The distribution of the maximum number of common neighbors in the random graph

I.V. Rodionov[1], M.E. Zhukovskii[2]

Abstract

Let $\Delta_{k;n}$ be the maximum number of common neighbors of a set of k vertices in $G(n,p)$. In this paper, we find a_n and σ_n such that $\frac{\Delta_{k;n}-a_n}{\sigma_n}$ converges in distribution to a random variable having the standard Gumbel distribution.

Keywords: binomial random graph, maximum degree, common neighbors, Gumbel distribution

1   Introduction

In 1980 [5], B. Bollobás studied the asymptotical behavior of the maximum degree Δ_n of the binomial random graph $G(n,p)$ ([4, 13]) for fixed $p \in (0,1)$. The main result of that paper is the following. Let, for $n \in \mathbb{N}$, $n \geq 2$,

$$a_n = pn + \sqrt{2p(1-p)n \ln n}\left(1 - \frac{\ln \ln n}{4 \ln n} - \frac{\ln(2\sqrt{\pi})}{2 \ln n}\right), \tag{1}$$

$$\sigma_n = \sqrt{\frac{p(1-p)n}{2 \ln n}}. \tag{2}$$

Then the shifted random variable $\frac{\Delta_n - a_n}{\sigma_n}$ converges in distribution to a random variable having the standard Gumbel distribution. Further in his paper, Bollobás considered the m-th highest degree Δ_n^m of $G(n,p)$ (in particular, $\Delta_n^1 = \Delta_n$) and proved a similar result: for every $y \in \mathbb{R}$,

$$\mathsf{P}\left(\frac{\Delta_n^m - a_n}{\sigma_n} \leq y\right) \to e^{-e^{-y}} \sum_{j=0}^{m-1} \frac{e^{-jy}}{j!} \text{ as } n \to \infty.$$

[1]Trapeznikov Institute of Control Sciences RAS, Moscow, Russian Federation
vecsell@gmail.com

[2]Moscow Institute of Physics and Technology (State University), laboratory of advanced combinatorics and network applications, Dolgoprodny, Moscow Region, Russian Federation; Adyghe State University, Caucasus mathematical center, Maykop, Republic of Adygea, Russian Federation; The Russian Presidential Academy of National Economy and Public Administration, Moscow, Russian Federation.
zhukmax@gmail.com

Note that elements ξ_i of a sequence of degrees of $G(n, p)$ have the binomial distribution with the parameters $n - 1, p$. For sequences of series of independent binomial random variables, the asymptotical distribution of maximums was studied by S. Nadarajah and K. Mitov in 2002 [16]. Their result states that the maximum D_n of n independent binomial random variables $\xi_1^{(n)}, \ldots, \xi_n^{(n)}$, where the upper index denotes the number of series, with parameters $M = M(n) = \omega(\ln^3 n)$ and $p = \text{const}$ (here and below we use the usual notation $g_1 = \omega(g_2)$ for two sequences $g_1(n)$ and $g_2(n)$ such that $g_2 = o(g_1)$) obeys the following asymptotical law: for $y \in \mathbb{R}$,

$$\mathsf{P}\left(D_n \leq pM + \sqrt{2p(1-p)M \ln n}\left[1 - \frac{\ln \ln n}{4 \ln n} - \frac{\ln(2\sqrt{\pi})}{2 \ln n} + \frac{y}{2 \ln n}\right]\right) \to e^{-e^{-y}} \text{ as } n \to \infty. \tag{3}$$

It is easy to see that, for $M = n - 1$, the result of Nadarajah and Mitov gives the same normalization functions a_n and σ_n and the same asymptotical distribution as the result of Bollobás (but for *dependent* random variables). Clearly, a simple substitution gives slightly different functions $\tilde{a}_n$ and $\tilde{\sigma}_n$. However, the convergence of $\mathsf{P}(\Delta_n \leq \tilde{a}_n + y\tilde{\sigma}_n)$ implies the convergence of $\mathsf{P}(\Delta_n \leq a_n + y\sigma_n)$ to the same limit since $a_n = \tilde{a}_n + O(1)$, $\sigma_n = \tilde{\sigma}_n + O((n \ln n)^{-1})$, and $\sigma_n = \Theta(\sqrt{n}/\sqrt{\ln n})$ (the relation $g_1 = \Theta(g_2)$ for two sequences $g_1(n)$ and $g_2(n)$ means that $g_1 = O(g_2)$ and $g_2 = O(g_1)$).

Results of such kind belong to the extreme value theory. A general result of this theory, the Fisher–Tippett–Gnedenko theorem [10] (see also [8], page 205) states the following. Let $\xi_1, \xi_2, \ldots$ be independent and identically-distributed random variables, $\xi^{(n)} = \max\{\xi_1, \ldots, \xi_n\}$. If there exist $a_n \in \mathbb{R}$, $\sigma_n > 0$ and a non degenerate distribution F such that $\frac{\xi^{(n)} - a_n}{\sigma_n}$ converges to an $\eta \sim F$ in distribution, then F belongs to either the Gumbel, the Fréchet or the Weibull family. In this way, three extremal types of distributions of ξ_i are distinguished. A comprehensive account of necessary and sufficient conditions for a distribution to belong to one of the extremal types is given in [15]. For further results of the extreme value theory and its applications see, e.g., [3, 11]. However, mentioned results can not be applied to the above case (when $\xi_1, \ldots, \xi_n$ are identically distributed but their distribution depends on n). Such *triangular arrays* of random variables were studied, e.g., in [2, 9, 16].

The result of Nadarajah and Mitov easily follows from a theorem about large deviations for Binomial random variables and certain properties of the normal distribution function. But how can (3) be obtained for dependent random variables $\xi_1, \ldots, \xi_n$? Bollobás, in his proof, introduced the random variable X being the number of vertices having degree greater than $y\sigma_n + a_n$, where a_n and σ_n are defined in (1) and (2) respectively. The result follows from the fact that, for every $j \in \mathbb{N}$, the j-th factorial moment of X converges to the j-th factorial moment of a Poisson random variable with the parameter e^{-y}. Obviously, the same

idea may be used to prove the result of Nadarajah and Mitov and the respective result for identically distributed (not depending on n) independent random variables under certain conditions (see [15], Chapter 2).

Let k be an arbitrary positive integer. In our paper, we solve the problem of finding an asymptotical distribution (precisely, the normalizing sequences $a_{k;n}$, $\sigma_{k;n}$ and an extremal type) of the maximum $\Delta_{k;n}$ of the number of common neighbors of k vertices in $G(n, p)$ (the case $k = 1$ is already solved by Bollobás since $\Delta_{1;n} = \Delta_n$). Notice that, in our paper, we consider not only constant p but p depending on n. In [12], it is proven that the result of Bollobás holds true for $p = p(n) \to 0$ as $n \to \infty$ such that $\frac{pn}{\ln^3 n} \to \infty$ as $n \to \infty$ (convergence to the Gumbel distribution but for *other* functions a_n, σ_n holds true even if $\frac{pn}{\ln n} \to \infty$, and the latter condition is optimal). In our paper, we obtain a similar result for $\Delta_{k;n}$ in the same most general settings.

For $k \geq 2$, this problem differs a lot from both mentioned problems (the case of independent binomial random variables and degrees of the random graph). The main difference is that the variance of an analogue of the random variable X defined above may approach infinity (e.g., this happens when $k = 2$ and $p > 1/2$). In particular, this fact makes it impossible to apply the method of factorial moments directly.

More formally, let $v_1, \ldots, v_k \in [n] := \{1, \ldots, n\}$ be distinct vertices of $G(n, p)$. Let $N_n(v_1, \ldots, v_k) \subseteq [n] \backslash \{v_1, \ldots, v_k\}$ be the set of all common neighbors of $v_1, \ldots, v_k$ in $G(n, p)$ ($u \in N_n(v_1, \ldots, v_k)$ if and only if, for every $i \in [k]$, $u \sim v_i$, i.e., u is adjacent to v_i). Set

$$\Delta_{k;n} = \max_{v_1, \ldots, v_k} |N_n(v_1, \ldots, v_k)|,$$

where the maximum is over all distinct vertices $v_1, \ldots, v_k \in [n]$. Moreover, let $\Delta_{k;n}^m$ be the m-th highest value among $|N_n(v_1, \ldots, v_k)|$ (in particular, $\Delta_{k;n} = \Delta_{k;n}^1$). The main result of our paper is given below.

Theorem 1 *Fix $y \in \mathbb{R}$ and $k, m \in \mathbb{N}$. Assume $p = p(n) \in (0, 1)$ is such that*

$$p^k \gg \frac{\ln^3 n}{n}, \quad 1 - p \gg \sqrt{\frac{\ln \ln n}{\ln n}}. \tag{4}$$

Let

$$a_{k;n} = np^k + \sqrt{2kp^k(1 - p^k)n \ln n} \left(1 - \frac{\ln[k!]}{2k \ln n} - \frac{\ln[4\pi k \ln n]}{4k \ln n}\right),$$

$$\sigma_{k;n} = \sqrt{\frac{p^k(1 - p^k)n}{2k \ln n}}.$$

3

Then

$$\mathsf{P}\left(\frac{\Delta_{k;n}^m - a_{k;n}}{\sigma_{k;n}} \leq y\right) \to e^{-e^{-y}} \sum_{i=0}^{m-1} \frac{e^{-yi}}{i!} \quad as \ n \to \infty.$$

Remark. The second condition may be strengthened: $1 - p \geq \sqrt{\frac{32 \ln \ln n}{k^3 \ln n}}(1 + o(1))$. For such p, the same techniques work. However, we give a proof in a weaker form to avoid some annoying technical details.

