_{G,k} \)) would imply that \( \Pr[X_{G,k} = \ell] = o(1) \). Otherwise, if \( \mu_1, \mu_2 \leq 1.01\ell \), we obtain the desired upper bound on \( \text{Var}[X_{G,k} - Z] \).

\[\square\]

**Claim 3.6.** Assume that \( \ell = \omega(\log^3 k) \). Then there exists some deterministic \( d = d(G, k, \ell) \) such that \( \mathcal{E} \subseteq \mathcal{D}_d \).

**Proof.** By Claim 3.4, we can assume that there are at most \( 5\ell^{2/3}n/k \) heavy vertices in \( G \), since otherwise \( \Pr[\mathcal{E}] = o(1) \) and then \( \mathcal{E} \subseteq \mathcal{D}_0 \) (say) holds trivially.

As in the statement of Claim 3.5, let \( Z := \sum_{v \in A \cap V_{\text{light}}} e(v, A) \). Again, since we are done when \( \Pr[\mathcal{E}] = o(1) \), we can assume that

\[
\text{Var}[X_{G,k} - Z] \leq 30\ell^{5/3}, \tag{5}
\]

using Claim 3.5.

We denote by \( D \) the random variable corresponding to the most frequent degree in \( G[A] \) (with ties broken arbitrarily). We first show that \( \mathcal{E} \) is essentially contained in each of the following events:

- \( \mathcal{F}_1 := \{ \text{every } v \in A \cap V_{\text{light}} \text{ satisfies } e(v, A) \leq 2\ell^{1/3} \} \),
- \( \mathcal{F}_2 := \{ \text{every } v \in A \cap V_{\text{heavy}} \text{ satisfies } e(v, A) \geq \ell^{1/3}/2 \} \),
- \( \mathcal{F}_3 := \{ X_{G,k} = Z + \mu + w_k\ell^{5/6} \} \), where \( \mu = \mathbb{E}[X_{G,k} - Z] \),
- \( \mathcal{F}_4 := \{ Z = kD \pm 3w_k\ell^{5/6} \} \).

Since \( \ell^{1/3} = \omega(\log k) \), the Chernoff bounds easily imply \( \Pr[\mathcal{F}_1 \cap \mathcal{F}_2] = 1 - o(1) \), so \( \mathcal{E} \subseteq \mathcal{F}_1 \) and \( \mathcal{E} \subseteq \mathcal{F}_2 \) hold trivially. For \( \mathcal{F}_3 \), note that using (5), Chebyshev’s inequality gives \( \Pr[\mathcal{F}_3] \leq O(1/w_k^2) = o(1) \), thus we have \( \mathcal{E} \subseteq \mathcal{F}_3 \) as well.

By Claim 3.3, we further know that \( \mathcal{E} \subseteq \mathcal{D}_d \), and therefore \( \mathcal{E} \subseteq \mathcal{E} \cap \mathcal{D}_d \cap \mathcal{F}_1 \cap \mathcal{F}_2 \). To prove that \( \mathcal{E} \subseteq \mathcal{F}_4 \), it is thus enough to show that \( \mathcal{E} \cap \mathcal{D}_d \cap \mathcal{F}_1 \cap \mathcal{F}_2 \subseteq \mathcal{F}_4 \) (note that this is a deterministic statement).

