Acta Math. Univ. Comenianae
Vol. LXXXVIII, 3 (2019), pp. 423–427

MAXIMUM INDUCED SUBGRAPHS OF THE BINOMIAL RANDOM GRAPH

J. BALOGH AND M. ZHUKOVSKII

ABSTRACT. We prove that a.a.s. the maximum size of an induced tree in the binomial random graph $G(n,p)$ is concentrated in four consecutive points. We also consider the following problem. Given $e(k)$, what is the maximum k such that $G(n,p)$ has an induced subgraph with k vertices and $e(k)$ edges? For $e = o(\frac{k \ln k}{\ln \ln k})$, we prove that a.a.s. this maximum size is concentrated in two consecutive points. In contrast, for $e(k) = p\binom{k}{2} + O(k)$, we prove that this size is not concentrated in any finite set. Moreover, we prove that for an $\omega_n \to \infty$, a.a.s. the size of the concentration set is smaller than $\omega_n \sqrt{n/\ln n}$. Otherwise, for an arbitrary constant $C > 0$, a.a.s. it is bigger than $C\sqrt{n/\ln n}$.

1. General framework

Consider a sequence $\mathcal{F}_k$ of sets of graphs on k vertices (i.e., for every $k \in \mathbb{N}$, $\mathcal{F}_k$ is a set of graphs on k vertices). Let X_n be the maximum k such that there exists $F \in \mathcal{F}_k$ and an induced subgraph H in the binomial random graph $G(n,p)$ [**1, 7, 12**] (edges in this graph appear independently with a constant probability $p \in (0,1)$) such that H and F are isomorphic. Below, we briefly discuss the main results on an asymptotical behaviour of X_n.

The first known related result describes an asymptotical behaviour of the independence number (the maximum size of an independent set) and the clique number (the maximum size of a clique) of $G(n,p)$ [**3, 9**]. It states that, for arbitrary constant $p \in (0,1)$, there exists $f(n)$ such that asymptotically almost surely (a.a.s.) the clique number of $G(n,p)$ belongs to $\{f(n), f(n)+1\}$ (below, in such situations we say that there is a *2-point concentration*) (certain generalizations of this result can be found in [**8, 11**]). By symmetry reasons, the same is true for the independence number.

Clearly, the above concentration results are special cases of the considered general problem. Indeed, X_n is the independence number (the clique number), if each $\mathcal{F}_k$ contains only the empty (complete) graph.

A natural question to ask is, what happens in the case of 'common' graph sequences, such as paths, cycles, etc.?

Received May 30, 2019.
2010 *Mathematics Subject Classification.* Primary 05C80; Secondary 05C35,05D99.

Let, for $k \in \mathbb{N}$, $\mathcal{F}_k = \{F_k\}$. In [**4**], 2-point concentration results are obtained for $F_k = P_k$ (simple path on k vertices) and $F_k = C_k$ (simple cycle on k vertices).

Let us turn to larger graph families $\mathcal{F}_k$. The following families were considered by several researchers: trees, regular graphs, complete bipartite graphs and complete multipartite graphs. For all these families, it was unknown, if there is a 2-point concentration, or even an m-point concentration for some fixed number m. In 1983, Erdős and Palka [**5**] proved that, for trees (i.e., $\mathcal{F}_k$ consists of all trees on k vertices), $\frac{X_n}{\ln n} \xrightarrow{\mathsf{P}} \frac{2}{\ln[1/(1-p)]}$ as $n \to \infty$ (hereinafter, $\xrightarrow{\mathsf{P}}$ denotes the convergence in probability). In 1987, Ruciński [**13**] obtained a similar law of large numbers type general result for a respectively wide class of graph families $\mathcal{F}_k$. In particular, from his result follows that: if $\mathcal{F}_k$ are sets of $ck(1+o(1))$-regular graphs, then $\frac{X_n}{\ln n} \xrightarrow{\mathsf{P}} \frac{2}{c\ln[1/p]+(1-c)\ln[1/(1-p)]}$ as $n \to \infty$. For several families of complete bipartite and multipartite graphs, similar results were obtained in [**10, 13**].

2. Induced trees

In [**4**], the authors ask, is it true that, for trees, the 2-point concentration result holds. We do not have an answer. Nevertheless, we prove the 4-point concentration result. Let X_n be the maximum size of an induced tree in $G(n,p)$.

Theorem 1. *Let $k = k(n) \geq 1$ be such that*

$$k\ln n - \frac{5}{2}\ln k + k - \binom{k}{2}\ln[1/(1-p)] + (k-1)\ln[p/(1-p)] - \frac{1}{2}\ln(2\pi) = 0.$$

Then a.a.s. $X_n \in \{\lceil k\rceil - 3, \lceil k\rceil - 2, \lceil k\rceil - 1, \lceil k\rceil\}$.

