Hamiltonicity of random graphs in the stochastic block model

Michael Anastos
Freie Universität Berlin
manastos@zedat.fu-berlin.de

Alan Frieze[*]
Carnegie Mellon University
alan@random.math.cmu.edu

Pu Gao[†]
University of Waterloo
pu.gao@uwaterloo.ca

Abstract

We study the Hamiltonicity of the following model of a random graph. Suppose that we partition $[n]$ into $V_1, V_2, \ldots, V_k$ and add edge $\{x, y\}$ to our graph with probability p if there exists i such that $x, y \in V_i$. Otherwise, we add the edge with probbability q. We denote this model by $\mathcal{G}(\mathbf{n}, p, q)$ and give tight results for Hamiltonicity, including a critical window analysis, under various conditions.

1 Introduction

The Hamiltonicity of various models of random graphs has been studied for many years. As far back as 1976, Komlós and Szemerédi announced their solution for the random graph $G_{n,m}$, although the published paper came out later in 1983 [5]. Since that time there have been many results on Hamiltonicity of random graphs, including but not restricted to, binomial random graphs, random regular graphs, binomial random graphs restricted to given minimum degrees, random k-out graphs, random percolation on given graphs, random graphs produced by (various types of) random graph processes, and also random hypergraphs. See a recent bibliography [3] by the second author which goes into great details.

In this paper we study Hamiltonicity of random graphs from the so-called *Stochastic Block Model*. This random graph model has been the subject of much research in the computer science community. It is a generative model for social networks consisting of distinct communities. The model generalises the Erdős-Rényi random graphs, where every pair of vertices is connected by an edge independently with the same probability. In the stochastic block model, the probability of connecting a pair of vertices depends on which communities they belong to. Research on the stochastic block model is mainly on inferring

the community membership given an instance sampled from the model. A recent paper by Abbe [1] surveys this aspect.

A formal definition of the stochastic block model is given as follows. Let P be a symmetric $k \times k$ matrix with nonnegative entries between 0 and 1, and $\mathbf{n} = (n_1, \ldots, n_k)$ be a vector of positive integers. Let $n = \sum_{i=1}^n n_i$. Let $\mathcal{G}(\mathbf{n}, P)$ be a random graph constructed as follows. The vertex set is $V = \cup_{i=1}^k V_i$ where $V_i = \{(i, j), j \in [n_i]\}$, and any two vertices (i_1, j_1) and (i_2, j_2) are adjacent with probability $P(i_1, i_2)$ mutually independently. In this paper we consider the special case where $k \geq 1$ is a fixed integer, and P has value $p = p(n)$ on the diagonal, and has value $q = q(n)$ off the diagonal. We denote $\mathcal{G}(\mathbf{n}, P)$ by $\mathcal{G}(\mathbf{n}, p, q)$ for this special P.

Unlike the Erdős-Rényi random graph, $\mathcal{G}(\mathbf{n}, p, q)$ is a non-homogeneous model where the distribution of the neighbourhood of vertex v depends on which V_i it belongs to. If $p = q$ then $\mathcal{G}(\mathbf{n}, p, q)$ reduces to $\mathcal{G}(n, p)$. If $p = 0$ then $\mathcal{G}(\mathbf{n}, p, q)$ reduces to a random k-partite graph. The closest previous results to this work are the cases of Hamiltonicity of Erdős-Rényi graphs by Komlós and Szemerédi [5], and of random bipartite graphs considered by the second author[2]. The present paper utilises and extends the proofs in these papers in a significant manner.

2 The main results

We call vertex sets V_i blocks, and an edge is called a *block edge* if its ends lie in the same block, and a *crossing edge* otherwise. Given a vertex $u \in V_i$, we say u has *partition index i*. We aim to determine when $\mathcal{G}(\mathbf{n}, p, q)$ is Hamiltonian.

We will assume the following set of conditions.

$$(\text{A1}) \quad \min_{1 \leq i \leq k} \{pn_i + (n - n_i)q - \log n_i\} = \log \log n + O(1);$$

$$(\text{A2}) \quad qn^2 = \omega(1);$$

$$(\text{A3}) \quad \max_{1 \leq i \leq k} n_i \leq n/2, \text{ if } p = O(1/n).$$

$$(\text{A4}) \quad \min_{1 \leq i \leq k} n_i = \Omega(n).$$

Note that if $\mathcal{G}(\mathbf{n}, p, q)$ is Hamiltonian then conditions (A2) and (A3) are necessary in general. If (A2) fails then with a non-vanishing probability there can be some V_i such that $E(V_i, V \setminus V_i) = \emptyset$. If (A3) fails then $\mathcal{G}(\mathbf{n}, p, q)$ cannot be Hamiltonian if $p = 0$. Condition (A4) can probably be relaxed, but it requires more delicate analysis. We will show that condition (A1) captures the critical window for Hamiltonicity of $\mathcal{G}(\mathbf{n}, p, q)$.

