Thresholds in Random Motif Graphs

Michael Anastos
Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA
http://www.math.cmu.edu/~manastos
manastos@andrew.cmu.edu

Peleg Michaeli
School of Mathematical Sciences, Tel Aviv University, Israel
http://www.math.tau.ac.il/~pelegm
peleg.michaeli@math.tau.ac.il

Samantha Petti
School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia, USA
http://people.math.gatech.edu/~spetti3
spetti@gatech.edu

Abstract

We introduce a natural generalization of the Erdős-Rényi random graph model in which random instances of a fixed motif are added independently. The *binomial random motif graph* $G(H, n, p)$ is the random (multi)graph obtained by adding an instance of a fixed graph H on each of the copies of H in the complete graph on n vertices, independently with probability p. We establish that every monotone property has a threshold in this model, and determine the thresholds for connectivity, Hamiltonicity, the existence of a perfect matching, and subgraph appearance. Moreover, in the first three cases we give the analogous hitting time results; with high probability, the first graph in the random motif graph process that has minimum degree one (or two) is connected and contains a perfect matching (or Hamiltonian respectively).

2012 ACM Subject Classification Mathematics of computing → Random graphs; Mathematics of computing → Paths and connectivity problems

Keywords and phrases Random graph, Connectivity, Hamiltonicty, Small subgraphs

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2019.66

Category RANDOM

Funding *Peleg Michaeli*: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement number 676970, RANDGEOM).
Samantha Petti: This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1650044.

Acknowledgements We thank Alan Frieze for helpful discussions and for connecting the authors.

1 Introduction

In the late 1950's Gilbert [11] and Erdős and Rényi [6] introduced two of the most fundamental models for generating random graphs: the *binomial random graph* $G(n, p)$, generated by independently adding an edge between each pair of vertices in the complete graph on n vertices with probability p, and the the *uniform random graph* $G(n, m)$, which is a uniformly chosen graph from all graphs on n vertices with m edges. Since, the extensive study of these simple constructions has influenced a variety of fields including combinatorics, computer science, and statistical physics (see [9, 4, 12] for surveys).

Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019).
Editors: Dimitris Achlioptas and László A. Végh; Article No. 66; pp. 66:1–66:19
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Detailed analysis of the model has led to the development of plethora of new techniques in probability for analyzing random processes, and the model has been used to verify the existence of structures with certain properties [1]. In computer science, the model has been used to analyze the performance of algorithms on an "average" case, showing that NP complete problems may be easier random instances.

The rise of data in the form of graphs (e.g. internet connections, biological networks, social networks) has further fueled the study of random graphs. In practice, the comparison of real world networks to the Erdős-Rényi model is a popular technique for highlighting the non-random aspects of a network's structure [20, 2, 17, 14]. Moreover, the model has inspired many other models which are designed to mirror some characteristic of real-world networks (e.g. Watts-Strogatz graphs have small diameter [18], Barabási-Albert preferential attachment graph exhibit a power law degree distribution [3]).

In this paper we consider a natural generalization of the Erdős-Rényi model in which random motifs are added rather than random edges. A *motif* is a fixed small subgraph, such as a triangle. The motifs that are overrepresented in a network are correlated to the function of the network [20, 2, 17, 14]. Analyzing random graphs formed as the union of many instances of a particular motif H will give insight into the structural properties of networks with many copies of the motif H.

We define the ***binomial*** *random motif graph* $G(H, n, p)$ as the random (multi)graph obtained by adding an instance of H on each of the $\binom{n}{|V(H)|} \cdot |V(H)|!/\operatorname{aut}(H)$ copies of H in the complete graph on n vertices K_n, independently with probability p. Here by $\operatorname{aut}(H)$ we denote the number of automorphisms of H. Note that if H is an edge, then this is exactly $G(n, p)$. Similarly, the ***uniform*** *random motif graph* $\bar{G}(H, n, m)$ is the random (multi)graph obtained by taking the union of m uniformly chosen copies of H in K_n without replacement.

Closely related to $\bar{G}(H, n, m)$ is the *random motif graph process* $\bar{G}_0(H, n), \bar{G}_1(H, n), ..., \bar{G}_N(H, n)$. $\bar{G}_0(H, n)$ is the empty graph on n vertices and for $0 \leq i \leq N = \binom{n}{|V(H)|}/\operatorname{aut}(H)$ the graph $\bar{G}_{i+1}(H, n)$ is generated by adding to $\bar{G}_i(H, n)$ a copy of H, H_{i+1}, chosen uniformly at random from all the copies of H except those in $\{H_1, H_2, ..., H_i\}$ i.e. those that have been added to $\bar{G}_0(H, n)$ so far. Clearly $\bar{G}_m(H, n)$ has the same law as $\bar{G}(H, n, m)$. In addition, by setting H to be an edge we retrieve the random graph process introduce by Erdős and Rényi [7]. By considering the random motif graph process in place of the uniform random motif graph model we can phrase results in a finer way (see for example Theorem 3).

In this work we show that every monotone graph property has a threshold in the binomial random motif graph $G(H, n, p)$. Then we determine the thresholds for connectivity, existence of a perfect matching, Hamiltoncity and subgraph appearance. In the first three cases we also show a hitting time result, according to which w.h.p.[1] the first graph in the random motif graph process that has minimum degree one (or two) is connected (or Hamiltonian respectively).

1.1 Notation

Throughout we assume the motif H has no isolated vertices. For an integer $r \geq 0$, denote by $m_r(H)$ the number of its copies in K_n which intersect the set $[r]$. For an integer $d \geq 0$ we define the quantities $\delta_d(H)$ and $p_d(H)$ by

$$\delta_d(H) := \lceil d/\delta(H) \rceil - 1 \qquad \text{and} \qquad p_d^{\pm}(H) := \frac{\ln n + \delta_d(H) \ln \ln n \pm x(n)}{m_1(H)},$$

[1] That is, with probability tending to 1 as n tends to infinity.

where $x(n)$ is any function of n satisfying $1 \ll x(n) \ll \ln\ln n$. Note that the expected number of added instances of H in $G(H, n, p_1^{\pm}(H))$ is $m_n(H) \cdot p_1^{\pm}(H)$, which only depends on n and on $|V(H)|$.

1.2 Results

A function $p^* = p^*(n)$ is a threshold for a monotone increasing property $\mathcal{P}$ in the random graph $G(H, n, p)$ if

$$\lim_{n\to\infty} \mathsf{Pr}[G(H,n,p) \in \mathcal{P}] = \begin{cases} 0 & \text{if } p/p^* \to 0, \\ 1 & \text{if } p/p^* \to \infty, \end{cases}$$

as $n \to \infty$. Our first result is a generalization of a theorem by Bollobás and Thomason [5].

▶ **Theorem 1.** *Every non-trivial monotone graph property has a threshold.*

Given Theorem 1, a natural goal is to find the thresholds for various monotone properties. The remaining results of this paper are dedicated towards this goal; we determine the threshold for connectivity, the existence of a perfect matchings, Hamiltonicity, and subgraph appearance.

A first such result, which generalizes a result in [6], shows, in particular, that the expected number of motifs needed to make the random motif graph connected depends only on the number of (non-isolated) vertices of the motif.

▶ **Theorem 2.** *Let H be a fixed graph. Then*

$$\lim_{n\to\infty} \mathit{Pr}[G(H,n,p) \textit{ is connected }] = \begin{cases} 0 & p \le p_1^-(H), \\ 1 & p \ge p_1^+(H). \end{cases}$$

In fact, we show a hitting time result, according to which the hitting time of connectivity equals, w.h.p., the hitting time of minimum degree one. In other words, the random motif graph process becomes connected exactly when the last isolated vertex disappears, with high probability.

Fix an integer n and a graph H. Let $\tau_{\mathsf{c}} = \min\{i : \bar{G}_i(H,n) \text{ is connected}\}$, and for $d \ge 1$ denote $\tau_d = \min\{i : \delta(\bar{G}_i(H,n)) \ge d\}$.

▶ **Theorem 3.** *Let H be a fixed graph. Then w.h.p. $\tau_{\mathsf{c}} = \tau_1$.*

We remark that if the motif H is connected, every connectivity related question depends solely on the sets of vertices on which copies of H are added, and not on the way they are put there. Thus, we may model the question as a (binomial or uniform) random k-uniform hypergraph, where $k = |V(H)|$. In this case, Theorems 2 and 3 follow immediately from known results about (loose) connectivity in random hypergraphs (see, e.g.,[16]).

In the following two theorems we show that the existence of a perfect matching is also dependent on the number of non-isolated vertices of the motif.

▶ **Theorem 4.** *Let H be a fixed graph, and assume that n is even. Then,*

$$\lim_{n\to\infty} \mathit{Pr}[G(H,n,p) \textit{ has a perfect matching }] = \begin{cases} 0 & p \le p_1^-(H), \\ 1 & p \ge p_1^+(H). \end{cases}$$

Let $\tau_{\mathsf{M}} = \min\{i : \bar{G}_i(H,n) \text{ has a perfect matching}\}$. The analogue hitting time result is also true.