Note that $\Delta_{k;n}$ is the maximum over $\binom{n}{k}$ binomial random variables with parameters $n - k, p^k$. Therefore, our result duplicates the statement (3) (but for this special case of dependent random variables). This motivates the following question. How strong can be dependencies between binomial random variables until (3) fails? A partial answer on this question and other further questions are given in Section 3.

While Theorem 1 is a natural extension of the result of Bollobás, it is also motivated by the study of *extension counts*. This study was initiated by Spencer in [20]. He proved that, given a *strictly balanced grounded* pair of graphs (G, H) (where G has k vertices) and $\varepsilon > 0$, there exists $C = C(\varepsilon) > 0$ such that if $\mu > C \ln n$, then

$$\mathsf{P}\left(\max_{v_1,\ldots,v_k} |X(v_1,\ldots,v_k) - \mu| < \varepsilon\mu\right) \to 1 \quad as \ n \to \infty. \tag{5}$$

Here, $X(v_1,\ldots,v_k)$ is the number of (G, H)-*extensions* of the tuple $(v_1,\ldots,v_k)$ and $\mu = \mathsf{E}X(1,\ldots,k)$.

This result was recently strengthened by Šileikis and Warnke in [18]. They proved that there exist constants $c, C, \alpha > 0$ such that, for all $p = p(n) \in [0, 1]$ and $\varepsilon = \varepsilon(n) \in [n^{-\alpha}, 1]$, the limit probability in (5) equals 0 if $\mu < \frac{c}{\varepsilon^2} \ln n$, and equals 1 if $\mu > \frac{C}{\varepsilon^2} \ln n$. Theorem 1 implies, in particular, that, for p such that $p \gg \sqrt{\frac{\ln \ln n}{\ln n}}$ and $1 - p \gg \sqrt{\frac{\ln \ln n}{\ln n}}$ (the first restriction on p appears because we also need an asymptotical distribution of the minimum number of (G, H)-extensions, and it follows from Theorem 1 by considering the complement of $G(n, p)$ which is distributed as $G(n, 1 - p)$) and H having one more vertex than G adjacent to all the vertices of G, the threshold for the concentration result (5) is fully determined: the result of Šileikis and Warnke is true for any $c < 2k(1 - p^k)\left(1 - (1 - \delta)\frac{\ln \ln n}{2k \ln \ln n}\right)$, $C > 2k(1 - p^k)\left(1 - (1 + \delta)\frac{\ln \ln n}{2k \ln n}\right)$ and any $\alpha > 0$. Notice that c is bounded away from 0 only when p is bounded away from 1. However, it does not contradict the result of Šileikis and Warnke since our conclusion becomes non-trivial only when ε is close to $n^{-1/2}$ since, in our settings, $\mu = n^{1-o(1)}$. We hope that our methods can be further applied to get similar

4

results for other pairs of G and H.

The rest of the paper is organized as follows. In Section 2, we give a proof of Theorem 1. Section 3 is devoted to some discussions of our method and its possible applications to more general questions.

2 Proof of Theorem 1

Consider the random variables $X = X_n^k$ being the number of k-sets of vertices having more than

$$b = b_{k;n}(y) := a_{k;n} + y\sigma_{k;n}$$

common neighbors. The reason why $\mathrm{Var}\, X$ may approach infinity is that the major contribution to the variance is made by those k-sets that have proper subsets with large number of common neighbors. A.a.s., in $G(n, p)$ there are no such k-sets (for details, see Section 2.3). So, we are able to exclude such k-sets from X (in what follows, we denote the shifted random variable by $\tilde{X}$). In Section 2.3, we estimate the expectation and the second moment of $\tilde{X}$.

The second problem we face is that a direct implementation of the approach of Bollobás (i.e., estimation of all the factorial moments of $\tilde{X}$) requires heavy computations. Fortunately, we have shown that it is enough to prove that $\mathsf{E}\tilde{X}(\tilde{X}-1) \sim (\mathsf{E}X)^2$. This observation follows from the Janson-type inequality that we prove in Section 2.2 (in Section 3, we state it in a more general form and describe its possible applications to other problems related to extreme value theory). It is worth mentioning that the result of Bollobás would follow directly from the Janson inequality for general upsets ([17], Theorem 1, Inequality (3)), if, for distinct vertices u, v, the events A_u, A_v of having more than $b_{1;n}(y)$ neighbors of u and having more than $b_{1;n}(y)$ neighbors of v respectively were independent (indeed, for a fixed vertex v, the property A_v is *increasing*, i.e. is the upset). If the premise held, estimating the first moment would be enough to get the asymptotics of $\mathsf{P}(X_n^1 = 0)$. Unfortunately, any two A_u, A_v are dependent, and so, the mentioned Janson inequality implies $\mathsf{P}(X_n^1 = 0) \le e^{-e^{-y} + \frac{1}{2}e^{-2y} + o(1)}$.

Nevertheless, we prove Janson-type bounds for $\mathsf{P}(\tilde{X} = 0)$ for all k (Section 2.2, Lemma 1). It is easy to see that, for $k = 1$, the same result holds true for the original (not shifted) random variable X_n^1: $e^{-\mathsf{E}X_n^1 + o(1)} \le \mathsf{P}(X_n^1 = 0) \le e^{-\mathsf{E}X_n^1 + o\left(e^{\mathsf{E}X_n^1}\right)}$, and it immediately implies the result of Bollobás on Δ_n.

2.1 Notations and approximations of binomial distributions

In our proofs, we will frequently use the classical De Moivre–Laplace central limit theorem. For the sake of convenience, below, we verify that it gives (in our settings (4)) an approximation error which is sufficient for our goals.

Given $q = q(n) \in (0,1)$ and $x = x(n) \in \mathbb{R}$, assume that

$$\frac{\ln^3 n}{nq(1-q)} \to 0, \quad x^3 \sqrt{\frac{\ln^3 n}{nq(1-q)}} \to 0 \quad \text{as } n \to \infty. \tag{6}$$

Below, we find asymptotics for $\mathsf{P}(\xi_{n,q} = a)$ and $\mathsf{P}(\xi_{n,q} > a)$ where

$$a = \left\lfloor nq + x\sqrt{q(1-q)n\ln n} \right\rfloor$$

and $\xi_{n,q}$ is a binomial random variable with parameters n, q.

Using Stirling's formula, the relation $\ln(1+z) = z - \frac{z^2}{2} + O(z^3)$ and the definition of a, we get

$$\mathsf{P}(\xi_{n,q} = a) = \binom{n}{a} q^a (1-q)^{n-a} = \frac{1 + O(1/a) + O(1/(n-a))}{\sqrt{2\pi n}\sqrt{a/n}\sqrt{1-a/n}} \frac{q^a(1-q)^{n-a}}{(a/n)^a(1-a/n)^{n-a}}$$

$$= \frac{1 + O\left(x\sqrt{\ln n/(nq(1-q))} + 1/a + 1/(n-a)\right)}{\sqrt{2\pi nq(1-q)}} \exp\left\{-n\ln\left[\left(\frac{a/n}{q}\right)^{a/n}\left(\frac{1-a/n}{1-q}\right)^{1-a/n}\right]\right\}$$

$$\sim \frac{1}{\sqrt{2\pi nq(1-q)}} \exp\left\{-nq\left(1 + x\sqrt{\frac{(1-q)\ln n}{qn}}\right)\ln\left(1 + x\sqrt{\frac{(1-q)\ln n}{qn}}\right)\right.$$

$$\left. -n(1-q)\left(1 - x\sqrt{\frac{q\ln n}{(1-q)n}}\right)\ln\left(1 - x\sqrt{\frac{q\ln n}{(1-q)n}}\right)\right\}$$

$$= \frac{1}{\sqrt{2\pi nq(1-q)}} \exp\left\{-\frac{x^2\ln n}{2} + O\left(x^3\sqrt{\frac{\ln^3 n}{nq(1-q)}}\right)\right\} \sim \frac{1}{\sqrt{2\pi nq(1-q)}} \exp\left\{-\frac{x^2\ln n}{2}\right\}. \tag{7}$$

Assume that (in addition to the above conditions)

$$x \geq 0, \quad x\sqrt{\ln n} \to \infty, \quad \frac{x^3}{(1-q)^2}\sqrt{\frac{\ln^3 n}{n}} \to 0 \quad \text{as } n \to \infty. \tag{8}$$

6

Denote $\hat{x} = \frac{2x}{\sqrt{1-q}}$. Let

$$\hat{a} = nq + \hat{x}\sqrt{q(1-q)n\ln n}, \quad a^+ = a, \ a^- = nq - x\sqrt{q(1-q)n\ln n}.$$