So assume that \( \mathcal{E} \cap \mathcal{D}_d \cap \mathcal{F}_1 \cap \mathcal{F}_2 \) holds. Since \( D \) is the most frequent degree in \( G[A] \), we see that \( \mathcal{E} \cap \mathcal{D}_d \) implies \( \ell = X_{G,k} \geq (k - o(k))D/2 \geq kD/3 \) for all sufficiently large \( k \). As \( \mathcal{F}_2 \) implies that every heavy vertex \( v \in A \) satisfies \( e(v, A) \geq \ell^{1/3}/2 \gg \ell/k \) (recall, \( \ell = o(k^{6/5}) \)), all of the at least \( k - w_k\ell^{1/3} \) vertices \( v \in A \) with \( e(v, A) = D \) are light. It follows that
\[
(k - w_k\ell^{1/3})D \leq Z \leq kD + 2w_k\ell^{5/6},
\]
where the upper bound is implied by \( \mathcal{F}_1 \). Therefore, using \( D \leq 3\ell/k \),
\[
kD - w_k\ell^{1/3} \cdot 3\ell/k \leq Z \leq kD + 2w_k\ell^{5/6}.
\]
Since \( \ell = o(k^{3/2}) \), we have \( \ell^{3/2}/k = o(\ell^{5/6}) \), so the above implies \( \mathcal{F}_4 \). It follows that \( \mathcal{E} \subseteq \mathcal{F}_4 \).

Finally, note that \( \mathcal{E} \cap \mathcal{F}_3 \cap \mathcal{F}_4 \) gives
\[
D = \ell - \frac{\mu}{k} \pm \frac{w_k}{k} \cdot O(\ell^{5/6}).
\]
By letting \( w_k \) be a sufficiently slowly diverging function, the error term in the right hand side is \( o(1) \) (using in addition \( \ell = o(k^{6/5}) \)), meaning there is only (at most) one possible integer value of \( D \) that can satisfy this. Let \( d \) be this value. Then \( \mathcal{E} \subseteq \mathcal{E} \cap \mathcal{D}_d \cap \mathcal{F}_3 \cap \mathcal{F}_4 \subseteq \mathcal{D}_d \), as desired. \( \square \)

Claims 3.2 and 3.6 imply that we have \( \Pr[X_{G,k} = \ell] \leq 1/e + o(1) \) for all \( \omega(\log^3 k) \leq \ell \leq o(k^{6/5}) \) (and we already proved the case \( 1 \leq \ell = o(k) \) in Section 2). Thus it only remains to prove Claim 3.3.
3.1. Proof of Claim 3.3. We now give the missing proof of Claim 3.3. Let \( m = k/\left(w_k^{1/3} \sqrt{\ell}\right) \). If \( w_k \) diverges sufficiently slowly, and using \( \ell = o(k^{6/5}) \), we have (say) \( m \geq w_k \). Observe that we can generate \( A \) by first choosing a random set \( S \) of size \( k - m \) and then choosing a random set \( Q \) of size \( m \) from the complement of \( S \). In terms of this process, we define the following events:

- \( \mathcal{E}_1 := \{e(Q) = 0\} \),
- \( \mathcal{E}_2 := \{e(S) + \sum_{v \in Q} e(v, S) = \ell\} \),
- \( \mathcal{E}_3 := \{\text{all but at most } w_k^{1/3} \text{ vertices in } Q \text{ have the same degree into } S\} \),
- \( \mathcal{E}_4 := \{\text{all but at most } w_k^{1/3} \text{ vertices in } Q \text{ have the same degree in } A\} \).

We prove that \( \mathcal{E} \) is essentially contained in each of these events, and then use this to conclude that \( \mathcal{E} \subseteq \mathcal{D}_s \).

We first prove that \( \mathcal{E} \subseteq \mathcal{E}_1 \). Since we can generate \( Q \) by first generating \( A \) and then choosing a random subset \( Q \subseteq A \) of size \( m \), we have

\[
\mathbb{E}[e(Q) \mid X_{G,k} = \ell] = \ell \cdot \binom{m}{2} / \binom{k}{2} = O(1/w_k^{2/3}),
\]

where the last inequality uses the definition of \( m \). Therefore, by Markov’s inequality,

\[
\mathbb{P}[X_{G,k} = \ell \text{ and } e(Q) \neq 0] \leq \mathbb{P}(X_{G,k} = \ell) \cdot O(1/w_k^{2/3}) = o(1),
\]

so \( \mathcal{E} \subseteq \mathcal{E}_1 \).

