Largest subgraph from a hereditary property in a random graph

Noga Alon[*] Michael Krivelevich[†] Wojciech Samotij[‡]

Abstract

Let $\mathcal{P}$ be a non-trivial hereditary property of graphs and let k be the minimum chromatic number of a graph that does not belong to $\mathcal{P}$. We prove that, for every fixed $p \in (0, 1)$, the maximum possible number of edges in a subgraph of the random graph $G(n, p)$ which belongs to $\mathcal{P}$ is, with high probability,

$$\left(1 - \frac{1}{k-1} + o(1)\right) p \binom{n}{2}.$$

1 Introduction

Let $\mathcal{P}$ be an arbitrary hereditary property of graphs, that is, a family of graphs that is closed under isomorphism and the operation of taking induced subgraphs. We assume throughout that $\mathcal{P}$ is non-trivial, i.e., it contains all edgeless graphs and misses some graph. For a graph G, let $\mathrm{ex}(G, \mathcal{P})$ denote the maximum number of edges of a subgraph of G that belongs to $\mathcal{P}$; the above definition of non-triviality guarantees that this number is well-defined. In this note we determine, for every fixed edge probability $p \in (0, 1)$, the typical asymptotic value of $\mathrm{ex}(G, \mathcal{P})$ for the random graph $G = G(n, p)$ as n tends to infinity. This is stated in the following theorem.

Theorem 1.1. *Let $\mathcal{P}$ be a non-trivial hereditary property of graphs and let $k = k(\mathcal{P}) \geq 2$ denote the minimum chromatic number of a graph that does not belong to $\mathcal{P}$. Then, for every fixed $p \in (0, 1)$, the random graph $G = G(n, p)$ satisfies, with high probability:*

$$\mathrm{ex}(G, \mathcal{P}) = \left(1 - \frac{1}{k-1} + o(1)\right) p \binom{n}{2},$$

where the $o(1)$-term tends to 0 as n tends to infinity.

[*]Department of Mathematics, Princeton University, Princeton, NJ 08544, USA and Schools of Mathematics and Computer Science, Tel Aviv University, Tel Aviv 6997801, Israel. Email: nalon@math.princeton.edu. Research supported in part by NSF grant DMS-2154082 and by USA-Israel BSF grant 2018267.

[†]School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997801, Israel. Email: krivelev@tauex.tau.ac.il. Research supported in part by USA-Israel BSF grant 2018267.

[‡]School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997801, Israel. Email: samotij@tauex.tau.ac.il. Research supported in part by the Israel Science Foundation grant 2110/22.

In fact, our argument yields the assertion of the theorem under the weaker assumption that $n^{-c} \leq p = p(n) \leq 1 - n^{-c}$ for some positive constant c that depends only on the property $\mathcal{P}$, see Proposition 2.1 below for the sparse case, the dense case follows in the same way.

In the statement above and throughout the rest of this note, the term "with high probability" (whp, for short) means, as usual, with probability tending to 1 as n tends to infinity.

The assertion of Theorem 1.1 for principal monotone properties (that is, properties defined by avoiding a single graph) is known in a strong form, and the precise range of the probability $p = p(n)$ for which it holds has been determined in [5, 7], following a considerable number of earlier results establishing special cases. There are, however, far more hereditary properties than monotone ones. A few arbitrarily chosen examples are perfect graphs, graphs with no induced hole of length 57, graphs containing no set of 10 vertices that span exactly 34 edges, or intersection graphs of discs in the plane.

The (short) proof of the theorem is presented in the next section. The third and final section contains some concluding remarks and open problems. Throughout the rest of this note, we systematically omit all floor and ceiling signs.

2 The proof

In this section, we prove Theorem 1.1. Let $\mathcal{P}$ be a hereditary property and put $k = k(\mathcal{P})$ (≥ 2). For the lower bound, note that every graph G contains a $(k-1)$-colorable subgraph with at least $(1 - \frac{1}{k-1})|E(G)|$ edges. Since every such subgraph belongs to $\mathcal{P}$ by the definition of $k = k(\mathcal{P})$, when $G = G(n,p)$ with[1] $p \gg n^{-2}$, then the assertion that

$$\mathrm{ex}(G, \mathcal{P}) \geq \left(1 - \frac{1}{k-1} - o(1)\right) p \binom{n}{2} \tag{1}$$

holds whp follows from the (standard and easy) fact that $|E(G)|$ is concentrated around its expectation as long as pn^2 tends to infinity.