Then, approximating the sum by an integral, we have

$$P(\xi_{n,q} > a^+) =$$

$$\sum_{j=a+1}^{\hat{a}} \left(1 + O\left(\frac{x^3}{(1-q)^2}\sqrt{\frac{\ln^3 n}{nq}}\right)\right)\frac{\exp\left(-(j-nq)^2/[2nq(1-q)]\right)}{\sqrt{2\pi nq(1-q)}} + P(\xi_{n,q} > \hat{a}) =$$

$$\int_{x\sqrt{\ln n}}^{\hat{x}\sqrt{\ln n}} \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt\left(1 + O\left(\frac{x^3}{(1-q)^2}\sqrt{\frac{\ln^3 n}{nq}}\right)\right) + P(\xi_{n,q} > \hat{a}) =$$

$$\left(1 - \Phi\left(x\sqrt{\ln n}\right)\right)(1 + o(1)) - \left(1 - \Phi\left(\hat{x}\sqrt{\ln n}\right)\right)(1 + o(1)) + P(\xi_{n,q} > \hat{a}),$$

where $\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz$. It is also worth to be noted that the O-term inside the integral on the second line of the latter relation converges to 0 by (6) if q is bounded away from 1 and by (8) if $q \to 1$. Applying the well-known relation

$$1 - \Phi(t) \sim \frac{1}{\sqrt{2\pi}t}e^{-t^2/2} \quad \text{as } t \to \infty \tag{9}$$

(see $(1')$ in [6]), we get

$$P(\xi_{n,q} > a^+) = \frac{\exp\left[-x^2\ln n/2\right]}{x\sqrt{2\pi\ln n}}(1 + o(1)) - \frac{\exp\left[-\hat{x}^2\ln n/2\right]}{\hat{x}\sqrt{2\pi\ln n}}(1 + o(1)) + P(\xi_{n,q} > \hat{a}) =$$

$$\frac{\exp\left[-x^2\ln n/2\right]}{x\sqrt{2\pi\ln n}}(1 + o(1)) + P(\xi_{n,q} > \hat{a}).$$

Next, by the Chernoff bound ([13], Theorem 2.1), for large n,

$$P\left(|\xi_{n,q} - nq| > \hat{a} - nq\right) \leq 2\exp\left[-\frac{\hat{x}^2 q(1-q)n\ln n}{2\left(nq + \hat{x}\sqrt{q(1-q)n\ln n}/3\right)}\right] \leq 2\exp\left[-\frac{4x^2\ln n}{3}\right].$$

Since

$$\exp\left[-\frac{4x^2\ln n}{3}\right] = o\left(\frac{\exp\left[-x^2\ln n/2\right]}{x\sqrt{2\pi\ln n}}\right),$$

7

we finally get

$$\mathsf{P}(\xi_{n,q} > a^+) \sim \frac{\exp\left[-x^2 \ln n/2\right]}{x\sqrt{2\pi \ln n}}. \tag{10}$$

In the same way,

$$\mathsf{P}(\xi_{n,q} < a^-) \sim \frac{\exp\left[-x^2 \ln n/2\right]}{x\sqrt{2\pi \ln n}}. \tag{10'}$$

In the remaining parts of the proof, we use the following notations. For every $\ell \in [k-1]$, denote

$$\Gamma_\ell = np^\ell + \sqrt{2\ell}\sqrt{np^\ell(1-p^\ell)\ln n}.$$

Let $\tilde{X} = \tilde{X}_n^k$ be the number of k-sets of vertices $U_k := \{u_1, \ldots, u_k\}$ such that, for every $\ell \in [k-1]$ and every distinct $i_1, \ldots, i_\ell \in [k]$, the following conditions hold:

$$|N_n(u_{i_1}, \ldots, u_{i_\ell})| \leq \Gamma_\ell \text{ and } |N_n(u_1, \ldots, u_k)| > b.$$

Fix a k-set $U = \{u_1, \ldots, u_k\} \in \binom{[n]}{k}$. For $W \subset [n] \setminus U$ and $W_i \subset [n] \setminus \{u_i\}$, $i \in [k]$, consider the following events

$$B_{U,W} = \{W = N_n(U)\}, \quad B_{U,W_1,\ldots,W_k} = \{W_1 = N_n(u_1), \ldots, W_k = N_n(u_k)\}.$$

Let

$$B_U = \bigvee_{W \subset [n]\setminus U:\, |W|>b} B_{U,W}, \quad \tilde{B}_U = \bigvee B_{U,W_1,\ldots,W_k},$$

where the second union is over sets $W_i \subset [n] \setminus \{u_i\}$, $i \in [k]$, such that $|W_1 \cap \ldots \cap W_k| > b$ and, for every $\ell \in [k-1]$ and every distinct $i_1, \ldots, i_\ell \in [k]$, $|W_{i_1} \cap \ldots \cap W_{i_\ell}| \leq \Gamma_\ell$. Thus, B_U is the event $\{N_n(U) > b\}$ and $\tilde{B}_U$ is the event $\{|N_n(U)| > b$ and, for every $\ell \in [k-1]$ and all distinct $i_1, \ldots, i_\ell \in [k]$, $|N_n(u_{i_1}, \ldots, u_{i_l})| \leq \Gamma_\ell\}$. In addition, note that

$$\{X = 0\} = \bigwedge_U \neg B_U \text{ and } \{\tilde{X} = 0\} = \bigwedge_U \neg \tilde{B}_U.$$

Recall that y is a constant. For a fixed U, after substituting

$$q = p^k, \quad a^+ = b = np^k + x\sqrt{p^k(1-p^k)n \ln n}, \quad x = \sqrt{2k}\left(1 - \frac{\ln \ln n}{4k \ln n} - \frac{\ln\left[k!\sqrt{4\pi k}\right] - y}{2k \ln n}\right), \tag{11}$$

8

in (10), we get

$$\mathsf{P}(B_U) = \mathsf{P}(\xi_{n-k,p^k} > b) \sim \frac{1}{2\sqrt{k\pi \ln n}} \exp \left[-k \ln n \left(1 - \frac{\ln \ln n}{4k \ln n} - \frac{\ln \left[k!\sqrt{4\pi k} \right] - y}{2k \ln n} \right)^2 \right]$$

$$\sim \frac{1}{2\sqrt{k\pi \ln n}} \exp \left[-k \ln n + \ln \sqrt{\ln n} + \ln \left[k!\sqrt{4\pi k} \right] - y \right] = \frac{k!}{n^k} e^{-y}. \tag{12}$$

2.2 The Janson-type inequality

Denote

$$\lambda = \sum_{U \in \binom{[n]}{k}} \mathsf{P}(B_U), \quad \tilde{\lambda} = \sum_{U \in \binom{[n]}{k}} \mathsf{P}(\tilde{B}_U), \tag{13}$$

$$\Delta = \sum_{U_1, U_2 \in \binom{[n]}{k} : U_1 \cap U_2 \neq \varnothing} \mathsf{P}(|N_n(U_1 \cap U_2)| \leq \Gamma_{|U_1 \cap U_2|}, |N_n(U_1)| > b, |N_n(U_2)| > b). \tag{14}$$

Lemma 1 *Under the conditions of Theorem 1, the following bounds hold:*

$$\exp \left[-\lambda + o(1) \right] \leq \mathsf{P}(X = 0) \leq \mathsf{P}(\tilde{X} = 0) \leq \exp \left[-(1 + o(1))\tilde{\lambda} + (1 + o(1))e^{\lambda}\Delta \right]. \tag{15}$$

Proof. The inequality $\mathsf{P}(X = 0) \leq \mathsf{P}(\tilde{X} = 0)$ follows from the definition of these random variables.

The inequality $e^{-\lambda + o(1)} \leq \mathsf{P}(X = 0)$ follows from [1, Theorem 6.3.3] which is itself a direct corollary of the well known FKG inequality [1, Theorem 6.2.1] (note that, in the case $p = 1/2$, this corollary is also known as Kleitman's Lemma [14], see also [1, Proposition 6.3.1]). Thus, since the indicator random variables of the events $\neg B_U$ are decreasing functions of the edges of the random graph, we get

$$\mathsf{P}(X = 0) = \mathsf{P} \left(\bigwedge_U \neg B_U \right) \geq \prod_U (1 - \mathsf{P}(B_U)) = e^{\sum_U \ln(1 - \mathsf{P}(B_U))}.$$

From (12), we get

$$\mathsf{P}(X = 0) \geq e^{-\lambda + o(1)}. \tag{16}$$

The proof of the remaining inequality is close to the proof of Janson's inequality proposed by Boppona and Spencer [7] (as well as to the proof of Suen's inequality proposed by

Spencer [19]). However, it is harder since we need to overcome two difficulties. First, any two of B's are not independent and, second, even after getting through the first barrier, we can not apply here the FKG inequality directly.

Let us consider an arbitrary ordering $\tilde{B}_1, \ldots, \tilde{B}_{\binom{n}{k}}$ of the events $\tilde{B}_U$. Then

$$\mathsf{P}(\tilde{X} = 0) = \mathsf{P}\left(\bigwedge_{i=1}^{\binom{n}{k}} \neg \tilde{B}_i\right) = \prod_{i=1}^{\binom{n}{k}} \left[1 - \mathsf{P}(\tilde{B}_i | \neg \tilde{B}_1 \wedge \ldots \wedge \neg \tilde{B}_{i-1})\right]. \tag{17}$$

Fix $i \in \left[\binom{n}{k}\right]$. Unfortunately, each event of $\tilde{B}_1, \ldots, \tilde{B}_{i-1}$ is not independent of $\tilde{B}_i$. Nevertheless, we may consider some slight modifications of them such that they are independent of $\tilde{B}_i$. Indeed, let U define $\tilde{B}_i$ (i.e., $\tilde{B}_i = \tilde{B}_U$). Consider all the events (say, $\tilde{B}_1, \ldots, \tilde{B}_d$) among $\tilde{B}_1, \ldots, \tilde{B}_{i-1}$ (they are defined by $U_1, \ldots, U_d$ respectively) such that each of $U_1, \ldots, U_d$ has an empty intersection with U. Note that $d = d(i)$ depends on i. However, we write d instead of $d(i)$ for shortening until the opposite is required. Let $j \in [d]$, and $\tilde{B}_j$ be defined by $U_j = \{u_1^j, \ldots, u_k^j\}$. In what follows, for $V \subset [n]$, we denote by $\hat{N}_n(V)$ the set of all common neighbors of V in $\mathcal{N} := [n] \setminus U$.

