A note on the distribution of the extreme degrees of a random graph via the Stein-Chen method

Yaakov Malinovsky *
Department of Mathematics and Statistics
University of Maryland, Baltimore County, Baltimore, MD 21250, USA

April 13, 2022

Abstract

We offer an alternative proof, using the Stein-Chen method, of Bollobás' theorem concerning the distribution of the extreme degrees of a random graph. The same method also applies in a more general setting where the probability of every pair of vertices being connected by edges depends on the number of vertices.

Keywords: random graphs, extremes, positive dependence, Poisson approximation, total variation distance
MSC2020: 05C80; 05C07; 62G32

Consider a random graph with n labeled vertices $\{1, 2, \ldots, n\}$ in which each edge $E_{ij}, 1 \leq i < j \leq n$ is chosen independently and with a fixed probability p, $0 < p < 1$. Denote by d_i degree of the vertex $i, i = 1, \ldots, n$ of a graph $G \in \mathbb{G}(n, p)$, where $\mathbb{G}(n, p)$ is the probability space of graphs, and by $d_{1:n} \geq d_{2:n} \geq \cdots \geq d_{n:n}$ the degree sequence arranged in decreasing order. Bollobás (1980) found the asymptotic distribution of $d_{m:n}$ and proved:

Theorem 1 (Bollobás (1980))**.** *Suppose p is fixed, $0 < 1 < p$, $q = 1 - p$. Then*

$$\lim_{n\to\infty} P\left(\frac{d_{m:n} - np}{\sqrt{npq}} < a_n t + b_n\right) = e^{-e^{-t}} \sum_{k=0}^{m-1} \frac{e^{-tk}}{k!}, \tag{1}$$

where

$$a_n = (2\log n)^{-\frac{1}{2}}, \quad b_n = (2\log n)^{\frac{1}{2}} - \frac{1}{2}(2\log n)^{-\frac{1}{2}} (\log\log n + \log 4\pi).$$

*email: yaakovm@umbc.edu

We offer an alternative proof of Bollobás' theorem via the Stein-Chen method.

Denote normalized vertex degrees (zero expectation and unit variance) as $d_1^*, d_2^*, \ldots, d_n^*$ and their corresponding decreasing sequence as $d_{1:n}^* \geq d_{2:n}^* \geq \cdots \geq \ldots \geq d_{n:n}^*$.

Let $I_i^{(n)} = I(d_i^* > x_n(t))$, where we choose $x_n(t) = a_n t + b_n$, with a_n and b_n as defined in (1).

Set

$$W_n = \sum_{i=1}^{n} I_i^{(n)}, \;\; \pi_i^{(n)} = P(I_i^{(n)} = 1), \;\; \lambda_n = E(W_n) = \sum_{i=1}^{n} \pi_i^{(n)} = n\pi_1^{(n)}.$$

We will need three Assertions.

Assertion 1.

$$\pi_1^{(n)} = P\left(d_1^* > x_n(t)\right) \sim 1 - \Phi\left(x_n(t)\right).$$

Proof. Follows from Feller (1968)(Chapter VII.6). □

Assertion 2.

$$\lim_{n\to\infty} \lambda_n = \lim_{n\to\infty} n\pi_1^{(n)} = \lim_{n\to\infty} nP(d_1^* > x_n(t)) = e^{-t} \equiv \lambda(t).$$

Proof. Follows from *Assertion 1* combined with the result on page 374 of Cramér (1946). □

Assertion 3.

$$\lim_{n\to\infty} n^2(P(d_1^* > x_n(t), d_2^* > x_n(t))) = e^{-2t} = \lambda^2(t). \tag{2}$$

Proof. Let e_{ij} be the indicator random variable for the event $\{E_{ij} = 1\}$. Thus, $d_1 = e_{12} + e_{13} + \cdots + e_{1n}$ and $d_2 = e_{21} + e_{23} + \cdots + e_{2n}$. Hence, conditional on the event $e_{12} = k, k \in \{0, 1\}$, d_1 and d_2 are independent. Let $d_{1'} = e_{13} + \cdots + e_{1n}$ and $d_{2'} = e_{23} + \cdots + e_{2n}$, and denote by $d_{1'}^*, d_{2'}^*$ the corresponding normalized scores (zero expectation and unit variance). We then have,

$$\begin{aligned} &P(d_1^* > x_n(t), d_2^* > x_n(t) \,\big|\, e_{12} = k) = P(d_1^* > x_n(t) \,\big|\, e_{12} = k) P(d_2^* > x_n(t) \,\big|\, e_{12} = k) \\ &= P^2\left(d_{1'}^* > x_{n-1}(t) \frac{x_n(t)}{x_{n-1}(t)} \sqrt{\frac{n-1}{n-2}} - \frac{k-p}{\sqrt{(n-2)pq}}\right), \;\; k = 0, 1. \end{aligned} \tag{3}$$

From (3) it follows that $P(d_1^* > x_n(t), d_2^* > x_n(t) \,|\, e_{12} = k) \sim P\left(s_{1'}^* > x_{n-1}(t)\right) P\left(s_{2'}^* > x_{n-1}(t)\right)$ and by combining that with the formula of total probability, we obtain

$$P(d_1^* > x_n(t), d_2^* > x_n(t)) \sim P\left(d_{1'}^* > x_{n-1}(t)\right) P\left(d_{2'}^* > x_{n-1}(t)\right).$$