▶ **Theorem 5.** *Let H be a fixed graph, and assume that n is even. Then w.h.p.* $\tau_{\mathsf{M}} = \tau_1$.

Theorem 6 establishes that the thresholds for minimum degree 2 and for Hamiltonicity are the same. Theorem 7 shows the hitting time version of that result.

▶ **Theorem 6.** *Let H be a fixed graph. Then*

$$\lim_{n\to\infty} \mathsf{Pr}[G(H,n,p) \text{ is Hamiltonian }] = \begin{cases} 0 & p \leq p_2^-(H), \\ 1 & p \geq p_2^+(H). \end{cases}$$

Let $\tau_{\mathsf{H}} := \min\{i : \bar{G}_i(H,n) \text{ is Hamiltonian}\}$.

▶ **Theorem 7.** *Let H be a fixed graph. Then w.h.p.* $\tau_{\mathsf{H}} = \tau_2$.

Next, we describe the threshold for the appearance of a subgraph S. If S appears in a random motif graph, then S is a subgraph of some configuration of b copies of H whose union contains a vertices. For such an (a,b) covering of S, we call a subset of the covering containing b' copies of H whose union contains a' vertices an (a',b') subset. The threshold for the appearance of S depends on $\bar{\gamma}$, the maximum over all covering configurations of the minimum ratio a'/b' for all subsets of the covering configuration. Definition 15 formally describes $\bar{\gamma}$.

▶ **Theorem 8.** *Let H be a fixed graph, let S be a fixed graph, and set $v = |V(H)|$ and $\bar{\gamma} = \bar{\gamma}(S,H)$. Then*

$$\lim_{n\to\infty} \mathsf{Pr}\big[S \subseteq \bar{G}(H,n,m)\big] = \begin{cases} 0 & m \ll n^{v-\bar{\gamma}} \\ 1 & m \gg n^{v-\bar{\gamma}}. \end{cases}$$

The number of excess edges of a connected graph S, or simply its *excess*, is defined to be $\mathrm{exc}(S) = |E(S)| - |V(S)| + 1$. In particular, trees have excess 0. We say that S is *unicyclic* if its excess is 1, or *complex* if its excess is at least 2. The following theorem gives a simple description of $\bar{\gamma}$ when the motif H is a path, which allows us to deduce how the copies of H fit together to form a copy of S at the threshold when S first appears. If S is a tree, a minimal set of edge disjoint copies of H typically forms S. If S is complex, each copy of the path H typically contributes a single edge to S. If it is unicyclic, it may be formed by any edge disjoint configuration of paths H.

▶ **Theorem 9.** *Let H be a path of length $v-1$ and let S be a connected graph. Let β be the minimum number of edge-disjoint copies of H whose union contains S as a subgraph. Let $\eta = \min_{X\subseteq S} \frac{|V(X)|}{|E(X)|}$. Then*

$$\bar{\gamma} = \begin{cases} v-1+1/\beta & \mathrm{exc}(S) = 0, \\ v-1 & \mathrm{exc}(S) = 1, \\ v-2+\eta & \mathrm{exc}(S) \geq 2. \end{cases}$$

In the case where the motif is a long path, this result establishes a connection between the threshold for the appearance of subgraphs in random motif graphs and the threshold for the appearance of subgraphs in the trace of a random walk on the complete graph K_n (studied in [13]). Let S be a connected graph and β be the minimum number of paths in any edge-disjoint decomposition of S into paths. If H is longer than the maximum length path in such a minimum edge-disjoint path decomposition, then the threshold implied by Theorem 9 matches the threshold for the appearance of S in the trace of a random walk on the complete graph [13].

This should not come as a surprise; by noticing that when the motif is a long path, the random motif graph model approximates the trace model, in the following sense. One may sequentially "cut" the (lazy) simple random walk into chunks with buffers of length 1. We delete loops created by the trace of each chunk, and we enforce the condition that the remaining edges span a path of length ℓ (which is fixed but large). Hence the trace of each such chunk is an independent copy of a path of length ℓ. Thus we may couple the trace model and the random motif model such that the trace model will include the random motif model plus some loops plus a small number of buffer edges (which gets smaller as ℓ gets larger).

Viewing this analogy this way, we may use Theorems 8 and 9 to reprove the main theorems of [13] for the case where the base graph is complete.

2 Existence of thresholds for monotone properties

Proof of Theorem 1. Assume that $\mathcal{P}$ is a monotone increasing property and let $H_1, H_2, \ldots, H_{m_0(H)}$ be the copies of H that are spanned by K_n. Observe that

$$\mathsf{Pr}[G(H,n,p) \in \mathcal{P}] = \sum_{i=0}^{m_0(H)} \sum_{S \in \binom{m_0(H)}{i}} p^i (1-p)^{\binom{n}{|V(H)|}-i} \mathbb{I}\left(\bigcup_{j \in S} H_j \in \mathcal{P} \right)$$

is a polynomial in p. In addition, since $\mathcal{P}$ is increasing, it is increasing. Therefore we may define $p_{1/2}$ by

$$\mathsf{Pr}\left[G(H,n,p_{1/2}) \in \mathcal{P}\right] = \frac{1}{2}.$$

We will show that $p_{1/2}$ is a threshold for $\mathcal{P}$. For two random graphs G, G' we write $G \subseteq G'$ if G, G' can be coupled such that G is a subgraph of G'.

First let $p = \omega(n)p_{1/2}$ where $\omega(n) \to \infty$ as $n \to \infty$ and let $k \in \mathbb{N}$. Let $G_i(H,n,p_{1/2})$ be distributed as a $G(H,n,p_{1/2})$ for $i \in [k]$. Then, by considering the probability of no appearance of a fixed copy of H, we have that the graph $\cup_{i\in[k]} G_i(H,n,p_{1/2})$ is distributed as $G(H,n,(1-(1-p_{1/2})^k))$. Thereafter $1-(1-p_{1/2})^k \leq kp_{1/2}$ implies,

$$\bigcup_{i\in[k]} G_i(H,n,p_{1/2}) = G(H,n,(1-(1-p_{1/2})^k)) \subseteq G(H,n,kp_{1/2}).$$

Hence,

$$\begin{aligned}\mathsf{Pr}\left[G(H,n,\omega(n)p_{1/2}) \in \mathcal{P}\right] &= 1 - \mathsf{Pr}\left[G(H,n,\omega(n)p_{1/2}) \notin \mathcal{P}\right] \\ &\geq \lim_{k\to\infty} 1 - \mathsf{Pr}\left[G(H,n,kp_{1/2}) \notin \mathcal{P}\right] \\ &\geq 1 - \lim_{k\to\infty} \prod_{1=i}^{k} \mathsf{Pr}\left[G_i(H,n,p_{1/2}) \notin \mathcal{P}\right] = 1.\end{aligned}$$

Now assume that $p = p_{1/2}/\omega(n)$ for some $\omega(n) \to \infty$ as $n \to \infty$ and let $k \in \mathbb{N}$. Similarly to before, if we let $G_i(H,n,p_{1/2}/\omega(n))$ to be distributed as a $G(H,n,p_{1/2}/\omega(n))$ for $i \in [k]$ then, we have that

$$\begin{aligned}\bigcup_{i\in[k]} G_i(H,n,p_{1/2}/\omega(n)) &= G(H,n,(1-(1-p_{1/2}/\omega(n))^k)) \\ &\subseteq G(H,n,kp_{1/2}/\omega(n)) \subseteq G(H,n,p_{1/2}).\end{aligned}$$

Hence,

$$\begin{aligned}\frac{1}{2} = \Pr\big[G(H,n,p_{1/2}) \in \mathcal{P}\big] &= 1 - \Pr\big[G(H,n,p_{1/2}) \notin \mathcal{P}\big] \\ &\geq \lim_{k\to\infty} 1 - \Pr\big[G(H,n,kp_{1/2}/\omega(n)) \notin \mathcal{P}\big] \\ &\geq 1 - \lim_{k\to\infty} \prod_{1=i}^{k} \Pr\big[G_i(H,n,p_{1/2}/\omega(n)) \notin \mathcal{P}\big] \\ &= 1 - \Pr\big[G_i(H,n,p_{1/2}/\omega(n)) \notin \mathcal{P}\big]^k.\end{aligned}$$

Rearranging the above gives,

$$\Pr\big[G_i(H,n,p_{1/2}/\omega(n)) \notin \mathcal{P}\big] \geq \lim_{k\to\infty} \left(\frac{1}{2}\right)^{1/k} = 1. \qquad ◀$$

3 Connectivity

Proof of Theorem 2. If $p \leq p_1^-(H)$ then by Theorem 19 the minimum degree of $G(H,n,p)$ is w.h.p. 0, hence it is not connected.