For $\hat{U} = \{\hat{u}_1, \ldots, \hat{u}_k\} \in \binom{\mathcal{N}}{k}$, define

$$C_{\hat{U}} = \bigvee_{W_1, \ldots, W_k} \left\{\hat{N}_n(\hat{u}_1) = W_1, \ldots, \hat{N}_n(\hat{u}_k) = W_k\right\}.$$

where the union is over sets $W_i \subset \mathcal{N} \setminus \{\hat{u}_i\}$, $i \in [k]$, such that $|W_1 \cap \ldots \cap W_k| > b - k$ and, for every $\ell \in [k-1]$ and every distinct $i_1, \ldots, i_\ell \in [k]$, $|W_{i_1} \cap \ldots \cap W_{i_\ell}| \leq \Gamma_\ell$. Consider an arbitrary ordering $\hat{U}_1, \ldots, \hat{U}_{\binom{n-k}{k}}$ of the sets $\hat{U} \in \binom{\mathcal{N}}{k}$ such that $\hat{U}_j = U_j$, $j \in [d]$. For $j \in \{1, \ldots, \binom{n-k}{k}\}$, set $C_j = C_{\hat{U}_j}$. Clearly $C_1, \ldots, C_{\binom{n-k}{k}}$ do not depend on $\tilde{B}_i$ since $\tilde{B}_i$ is defined only by edges having vertices in U. Note also that, generally speaking, the events C_j depend on i by the definition of $\mathcal{N}$, but, hereinafter, we write C_j instead of $C_j(i)$ to prevent overloading with double indexations.

By the proven lower bound (16) (note that this lower bound still holds true if we replace b with $b + O(1)$ in the definition of X),

$$\mathsf{P}\left(\bigwedge_{i=1}^{\binom{n-k}{k}} \neg C_i\right) \geq \mathsf{P}\left(\text{there is no } \hat{U} \in \binom{\mathcal{N}}{k} \text{ in } G(n-k, p) \text{ with } |\hat{N}_n(\hat{U})| > b - k\right)$$

$$\geq \mathsf{P}\left(\text{there is no } \hat{U} \in \binom{[n]}{k} \text{ in } G(n, p) \text{ with } |N_n(\hat{U})| > b - k\right) \geq e^{-\lambda + o(1)}. \tag{18}$$

Clearly,

$$P(\tilde{B}_i \wedge \neg C_1 \wedge \ldots \wedge \neg C_d) \leq P(\tilde{B}_i \wedge \neg \tilde{B}_1 \wedge \ldots \wedge \neg \tilde{B}_d) \tag{19}$$

and

$$P(\neg C_1 \wedge \ldots \wedge \neg C_d) \geq P(\neg \tilde{B}_1 \wedge \ldots \wedge \neg \tilde{B}_d) - P\left(\exists V \in \binom{[n]}{k} \text{ s.t. } |N_n(V)| \in \{b-k+1, \ldots, b\}\right) -$$

$$\sum_{\ell=1}^{k-1} P\left(\exists V \in \binom{[n]}{\ell} \text{ s.t. } |N_n(V)| \in \{\Gamma_\ell + 1, \ldots, \Gamma_\ell + k\}\right) = P(\neg \tilde{B}_1 \wedge \ldots \wedge \neg \tilde{B}_d) - o\left(\frac{1}{\ln n}\right) \tag{20}$$

by the union bound (the remainder $o(1/\ln n)$ does not depend on i). Indeed, for every $a \in \{b-k+1, \ldots, b\}$ and every $V \in \binom{[n]}{k}$, from (7), we get (here, we apply (7) with the parameters q, a, x defined in (11); note that additive $O(1)$ terms that come with n and a here do not affect asymptotics)

$$P\left(|N_n(V)| = a\right) = P\left(\xi_{n-k,p^k} = a\right) \sim \frac{\sqrt{2k \ln n} k! e^{-y}}{n^k \sqrt{np^k(1-p^k)}} = o\left(\frac{1}{n^k \ln n}\right).$$

For every $\ell \in \{1, \ldots, k-1\}$, $V \in \binom{[n]}{\ell}$, $a \in \{\Gamma_\ell + 1, \ldots, \Gamma_\ell + k\}$, after substituting

$$q = p^\ell, \quad a = np^\ell + x\sqrt{p^\ell(1-p^\ell)n \ln n}, \quad x = \sqrt{2\ell} + o\left(n^{-1/(2k)}\right)$$

in (7), we get

$$P\left(|N_n(V)| = a\right) = P\left(\xi_{n-\ell,p^\ell} = a\right) \sim \frac{1}{\sqrt{2\pi np^\ell(1-p^\ell)}} e^{-\ell \ln n} = o\left(\frac{1}{n^\ell(\ln n)^{3/2}}\right).$$

Below, we use a standard tool from the proof of Janson's inequality:

$$P(\tilde{B}_i | \neg \tilde{B}_1 \wedge \ldots \wedge \neg \tilde{B}_{i-1}) \geq \frac{P(\tilde{B}_i \wedge \neg \tilde{B}_1 \wedge \ldots \wedge \neg \tilde{B}_d \wedge \neg \tilde{B}_{d+1} \wedge \ldots \wedge \neg \tilde{B}_{i-1})}{P(\neg \tilde{B}_1 \wedge \ldots \wedge \neg \tilde{B}_d)} =$$

$$P(\tilde{B}_i | \neg \tilde{B}_1 \wedge \ldots \wedge \neg \tilde{B}_d) P(\neg \tilde{B}_{d+1} \wedge \ldots \wedge \neg \tilde{B}_{i-1} | \tilde{B}_i \wedge \neg \tilde{B}_1 \wedge \ldots \wedge \neg \tilde{B}_d) =$$

$$\frac{P(\neg C_1 \wedge \ldots \wedge \neg C_d)}{P(\neg B_1 \wedge \ldots \wedge \neg \tilde{B}_d)} \times \frac{P(\tilde{B}_i \wedge \neg \tilde{B}_1 \wedge \ldots \wedge \neg \tilde{B}_d)}{P(\tilde{B}_i \wedge \neg C_1 \wedge \ldots \wedge \neg C_d)} \times P(\tilde{B}_i) \times$$

$$P(\neg \tilde{B}_{d+1} \wedge \ldots \wedge \neg \tilde{B}_{i-1} | \tilde{B}_i \wedge \neg \tilde{B}_1 \wedge \ldots \wedge \neg \tilde{B}_d) \tag{21}$$

since $\tilde{B}_i$ is independent of $\neg C_1 \wedge \ldots \wedge \neg C_d$. It remains to estimate the factor from the last line. For every $j \in \{d+1, \ldots, i-1\}$, from (18) and (19),

$$\mathsf{P}(\tilde{B}_j | \tilde{B}_i \wedge \neg \tilde{B}_1 \wedge \ldots \wedge \neg \tilde{B}_d) = \frac{\mathsf{P}(\tilde{B}_j \wedge \tilde{B}_i \wedge \neg \tilde{B}_1 \wedge \ldots \wedge \neg \tilde{B}_d)}{\mathsf{P}(\tilde{B}_i \wedge \neg \tilde{B}_1 \wedge \ldots \wedge \neg \tilde{B}_d)} \leq \frac{\mathsf{P}(\tilde{B}_j \wedge \tilde{B}_i \wedge \neg \tilde{B}_1 \wedge \ldots \wedge \neg \tilde{B}_d)}{\mathsf{P}(\tilde{B}_i) \mathsf{P}(\neg C_1 \wedge \ldots \wedge \neg C_d)} \leq$$

$$\frac{\mathsf{P}(\tilde{B}_j \wedge \tilde{B}_i)}{\mathsf{P}(\tilde{B}_i) \mathsf{P}(\neg C_1 \wedge \ldots \wedge \neg C_{\binom{n-k}{k}})} \leq \mathsf{P}(\tilde{B}_j | \tilde{B}_i) e^{\lambda + o(1)}.$$

Therefore,

$$\mathsf{P}(\neg \tilde{B}_{d+1} \wedge \ldots \wedge \neg \tilde{B}_{i-1} | \tilde{B}_i \wedge \neg \tilde{B}_1 \wedge \ldots \wedge \neg \tilde{B}_d) \geq$$

$$1 - \sum_{j=d+1}^{i-1} \mathsf{P}(\tilde{B}_j | \tilde{B}_i \wedge \neg \tilde{B}_1 \wedge \ldots \wedge \neg \tilde{B}_d) \geq 1 - e^{\lambda + o(1)} \sum_{j=d+1}^{i-1} \mathsf{P}(\tilde{B}_j | \tilde{B}_i). \tag{22}$$