We now turn to the proof of the upper bound on $\mathrm{ex}(G, \mathcal{P})$. We start by observing that, for every graph G and all $L \notin \mathcal{P}$, letting $\mathcal{F}_L$ denote the property of not containing L as an induced subgraph, we have $\mathcal{P} \subseteq \mathcal{F}_L$ and thus $\mathrm{ex}(G, \mathcal{P}) \leq \mathrm{ex}(G, \mathcal{F}_L)$. The following proposition establishes an optimal upper bound on $\mathrm{ex}(G, \mathcal{F}_L)$ in the case where G is a *sparse* binomial random graph. Recall that the 2-density of a nonempty graph L is defined by

$$m_2(L) = \max\left\{\frac{|E(K)| - 1}{|V(K)| - 2} : K \subseteq L \text{ with } |E(K)| \geq 2\right\},$$

when L has at least two edges, and $m_2(L) = 1/2$ when L has only one edge. This notion is defined so that $p \gg n^{-1/m_2(L)}$ precisely when, for every subgraph $K \subseteq L$ with at least two edges, the expected number of copies of K in $G(n,p)$ is asymptotically much

[1] When we write $f \ll g$ (resp. $f \gg g$), we mean $f = o(g)$ (resp. $g = o(f)$).

bigger than pn^2. In particular, it is not hard to see that the lower-bound assumption $p \gg n^{-1/m_2(L)}$ is necessary.

Proposition 2.1. *Let L be a nonempty graph. If $n^{-1/m_2(L)} \ll p = p(n) \ll 1$, then the random graph $G = G(n, p)$ whp satisfies*

$$\mathrm{ex}(G, \mathcal{F}_L) \le \left(1 - \frac{1}{\chi(L) - 1} + o(1)\right) p \binom{n}{2}.$$

We will derive the proposition from the following "supersaturated" version of the random analogue of Turán's theorem in $G(n, p)$ proved by Conlon and Gowers [5] (under an additional technical assumption on L) and by Schacht [7] (in full generality):

Theorem 2.2 ([5, 7])**.** *For every nonempty graph L and every $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds for every $p = p(n) \gg n^{-1/m_2(L)}$. With high probability, every subgraph G' of the random graph $G(n, p)$ with*

$$|E(G')| \ge \left(1 - \frac{1}{\chi(L) - 1} + \varepsilon\right) p \binom{n}{2}$$

contains at least $\delta n^{|V(L)|} p^{|E(L)|}$ copies of L.

Remark. Even though the above result is not explicitly stated in either [5] or [7], one obtains it easily by: (i) using the stronger conclusion of [5, Theorem 9.4] while deriving [5, Theorem 10.9]; (ii) replacing [7, Theorem 3.3] with its stronger version [7, Lemma 3.4] in the derivation of [7, Theorem 2.7]. A stronger version of Theorem 2.2, with optimal dependence of δ on ε and L, is stated and proved in [6, Theorem 1.10].

Proof of Proposition 2.1. It suffices to show that, for every fixed $\varepsilon > 0$, with high probability, every subgraph $G' \subseteq G$ with at least $(1 - \frac{1}{\chi(L)-1} + \varepsilon)p\binom{n}{2}$ edges contains more (not necessarily induced) copies of L than the total number of copies of all strict supergraphs of L (with the same vertex set) in G. (We note that a similar idea was used in [4] to derive upper bounds on induced Ramsey numbers.) On the one hand, Theorem 2.2 implies that each G' as above contains at least $\delta n^{|V(L)|} p^{|E(L)|}$ copies of L. On the other hand, a simple application of Markov's inequality gives that, with probability $1 - O(\sqrt{p})$, the total number of copies of all strict supergraphs of L is at most $O(n^{|V(L)|} p^{|E(L)|+1/2})$. Since we assume that $p \ll 1$, the latter of the above two quantities is much smaller. $\square$

The following statement is a straightforward corollary of the definition of k and Proposition 2.1 invoked with some $L \notin \mathcal{P}$ with $\chi(L) = k$.

Corollary 2.3. *For every $\varepsilon > 0$, there exists a graph H such that*

$$\mathrm{ex}(H, \mathcal{P}) < \left(1 - \frac{1}{k - 1} + \varepsilon\right) e(H).$$

We will now deduce the statement of Theorem 1.1 from Corollary 2.3 and the following standard probabilistic estimate.