Having this, it follows directly from the definitions that \( \mathcal{E} \subseteq \mathcal{E} \cap \mathcal{E}_1 \subseteq \mathcal{E}_2 \).

Next, we show that \( \mathcal{E}_2 \subseteq \mathcal{E}_3 \), which then implies \( \mathcal{E} \subseteq \mathcal{E}_3 \). Expose first only the set \( S \) and let \( d_{med} \) be the median of \( e(v, S) \) over all \( v \in V(G) \setminus S \). We consider two cases, depending on the properties of the set \( S \).

Case 1. All but at most \( w_k^{1/4}n/m \) vertices \( v \in V(G) \setminus S \) satisfy \( e(v, S) = d_{med} \). Clearly, the expected number of vertices \( v \in Q \) for which \( e(v, S) \neq d_{med} \) is then at most \( O(w_k^{1/4}) = o(w_k^{1/3}) \). Thus, by Markov’s inequality, we have \( \mathbb{P}[\mathcal{E}_3] = 1 - o(1) \), which implies \( \mathcal{E}_2 \subseteq \mathcal{E}_3 \) in this case.

Case 2. At least \( w_k^{1/4}n/m \) vertices \( v \in V(G) \setminus S \) satisfy \( e(v, S) \neq d_{med} \). We claim that in this case, we have \( \mathbb{P}[\mathcal{E}_2] = o(1) \). We can assume that at least \( w_k^{1/4}n/(2m) \) vertices \( v \in V(G) \setminus S \) satisfy, say, \( e(v, S) > d_{med} \) (the case in which at least \( w_k^{1/4}n/(2m) \) vertices \( v \in V(G) \setminus S \) satisfy \( e(v, S) < d_{med} \) is analogous). Let us denote the number of such vertices by \( t \) and let \( N := |V(G) \setminus S| = n - k + m \).

Note that we can generate the set \( Q \) in the following way. First, let \( v'_1, v'_2, \ldots, v'_{N-t} \) be a random permutation of the vertices \( v \in V(G) \setminus S \) with \( e(v, S) \leq d_{med} \), and let \( v''_1, v''_2, \ldots, v''_t \) be a random permutation of the vertices \( v \in V(G) \setminus S \) with \( e(v, S) > d_{med} \). Let \( I \) be the random variable corresponding to the number of red balls one obtains when drawing \( m \) balls without replacement from a population of size \( N \) containing \( N - t \) red balls and \( t \) blue balls (in other words, let \( I \) be a hypergeometric random variable with these parameters). Finally, let

\[
Q = \{v'_1, v'_2, \ldots, v'_{t}, v''_1, v''_2, \ldots, v''_{m-I}\}.
\]

Note that in this way, \( Q \) is really a uniformly random \( m \)-element subset of \( V(G) \setminus S \).

Now, in order for \( \mathcal{E}_2 \) to occur we need

\[
\sum_{v \in Q} e(v, S) = \ell - e(S).
\]

Observe that for every fixed choice of the permutations \( v'_1, v'_2, \ldots, v'_{N-t} \) and \( v''_1, v''_2, \ldots, v''_t \), there is at most one value of \( I \) that achieves this. However, since \( I \) is a hypergeometric random variable with population size \( N \) and sample size \( m \), which satisfy \( m^2/N \leq k^2/(n - k) = o(1) \) if \( n = \omega(k^2) \), and since \( t \geq w_k^{1/4}n/(2m) = \omega(N/m) \) and \( t \leq N/2 \) (as \( d_{med} \) is a median) imply...
\(m(t/N)(1 - t/N) = \omega(1)\), it follows from Lemma 3.1 that \(P[I = i] = o(1)\). Thus in this case, we have \(P[E_3] = o(1)\), from which \(E_2 \subseteq E_3\) follows trivially.

Having shown \(E \subseteq E_1\) and \(E \subseteq E_3\), it follows easily from the definitions that \(E \subseteq E_1 \cap E_3 \subseteq E_4\).