Estimating from below the first, second and fourth multipliers in the right-hand side of (21) by (20), (19) and (22), respectively, we derive the following lower bound

$$\mathsf{P}(\tilde{B}_i | \neg \tilde{B}_1 \wedge \ldots \wedge \neg \tilde{B}_{i-1}) \geq \left(1 - \frac{o(1/\ln n)}{\mathsf{P}(\neg \tilde{B}_1 \wedge \ldots \wedge \neg \tilde{B}_{\binom{n}{k}})} \right) \mathsf{P}(\tilde{B}_i) \left[1 - e^{\lambda + o(1)} \sum_{j=d(i)+1}^{i-1} \mathsf{P}(\tilde{B}_j | \tilde{B}_i) \right].$$

Thus, combining the latter and (17) and recalling that $\lambda = O(1)$, we finally get

$$\mathsf{P}(\tilde{X} = 0) \leq \prod_{i=1}^{\binom{n}{k}} \left(1 - \left(1 - \frac{o(1/\ln n)}{\mathsf{P}(\neg \tilde{B}_1 \wedge \ldots \wedge \neg \tilde{B}_{\binom{n}{k}})} \right) \mathsf{P}(\tilde{B}_i) \left[1 - e^{\lambda + o(1)} \sum_{j=d(i)+1}^{i-1} \mathsf{P}(\tilde{B}_j | \tilde{B}_i) \right] \right) \leq$$

$$\exp\left[-\left(1 - \frac{o(1/\ln n)}{\exp[-\lambda + o(1)]} \right) \left(\sum_{i=1}^{\binom{n}{k}} \mathsf{P}(\tilde{B}_i) - e^{\lambda + o(1)} \sum_{i=1}^{\binom{n}{k}} \sum_{j=d(i)+1}^{i-1} \mathsf{P}(\tilde{B}_j \wedge \tilde{B}_i) \right) \right] \leq$$

$$\exp\left[-(1 + o(1)) \left(\tilde{\lambda} - e^{\lambda + o(1)} \Delta \right) \right].$$

A noteworthy detail is that $\sum_{i=1}^{\binom{n}{k}} \sum_{j=d(i)+1}^{i-1} \mathsf{P}(\tilde{B}_j \wedge \tilde{B}_i)$ is strictly less than Δ, since, in the definition of Δ, we remove the restrictions on the cardinalities of sets of common neighbors of all proper subsets of both k-sets but the only subset which is the intersection of k-sets. So, the upper bound in (15) can be strengthened, but, for our purpose, this bound is more convenient. $\square$

2.3 The second moment

In this section, we prove the following

Lemma 2 *Let λ, $\tilde{\lambda}$ and Δ be defined by (13) and (14). Then under the conditions of Theorem 1,*

$$\tilde{\lambda} \sim \lambda \sim e^{-y}, \ \ and \ \Delta \to 0$$

as $n \to \infty$.

Proof. Let us start from estimating $\lambda, \tilde{\lambda}$ and showing that $\tilde{\lambda} = \lambda(1 + o(1))$. From (12),

$$\lambda = \sum_{U \in \binom{[n]}{k}} \mathsf{P}(B_U) = \binom{n}{k} \frac{k!}{n^k} e^{-y} (1 + o(1)) = e^{-y}(1 + o(1)). \tag{23}$$

2.3.1 Estimation of $\tilde{\lambda}$

Fix a k-set $U = \{u_1, \ldots, u_k\} \in \binom{[n]}{k}$.

$$\mathsf{P}(B_U) \geq \mathsf{P}(\tilde{B}_U) \geq \mathsf{P}(B_U) - \sum_{\ell=1}^{k-1} \sum_{V \in \binom{U}{\ell}} \mathsf{P}\left(|N_n(U)| > b, \ |N_n(V)| > \Gamma_\ell\right) =$$

$$\mathsf{P}(B_U) - \sum_{\ell=1}^{k-1} \binom{k}{\ell} \mathsf{P}\left(|N_n(U)| > b, \ |N_n(\{u_1, \ldots, u_\ell\})| > \Gamma_\ell\right). \tag{24}$$

Denote $U_\ell = \{u_1, \ldots, u_\ell\}$. Thus, we get

$$\mathsf{P}(|N_n(U)| > b, \ |N_n(U_\ell)| > \Gamma_\ell) \leq \sum_{i > \Gamma_\ell} \mathsf{P}(\xi_{n-\ell,p^\ell} = i) \mathsf{P}(\xi_{i,p^{k-\ell}} > b - (k - \ell))$$

$$\leq \sum_{\Gamma_\ell < i \leq np^\ell + \sqrt{2kp^\ell(1-p^\ell)n \ln n}} \mathsf{P}(\xi_{n-\ell,p^\ell} = i) \mathsf{P}(\xi_{i,p^{k-\ell}} > b - (k - \ell))$$

$$+ \mathsf{P}\left(\xi_{n-\ell,p^\ell} > np^\ell + \sqrt{2kp^\ell(1 - p^\ell)n \ln n}\right),$$

We aim at proving that the latter sum is $o(n^{-k})$. From relations (7) and (10) (since x in this case belongs to the interval $\left(\sqrt{2\ell}, \sqrt{2k}\right]$, the conditions (6) and (8) immediately follow from

(4), as usual),

$$\mathsf{P}(\xi_{n-\ell,p^\ell} = i) = \frac{\exp\left[-\frac{(np^\ell-i)^2}{2np^\ell(1-p^\ell)}\right]}{\sqrt{2\pi np^\ell(1-p^\ell)}}(1+o(1))$$

$$\text{uniformly over } i \in \left(\Gamma_\ell, np^\ell + \sqrt{2kp^\ell(1-p^\ell)n\ln n}\right], \quad (25)$$

$$\mathsf{P}\left(\xi_{n-\ell,p^\ell} > np^\ell + \sqrt{2kp^\ell(1-p^\ell)n\ln n}\right) = \frac{1}{2\sqrt{k\pi\ln n}}n^{-k}(1+o(1)). \quad (26)$$

It remains to estimate $\mathsf{P}(\xi_{i,p^{k-\ell}} > b - (k-\ell))$. Let us verify the conditions of (10). Since $k - \ell$ is constant, it is sufficient to prove that

$$1 \ll b - ip^{k-\ell} = \Theta\left(\sqrt{\frac{\ln n}{\ln i}}\right) \times \sqrt{p^{k-\ell}(1-p^{k-\ell})i\ln i}. \quad (27)$$

Indeed, we should check (6) and (8) for

$$n = i, \quad q = p^{k-\ell} \quad \text{and} \quad x = (b - ip^{k-\ell})/\sqrt{p^{k-\ell}(1-p^{k-\ell})i\ln i}.$$

From (27) it follows that $0 < x = \Theta(\sqrt{\ln n/\ln i}) = \Theta(1)$ which immediately implies the first two conditions in (8). The rest is straightforward due to the restrictions on p given in Theorem 1 and the relation $x = \Theta(1)$.

Now, let us prove (27). For i in the range,

$$b - ip^{k-\ell} \geq \sqrt{2kp^k(1-p^k)n\ln n}(1+o(1)) - \sqrt{2kp^{2k-\ell}(1-p^\ell)n\ln n} =$$

$$\sqrt{2kp^k n\ln n}\left(\sqrt{1-p^k}(1+o(1)) - \sqrt{p^{k-\ell}-p^k}\right) \geq \sqrt{2kp^k n\ln n}\sqrt{1-p^k}\left(1 - \sqrt{\ell/k} + o(1)\right),$$
$$(28)$$

where the latter inequality follows from

$$\sqrt{\frac{k(p^{k-\ell}-p^k)}{1-p^k}} < \sqrt{\ell}. \quad (29)$$

Indeed,

$$\frac{\partial}{\partial p}\left[\ell(1-p^k) - k(p^{k-\ell}-p^k)\right] = k(\ell-k)p^{k-\ell-1}(1-p^\ell) < 0$$

for $p \in (0,1)$, and $\ell(1-p^k) - k(p^{k-\ell}-p^k)\big|_{p=1} = 0$.

14

From (28), it follows that $b - ip^{k-\ell} \gg 1$. Moreover,

$$\frac{b - ip^{k-\ell}}{\sqrt{p^{k-\ell}(1 - p^{k-\ell})i \ln i}} \geq \sqrt{\frac{2k(1 - p^k) \ln n}{(1 - p^{k-\ell}) \ln i}} \left(1 - \sqrt{\ell/k} + o(1)\right) \geq \sqrt{\frac{\ln n}{\ln i}} \left(\sqrt{2k} - \sqrt{2\ell} + o(1)\right)$$

as needed.

Let us verify the upper bound:

$$\frac{b - ip^{k-\ell}}{\sqrt{p^{k-\ell}(1 - p^{k-\ell})i \ln i}} \leq \frac{\sqrt{2kp^k(1 - p^k)n \ln n}}{\sqrt{p^k(1 - p^{k-\ell})n \ln i}(1 + o(1))} = O\left(\sqrt{\frac{\ln n}{\ln i}}\right),$$

that completes the proof of (27).