Let a_n and b_n be two sequences of real numbers. We say $a_n = O(b_n)$ if there exists an absolute constant $C > 0$ such that $|a_n| \leq C|b_n|$ for every $n \geq 1$. We say $a_n = o(b_n)$ if $b_n > 0$ for all $n \geq 1$ and $\lim_{n \to \infty} a_n/b_n = 0$. If $a_n > 0$ for all $n \geq 1$, and $b_n = O(a_n)$ (or $b_n = o(a_n)$) then we write $a_n = \Omega(b_n)$ (or $b_n = \omega(a_n)$ respectively). We will consider a sequence of random graphs indexed by their order, denoted by n, and all constants in this paper do not depend on n. All asymptotics refer to $n \to \infty$. Given a graph property Γ, we say $\mathcal{G}(n, p, q) \in \Gamma$ asymptotically almost surely (a.a.s.) if $\lim_{n \to \infty}(\mathcal{G}(n, p, q) \in \Gamma) = 1$. Let `HAM` denote the class of Hamiltonian graphs. Our main result is the following.

Theorem 1. *Assume p and q and $\mathbf{n}$ satisfy assumptions (A1)–(A4).*

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{G}(\mathbf{n}, p, q) \in \mathtt{HAM}) = \exp\left(-\sum_{1\le i\le k} e^{-c_i}\right),$$

where $c_i = pn_i + (n - n_i)q - \log n_i - \log\log n$.

As Hamiltonicity is an increasing property, the following corollary follows immediately.

Corollary 2. *Assume p and q and $\mathbf{n}$ satisfy assumptions (A2)–(A4).*

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{G}(\mathbf{n}, p, q) \in \mathtt{HAM}) = \begin{cases} 0 & \text{if } \min_{1\le i\le k}\{pn_i + (n-n_i)q - \log n_i\} < \log\log n - \omega(1) \\ 1 & \text{if } \min_{1\le i\le k}\{pn_i + (n-n_i)q - \log n_i\} > \log\log n + \omega(1). \end{cases}$$

3   Small degrees

Let $\mathtt{D2}$ denote the class of graphs with minimum degree at least 2. Note that $G \notin \mathtt{D2}$ implies that $G \notin \mathtt{HAM}$. Thus, the following lemma immediately yields an upper bound on the probability that $\mathcal{G}(\mathbf{n}, p, q)$ is Hamiltonian.

Lemma 3. *Assume (A1) and (A4). Then,*

(a)

$$\lim_{n\to\infty} \mathbb{P}(\mathcal{G}(\mathbf{n}, p, q) \in \mathtt{D2}) = \exp\left(-\sum_{1\le i\le k} e^{-c_i}\right),$$

where $c_i = pn_i + (n - n_i)q - \log n_i - \log\log n$.

(b) *For constants $0 < \alpha < 1$, and $\beta > \alpha + \alpha\ln(1/\alpha)$. A.a.s. $\mathcal{G}(\mathbf{n}, p, q)$ contains at most n^β vertices whose degree is at most $\alpha\log n$.*

Proof. For part (a), let $X_j(i) = \sum_{v\in V_i} 1_{\{d(v)=j\}}$ be the number of vertices in V_i with degree j. Let W_1 and W_2 be two independent random variables with $W_1 \sim \mathbf{Bin}(n_i - 1, p)$ and $W_2 \sim \mathbf{Bin}(n - n_i, q)$. Let $j = O(\log n)$. By (A1) and (A4) we have $p^2 n = o(1)$, $pj = o(1)$, $j^2 = o(n_i)$, and $j^2 = o(n - n_i)$. Then,

$$\mathbb{E}X_j(i) = n_i \mathbb{P}(W_1 + W_2 = j)$$

$$= n_i \sum_{s=0}^{j} \binom{n_i - 1}{s} p^s (1-p)^{n_i - 1 - s} \binom{n - n_i}{j - s} q^{j-s}(1-q)^{n - n_i - j + s}$$

$$\sim n_i \exp(-pn_i - q(n - n_i)) \sum_{s=0}^{j} \frac{n_i^s (n - n_i)^{j-s}}{s!(j-s)!} p^s q^{j-s}$$

$$= n_i e^{-\phi_i} \frac{\phi_i^j}{j!}, \tag{1}$$

where $\phi_i = pn_i + q(n - n_i)$. By (A1), $\phi_i - (\log n_i + \log\log n) > C$ for some constant real C and for all $i \in [k]$. It follows immediately that a.a.s. $X_0(i) = 0$ for every i.