Suppose $p \geq p_1^+(H)$. In fact, for the argument below, we only assume that $p = (\ln n \pm o(\ln n))/m_1(H)$ (and the conclusion will follow by monotonicity). Let k denote the number of vertices of H. For $r = 1,\ldots,n/2$ denote by S_r the number of connected components of size r in $G(H,n,p)$. Note that for $r \geq k$, if a set of cardinality r is a connected component, then there exist $\lceil (r-1)/(k-1)\rceil$ copies of H inside the set which appear in $G(H,n,p)$, and there are no edges between it and its complement, so none of the $q = q_r(H)$ copies of H that intersect that set appear. By Lemma 17,

$$qp \sim r f_k(r/n)\cdot \ln n \geq (1+o(1))k\ln n.$$

Let $\eta = k!/\operatorname{aut}(H)$ and suppose $r \geq k$. By Lemma 18 and by the union bound there exist constants $c, c', C > 0$ depending only on H such that

$$\begin{aligned}\Pr[S_r > 0] &\leq \binom{n}{r}\binom{\eta\binom{r}{k}}{\lceil\frac{r-1}{k-1}\rceil} p^{\lceil\frac{r-1}{k-1}\rceil}(1-p)^q \leq \left(\frac{en}{r}\right)^r \left(\frac{e\eta\binom{r}{k}p}{\lceil\frac{r-1}{k-1}\rceil}\right)^{\lceil\frac{r-1}{k-1}\rceil} e^{-qp} \\ &\leq \left[C\cdot\frac{n}{r}\cdot r\cdot p^{(r-1)/(r(k-1))} n^{-(1+o(1))k/r}\right]^r \\ &= \left[C\cdot \operatorname{polylog} n \cdot n^{1/r-(1+o(1))k/r}\right]^r = o(1).\end{aligned}$$

It follows that

$$\begin{aligned}\Pr[G(H,n,p) \text{ is not connected}] &\leq \sum_{r=1}^{n/2} \Pr[S_r > 0] \\ &= \Pr[S_1 > 0] + \sum_{r=k}^{n/2}\Pr[S_r > 0] = \Pr[S_1 > 0] + o(1),\end{aligned}$$

but according to Theorem 19 (for $p \geq p_1^+(H)$), there are no isolated vertices w.h.p., and the result follows. ◀

Note that a consequence of this proof is that for $p = (\ln n \pm o(\ln n))/m_1(H)$, with high probability, every connected component is of cardinality 1 or at least $n/2$. This means that w.h.p. there exists a unique "giant" component of linear size, and the rest of the vertices are isolated. The next lemma, whose proof uses a simple second moment argument, estimates the number of these isolated vertices for $p_- = (\ln n - \ln\ln n)/m_1(H)$.

▶ **Lemma 10.** *The number of isolated vertices in $G(H, n, p_-)$ is w.h.p. at most $2\ln n$.*

Proof. Let D_0 be the number of isolated vertices in $G(H, n, p_-)$. First,

$$\mathsf{E}[D_0] = n(1-p_-)^{m_1(H)} \sim ne^{-p_- \cdot m_1(H)} = ne^{-\ln n + \ln\ln n} = \ln n.$$

Moreover,

$$\mathsf{E}\left[D_0^2\right] = \mathsf{E}[D_0] + n(n-1)(1-p_-)^{m_2(H)}.$$

Denote $L := 2m_1(H) - m_2(H)$. Thus

$$\mathsf{E}\left[D_0^2\right] \le \mathsf{E}[D_0] + \mathsf{E}[D_0]^2(1-p_-)^{-L},$$

and since $(1-p_-)^{-L} - 1 \sim Lp_-$, we have that

$$\mathsf{Var}[D_0] \le \mathsf{E}[D_0] + \mathsf{E}[D_0]^2((1-p_-)^{-L} - 1) \le \mathsf{E}[D_0] + (L+1)p_-\,\mathsf{E}[D_0]^2.$$

Thus, noting that $Lp_- = o(1)$,

$$\begin{aligned}\mathsf{Pr}[D_0 \ge 2\ln n] &= \mathsf{Pr}[|D_0 - \mathsf{E}[D_0]| \ge (1+o(1))\,\mathsf{E}[D_0]] \\ &\le (1+o(1))\left(\mathsf{E}[D_0]^{-1} + (L+1)p_-\right) = o(1).\end{aligned}$$ ◀

Proof of Theorem 3. Denote $p_\pm = (\ln n \pm \ln\ln n)/m_1(H)$ and $m_\pm = p_\pm \cdot m_n(H)$. By asymptotic equivalence of the binomial and the uniform models (see, e.g., [12]*Section 1.4) we have that w.h.p. $G(H, n, m_-)$ has a unique giant component, and the rest of the connected components are isolated vertices, whose number is at most $2\ln n$. Denote the set of these isolated vertices by V_0. Together with Theorem 2 we also conclude that w.h.p.

$$m_- \le \tau_1 \le \tau_\mathsf{c} \le m_+.$$

We may thus couple $\bar{G}(H, n, m_-)$, $\bar{G}(H, n, \tau_1)$, $\bar{G}(H, n, \tau_\mathsf{c})$ and $\bar{G}(H, n, m^+)$ such that

$$\bar{G}(H, n, m_-) \subseteq \bar{G}(H, n, \tau_1) \subseteq \bar{G}(H, n, \tau_\mathsf{c}) \subseteq \bar{G}(H, n, m_+),$$

by starting with $\bar{G}(H, n, m_-)$ and adding $M = m_+ - m_-$ random copies of H to create $\bar{G}(H, n, m_+)$. Note that if none of these M edges is fully contained in V_0 (and the coupling succeeds) then $\tau_1 = \tau_\mathsf{c}$. Thus, there exist positive constants C_1, C_2 such that,

$$\mathsf{Pr}[\tau_1 < \tau_\mathsf{c}] \le o(1) + M \cdot \frac{C_1\binom{|V_0|}{k}}{m_n(H) - m_+} \le o(1) + C_2 \cdot \frac{m_n(H)\ln\ln n}{m_1(H)} \cdot \frac{\ln^2 n}{m_n(H)} = o(1).$$ ◀

4 Hamiltoncity and Perfect Matchings

The proof of Theorems 7 and 5 can be given in parallel, using the same techniques and tools. For clarity though, in this section we focus mainly on proving Theorem 7 and we give a sketch of the proof of Theorem 5 in the appendix.

For proving our Hamiltonicity result we use the standard technique of Posa's rotations. We define Small to be the vertices of significantly smaller degree than the expected one and we set Large to be the rest of the vertices. We first show that small to medium subsets of Large expand and that the vertices in Small are well spread. This is done in the context of Lemmas 11 and 12, 13 respectively. We use these properties of Small and Large in order to prove all the the ingredients needed to apply the Posa's rotations, which we gather in Lemma 14.

Let $p_0 := (\ln n - 2\ln\ln n)/m_1(H)$ and recall that $p_2^{\pm} = (\ln n + r_2 \ln\ln n \pm \omega(1))/m_1(H)$, $r_2 = \lfloor 2/\delta(H) - 1 \rfloor$. W.h.p. (see [9]) we can couple $G(H,n,p_0), G(H,n,p_2^-), \bar{G}(H,n,\tau_2)$ and $G(H,n,p_2^+)$ such that

(i) $G(H,n,p_0) \subset G(H,n,p_2^-) \subset \bar{G}(H,n,\tau_2) \subset G(H,n,p_2^+)$ and

(ii) there are $(1+o(1))(p_2^- - p_0)\frac{r!}{\mathrm{aut}(H)}\binom{n}{r} > n\ln\ln n/2r$ copies of H in $G(H,n,p_2^-)$, hence in $\bar{G}(H,n,\tau_2)$, that are not present in $G(H,n,p_0)$.

Observe that the above coupling and Theorem 7 imply Theorem 6. In addition a similar coupling and Theorem 5 imply Theorem 4.

We now define the sets Small, Large based on the degrees of the vertices in $G(H,n,p_0)$. Let Large $= \{v \in V : v$ intersects at least $\ln\ln n$ copies of H in $G(H,n,p_0)\}$ and Small $= V \setminus$ Large.

▶ **Lemma 11.** *W.h.p. every $S \subset$ Large of size at most $n/30r$ satisfies $|N(S)| \geq 10|S|$.*

▶ **Lemma 12.** *W.h.p. for every pair $u,v \in$ Small there do not exist $\ell \leq 6$ copies of H in $G(H,n,p_2^+)$ that span a connected subgraph containing both u,v. Hence w.h.p. every pair $u,v \in$ Small is at distance at least 7 in $G(H,n,p_2^+)$.*

▶ **Lemma 13.** *W.h.p. for every $v \in V$ there exists at most one copy of H in $G(H,n,p_2^+)$, hence in $\bar{G}(H,n,\tau_2)$, that intersect both $\{v\}$ and Small $\setminus \{v\}$.*

Now we generate $\bar{G}(H,n,\tau_2)$ as follows. We first generate $G'_0 = G(H,n,p_0)$. Then we randomly permute the copies of H not appearing in G'_0, let them be $H_1, H_2, ...$ We also let $S_0 = \emptyset$. We define the sequences $G'_0, G'_1, ...$ and $S_0, S_1, ...$ in the following way. At step $i \in \mathbb{N}$ we query H_i whether it is incident to a vertex in Small. If it is then we set $S_i = S_{i-1}$ and $G'_i = G'_{i-1} \cup H_i$. Otherwise we set $S_i = S_{i-1} \cup \{H_i\}$ and $G'_i = G'_{i-1}$. Let $t^* = \min\{i : \delta(G'_i) = 2\}$ and $S_{t^*} = \{H_{i_1}, H_{i_2}, ..., H_{i_w}\}$.