Therefore, we may apply (10) for $\mathsf{P}(\xi_{i,p^{k-\ell}} > b - (k - \ell))$ as well:

$$\mathsf{P}(\xi_{i,p^{k-\ell}} > b - (k - \ell)) = \frac{\sqrt{ip^{k-\ell}(1 - p^{k-\ell})}}{\sqrt{2\pi}(b - ip^{k-\ell})} \exp\left[-\frac{(b - ip^{k-\ell})^2}{2ip^{k-\ell}(1 - p^{k-\ell})}\right](1 + o(1)) \leq \quad (30)$$

$$\frac{\sqrt{1 - p^{k-\ell}}}{2\sqrt{\pi k \ln n}\left(\sqrt{1 - p^k} - \sqrt{p^{k-\ell} - p^k}\right)} \exp\left[-\frac{(b - ip^{k-\ell})^2}{2ip^{k-\ell}(1 - p^{k-\ell})}\right](1 + o(1)).$$

Putting it all together, we get

$$\mathsf{P}(|N_n(U)| > b, |N_n(U_\ell)| > \Gamma_\ell) \leq$$

$$\left(\frac{\sqrt{1 - p^{k-\ell}}}{2\pi\sqrt{2kp^\ell(1 - p^\ell)n \ln n}\left(\sqrt{1 - p^k} - \sqrt{p^{k-\ell} - p^k}\right)} \sum e^{-f(i)} + \frac{1}{2\sqrt{k\pi \ln n}}n^{-k}\right)(1 + o(1)),$$

(31)

where the summation is over $i \in \left(\Gamma_\ell, np^\ell + \sqrt{2kp^\ell(1 - p^\ell)n \ln n}\right]$ and

$$f(i) = \frac{\left(np^\ell - i\right)^2}{2np^\ell(1 - p^\ell)} + \frac{\left(b - ip^{k-\ell}\right)^2}{2ip^{k-\ell}(1 - p^{k-\ell})}.$$

Let $i > \Gamma_\ell$. Denote $i = np^\ell + x\sqrt{np^\ell(1 - p^\ell) \ln n}$. Then, omitting direct, but tedious calculations, we have

$$f(i) = \frac{x^2(1 - p^k) - 2\sqrt{2k}x\sqrt{(p^{k-\ell} - p^k)(1 - p^k)} + 2k(1 - p^k)}{2(1 - p^{k-\ell})} \ln n +$$

15

$$\frac{x\sqrt{(p^{k-\ell} - p^k)(1 - p^k)} - \sqrt{2k}(1 - p^k)}{2\sqrt{2k}(1 - p^{k-\ell})} \ln\ln n(1 + o(1)), \tag{32}$$

From the definition of Γ_ℓ, $x > \sqrt{2\ell}$. On the other hand, the function of x in the numerator of the first summand in the right-hand side of (32) (denote it by $\hat{f}(x)$) achieves its minimum at $\sqrt{\frac{2k(p^{k-\ell} - p^k)}{1 - p^k}} < \sqrt{2\ell}$ due to (29). Therefore, $\hat{f}(x) > \hat{f}(\sqrt{2\ell})$.

Next, let us prove that the ratio before $\ln\ln n$ in the right-hand side of (32) is negative and bounded. Since $x \leq \sqrt{2k}$, we have

$$\frac{x\sqrt{(p^{k-\ell} - p^k)(1 - p^k)} - \sqrt{2k}(1 - p^k)}{2\sqrt{2k}(1 - p^{k-\ell})} \leq \frac{\sqrt{2k(1 - p^k)}\left(\sqrt{p^{k-\ell} - p^k} - \sqrt{1 - p^k}\right)}{2\sqrt{2k}(1 - p^{k-\ell})} < 0.$$

On the other hand,

$$\frac{x\sqrt{(p^{k-\ell} - p^k)(1 - p^k)} - \sqrt{2k}(1 - p^k)}{2\sqrt{2k}(1 - p^{k-\ell})} > -\frac{(1 - p^k)}{2(1 - p^{k-\ell})} > -\frac{\frac{k}{k-\ell}(1 - p^{k-\ell})}{2(1 - p^{k-\ell})} = -\frac{k}{2(k - \ell)},$$

where the second inequality follows from the relation $(k - \ell)(1 - p^k) < k(1 - p^{k-\ell})$ which can be proved similarly to (29).

To prove (35), we need to make the bound (29) tighter. Namely, we want to show that

$$k\frac{p^{k-\ell} - p^k}{1 - p^k} < \ell p^{(k-\ell)/2}. \tag{33}$$

Indeed, it is equivalent to show that

$$h(p) := \ell(1 - p^k) - k(p^{(k-\ell)/2} - p^{(k+\ell)/2}) > 0.$$

Since $h(1) = 0$, it is sufficient to prove that $h'(p) < 0$ for $p \in (0, 1)$. We have,

$$h'(p) = -\ell k p^{k-1} - k(k - \ell)p^{(k-\ell)/2-1}/2 + k(k + \ell)p^{(k+\ell)/2-1}/2 < 0,$$

or, equivalently,

$$2\ell p^{(k+\ell)/2} + (k - \ell) - (k + \ell)p^\ell > 0. \tag{34}$$

Its derivative $\ell(k + \ell)p^{(k+\ell)/2-1} - \ell(k + \ell)p^{\ell-1} < 0$ for $p \in (0, 1)$ and $2\ell p^{(k+\ell)/2} + (k - \ell) - (k + \ell)p^\ell|_{p=1} = 0$, thus (34) and, consequently, (33) hold.

Thus, after plugging in the numerator of the first summand in the right-hand side of (32) $\sqrt{2\ell}$ and using the relation (33), we get that

$$
\begin{aligned}
f(i) &= \left[k + \frac{(1 - p^k)\left(x - \frac{\sqrt{2k(p^{k-\ell} - p^k)}}{\sqrt{1-p^k}} \right)^2}{2(1 - p^{k-\ell})} \right] \ln n - O(\ln\ln n) \\
&> \left[k + \frac{1}{2}\left(x - \frac{\sqrt{2k(p^{k-\ell} - p^k)}}{\sqrt{1 - p^k}} \right)^2 \right] \ln n - O(\ln\ln n) \\
&> \left[k + \left(\sqrt{2\ell} - \sqrt{2\ell p^{(k-\ell)/2}} \right)^2 / 2 \right] \ln n - O(\ln\ln n) \\
&= \left[k + \ell(1 - p^{(k-\ell)/4})^2 \right] \ln n - O(\ln\ln n). \tag{35}
\end{aligned}
$$

Using the assumption $1 - p \gg \sqrt{\frac{\ln\ln n}{\ln n}}$, we have $(1 - p^{(k-\ell)/4})^2 \gg \frac{\ln\ln n}{\ln n}$. Thus,

$$
f(i) > \ln n (k + \omega(\ln\ln n / \ln n)) - O(\ln\ln n) = k \ln n + \omega(\ln\ln n).
$$

Thus,

$$
\sum_{i \in \left(\Gamma_\ell, np^\ell + \sqrt{2kp^\ell(1-p^\ell)n \ln n} \right]} e^{-f(i)} < \sqrt{2kp^\ell(1 - p^\ell)n \ln n} \, n^{-k}. \tag{36}
$$

From (24), (31) and (36), we get

$$
\mathsf{P}(B_U) \geq \mathsf{P}\left(\tilde{B}_U \right) \geq \mathsf{P}(B_U) - o(n^{-k}).
$$

Finally, from (12) and (23),

$$
e^{-y} + o(1) = \binom{n}{k} \left(\mathsf{P}(B_U) - o(n^{-k}) \right) \leq \binom{n}{k} \mathsf{P}\left(\tilde{B}_U \right) = \tilde{\lambda} \leq \lambda = \binom{n}{k} \mathsf{P}(B_U) = e^{-y} + o(1).
$$

2.3.2 Estimation of Δ

It remains to estimate from above Δ and prove its convergence to 0. By the definition,

$$
\Delta = \sum_{\ell=1}^{k-1} \sum_{V \in \binom{[n]}{\ell}, \, U_1, U_2 \in \binom{[n]}{k} : U_1 \cap U_2 = V} \mathsf{P}(|N_n(V)| \leq \Gamma_\ell, |N_n(U_1)| > b, |N_n(U_2)| > b).
$$

17

Fix k-sets U_1, U_2 such that $|U_1 \cap U_2| = \ell$, $1 \le \ell \le k - 1$. Set $V = U_1 \cap U_2$. Let

$$A^\ell_{U_1,U_2} = \{|N_n(V)| \le \Gamma_\ell, |N_n(U_1)| > b, |N_n(U_2)| > b\}.$$

Then

$$\Delta = \sum_{\ell=1}^{k-1} \binom{n}{\ell}\binom{n-\ell}{k-\ell}\binom{n-k}{k-\ell} \mathsf{P}(A^\ell_{U_1,U_2}) = \sum_{\ell=1}^{k-1} O\left(n^{2k-\ell}\mathsf{P}(A^\ell_{U_1,U_2})\right). \tag{37}$$

So it is sufficient to show that $\mathsf{P}(A^\ell_{U_1,U_2}) = o(n^{-(2k-\ell)})$. Obviously,

$$\mathsf{P}(A^\ell_{U_1,U_2}) \le \sum_{i \le \Gamma_\ell} \mathsf{P}(\xi_{n,p^\ell} = i)[\mathsf{P}(\xi_{i,p^{k-\ell}} > b - (k - \ell))]^2 \le$$

$$\sum_{np^\ell - 2\sqrt{kp^\ell(1-p^\ell)n \ln n} < i \le \Gamma_\ell} \mathsf{P}(\xi_{n,p^\ell} = i)[\mathsf{P}(\xi_{i,p^{k-\ell}} > b - (k - \ell))]^2 +$$

$$\mathsf{P}\left(\xi_{n,p^\ell} \le np^\ell - 2\sqrt{kp^\ell(1-p^\ell)n \ln n}\right).$$