3. Fixed number of edges

In [**6**], families of graphs having special edge conditions are considered. More formally, given a sequence $e = e(k)$, $\mathcal{F}_k = \mathcal{F}_k(e)$ is a set of all graphs on k vertices having *at most* $e(k)$ edges. The main result of [**6**] states, in particular, the following. Let $e = e(k) = o(\frac{k\ln k}{\ln\ln k})$ be a sequence of non-negative integers. Then there is a function $f(n)$ such that a.a.s. $X_n \in \{f(n), f(n)+1\}$. We state that the same is true for families of graphs having *exactly* e edges.

Theorem 2. *Let $e(k)$ be a sequence of non-negative integers such that $e(k) = o(\frac{k\ln k}{\ln\ln k})$ and $|t(k+1)-t(k)| = o(k/\ln k)$. Let $\mathcal{F}_k = \mathcal{F}_k(e)$ be the set of all graphs on k vertices with exactly $e(k)$ edges. Then there exists $\hat{k} = \frac{2}{\ln(1/(1-p))}\ln n(1+o(1))$ such that a.a.s. $X_n \in \{\hat{k}-1, \hat{k}\}$.*

For the sake of convenience, let us denote the latter random variable X_n by $\mathcal{X}_n[e]$ (i.e., $\mathcal{X}_n[e]$ is the maximum k such that $G(n,p)$ contains an induced subgraph with k vertices and $e(k)$ edges).

Clearly, the 2-point concentration result is also true when $e(k) = \binom{k}{2} - o(\frac{k\ln k}{\ln\ln k})$. The natural question to ask: is the same true for $e(k)$ close to the average number of edges $p\binom{k}{2}$? We give the following negative answer on this question.

Theorem 3. *Let $e(k) = \binom{k}{2}p + O(k)$ be a sequence of non-negative integers.*

(i) *There exists $t > 0$ such that, for $c > t$ and $C > 2c + t$, we have*

$$0 < \liminf_{n\to\infty} \mathsf{P}\left(n - C\sqrt{\frac{n}{\ln n}} < \mathcal{X}_n(e) < n - c\sqrt{\frac{n}{\ln n}}\right)$$
$$\le \limsup_{n\to\infty} \mathsf{P}\left(n - C\sqrt{\frac{n}{\ln n}} < \mathcal{X}_n(e) < n - c\sqrt{\frac{n}{\ln n}}\right) < 1.$$

(ii) *Let, for a sequence $m_k = O(\sqrt{k/\ln k})$ of non-negative integers, the following smoothness condition hold:*

$$\left|\left(e(k) - \binom{k}{2}p\right) - \left(e(k - m_k) - \binom{k - m_k}{2}p\right)\right| = o(k).$$

Then, for every $\varepsilon > 0$, there exist c, C such that

$$\liminf_{n\to\infty} \mathsf{P}\left(n - C\sqrt{\frac{n}{\ln n}} < \mathcal{X}_n(e) < n - c\sqrt{\frac{n}{\ln n}}\right) > 1 - \varepsilon.$$

Remark. The first part of Theorem 3 implies that there is no m such that $\mathcal{X}_n(e)$ is concentrated in m points. Moreover, the size of the concentration set is $O(\sqrt{\frac{\ln n}{n}})$, and this asymptotical bound is best possible.

The smoothness condition in (ii) holds, in particular, for all $e(k) = \binom{k}{2}p + o(k)$.

To prove Theorems 1 and 2, we, as usual, use the so-called second moment method.

Let us briefly discuss the scheme of our proof of Theorem 3. Denote $f(k) = e(k) - \binom{k}{2}p$. Assume that $Q \in \mathbb{R}$ is such that $-Qk \le f(k) \le Qk$ for all k. Fix real numbers $a_1 < b_1 < a_2 < b_2$ such that $a_1 > 0$, $b_1 > a_1 + 2Q$, $a_2 > 2b_1$, $b_2 > a_2 + 2Q$. Consider the sets $I_n^j = \left(p\binom{n}{2} + (a_j + Q)n,\ p\binom{n}{2} + (b_j - Q)n\right)$, $j \in \{1, 2\}$. Let $\gamma > 0$ be such that, for n large enough,

$$\min\left\{\mathsf{P}(e(G(n,p)) \in I_n^1),\ \mathsf{P}(e(G(n,p)) \in I_n^2)\right\} > \gamma. \tag{1}$$

Moreover, for every $\varepsilon > 0$, consider $a = a(\varepsilon)$ and $b = b(\varepsilon)$ such that, for n large enough, $\mathsf{P}\left(e(G(n,p)) \in I_n(\varepsilon)\right) > 1 - \varepsilon$, where

$$I_n(\varepsilon) = \left(p\binom{n}{2} - (b - Q)n, p\binom{n}{2} + (b - Q)n\right) \setminus [e(n) - an, e(n) + an]\,. \tag{2}$$