Given the sequence $G'_0, G'_1, ..., G'_{t^*}$ and the set $S_{t^*} = \{H_{i_1}, H_{i_2}, ..., H_{i_w}\}$ we define the graph sequence $F_0, ..., F_w$ by $F_0 = G'_{t^*}$ and $F_j = F_{j-1} \cup H_{i_j}$ for $1 \leq j \leq w$. Observe that S_{t^*} consists of all copies of H in $\{H_1, ..., H_{t^*}\}$ that have not been added to G'_0, equivalently the copies of H that are not incident to Small. Thus $F_w = G'_{t^*} \cup \left(\bigcup_{j=1}^{w} H_{i_j}\right) = G'_0 \cup \left(\bigcup_{i=1}^{t^*} H_i\right) = \bar{G}(H,n,\tau_2)$.

▶ **Lemma 14.** *W.h.p. the following hold:*

i) $w \geq n\ln\ln n/2r - n$,

ii) *every $S \subset V$ of size at most $n/30r$ satisfies $|N(S)| \geq 2|S|$ in F_0,*

iii) *F_0 is connected,*

iv) *for every $1 \leq j \leq w$, $\epsilon > 0$, and every set Q_j consisting of ϵn^2 edges not present in F_j there exist a constant $C_\epsilon > 0$ such that the probability that Q_j intersects $E(H_{i_{j+1}})$ is at least C_ϵ.*

We are now ready to apply Posa's rotations . For that assume that F_j is not Hamiltonian and consider a longest path in F_j, P_j, $j \geq 0$. Let x, y be the end-vertices of P_j. Given yv where v is an interior vertex of P_j we can obtain a new longest path $P'_j = x..vy..w$ where w is the neighbor of v on P_j between v and y. In such a case we say that P'_j is obtained from P_j by a rotation with the end-vertex x being the fixed end-vertex.

Let $\textsc{End}_j(x;P_j)$ be the set of end-vertices of longest paths of F_j that can be obtained from P_j by a sequence of rotations that keep x as the fixed end-vertex. Thereafter for $z \in \textsc{End}_j(x;P_j)$ let $P_j(x,z)$ be a path that has end-vertices x,z and can be obtain form P_j by a sequence of rotations that keep x as the fixed end-vertex. Observe that for $z \in \textsc{End}_j(x;P_j)$ and $z' \in \textsc{End}_j(z;P_j(x,z))$ there exists a z-z' path $P_{z,z'}$ of length $|P_j|$ that can be obtained from P_j via a sequence of Posa rotations. Thus we can conclude that $\{z,z'\}$ does not belong to F_j. Indeed assume that $\{z,z'\} \in E(G_i)$. Then we can close $P_{z,z'}$ into a cycle $C_{z,z'}$ that is not Hamiltonian. Since F_j is connected there is an edge e spanned by $V(C_{z,z'}) \times V \setminus V(C_{z,z'})$. $E(C_{z,z'}) \cup \{e\}$ spans a path of length $|P_j|+2$ contradicting the maximality of P_j. Similarly if $\{z,z'\} \in E(H_{i_{j+1}})$ then F_{j+1} is either Hamiltonian or it contains a path that is longer than P_j. At the same time it follows (see [9]*Corollary 6.7) that

$$|N(\textsc{End}(x,P_j))| < 2|\textsc{End}(x,P_j)|.$$

Moreover for every $z \in \textsc{End}_j(x;P_j)$

$$|N(\textsc{End}(z,P_j(x,z)))| < 2|\textsc{End}(z,P_j(x,z))|.$$

As a consequence of Lemma 11, we have that $|\textsc{End}(x,P_j)| \geq n/30r$ and $|\textsc{End}(z,P_j(x,z))| \geq n/30r$ for every $z \in \textsc{End}_j(x;P_j)$. Let $E_j = \{\{z,z'\} : z \in \textsc{End}_j(x;P_j)$ and $z' \in \textsc{End}_j(z;P_j(x,z))\}$. Then $|E_j| \geq (n/30r)^2/2$.

Now let Y_j be the indicator of the event $\{E_j \cap E(H_{i_{j+1}}) \neq \emptyset\}$ and set $Z = \sum_{j=1}^{w} Y_i$. From Lemma 14 iv) we have $\mathsf{Pr}[Y_j = 1] \geq C_\epsilon$ (here $\epsilon = 1/2(30r)^2$). In the event that G_w is not Hamiltonian, $Z \leq n$ while Y_j is a $\mathsf{Bernoulli}(C_\epsilon)$ random variable for $1 \leq j \leq w$. Since $w \geq n \ln\ln n/2r - n$ we have $\mathsf{Pr}[\mathsf{Bin}(w, C_\epsilon) \leq n] = o(1)$. Hence w.h.p. $F_w = \bar{G}(H,n,\tau_2)$ is Hamiltonian and the hitting time for Hamiltonicity equals the hitting time for minimum degree 2.

5 Subgraph appearance

In $G(n,p)$ there is only one way for a specified subgraph to appear on a fixed set of vertices: all the edges in the subgraph must be present. In the case of random motif graphs, there are multiple ways to place motifs so that a specified subgraph appears on a fixed set of vertices. For example, in a random two-path graph, a triangle may appear on $\{1,2,3\}$ if (i) the paths $(1,2,3)$ and $(3,1,z)$ are present or (ii) the paths $(1,2,x)$, $(2,3,y)$ and $(3,1,z)$ are present. In order to pin down the threshold for subgraph appearance, it is necessary to understand the various motif configurations that cause the subgraph to appear and their relative probabilities. The following definition provides the notation to describe such configurations.

▶ **Definition 15.** *Let V be a set of vertices. Let S be a fixed graph on a subset of the vertices of V. Let $H_1, H_2, \ldots H_b$ be copies H also defined on subsets of vertices of V.*

(a) *We say $\{H_1, H_2, \ldots H_b\}$ is an (a,b) covering of S if (i) $S \subseteq \bigcup_{j=1}^{b} H_j$, (ii) $|V(\bigcup_{j=1}^{b} H_j)| = a$, and (iii) for each $\ell \in [b]$, $S \not\subseteq \bigcup_{j=1}^{b} H_j \setminus H_\ell$.*

(b) *Let $k(a,b)$ be the number of unique configurations of (a,b) coverings, i.e. the number of ways to place b copies of H on a vertices such that conditions (i)-(iii) of (a) hold. Enumerate the possible configurations of (a,b) coverings with values in $[k(a,b)]$. For $i \in [k(a,b)]$, an (a,b,i) covering of S is an (a,b) covering with configuration i.*

(c) *We say the set $\{F_1, F_2, \ldots F_{b'}\}$ (with precisely b' elements) is an (a',b') subset of an (a,b,i) covering $\{H_1, H_2, \ldots H_b\}$ if (i) $\{F_1, F_2, \ldots F_{b'}\} \subseteq \{H_1, H_2, \ldots H_b\}$, and (ii) $|V(\bigcup_{\ell=1}^{b'} F_\ell)| = a'$.*

(d) *Let* $\mathcal{I}(S,H) = \{(a,b,i) \mid$ *there exists an* (a,b) *covering of* S *by* H *and* $i \in [k(a,b)]\}$.
(e) *For* $(a,b,i) \in \mathcal{I}(S,H)$, *let*

$$\mathcal{D}(a,b,i) = \{(a',b') \mid \textit{there exists an } (a',b') \textit{ subset of the } (a,b,i) \textit{ covering}\}.$$

(f) *For* $(a,b,i) \in \mathcal{I}(S,H)$, *let* $\gamma(a,b,i) = \min_{(a',b')\in\mathcal{D}(a,b,i)} \frac{a'}{b'}$ *and denote*

$$\bar{\gamma} = \max_{(a,b,i)\in\mathcal{I}(S,H)} \gamma(a,b,i).$$

Proof of Theorem 8. Let $G \sim \bar{G}(H,n,m)$. We say that an instance of the subgraph S in G is an (a,b,i) instance if the placed graphs $H_1, \dots H_b$ that contribute at least one edge to S form an (a,b,i) covering of S. Let X_i^{ab} denote the number of (a,b,i) instances of S in G. Let $Z = \sum_{(a,b,i)\in\mathcal{I}(S,H)} X_i^{ab}$ be the total number of instances of the subgraph S in G.

First we use the first moment method to show that if $m \ll n^{v-\bar{\gamma}}$, then the probability that S occurs as a subgraph is $o(1)$. It suffices to show that for all $(a,b,i) \in \mathcal{I}(S,H)$, $\mathsf{E}\left[X_i^{ab}\right] = o(1)$ since

$$\mathsf{Pr}[Z > 0] \leq \mathsf{E}[Z] = \sum_{(a,b,i)\in\mathcal{I}(S,H)} X_i^{ab},$$

and $|\mathcal{I}(S,H)|$ is a constant independent of n.