Lemma 2.4. *Let H be a fixed graph, let $\varepsilon > 0$ be a small positive real and let $p \in (0,1)$ be a fixed real. Let $G = G(n,p)$ be the random graph and let $\mu = \mu(H,n,p)$ be the expected number of induced copies of H in G that contain a fixed edge of G. Then, whp, for every edge e of G the number of induced copies of H in G containing e is at least $(1-\varepsilon)\mu$ and at most $(1+\varepsilon)\mu$.*

Proof. Fix an edge e of K_n, assume it belongs to $E(G)$, and apply the edge exposure martingale to the random variable X_e counting the number of induced copies of H in G that contain e. As p and H are fixed, the expectation of this random variable is $\mu = \Theta(n^{h-2})$, where h is the number of vertices of H and the hidden constant in the Θ-notation is a function of p and H. The existence or nonexistence of each of the potential $2n-4$ edges of G (besides e) incident with e can change the value of X_e by at most $O(n^{h-3})$. Similarly, each of the $\binom{n-2}{2}$ other edges can change the value of X_e by at most $O(n^{h-4})$. Therefore, by Azuma's Inequality (see, e.g., [2, Chapter 7]), the probability that X_e deviates from its expectation by

$$\lambda \cdot \left[(2n-4)O(n^{2h-6}) + \binom{n-2}{2} O(n^{2h-8})\right]^{1/2} = \lambda \cdot O(n^{h-5/2}) = \lambda \cdot O\left(\frac{\mu}{\sqrt{n}}\right)$$

is at most $2e^{-\lambda^2/2}$. In particular, letting $\lambda = a\sqrt{\log n}$, we deduce that the probability that X_e deviates from its expectation μ by $a \cdot \mu \frac{\sqrt{\log n}}{\sqrt{n}}$ is at most $n^{-\Omega(a^2)}$. Taking a sufficiently large constant a, this implies the desired result by the union bound. □

Proof of Theorem 1.1. Let $\mathcal{P}$ and $k = k(\mathcal{P})$ be as in the statement of the theorem. Fix a small real $\varepsilon > 0$ and let H be the graph from Corollary 2.3. Fix $p \in (0,1)$ and let $G = G(n,p)$ be the binomial random graph. Let $\mathcal{H} = \{H_i\}_{i \in I}$ be the collection of all induced copies of H in G. By Lemma 2.4, whp the number of subgraphs H_i containing any edge of G is at least $(1-\varepsilon)\mu$ and at most $(1+\varepsilon)\mu$, where $\mu > 0$ is the expected number of such subgraphs. Assuming this holds, let G' be a subgraph of G that belongs to $\mathcal{P}$. We need to estimate the number of edges of G' from above. We will do so by comparing two bounds on the size of the set S of all ordered pairs (e, H_i), where e is an edge of G', $H_i \in \mathcal{H}$, and e is also an edge of H_i. Since every edge of G' is contained in at least $(1-\varepsilon)\mu$ of the graphs H_i, we have:

$$|S| \geq |E(G')|(1-\varepsilon)\mu. \tag{2}$$

On the other hand, since H satisfies the assertion of Lemma 2.3,

$$|S| \leq |\mathcal{H}| \left(1 - \frac{1}{k-1} + \varepsilon\right) |E(H)| \tag{3}$$

as every H_i can contain at most $(1 - \frac{1}{k-1} + \varepsilon)|E(H)|$ edges of G'. Indeed,

$$(V(H_i), E(G') \cap E(H_i))$$

is an induced subgraph of G', and as G' lies in $\mathcal{P}$ which is hereditary, so does this graph. Therefore, each $H_i \in \mathcal{H}$ can contain at most $(1 - \frac{1}{k-1} + \varepsilon)|E(H)|$ edges of G'. Finally, since no edge of G lies in more than $(1+\varepsilon)\mu$ members of $\mathcal{H}$, it follows that

$$|\mathcal{H}||E(H)| \leq |E(G)|(1+\varepsilon)\mu. \tag{4}$$

Combining (2), (3), and (4), we conclude that

$$\begin{aligned} |E(G')|(1-\varepsilon)\mu \leq |S| \leq |\mathcal{H}| \left(1 - \frac{1}{k-1} + \varepsilon\right) |E(H)| \\ \leq \left(1 - \frac{1}{k-1} + \varepsilon\right) |E(G)|(1+\varepsilon)\mu. \end{aligned}$$