Combining this with *Assertion 2*, we obtain (2). $\square$

The indicators $\left(I_1^{(n)}, I_2^{(n)}, \ldots, I_n^{(n)}\right)$ are increasing functions of the independent edge indicators $\{e_{ij}, 1 \le i < j \le n\}$. As such, by Theorem 2.G, and hence by Corollary 2.C.4. in Barbour *et al.* (1992) (see also related discussion in section 5.2 there), we obtain

Assertion 4.

$$\begin{aligned} d_{TV}\left(L(W_n), Poi(\lambda_n)\right) &\le \frac{1-e^{\lambda_n}}{\lambda_n}\left(Var(W_n) - \lambda_n + 2\sum_{i=1}^{n}\left(\pi_i^{(n)}\right)^2\right) \\ &= \frac{1-e^{\lambda_n}}{\lambda_n}\left(\sum_{i=1}^{n}\left(\pi_i^{(n)}\right)^2 + \sum_{i\neq j} Cov\left(I_i^{(n)}, I_j^{(n)}\right)\right), \end{aligned} \tag{4}$$

where $d_{TV}\left(L(W_n), Poi(\lambda_n)\right)$ is the total variation distance between distributions of W_n and the Poisson distribution with mean λ_n.

In our case, since $d_1^*, \ldots, d_n^*$ are identically distributed, $\sum_{i=1}^{n}\left(\pi_i^{(n)}\right)^2 = nP(d_1^* > x_n)P(d_1^* > x_n)$, and $\sum_{i\neq j} Cov\left(I_i^{(n)}, I_j^{(n)}\right) = n(n-1)\left[P\left(d_1^* > x_n(t), d_2^* > x_n(t)\right) - P\left(d_1^* > x_n(t)\right)P\left(d_2^* > x_n(t)\right)\right]$. Hence, from Assertion 2 it follows that

$$\lim_{n\to\infty}\sum_{i=1}^{n}\left(\pi_i^{(n)}\right)^2 = 0. \tag{5}$$

and from Assertion 2 and 3 it follows that

$$\lim_{n\to\infty}\sum_{i\neq j} Cov\left(I_i^{(n)}, I_j^{(n)}\right) = 0. \tag{6}$$

Then, from (5) and (6) it follows that $\lim_{n\to\infty} d_{TV}\left(L(W_n), Poi(\lambda_n)\right) = 0$, and we obtain the following result:

Theorem 2. *For $p \in (0,1)$ and a fixed value of k,*

$$\lim_{n\to\infty} P(S_n = k) = e^{-\lambda(t)}\frac{\lambda(t)^k}{k!}, \;\; \lambda(t) = e^{-t}.$$

Noticing that $P(d^*_{m:n} \le x_n(t)) = P(S_n \le m-1)$, and applying Theorem 2, we obtain

$$\lim_{n\to\infty} P\left(\frac{d_{m:n}-(n-1)p}{\sqrt{(n-1)pq}} \le a_n t + b_n\right) = e^{-e^{-t}} \sum_{k=0}^{m-1} \frac{e^{-tk}}{k!}. \tag{7}$$

Comment 1. *The method and the results can be extended to the case where p depends on n. In such case, Assertion 1 holds if $p(1-p)n \to \infty$ as $n \to \infty$ and $x_n(t)(p(1-p)n)^{1/2} = o\{(p(1-p)n)^{2/3}\}$ (see for example Bollobás (2001), Theorem 1.6). Since, $x_n(t) \sim 2\log(n))$, the above conditions are satisfied if $p(1-p)n/(\log n)^3 \to \infty$ as $n \to \infty$, which coincides with the condition in Theorem 3.3' Bollobás (2001), in which the same result was obtained in the case where p is a function of n but by different method.*

Comment 2. *If $\xi_1, \dots, \xi_n$ are independent and identically distributed standard normal random variables, with corresponding $\xi_{1:n} \ge \xi_{2:n} \ge \cdots \ge \xi_{n:n}$ sequence arranged in decreasing order, then the limit distribution $P(\xi_{m:n} \le a_n t + b_n)$ is identical to the RHS of (1), with the same a_n, b_n (see for example Galambos (1987)). However, (1) is not obvious since $d_1, \dots, d_n$ are dependent and their joint distribution depends on n.*

Comment 3. *The same asymptotic distribution as in (1), for the ordered normalized scores, holds for a round-robin tournament model with n players (Malinovsky, 2021). The difference is that the round-robin tournament is a complete directed graph and the total scores (degrees) of the players are negatively correlated.*

Acknowledgement

This research was supported by grant no 2020063 from the United States-Israel Binational Science Foundation (BSF).

References

Barbour, A. D., Holst, L., Janson, S. (1992). Poisson approximation. Oxford Studies in Probability. The Clarendon Press, Oxford, New York.

Bollobás, B. (1980). The distribution of the maximum degree of a random graph. Discrete Math. **3**2, 201–203.

Bollobás, B. (2001). Random graphs. Second edition. Cambridge University Press, Cambridge.

Cramér, H. (1946). Mathematical Methods of Statistics. Primceton University Press.

Feller, W. (1968). An introduction to probabilty theory and its applications. Vol. I. Third edition. New York-London-Sydney: Wiley.

Galambos, J. (1987). The asymptotic theory of extreme order statistics. Second edition. Robert E. Krieger Publishing Co., Inc., Melbourne, FL.

Malinovsky, Y. (2021). On the distribution of winners' scores in a round-robin tournament. Probability in the Engineering and Informational Sciences. In press. Corrected version: https://arxiv.org/abs/2201.05018