We now compute $\mathsf{E}\left[X_i^{ab}\right]$ for a fixed triple $(a,b,i) \in \mathcal{I}(S,H)$. Let $\{F_1, \dots F_{b'}\}$ be an (a',b') subset of the configuration (a,b,i) with $a'/b' = \gamma(a,b,i)$. Let Y be the number of instances of $F = \bigcup_{i=1}^{b'} F_{b'}$ in G formed by the configuration $\{F_1, \dots F_{b'}\}$. Since an (a,b,i) instance of S contains an instance of the configuration $\{F_1, \dots F_{b'}\}$, $X_i^{ab} \leq Y$. The number of ways to select a' vertices is at most $n^{a'}$. The probability that a labeled copy of H is placed on a specified set of vertices is m/n^v. We compute

$$\mathsf{E}\left[X_i^{ab}\right] \leq \mathsf{E}[Y] \leq cn^{a'}\left(\frac{m}{n^v}\right)^{b'} = c\left(n^{\gamma(a,b,i)-v}m\right)^{b'} \leq c\left(n^{\bar{\gamma}-v}m\right)^{b'},$$

where c is a constant depending only on the number of automorphisms of S and the number of automorphisms of the configuration $\{F_1, \dots F_{b'}\}$. It follows that for $m \ll n^{\bar{\gamma}-v}$, $\mathsf{E}\left[X_i^{ab}\right] = o(1)$, as desired.

Next we use the second moment method to show that if $m \gg n^{v-\bar{\gamma}}$ then S appears as a subgraph almost surely. It suffices to show that there exists some $(a,b,i) \in \mathcal{I}(S,H)$ such that X_i^{ab} is almost surely positive. Let (a,b,i) be such that $\bar{\gamma} = \gamma(a,b,i)$. We apply Corollary 4.3.5 of [1] to show that X_i^{ab} is almost surely positive. Let $X_i^{ab} = \sum_j A_j$ where A_j is an indicator random variable for the event that there is an (a,b,i) instance of S formed by a configuration of $H_1, H_2, \dots H_b$ each present on a specified set of vertices. Fix A_ℓ, and let

$$\Delta^* = \sum_{j\sim\ell} \mathsf{Pr}[A_j \mid A_\ell],$$

where $j \sim \ell$ indicates that A_j and A_ℓ are not independent. By 4.3.5 of [1], if $\mathsf{E}\left[X_i^{ab}\right] \to \infty$ and $\Delta^* = o(\mathsf{E}\left[X_i^{ab}\right])$, then $X_i^{ab} > 0$ almost surely.

First we show that $\mathsf{E}\left[X_i^{ab}\right] \to \infty$. We compute as above

$$\mathsf{E}\left[X_i^{ab}\right] \geq c'n^a\left(\frac{m}{n^v}\right)^b = c'\left(n^{a/b-v}m\right)^b \geq c'\left(n^{\bar{\gamma}-v}m\right)^b$$

where c' is a constant depending only on the number of automorphisms of S and the number of automorphisms of the configuration $\{H_1, \dots H_b\}$. It follows that if $m \gg n^{v-\bar{\gamma}}$ then $\mathsf{E}\left[X_i^{ab}\right] \to \infty$.

Finally, we show $\Delta^* = o(\mathsf{E}\left[X_i^{ab}\right])$. Observe that under the assumption $m \gg n^{v-\bar{\gamma}}$,

$$\Delta^* = \sum_{(a',b')\in\mathcal{D}(a,b,i)} cn^{a-a'}\left(\frac{m}{n^v}\right)^{b-b'} = \sum_{(a',b')\in\mathcal{D}(a,b,i)} c\,\mathsf{E}\left[X_i^{ab}\right]\left(n^{-a'/b'+v}m^{-1}\right)^{b'}$$
$$\leq c'\,\mathsf{E}\left[X_i^{ab}\right]\left(n^{-\gamma(a,b,i)+v}m^{-1}\right)^{b} = c'\,\mathsf{E}\left[X_i^{ab}\right]\left(n^{v-\bar{\gamma}}m^{-1}\right)^{b} = o\left(\mathsf{E}\left[X_i^{ab}\right]\right). \quad \blacktriangleleft$$

6 Conclusion

6.1 The value of the random motif model

The study of random motif graphs has the potential to strengthen the impact of the Erdős-Rényi construction. In the context of analyzing real-world networks with an overrepresented motif, random motif graphs may be a more insightful null hypothesis model to compare against to identify non-random structure. For instance by studying subgraphs counts of random H motif graphs one can determine if some larger motif pattern is a byproduct of having many copies of H or is itself some novel aspect of the network structure. Moreover, it is possible that a random motif graph may be used to establish the existence of a graph with some extremal property of interest. Finally, random motif graphs can be used as an alternate definition of average case for analyzing algorithms under the assumption that the input has some motif structure.

6.2 Future directions: understanding threshold behavior more broadly

We have established that random motif graphs behave similarly to traditional Erdős-Rényi random graphs with respect to thresholds and hitting times for monotone properties. Does similar behavior appear when we consider random graphs formed by randomly adding primitive subgraphs H whose size scales with n, the number of vertices of the random graph? Instead of taking H to be a fixed motif, H could be a path, cycle, matching or clique whose size depends on n, for example. Some of these cases were in fact studied in several contexts. For example, the union of $d \geq 3$ random perfect matchings is contiguous to the random d-regular graph, and is sometimes easier to analyze [19]. Moreover, we can consider the class of models where H itself is chosen from some probability distribution. In several cases, this has been studied as well. For instance, [10] and [8] consider the case when H is the uniform spanning tree, and [15] considers the case when H is an Erdős-Rényi random graph with constant density and size dependent on n. Further study of these models is a first step toward delineating a larger family of random graphs that exhibit Erdős-Rényi like threshold and hitting time behaviors.

References

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

2 Uri Alon. Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, 8(6):450–461, 2007.

3 Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999. `doi:10.1126/science.286.5439.509`.

4 Béla Bollobás. *Random graphs*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London, 1985.

5 Béla Bollobás and Andrew Thomason. Hereditary and monotone properties of graphs. In *The mathematics of Paul Erdős, II*, volume 14 of *Algorithms Combin.*, pages 70–78. Springer, Berlin, 1997. doi:10.1007/978-3-642-60406-5_7.
6 Paul Erdős and Alfréd Rényi. On random graphs. I. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.
7 Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.
8 Alan Frieze, Navin Goyal, Luis Rademacher, and Santosh Vempala. Expanders via random spanning trees. *SIAM Journal on Computing*, 43(2):497–513, 2014. doi:10.1137/120890971.
9 Alan Frieze and Michał Karoński. *Introduction to random graphs*. Cambridge University Press, Cambridge, 2016. doi:10.1017/CBO9781316339831.
10 Alan Frieze, Michał Karoński, and Luboš Thoma. On perfect matchings and Hamilton cycles in sums of random trees. *SIAM Journal on Discrete Mathematics*, 12(2):208–216, 1999. doi:10.1137/S0895480196313790.
11 Edgar N. Gilbert. Random graphs. *Annals of Mathematical Statistics*, 30:1141–1144, 1959. doi:10.1214/aoms/1177706098.
12 Svante Janson, Tomasz Łuczak, and Andrzej Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000. doi:10.1002/9781118032718.
13 Michael Krivelevich and Peleg Michaeli. Small subgraphs in the trace of a random walk. *Electronic Journal of Combinatorics*, 24(1):Paper 1.28, 22, 2017.
14 Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri B. Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.
15 Samantha Petti and Santosh S. Vempala. Approximating Sparse Graphs: The Random Overlapping Communities Model. *arXiv e-prints*, February 2018. arXiv:1802.03652.
16 Daniel Poole. On the strength of connectedness of a random hypergraph. *Electronic Journal of Combinatorics*, 22(1):Paper 1.69, 16, 2015.
17 Sen Song, Per Jesper Sjöström, Markus Reigl, Sacha B. Nelson, and Dmitri B. Chklovskii. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS biology*, 3(3):e68, 2005.
18 Duncan J. Watts and Steven H. Strogatz. Collective dynamics of "small-world" networks. *Nature*, 393(6684):440–442, 1998. doi:10.1038/30918.
19 Nicholas C. Wormald. Models of random regular graphs. In *Surveys in combinatorics, 1999 (Canterbury)*, volume 267 of *London Math. Soc. Lecture Note Ser.*, pages 239–298. Cambridge Univ. Press, Cambridge, 1999.
20 Esti Yeger-Lotem, Shmuel Sattath, Nadav Kashtan, Shalev Itzkovitz, Ron Milo, Ron Y. Pinter, Uri Alon, and Hanah Margalit. Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction. *Proceedings of the National Academy of Sciences of the United States of America*, 101(16):5934–5939, 2004.