Therefore

$$|E(G')| \leq \left(1 - \frac{1}{k-1} + \varepsilon\right) |E(G)| \frac{1+\varepsilon}{1-\varepsilon} < \left(1 - \frac{1}{k-1} + 4\varepsilon\right) |E(G)|,$$

where here we used that $\varepsilon > 0$ is small. Since ε can be chosen to be arbitrarily small, this shows that whp

$$\mathrm{ex}(G, \mathcal{P}) \leq \left(1 - \frac{1}{k-1} + o(1)\right) |E(G)| = \left(1 - \frac{1}{k-1} + o(1)\right) p\binom{n}{2}.$$

This, together with (1), completes the proof of the theorem. □

3 Concluding remarks

- Theorem 1.1 can be proved in a more self-contained way, using Szemerédi's Regularity Lemma and following the approach in [1, Section 4.4]. Since unlike the proof described here, this proof does not work for $p \leq n^{-\varepsilon}$ when $\varepsilon > 0$ is fixed, we omit the details.

- It may be interesting to characterize all edge probabilities p for which the assertion of Theorem 1.1 holds. It is not difficult to describe hereditary properties (even monotone ones) for which the fraction of edges of $G(n,p)$ that lie in a maximum subgraph that belongs to the property changes, whp, several times as $p = p(n)$ increases from 0 to 1.

- Our main result, Theorem 1.1, shows that if $\mathcal{P}$ misses a bipartite graph, then for $G = G(n,p)$ we have whp $\mathrm{ex}(G, \mathcal{P}) = o(n^2)$. It would be interesting to provide a more accurate estimate for this "bipartite" case. It seems plausible that in this case $\mathrm{ex}(G, \mathcal{P}) \leq n^{2-\varepsilon}$ for some $\varepsilon = \varepsilon(\mathcal{P}) > 0$.

- The typical edit distance of a random graph from a hereditary property is very different from the typical minimum number of edges that have to be deleted from it to get a graph that belongs to the property. The latter quantity is the one studied here, the former is treated, for example, in [3]. An example illustrating the difference is that of the property of avoiding an induced copy of a long even cycle C_{2k}. Theorem 1.1 shows that for, say, $p = 1/2$, whp one has to delete nearly all $(1/4 + o(1))n^2$ edges of $G(n, 1/2)$ to get a subgraph that contains no induced copy of C_{2k}. On the other hand, since the vertices of C_{2k} cannot be covered by $k-1$ cliques it suffices, whp, to add to $G(n, 1/2)$ only $(1/4+o(1))n^2/(k-1)$ edges in order to cover all its vertices by $k-1$ cliques, ensuring that the resulting graph will not contain an induced copy of C_{2k}.

Acknowledgment

The initial results in this note were obtained when the second author visited the Department of Mathematics of Princeton University. He would like to thank the department for the hospitality.

References

[1] Noga Alon, Anna Gujgiczer, János Körner, Aleksa Milojević, and Gábor Simonyi, *Structured Codes of Graphs*, SIAM J. Discrete Math. **37** (2023), no. 1, 379–403.

[2] Noga Alon and Joel H. Spencer, *The probabilistic method*, fourth ed., Wiley Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., Hoboken, NJ, 2016.

[3] Noga Alon and Uri Stav, *What is the furthest graph from a hereditary property?*, Random Structures Algorithms **33** (2008), no. 1, 87–104.

[4] David Conlon, Domingos Dellamonica, Steven La Fleur, Vojtěch Rödl, and Mathias Schacht, *A note on induced Ramsey numbers*, A journey through discrete mathematics, Springer, 2017, pp. 357–366.

[5] David Conlon and William Timothy Gowers, *Combinatorial theorems in sparse random sets*, Ann. of Math. (2) **184** (2016), no. 2, 367–454.

[6] David Conlon, William Timothy Gowers, Wojciech Samotij, and Mathias Schacht, *On the KŁR conjecture in random graphs*, Israel J. Math. **203** (2014), no. 1, 535–580.

[7] Mathias Schacht, *Extremal results for random discrete structures*, Ann. of Math. (2) **184** (2016), no. 2, 333–365.