From relations (7) and (10′), we get

$$\mathsf{P}(\xi_{n,p^\ell} = i) = \frac{\exp\left[-\frac{(np^\ell - i)^2}{2np^\ell(1-p^\ell)}\right]}{\sqrt{2\pi np^\ell(1-p^\ell)}}(1+o(1)) \text{ uniformly over } i \in \left(np^\ell - 2\sqrt{kp^\ell(1-p^\ell)n \ln n}, \Gamma_\ell\right],$$

$$\mathsf{P}\left(\xi_{n,p^\ell} \le np^\ell - 2\sqrt{kp^\ell(1-p^\ell)n \ln n}\right) = \frac{1}{2\sqrt{2k\pi \ln n}}n^{-2k}(1 + o(1)) = o(n^{-2k}).$$

As in Section 2.3.1, to estimate $\mathsf{P}(\xi_{i,p^{k-\ell}} > b - (k - \ell))$, we should verify the conditions of (10), namely, the assumptions (6) and (8). For $i \in \left(np^\ell - 2\sqrt{kp^\ell(1-p^\ell)n \ln n}, \Gamma_\ell\right]$, the value of $b - ip^{k-\ell}$ is even bigger than in (28) and, therefore, here, (28) holds as well. Moreover,

$$\frac{b - ip^{k-\ell}}{\sqrt{p^{k-\ell}(1-p^{k-\ell})i \ln i}} \le \frac{\sqrt{2kp^k(1-p^k)n \ln n} + 2\sqrt{kp^k(p^{k-\ell}-p^k)n \ln n}}{\sqrt{p^k(1-p^{k-\ell})n \ln i}(1 + o(1))} = O\left(\sqrt{\frac{\ln n}{\ln i}}\right).$$

Therefore, similarly to corresponding argument in Section 2.3.1, we may apply (10) and get, in the same way as in (30),

$$\mathsf{P}(\xi_{i,p^{k-\ell}} > b - (k-\ell)) \le \frac{\sqrt{1 - p^{k-\ell}}}{\sqrt{2\pi \ln n}\left(\sqrt{2k(1-p^k)} - \sqrt{2\ell(p^{k-\ell}-p^k)}\right)}e^{-\frac{(b-ip^{k-\ell})^2}{2ip^{k-\ell}(1-p^{k-\ell})}}(1+o(1)).$$

Therefore, we get from above

$$\mathsf{P}(A^\ell_{U_1,U_2}) \leq$$

$$(1+o(1))\frac{1-p^{k-\ell}}{(2\pi)^{3/2}\ln n\sqrt{np^\ell(1-p^\ell)}\left(\sqrt{2k(1-p^k)}-\sqrt{2\ell p^{k-\ell}(1-p^\ell)}\right)^2}\sum e^{-g(i)}+o(n^{-2k}),$$

$$(38)$$

where the summation is over $i \in \left(np^\ell - 2\sqrt{kp^\ell(1-p^\ell)n\ln n}, \Gamma_\ell\right]$ and

$$g(i) = \frac{(np^\ell - i)^2}{2np^\ell(1-p^\ell)} + \frac{(ip^{k-\ell} - b)^2}{ip^{k-\ell}(1-p^{k-\ell})}.$$

As in Section 2.3.1, denote $i = np^\ell + x\sqrt{np^\ell(1-p^\ell)\ln n}$. Notice that $x \in (-2\sqrt{k}, \sqrt{2\ell}]$. We get

$$g(i) = \frac{x^2(1+p^{k-\ell}-2p^k) - 4\sqrt{2k}\sqrt{(p^{k-\ell}-p^k)(1-p^k)}x + 4k(1-p^k)}{2(1-p^{k-\ell})}\ln n+$$

$$\frac{x\sqrt{(p^{k-l}-p^k)(1-p^k)} - \sqrt{2k}(1-p^k)}{\sqrt{2k}(1-p^{k-l})}\ln\ln n(1+o(1)) =: \tilde{g}_p(x)\ln n + \hat{g}_p(x)\ln\ln n(1+o(1)).$$

We note that $\hat{g}_p(x)/2$ is equal to the ratio before $\ln\ln n$ in the right-hand side of (32). It can be proved similarly as in Section 2.3.1 that $\hat{g}_p(x)$ is negative and bounded from below.

As in Section 2.3.1, in order to show $\mathsf{P}(A^\ell_{U_1,U_2}) = o(n^{-(2k-\ell)})$, we prove $\tilde{g}_p(x) \geq 2k - \ell + \omega(\ln\ln n/\ln n)$. The minimum of $\tilde{g}_p(x)$ equals $\tilde{g}_p(x_0) = \frac{2k(1-p^k)}{1+p^{k-\ell}-2p^k}$, where

$$x_0 = \frac{2\sqrt{2k}\sqrt{(p^{k-\ell}-p^k)(1-p^k)}}{1+p^{k-\ell}-2p^k}.$$

Note that $\tilde{g}_p(x_0)$ decreases in $p \in (0,1)$ since

$$\frac{2k}{\tilde{g}_p(x_0)} = 2 - \frac{1-p^{k-\ell}}{1-p^k},$$

and $(1-p^{k-\ell})/(1-p^k)$ decreases in $p \in (0,1)$ as well. Let $\varepsilon \in \left(0, \frac{2}{k-1}\right)$ and $p_0 \in (0,1)$ be the root of the equation $\frac{1-p^k}{p^{k-\ell}-p^k} = \frac{2k-\ell}{\ell} + \varepsilon$. Then we have

$$\tilde{g}_{p_0}(x_0) = 2k\left(\frac{p_0^{k-\ell}-p_0^k}{1-p_0^k}+1\right)^{-1} = 2k\left(\frac{\ell}{2k-\ell(1-\varepsilon)}+1\right)^{-1} = 2k\frac{2k-(1-\varepsilon)\ell}{2k+\varepsilon\ell}.$$

Thus,
$$\tilde{g}_p(x_0) > 2k\frac{2k - (1 - \varepsilon)\ell}{2k + \varepsilon\ell} = 2k - \ell + \frac{\varepsilon\ell^2}{2k + \varepsilon\ell} > 2k - \ell$$

for all $p < p_0$. For such p, we immediately get the required relation $\mathsf{P}(A^\ell_{U_1,U_2}) = o(n^{-(2k-\ell)})$.

If $p \geq p_0$, then $\frac{1-p^k}{p^{k-\ell}-p^k} \leq \frac{2k-\ell}{\ell} + \varepsilon$, since the function $\frac{1-p^k}{p^{k-\ell}-p^k}$ decreases for $p \in (0,1)$. Therefore, since the function $x_0 = x_0(p)$ is increasing for $p \in (0,1)$, we get

$$x_0 = \frac{2\sqrt{2k}\sqrt{(p^{k-\ell} - p^k)(1 - p^k)}}{1 + p^{k-\ell} - 2p^k} \geq x_0(p_0) = 2\sqrt{2k}\frac{\sqrt{\ell(2k - (1 - \varepsilon)\ell)}}{2k + \varepsilon\ell} > \sqrt{2\ell}, \qquad (39)$$

where the last inequality follows from the relation $4k(2k - (1 - \varepsilon)\ell) > (2k + \varepsilon\ell)^2$ which can be shown directly using $\ell \leq k - 1$ and $\varepsilon < 2/(k-1)$. Moreover,

$$\tilde{g}_p\left(\sqrt{2\ell}\right) = \frac{\ell(1 + p^{k-\ell} - 2p^k) - 2\sqrt{2k}\sqrt{2\ell}\sqrt{(p^{k-\ell} - p^k)(1 - p^k)} + 2k(1 - p^k)}{1 - p^{k-\ell}} =$$

$$\frac{\left[\sqrt{2\ell(1 - p^k)} - \sqrt{2k(p^{k-\ell} - p^k)}\right]^2 - \ell(1 - p^{k-\ell}) + 2k(1 - p^{k-\ell})}{1 - p^{k-\ell}} = 2k - \ell + \omega\left(\frac{\ln\ln n}{\ln n}\right)$$

since $\left[\sqrt{2\ell(1 - p^k)} - \sqrt{2k(p^{k-\ell} - p^k)}\right]^2/(1 - p^{k-\ell}) = \omega\left(\frac{\ln\ln n}{\ln n}\right)$ for $1 - p \gg \sqrt{\frac{\ln\ln n}{\ln n}}$, see (35) and the argument below (35). As $x \leq \sqrt{2\ell} < x_0$, we get $\tilde{g}_p(x) \geq \tilde{g}_p\left(\sqrt{2\ell}\right)$. Since $\hat{g}_p(x) = O(1)$,

$$\sum_{i \in \left(np^\ell - 2\sqrt{kp^\ell(1-p^\ell)n\ln\ln n}, \Gamma_\ell\right]} e^{-g(i)} < 4\sqrt{kp^\ell(1 - p^\ell)n\ln\ln n}\, n^{-(2k-\ell)}. \qquad (40)$$

Relations (37), (38), (40) imply the desired convergence $\Delta \to 0$. $\square$

2.4   Final steps

From Lemma 1 and Lemma 2 we immediately get the statement of Theorem 1 for $m = 1$ since $\mathsf{P}(\Delta^1_{k;n} \leq b) = \mathsf{P}(X_n = 0)$.