A Estimates for useful functions

▶ **Lemma 16.** *For $r = r(n)$, if $k = |V(H)|$ and $\alpha = r/n$ then $m_r(H) \sim rm_1(H) \cdot \frac{1-(1-\alpha)^k}{k\alpha}$.*

Proof. Observe that for $r \geq 0$,

$$m_r(H) = \left(\binom{n}{k} - \binom{n-r}{k}\right) \cdot \frac{k!}{\mathrm{aut}(H)},$$

thus

$$\frac{m_r(H)}{rm_1(H)} = \frac{\binom{n}{k} - \binom{n-r}{k}}{r\left(\binom{n}{k} - \binom{n-1}{k}\right)} \sim \frac{n^k - (n-r)^k}{r(n^k - (n-1)^k)} = \frac{1-(1-\alpha)^k}{r(1-(1-n^{-1})^k)} \sim \frac{1-(1-\alpha)^k}{k\alpha}. \blacktriangleleft$$

For $r \geq 1$, denote by $q_r(H)$ the number of copies of H that intersect $[r]$ but that are not contained in $[r]$.

▶ **Lemma 17.** *For $r = r(n)$, if $k = |V(H)|$ and $\alpha = r/n$ then $q_r(H) \sim rm_1(H) \cdot \frac{1-(1-\alpha)^k-\alpha^k}{k\alpha}$.*

Proof. Observe that for $r \geq 0$,

$$q_r(H) = \left(\binom{n}{k} - \binom{n-r}{k} - \binom{r}{k}\right) \cdot \frac{k!}{\mathrm{aut}(H)},$$

thus

$$\frac{q_r(H)}{rm_1(H)} = \frac{\binom{n}{k} - \binom{n-r}{k} - \binom{r}{k}}{r\left(\binom{n}{k} - \binom{n-1}{k}\right)} \sim \frac{n^k - (n-r)^k - r^k}{r(n^k - (n-1)^k)}$$
$$= \frac{1-(1-\alpha)^k-\alpha^k}{r(1-(1-n^{-1})^k)} \sim \frac{1-(1-\alpha)^k-\alpha^k}{k\alpha}. \quad ◀$$

For convenience we define for $\alpha \in [0,1]$ and $k \geq 1$,

$$f_k(\alpha) = \frac{1-(1-\alpha)^k-\alpha^k}{k\alpha}.$$

▶ **Lemma 18.** *For $2 \leq k \leq r$ we have that $rf_k(r/n) \geq (1+o(1))k$.*

Proof. Write $g_k(\alpha) = f_k(\alpha) \cdot k\alpha = 1-(1-\alpha)^k-\alpha^k$. Observe that it is strictly increasing in $(0, 1/2)$. Note also that

$$n \cdot g_k\left(\frac{k}{n}\right) = n - ne^{-k^2/n} - o(1) \sim k^2.$$

It follows that

$$\frac{kr}{n} \cdot f_k\left(\frac{r}{n}\right) = g_k\left(\frac{r}{n}\right) \geq g_k\left(\frac{k}{n}\right) \sim \frac{k^2}{n},$$

so $rf_k(r/n) \geq (1+o(1))k$. ◀

B Minimum degree

▶ **Theorem 19.** *With high probability*

$$\delta(G(H,n,p_d^-)) < d \qquad \text{and} \qquad \delta(G(H,n,p_d^+)) \geq d.$$

Proof. Let $\delta = \delta(H)$. It suffices to show that with high probability for $\ell \in \mathbb{Z}_{\geq 0}$

$$\Pr\left[\delta(G(H,n,p_{\ell\cdot\delta}^-)) > (\ell-1)\delta\right] = o(1) \tag{1}$$

and

$$\Pr\left[\delta(G(H,n,p_{\ell\cdot\delta}^+)) < \ell\delta\right] = o(1). \tag{2}$$

Proof of (1): Let $p = p_{\ell\cdot\delta}^-$. For $v \in V$ let $I_v = \mathbb{I}\{d(v) = (\ell-1)\delta\}$ and $Z = \sum_{v\in V} I_v$.

$$\begin{aligned}\mathsf{E}[Z] &\geq (1-o(1))n\binom{n-1}{v_H-1}^{\ell-1} p^{\ell-1}(1-p)^{m_1(H)-\ell+1} \\ &\geq C_1 n(pn^{(v_H-1)})^{\ell-1}e^{-(p+4p^2)(m_1(H)-\ell+1)} \\ &\geq C_2 n(\log n)^{\ell-1}e^{-\log n-(\ell-1)\log\log n+\omega(1)} \geq e^{\omega(1)/2}.\end{aligned}$$

In addition,

$$\begin{aligned}
\mathsf{E}\left[Z^2\right] &= \sum_{u,v\in V} \mathsf{Pr}[I_v \wedge I_u] \\
&\le \mathsf{E}[Z]^2 + \sum_{u\neq v\in V} \mathsf{Pr}[I_u \wedge I_v \wedge \{u, v \text{ lie on the same copy of } H\}] \\
&\le \mathsf{E}[Z]^2 + \binom{n}{2}\binom{n-2}{r-2}\frac{r!}{\mathrm{aut}(H)}p(1-p)^{(1-o(1))2m_1} \\
&= \mathsf{E}[Z]^2 + nm_1p(1-p)^{m_1-1}C_3(1-p_2^-)^{(1-o(1))m_1} = \mathsf{E}[Z]^2 + o(1)\,\mathsf{E}[Z] \\
&= (1+o(1))\,\mathsf{E}[Z]^2.
\end{aligned}$$

Chebyshev's inequality give us,

$$\mathsf{Pr}[|Z - \mathsf{E}[Z]| \ge \mathsf{E}[Z]/2] \le \frac{\mathsf{E}\left[Z^2\right] - \mathsf{E}[Z]^2}{0.25\,\mathsf{E}[Z]^2} = o(1).$$

Hence with high probability there exist vertices of degree $(\ell-1)\delta$.

Proof of (2): Let $p = p^+_{\ell\cdot\delta}$. Let $\mathcal{E}_1$ be the event that in $G(H, n, p)$ there exists a vertex of degree $d \le \ell\delta$ that lies on more than ℓ copies of H. In the event $\mathcal{E}_1$ there exists a vertex v and a vertex set S of size d such that all the neighbors of v lie in S and at least $\ell+1$ copies of H intersect $S \cup \{v\}$, each in at least $\delta+1$ vertices. Therefore,

$$\begin{aligned}
\mathsf{Pr}[\mathcal{E}_1] &\le n\binom{n}{d}[1-p]^{\binom{n-d-1}{v_H-1}}\left(\binom{d+1}{\delta+1}\binom{n-\delta-1}{v_H-\delta-1}\right)^{\ell+1} p^{\ell+1} \\
&\le e^{-p\cdot\binom{n-d-1}{v_H-1}} n^{d+1-\delta(\ell+1)}[n^{v_H-1}p]^{\ell+1} \\
&\le e^{-(1+o(1))p\cdot m_1(H)}(\log^2 n)^{\delta_d(H)+1} = o(1).
\end{aligned}$$

In the event $\neg\mathcal{E}_1$ the number of vertices of degree less than $\ell\delta$ is bounded by the number of vertices that are covered by at most $\ell-1$ copies of H. Thus

$$\begin{aligned}
\mathsf{Pr}\left[\delta(G(H, n, p^+_{\ell\cdot\delta})) < \ell\delta\right] &\le \mathsf{Pr}[\mathcal{E}_1] + n\sum_{i=0}^{\ell-1}\binom{m_1(H)}{i}p^i(1-p)^{m_1(H)-i} \\
&\le \ell n(m_1(H)p)^{\ell-1}e^{-pm_1(H)+p\ell} + o(1) \\
&\le \ell n[2\log n]^{\ell-1}e^{-\log n-(\ell-1)\log\log n-\omega(1)} + o(1) = o(1). \qquad \blacktriangleleft
\end{aligned}$$

C Proofs of lemmas for Hamiltoncity

Proof of Lemma 11. If there exists $S \subset$ LARGE of size $n^{19/20} \le |S| \le n/30r$ such that $|N(S)| < 10|S|$ then there exist sets A, B of size $n^{19/20} \le s \le n/30r$ and $n - 11s$ respectively such that no copy of H, H' satisfies $|A \cap H'| = 1$ and $|B \cap H'| = r - 1$ (take $S = A$ and B to be any subset of $V \setminus (S \cup N(S))$ of size $n - 11s$). The probability of such event occurring is bounded above by

$$\begin{aligned}
&\sum_{s=n^{19/20}}^{n/30r}\binom{n}{s}\binom{n-s}{10s}(1-p_0)^{\frac{r!}{\mathrm{aut}(H)}\cdot s\binom{n-11s}{r-1}} \\
&\le \sum_{s=n^{19/20}}^{n/30r}\left[\frac{en}{s}\cdot\left(\frac{en}{10s}\right)^{10} e^{-p_0\frac{r!}{\mathrm{aut}(H)}\cdot\binom{n-11s}{r-1}}\right]^s
\end{aligned}$$

$$\dots \leq \sum_{s=n^{19/20}}^{n/30r} \left[\left(\frac{n}{s}\right)^{11} e^{-\frac{\ln n - 2\ln\ln n}{\binom{n-1}{r-1}} \cdot \binom{n-11s}{r-1}} \right]^s$$
$$\leq \sum_{s=n^{19/20}}^{n/30r} \left[\left(\frac{n}{s}\right)^{11} \left(\frac{\ln^2 n}{n}\right)^{(1-\frac{11s}{n})\cdots(1-\frac{11s-r+2}{n-r+2})} \right]^s$$
$$\leq \sum_{s=n^{19/20}}^{n/30r} \left[\left(\frac{n}{s}\right)^{11} \left(\frac{\ln^2 n}{n}\right)^{1-\frac{12sr}{n}} \right]^s \leq \sum_{s=n^{19/20}}^{n/30r} \left[n^{11/20} \left(\frac{\ln^2 n}{n}\right)^{18/30} \right]^s = o(1).$$