Lastly, we show that \(E \subseteq D_\ast\), which completes the proof. Suppose that \(A\) is such that \(D_\ast\) does not occur. We show that, conditioning on this event (but leaving the subset \(Q \subseteq A\) random), the probability of \(E_4\) is \(o(1)\). For this, let \(d\) be the median degree in \(G[A]\). Then at least \(w_k^{1/2} \sqrt{t}/2\) vertices have degree, say, larger than \(d\) in \(G[A]\) (the case where \(w_k^{1/2} \sqrt{t}/2\) vertices have degree smaller than \(d\) is analogous). Let \(t\) be the number of such vertices in \(A\) and let \(X_t\) be the random variable denoting the number of such vertices in \(Q\) (which, recall, is a random subset of \(A\) of size \(m\)). Then since \(m = k/(w_k^{1/3} \sqrt{t})\), we have

\[
E[X_t] = t \cdot \frac{m}{k} \geq w_k \cdot \frac{m \sqrt{t}}{2k} = w_k^{2/3} / 2 = \omega(1),
\]

and \(\sigma(X_t) = O(\sqrt{tm/k})\). Therefore, by Chebyshev’s inequality, w.h.p. we have \(w_k^{1/3} \leq X_t\). On the other hand, as \(t \leq k/2\) (recall, \(d\) is a median), we also have w.h.p. \(X_t \leq (1/2 + o(1))m\).

Since \(w_k \ll m\), these two inequalities imply that there is no set of \(m - w_k^{1/4}\) vertices in \(Q\) which have the same degree in \(A\). Consequently, \(P(E_4 | D_\ast) = o(1)\), which implies \(P(E_4 \setminus D_\ast) = o(1)\), as desired. \(\square\)

**References**

[1] N. Alon, D. Hefetz, M. Krivelevich, and M. Tyomkin. Edge-statistics on large graphs. *arXiv preprint arXiv:1805.06848*, 2018.

[2] J. Balogh, P. Hu, B. Lidický, and F. Pfender. Maximum density of induced 5-cycle is achieved by an iterated blow-up of 5-cycle. *European Journal of Combinatorics*, 52:47–58, 2016.

[3] J. Fox and L. Sauer. A completion of the proof of the Edge-statistics Conjecture. *arXiv preprint arXiv:1809.01352*, 2018.

[4] H. Hatami, J. Hirst, and S. Norine. The inducibility of blow-up graphs. *Journal of Combinatorial Theory, Series B*, 109:196–212, 2014.

[5] D. Hefetz and M. Tyomkin. On the inducibility of cycles. *Journal of Combinatorial Theory, Series B*, 2018.

[6] D. Král’, S. Norin, and J. Volec. A bound on the inducibility of cycles. *Journal of Combinatorial Theory, Series A*, 161:359–363, 2019.

[7] M. Kwan, B. Sudakov, and T. Tran. Anticoncentration for subgraph statistics. *Journal of the London Mathematical Society*, 99(3):757–777, 2019.

[8] N. Pippenger and M. C. Golumbic. The inducibility of graphs. *Journal of Combinatorial Theory, Series B*, 19(3):189–203, 1975.

[9] R. Yuster. On the exact maximum induced density of almost all graphs and their inducibility. *Journal of Combinatorial Theory, Series B*, 136:81–109, 2019.

**Anders Martinsson, Department of Computer Science, ETH Zürich, 8092 Zurich, Switzerland**

Email address: anders.martinsson@inf.ethz.ch

**Frank Mousset, School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997801, Israel**

Email address: moussetfrank@gmail.com

**Andreas Noever, Department of Computer Science, ETH Zürich, 8092 Zurich, Switzerland**

Email address: andreas.noever@gmail.com

**Mišo Trujč, Department of Computer Science, ETH Zürich, 8092 Zurich, Switzerland**

Email address: mtrujic@inf.ethz.ch