Consider a sequence of integers $m = m(n) \le \frac{c}{\sqrt{2p(1-p)}}\sqrt{\frac{n}{\ln n}}$. Denote $M = M(m) = \binom{m}{2} + m(n - m)$ the maximum possible *degree of an m-set.* Then, for a fixed m-set, the expected value of its degree equals pM. Consider the random variable $Y_m = \max_{U\in\binom{V_n}{m}} \delta(U)$. Since $Y_1 < pn + \sqrt{2p(1-p)n\ln n}$ holds a.a.s. [**2**], we immediately get that, a.a.s.

$$Y_m \le mY_1 < mpn + m\sqrt{2p(1-p)n\ln n}$$
$$= Mp + m\sqrt{2p(1-p)n\ln n} + o(n) \le Mp + cn + o(n).$$

Under the assumption that $e(G(n,p)) > p\binom{n}{2} + (a_i + Q)n$, we should "kill" at least $a_i n$ extra edges to obtain at most Qn edges more than the average value. Thus, if $c < a_i$, a.a.s. we cannot reach the desired number of edges by removing an m-set. Therefore, for every $\delta > 0$, from (1), we get that

$$\mathsf{P}\left(\mathcal{X}_n(e) < n - \frac{a_i(1-\delta)}{\sqrt{2p(1-p)}}\sqrt{\frac{n}{\ln n}}\right) > \gamma \tag{3}$$

for all large enough n and $i \in \{1, 2\}$.

Since $|f(n) - f(n-m)| = o(n)$, in the same way, from (2), we get that, for all large enough n, with a probability greater than $1 - \varepsilon$, $\mathcal{X}_n(e)$ is bounded from above by $n - \frac{a(1-\delta)}{\sqrt{2p(1-p)}}\sqrt{\frac{n}{\ln n}}$. This finishes the proof of the upper bounds.

The overall idea of the proof of the lower bounds is to remove a small set of vertices from $G(n,p)$ and get it back after the major part of extra edges is destroyed. More precisely, having $(b + Q)n$ edges more than the average, we can easily destroy extra bn edges by removing a set of $O(\sqrt{n/\ln n})$ vertices. But this is far from what we need since f may differ a lot from its bound Q. Using a half of the small set, we can reduce the number of extra edges up to $O(\sqrt{n \ln n})$. The second half is used to get the precise number of edges.

Acknowledgment. The first author's research is partially supported by NSF Grant DMS-1500121 and DMS-1764123, Arnold O. Beckman Research Award (UIUC Campus Research Board RB 18132) and the Langan Scholar Fund (UIUC). The second authors's research is supported by the grant 16-11-10014 of Russian Science Foundation.

References

1. Bollobás B., *Random Graphs*, Cambridge University Press, 2001.
2. Bollobás B., *The distribution of the maximum degree of a random graph*, Discrete Math. **32** (1980), 201–203.
3. Bollobás B. and Erdős P., *Cliques in random graphs*, Math. Proc. Camb. Phil. Soc. **80** (1976), 419–427.
4. Dutta K. and Subramanian C. R., *On induced paths, holes and trees in random graphs*, in: Proccedings ANALCO, 2018, 168–177.
5. Erdős P. and Palka Z., *Trees in random graphs*, Discrete Math. **46** (1983), 145–150.
6. Fountoulakis N., Kang R. J. and McDiarmid C., *Largest sparse subgraphs of random graphs*, European J. Combin. **35** (2014), 232–244.
7. Janson S., Łuczak T. and Ruciński A., *Random Graphs*, New York, Wiley, 2000.
8. Krivelevich M., Sudakov B., Vu V. H. and Wormald N. C., *On the probability of independent sets in random graphs*, Random Structures Algorithms **22** (2003), 1–14.
9. Matula D., *The Largest Clique Size in a Random Graph*, Tech. Rep. Dept. Comp. Sci., Southern Methodist University, Dallas, Texas, 1976.
10. Palka Z., *Bipartite complete induced subgraphs of a random graph*, Annals of Discrete Mathematics **28** (1985), 209–219.
11. Raigorodskii A. M., *On the stability of the independence number of s random subgraph*, Dokl. Math. **96** (2017), 628–630.
12. Raigorodskii A. M. and Zhukovskii M. E., *Random graphs: models and asymptotic characteristics*, Russian Math. Surveys **70** (2015) 33–81.

13. Ruciński A., *Induced subgraphs in a random graph*, Annals of Discrete Mathematics **33** (1987), 275–296.

J. Balogh, Department of Mathematical Sciences, University of Illinois at Urbana-Champaign, Urbana, USA,
Moscow Institute of Physics and Technology (State University), Russian Federation,
e-mail: jobal@math.uiuc.edu

M. Zhukovskii, Moscow Institute of Physics and Technology (State University), 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,
e-mail: zhukmax@gmail.com