Now assume that there exists a set $S \subset \textsc{Large}$ of size at most $n^{19/20}$ that satisfies $|N(S)| < 10|S|$. Since every vertex in S is in at least $\ln\ln n$ copies of H and every copy of H covers r vertices we have that S intersects at least $|S|\ln\ln n/11$ copies of H. Each of those copies is spanned by $S \cup N(S)$. Therefore there exists a set $W \supseteq S \cup N(S)$ of size $w = |W| = 11|S| \leq 11n^{19/20}$ that intersects at least $\frac{|W|\ln\ln n}{11r}$ copies of H each, in at least 2 vertices. Since every vertex in $\textsc{Large}$ has $\ln\ln n$ neighbors $|W| \geq \ln\ln n$. The probability that such a set exists is bounded by

$$\sum_{w=\ln\ln n}^{11n^{19/20}} \binom{n}{w} \binom{r!\binom{w}{2}\binom{n}{r-2}}{w\ln\ln n/11r} p_0^{w\ln\ln n/11r}$$
$$\leq \sum_{w=\ln\ln n}^{11n^{19/20}} n^w \left(\frac{11er^3wn^{r-2}}{\ln\ln n}\right)^{w\ln\ln n/11r} p_0^{w\ln\ln n/11r}$$
$$\leq \sum_{w=\ln\ln n}^{11n^{19/20}} \left[n^{11r/\ln\ln n} \cdot \frac{11er^3wn^{r-2}}{\ln\ln n} \cdot p_0 \right]^{w\ln\ln n/11r}$$
$$\leq \sum_{w=\ln\ln n}^{11n^{19/20}} \left(n^{11r/\ln\ln n} \cdot \frac{w\log n}{n} \right)^{w\ln\ln n/11r} = o(1). \qquad \blacktriangleleft$$

Proof of Lemma 12. For $u \in V$ and $Q \subset V$ let $S(u,Q)$ be the event that in $G(H,n,p_0)$ u intersects at most $\ln\ln n$ copies of H that do not intersect Q. For $0 \leq |Q| \leq 6$,

$$\Pr[S(u,Q)] \leq \Pr\left[\mathsf{Bin}\left(\frac{r!}{\mathrm{aut}(H)}\binom{n-7}{r-1}, p_0\right) \leq \ln\ln n\right] \leq n^{-0.9}.$$

Let $\mathcal{B}$ be the event that for some $u, v \in \textsc{Small}$ there exist $\ell \leq 6$ copies of H in $G(H,n,p_2^+)$ that span a connected subgraph containing both u, v. If $\mathcal{B}$ occurs then we can find a set $Q = \{v = v_0, v_1, ..., v_{\ell-1}, v_\ell = u\}$ such that i) the events $S(v, Q \setminus \{v\})$, $S(u, Q \setminus \{u\})$ occur and ii) there exist $H_1, ..., H_\ell$ in $G(H,n,p_2^+)$ such that $H_i \cap Q = \{v_{i-1}, v_i\}$. Since all the aforementioned events are independent

$$\Pr[\mathcal{B}] \leq \sum_{\ell=1}^{6} \sum_{Q=\{v_0,v_1,\dots,v_\ell\}} \Pr[S(v_0, Q\setminus\{v_0\}] \cdot \left(\binom{n-2}{r-2}\frac{r!}{\mathrm{aut}(H)}p_2^+\right)^\ell \cdot \Pr[S(v_\ell, Q\setminus\{v_\ell\}]$$
$$\leq \sum_{\ell=1}^{6} n^{\ell+1} \cdot n^{-0.9} \cdot \left(\frac{C_3 \ln n}{n}\right)^\ell \cdot n^{-0.9} = o(1). \qquad \blacktriangleleft$$

Proof of Lemma 13. Lemma 12 implies that w.h.p. there do not exist $v \in V$ and $u, w \in$ Small, $u \neq w$ such that in $G(H, n, p_2^+)$ v and u are in a copy of H and v and w are in a copy of H. The probability that there exist $v \in V$, $u \in \text{Small} \setminus \{v\}$ that are both contained in more than one copy of H in $G(H, n, p_2^+)$ is bounded by

$$\sum_{v,u \in V} \Pr[S(u, \{v\})] \left(\binom{n-2}{r-2} \frac{r!}{\text{aut}(H)} p_2^+ \right)^2 \leq C_4 n^{-0.9} \log^2 n = o(1). \qquad \blacktriangleleft$$

Proof of Lemma 14.

1. Recall that we can couple $G(H, n, p_0), \bar{G}(H, n, \tau_2)$ such that $G(H, n, p_0) \subset \bar{G}(H, n, \tau_2)$ w.h.p. and there are at least $n \ln \ln n / 2r$ copies of H in $\bar{G}(H, n, \tau_2)$ that are not present in $G(H, n, p_0)$. From Lemma 13 it follows that w.h.p. each of those copies that spans a vertex in Small also spans a unique vertex in $V \setminus \text{Small}$. Hence $w \geq n \ln \ln n / 2r - n$.
2. Let $S \subset V$, $|S| \leq n/30r$ and set $S_s = S \cap \text{Small}$, $S_L = S \cap \text{Large}$. Lemma 11 implies that $|N(S_L)| \geq 10|S_L|$. In the case $|S_L| \geq |S_s|$ we have

   $$|N(S)| \geq |N(S_L) \setminus S_s| \geq 10|S_L| - |N(S_L) \cap S_s| \geq 10|S_L| - |S_s| \geq 9|S_L| \geq 2|S|.$$

   Next assume $|S_L| < |S_s|$. Lemma 12 implies that no two vertices in Small are within distance 2 in $G(H, n, p_2^+)$, hence their neighborhoods are disjoint. Also F_0 has minimum degree 2. Therefore $|N(S_s)| \geq 2|S_s|$. Now let $S_L = S_1 \cup S_2$ where S_2 consists of all the vertices in S_L that are within distance 2 from S_s and $S_1 = S_L \setminus S_2$. If $|S_1| \geq |S_2|$ then since S_s and S_1 have disjoint neighborhoods we have that

   $$|N(S)| \geq |N(S_s) \setminus S_2| + |N(S_1) \setminus S_2| \geq 2|S_s| + 10|S_1| - 2|S_2| \geq 2|S|.$$

   Otherwise $|S_s| > |S_L|$ and $|S_2| > |S_1|$. For $v \in S_s$ let $N_{S_2}(v)$ be the set of vertices in S_2 that are within distance 2 from v, hence $\cup_{v \in S_s} N_{S_2}(v) = |S_2|$. Lemma 12 states that no two vertices in Small are within distance 6, thus for $v, u \in S_s, v \neq u$ the sets $N(N_{S_2}(v)), N(N_{S_2}(u))$ are disjoint. In addition since $N_{S_2} \subset S_L$ and $|S_L| \leq |S| \leq n/30r$, Lemma 11 implies that $|N(N_{S_2}(v))| \geq 10|N_{S_2}(v)|$ for all $v \in S_s$. Thus

   $$\begin{aligned} |N(S)| &\geq \sum_{v \in S_s} |N(N_{S_2}(v) \cup \{v\})| \\ &\geq \sum_{v \in S_s} [10|N_{S_2}(v)| - |\{v\}|] \cdot \mathbb{I}_{N_{S_2}(v) \neq \emptyset} + |N(v)| \mathbb{I}_{N_{S_2}(v) = \emptyset} \\ &\geq \sum_{v \in S_s} 2 = 2|S_s| \geq |S|. \end{aligned}$$