The very last step is to find the limit of $\mathsf{P}(\Delta^m_{k;n} \leq b) - \mathsf{P}(\Delta^{m-1}_{k;n} \leq b) = \mathsf{P}(X_n = m - 1)$ for $m \geq 2$.

Note that the probability δ_n of the existence of two distinct overlapping k-sets $U_1, U_2 \in \binom{[n]}{k}$ such that $|N_n(U_j)| > b$ for both $j = 1$ and $j = 2$ is at most $\Delta + o(1)$ as, for every

$\ell \in [k-1]$, a.a.s. there are no ℓ-sets with more than Γ_ℓ common neighbors in $G(n,p)$. Therefore, $\delta_n = o(1)$. Thus, $\mathsf{P}(X_n = m-1)$ is equal to $h_{m-1} + o(1)$ where h_{m-1} is the probability that the maximum i such that there exist i disjoint k-sets $U_1, \ldots, U_i \in \binom{[n]}{k}$ having $|N_n(U_j)| > b$ for all $j \in [i]$ equals $m-1$. But this probability is much easier to estimate. Indeed, set

$$G(m-1) = \frac{1}{(m-1)!} \prod_{j=1}^{m-1} \binom{n - k(j-1)}{k}, \quad n_{m-1} = n - k(m-1), \quad b_{m-1} = b - k(m-1).$$

Fix disjoint k-sets $U_1, \ldots, U_{m-1}$. For a k-set U, denote by $\hat{N}_n(U)$ the number of common neighbors of U in $\mathcal{N}_{m-1} = [n] \setminus [U_1 \sqcup \ldots \sqcup U_{m-1}]$. Then

$$G(m-1) \left(\mathsf{P}(|\hat{N}_n(U_1)| > b) \right)^{m-1} \mathsf{P}\left(\Delta^1_{k;n_{m-1}} \le b \right) \le h_{m-1} \le$$

$$G(m-1) \left(\mathsf{P}(|\hat{N}_n(U_1)| > b_{m-1} + k) \right)^{m-1} \mathsf{P}\left(\Delta^1_{k;n_{m-1}} \le b_{m-1} \right).$$

These bounds, in particular, follow from the fact that all U_i are disjoint and the numbers of their common neighbors in $\mathcal{N}_{m-1}$ are not greater than $|N_n(U_i)|$ but not less than $|N_n(U_i)| - k(m-1) + k$. As $b_{m-1} = b + O(1)$ and $n_{m-1} = n + O(1)$, both lower and upper bounds are equal to $\frac{1}{(m-1)!} \left(e^{-y} \right)^{m-1} e^{-e^{-y}} (1 + o(1))$. Therefore, reminding that y and m are constants, we finally derive

$$\mathsf{P}(\Delta^m_{k;n} \le b) = \sum_{i=0}^{m-1} \mathsf{P}(X_n = i) \sim \sum_{i=0}^{m-1} h_i \sim e^{-e^{-y}} \sum_{i=0}^{m-1} \frac{1}{i!} \left(e^{-y} \right)^i.$$

3 Discussions and further questions

In Section 1, we have mentioned that the dependencies between degrees of $G(n,p)$ are weak enough, and so, the result of Bollobás does not essentially differ from the result of Nadarajah and Mitov (for independent binomial random variables). This motivates the following question. How strong can be dependencies between the binomial random variables until (3) fails?

Let us formalize this question in the following way. Let $\xi_1, \xi_2, \ldots$ be independent Bernoulli random variables with parameter p. For every $n \in \mathbb{N}$, consider an $M = M(n)$-element set Σ_n of n-vectors $(\xi_{i_1}, \ldots, \xi_{i_n})$ having components in the given sequence, $M \gg (\ln n)^3$. Let $D_n = \max_{(\xi_{i_1}, \ldots, \xi_{i_n}) \in \Sigma_n} (\xi_{i_1} + \ldots + \xi_{i_n})$. Assume that $m = m(n)$ is such that any two vectors

from Σ_n have at most m common components. For $m = 0$, we have (3), and the result of Bollobás relates to $m = 1$. Can we guarantee the same for larger m?

Below, we state a generalization of our Janson-type inequality given in Section 2.2, that has an analogous proof and immediately implies the following answer on the above question. If $m = o(\sqrt{n/\ln n})$, then (3) is true.

Lemma 3 *For every $n \in \mathbb{N}$, consider a sequence of independent random variables $\xi^n = (\xi_1^n, \xi_2^n, \ldots)$, an $M(n)$-element set $\{\eta_1^n, \ldots, \eta_M^n\}$ of vectors having components in ξ^n and Borel sets $A_1^n, \ldots, A_M^n$, where $A_i^n \subset \mathbb{R}^{k_i}$, k_i is the dimension of η_i^n. Let, for every $i \in \{1, \ldots, M\}$, $J(i) \subset \{1, \ldots, M\}$ be such that, for every $J \subseteq J(i)$,*

$$\mathsf{P}(\wedge_{j \in J}\{\eta_j^n \notin A_j^n\}|\eta_i^n) = \mathsf{P}(\wedge_{j \in J}\{\eta_j^n \notin A_j^n\}) + o(1) \quad a.s.$$

uniformly over all i and J. Let

$$X_n = \sum_{i=1}^M I(\eta_i^n \in A_i^n), \quad \Delta_n = \sum_{i=1}^n \sum_{j \in [M] \setminus J(i)} \mathsf{P}(\eta_i^n \in A_i^n, \eta_j^n \in A_j^n).$$

If $\mathsf{P}(X_n = 0) \geq e^{-\mathsf{E}X_n + o(1)}$, then

$$\mathsf{P}(X_n = 0) \leq \exp\left[-(1 + o(1))\mathsf{E}X_n - (1 + o(1))e^{\mathsf{E}X_n}\Delta_n\right].$$

It is also of interest to consider the case $k = k(n)$ as well as $1 - p$ approaching 0 faster than $\sqrt{\frac{\ln \ln n}{\ln n}}$ and to prove (or disprove) an analogue of Theorem 1.

Finally, in Section 1, we mentioned that Theorem 1 can be formulated in terms of the maximum number of (G, H)-extensions where $|V(H)| - |V(G)| = 1$ and $|E(H)| - |E(G)| = |V(G)|$. We expect that our techniques may give analogous results for a wider class of grounded strictly balanced pairs (G, H).

4 Acknowledgement

References

[1] N. Alon, J.H. Spencer, *The Probabilistic Method*, Third Edition, John Wiley & Sons, 2008.

[2] C.W. Anderson, S.G. Coles, J. Hűsler, *Maxima of Poisson-like variables and related triangular arrays*, Ann. Appl. Probab., **7**:4 (1997), 953–971.

[3] J. Beirlant, Y. Goegebeur, J. Segers, J.L. Teugels, *Statistics of Extremes: Theory and Applications*, John Wiley & Sons, 2004.

[4] B. Bollobás, *Random Graphs*, 2nd Edition, Cambridge University Press, 2001.

[5] B. Bollobás, *The distribution of the maximum degree of a random graph*, Discrete Mathematics, **32** (1980), 201–203.

[6] B. Bollobás, *Degree sequences of random graphs*, Discrete Mathematics, **33** (1981), 1–19.

[7] R. Boppona, J. Spencer, *A useful elementary correlation inequality*, J. Combin. Theory Ser. A, **50** (1989), 305–307.

[8] H.A. David, *Order Statistics*, New York: John Wiley & Sons, 1970.

[9] P. Sielenou Dkengne, N. Eckert, P. Naveau, *A limiting distribution for maxima of discrete stationary triangular arrays with an application to risk due to avalanches*, Extremes, **19**:1 (2016), 25–40.

[10] B. Gnedenko, *Sur la Distribution Limite du Terme Maximum d'une Série Aléatoire*, Annals of Mathematics, **44** (1943), 423–453.

[11] L. Haan, A. Ferreira, *Extreme Value Theory. An Introduction*, Springer, 2006.

[12] G.I. Ivchenko, *On the asymptotic behavior of degrees of vertices in a random graph*, Theory Probab. Appl., **18**:1 (1973), 188–195.

[13] S. Janson, T. Łuczak, A. Rucinski, *Random Graphs*, New York, Wiley, 2000.

[14] D.J. Kleitman, *Families of non-disjoint subsets*, J. Combinatorial Theory, **1** (1966), 153–155.

[15] M.R. Leadbetter, G. Lindgren, H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes*, New York: Springer Verlag, 1983.

[16] S. Nadarajah, K. Mitov, *Asymptotics of Maxima of Discrete Random Variables*, Extremes, **5** (2002), 287–294.

[17] O. Riordan, L. Warnke, *The Janson inequalities for general up-sets*, Random Structures & Algorithms, **46**:2 (2015), 391–395.

[18] M. Šileikis, L. Warnke, *Counting extensions revisited*, 2019, https://arxiv.org/pdf/1911.03012.pdf.

[19] J. Spencer, *A useful elementary correlation inequality, II*, J. Combin. Theory Ser. A, **84** (1998), 95–98.

[20] J.H. Spencer, *Counting extensions*, J. of Comb. Th. Ser A, **55** (1990), 247–255.