3. Assume that there exists a set $S \subset V$ such that S is a connected component of F_0 and let $s = |S|$. F_0 has minimum degree 2 therefore $s \geq 3$. Let $S_L = S \cap \text{Large}$ and $S_s = S \cap \text{Small}$. Lemma 13 implies that every vertex in S_L can be adjacent to at most 1 vertex in Small hence $|S_L| \geq |S_s|$. Thereafter Lemma 11 implies that $|S| > n/30r$ since otherwise

   $$|N(S)| \geq |N(S_L)| - |S_s| \geq 10|S_L| - |S_L| > 0.$$

   Finally the probability that there exists a connected component of size $n/30r \leq s \leq n/2$ in $G(H, n, p_0) \subset F_0$ is bounded by

   $$\sum_{s=n/30r}^{0.5n} \binom{n}{s} (1 - p_0)^{\frac{r!}{\text{aut}(H)} \cdot s \binom{n-s}{r-1}} \leq \sum_{s=n/30r}^{0.5n} \left[\frac{en}{s} \cdot e^{-C_5 \ln n} \right]^s = o(1).$$

4. First we show that w.h.p. $|\textsc{Small}| \leq n^{0.1}$. Indeed by Markov's inequality,

$$\Pr\left[|\textsc{Small}| > n^{0.1}\right] \leq n^{-0.1} \cdot n \Pr\left[\mathsf{Bin}\left(\frac{r!}{\mathrm{aut}(H)}\binom{n-1}{r-1}, p_0\right) \leq \ln\ln n\right] = o(1).$$

Now let Q_j be a set of ϵn^2 edges not present in F_j and Q'_j be the subset of Q_j consisting of the edges that are not incident to Small. Then w.h.p. $|Q'_j| = (1+o(1))\epsilon n^2$. Every edge in Q'_j belongs to $C_6 n^{r-2}$ copies of H that are no present in F_j and every copy of H may cover at most $\binom{r}{2}$ edges in Q'_j. Therefore there exists a set W_i consisting of at least $C_6 n^{r-2} \cdot (1+o(1))\epsilon n^2/\binom{r}{2}$ distinct copies of H that intersect Q'_j. $H_{i_{j+1}}$ is uniformly distributed among the copies of H that are not present in F_j and are not incident to a vertex in Small. Thus

$$\Pr[iv] = \Pr[H_i \in W_i] \geq \frac{C_6 n^{r-2} \cdot (1+o(1))\epsilon n^2/\binom{r}{2}}{n^r} \geq C_7\epsilon = C_\epsilon. \qquad \blacktriangleleft$$

D Proof sketch of Theorems 4 and 5

To prove Theorem 5 we first indicate the edge set Q_1, consisting of the edges that are incident to vertices of degree 1. Then we delete these edges and the vertex set U_1 consisting of the vertices incident to them. Thereafter we use exactly the same techniques as above in order to find a Hamilton cycle in the remaining graph. We use half of the edges of that cycle and the edges in Q_1 to form a perfect matching.

Given the above, the only substantial difference is that while generating $\bar{G}(H, n, \tau_1)$ (in place of $\bar{G}(H, n, \tau_2)$) we stop at time $t^* = \min\{i : \delta(G'_i) = 1\}$. The proofs of all Lemmas with exception the proof of Lemma 14, follow in exactly the same way. For the proof of Lemma 14 we have to be slightly more cautious as we want to prove the corresponding statements for the subgraph that is spanned by $V \setminus U_1$. Thus we have to use $\textsc{Small} \setminus U_1$ and $\textsc{Large} \setminus U_1$ in place of Small and Large respectively.

E Proof of Theorem 9

Before proving Theorem 9, we derive an expression for a'/b' and establish the following upper bound on $\gamma(a, b, i)$.

▶ **Lemma 20.** *Consider an (a, b, i) covering of S by a path of length $v-1$ and an (a', b') subcovering with c' connected components. Let S_j be the subgraph of S covered by j^{th} connected component of the (a', b') subcovering. Let $f_j = |E(S_j)| - |V(S_j)| + 1$ and $f' = \sum_{j=1}^{c'} f_j$. Let k be the number of duplicate edges in the (a', b') subcovering, i.e. k is the smallest integer such that removing k edges from multigraph union of b' copies of H can yield a simple graph. Then*

$$\frac{a'}{b'} = v - 1 + \frac{c' - f' - k}{b'} \tag{3}$$

and

$$\gamma(a, b, i) \leq \begin{cases} v - 1 + \frac{1-f}{b} & (a, b, i) \text{ is edge-disjoint} \\ v - 1 - \frac{f}{b} & (a, b, i) \text{ is not edge-disjoint} \end{cases} \tag{4}$$

Proof. We compute a'. Note that each of the b' copies of H contributes v vertices, however vertices may be counted multiple times. We compute

$$a' = b'v - \left(b' - \sum_{j=1}^{c'} 1 - f_j \right) - k = b'(v-1) + c' - f' - k,$$

where the first term subtracted corresponds to doubling counting vertices in each connected component and subtracting k corresponds to removing double counting for vertices adjacent to edges of S that are covered multiple times.

By definition, $\gamma(a, b, i) \leq a/b$. For the $(a', b') = (a, b)$ subcover that is the entire (a, b, i) cover, $c' = 1$, $f' = f$ and $k = 0$ if (a, b, i) is edge-disjoint and $k \geq 1$ if (a, b, i) is not edge-disjoint. Thus, Equation (4) follows directly from Equation (3). ◀

Proof of Theorem 9. We consider each case separately.

Case: $f = 0$**.** Consider an (a, b, i) covering. If (a, b, i) is edge-disjoint, then $b \geq \gamma$. It follows from Equation (4) that

$$\gamma(a, b, i) \leq \begin{cases} v - 1 + \frac{1}{\beta} & (a, b, i) \text{ is edge-disjoint} \\ v - 1 & (a, b, i) \text{ is not edge-disjoint.} \end{cases}$$

Thus $\bar{\gamma} = \max_{(a,b,i) \in \mathcal{I}(S,H)} \gamma(a, b, i) \leq v - 1 + 1/\beta$.
Next consider an edge-disjoint cover of S by β copies of H, (a, β, i). By Equation (3), for any (a', b') subcover of the (a, β, i) cover,

$$\frac{a'}{b'} = v - 1 + \frac{c'}{b'}.$$

This value is minimized with $c' = 1$ and $b' = \beta$, which is achieved by the (a, β) subcover which is the whole cover. Thus $\gamma(a, \beta, i) = v - 1 + 1/\beta$, and so $\bar{\gamma} \geq v - 1 + 1/\beta$.

Case: $f = 1$**.** By Equation (4), $\gamma(a, b, i) \leq v - 1$ for all (a, b, i) and so it follows that $\bar{\gamma} \leq v - 1$.
Next consider an edge-disjoint cover of S, (a, b, i). By Equation (3), for any (a', b') subcover of the (a, β, i) cover,

$$\frac{a'}{b'} = v - 1 + \frac{c' - 1}{b'}.$$

This value is minimized with $c' = 1$, which is achieved by the (a, b) subcover which is the whole cover. Thus $\gamma(a, b, i) = v - 1$, and so $\bar{\gamma} \geq v - 1$.

Case: $f \geq 2$**.** Consider an (a, b, i) cover. By Equation (3),

$$\gamma(a, b, i) = \min_{(a',b') \in \mathcal{D}(a,b,i)} \frac{a'}{b'} = \min_{a',b',c',k} v - 1 + \frac{c' - f' - k'}{b'}.$$

Let t' and e' be the number of edges and vertices of S covered by the subcover, so $e' = t' - c' + f' + k$. It follows

$$\gamma(a, b, i) = \min_{t',e',b'} v - 1 + \frac{t' - e'}{b'}. \tag{5}$$

To give an upper bound on $\gamma(a, b, i)$, we construct a subcover of the (a, b, i) cover as follows. Let X be a subgraph of S with t^* vertices and e^* edges such that $t^*/e^* = \eta$. Let t', e', b' correspond to the subcover that minimally covers X, and let C be the subgraph of S covered by this subcover (so X is a subgraph of C).

We claim that $t'-e' \leq t^*-e^*$. Note that $t'-t^* = |V(C) \setminus V(X)|$ and $e'-e^* = |E(C) \setminus E(X)|$. In each component of $C \setminus E(X)$, at least one vertex is included in $V(X)$. Since the number of vertices in a connected component is at least the number of edges in the connected component minus one, and at least one vertex in each connected component is not included in $V(C) \setminus V(X)$, it follows that $|V(C) \setminus V(X)| \geq |E(C) \setminus E(X)|$. Thus $t' - t^* \leq e' - e^*$ and the claim follows.
By considering this subcover with parameters t', e', b', we obtain

$$\gamma(a, b, i) \leq v - 1 + \frac{t' - e'}{b'} \leq v - 1 + \frac{t^* - e^*}{e^*} = v - 2 + \eta$$

since $b' \leq e^*$ and $t^* - e^* \leq 0$. It follows that $\bar{\gamma} \leq v - 2 + \eta$.
Finally to lower bound $\bar{\gamma}$ we consider a cover in which there are $b = |E(S)|$ copies of H and each copy covers precisely one edge of S. In this case in all subcovers $b' = e'$. By Equation (5)

$$\gamma(a, b, i) = \min_{t', e', b'} v - 1 + \frac{t' - e'}{b'} = \min_{t', e'} v - 2 + \frac{t'}{e'} = v - 2 - \eta.$$

Thus $\bar{\gamma} \geq v - 2 + \eta$. ◀