Recall that $c_i = \phi_i - (\log n_i + \log \log n)$. Let $X_1 = \sum_{i \in [k]} X_1(i)$. Then, $\mathbb{E}X_1 \sim \sum_{i \in [k]} e^{-c_i}$. By (A1), $\mathbb{E}X_1 = \Theta(1)$. Using the standard method of moments (we omit the tedious calculations), it is easy to prove that X_1 is asymptotically Poisson. Hence,

$$\mathbb{P}(X_1 = 0) \sim \exp\left(-\sum_{i \in [k]} e^{-c_i}\right). \tag{2}$$

The lemma follows by (2) and the fact that a.a.s. $X_0(i) = 0$ for every $i \in [k]$.

For part (b), from (1) we have

$$\sum_{j \leq \alpha \log n} \sum_{i \in [k]} \mathbb{E}X_j(i) < (1 + o(1)) \sum_{j \leq \alpha \log n} \sum_{i \in [k]} n_i \exp(-\log n_i) \frac{(\log n_i)^j}{j!}$$

$$< k \sum_{j \leq \alpha \log n} \left(\frac{e \log n}{j}\right)^j = (1 + o(1)) k n^{\rho(\alpha)},$$

where $\rho(\alpha) = \alpha + \alpha \log(1/\alpha)$. Part (b) follows by the Markov inequality. ∎

4 Vertex expansion and connectivity

Let G be a graph and $S \subseteq V(G)$, define

$$\begin{aligned}
\mathcal{N}_G(S) &= \{j \in V(G) \setminus S : \exists i \in S, i \sim j\} \\
N_G(S) &= |\mathcal{N}_G(S)| \\
n_1(G) &= \sum_{i \in V(G)} 1_{\{d(i) \leq 1\}}.
\end{aligned}$$

I.e. $\mathcal{N}_G(S)$ is the set of vertices not in S which are adjacent to some vertex in S in graph G. We may drop G from the subscript if the underlying graph G is clear from the context.

Definition 4. *A vertex in G is called small, if its degree is less than $\log n / 10$. A vertex with degree at least $\log n / 10$ is called large.*

We say G has property EXPN, if: there exists a constant $\epsilon_0 > 0$ such that

for every $S \subseteq V(G)$ where $|S| \leq \epsilon_0 n$, $|N_G(S)| \geq 2|S| \cdot 1_{\{n_1(G)=0\}}$.

If F is a subset of edges in G, we use $G - F$ to denote the subgraph of G obtained by deleting edges in F. We say G has property SEXPN, if the following holds: For any $F \subseteq E(G)$, such that $|F \cap \mathcal{N}_G(v)| = 0$ if v is small, and is at most $\log n / 100$ if v is large, we have that $G - F$ is connected and $G - F \in$ EXPN.

Lemma 5. *Assume (A1), (A2) and (A4). Then a.a.s. $\mathcal{G}(\mathbf{n}, p, q) \in$ SEXPN.*

Its technical proof is postponed till Section 9.

5 Overview of the proof of Theorem 1

Since $G \notin \mathtt{D2}$ implies that $G \notin \mathtt{HAM}$, the upper bound of $\mathbb{P}(\mathcal{G}(\mathbf{n}, p, q) \in \mathtt{HAM})$ is implied by Lemma 3(a). Next we prove the lower bound. Let $\mathtt{LC}$ denote the set of graphs where a longest path contains the same number of vertices as in a longest cycle, and let $\mathtt{CNT}$ denote the class of connected graphs. Note that if $G \in \mathtt{LC} \cap \mathtt{CNT}$ then G must be Hamiltonian, since otherwise, by connectivity it is always possible to extend a longest cycle into a path which contains more vertices than the cycle we start with, contradicting with $G \in \mathtt{LC}$. It follows then that

$$\mathbb{P}(\mathcal{G}(\mathbf{n}, p, q) \in \mathtt{HAM}) \geq \mathbb{P}(\mathcal{G}(\mathbf{n}, p, q) \in \mathtt{LC} \cap \mathtt{CNT}) \geq \mathbb{P}(\mathcal{G}(\mathbf{n}, p, q) \in \mathtt{LC} \cap \mathtt{CNT} \cap \mathtt{D2}). \quad (3)$$

Our goal is to prove that $\mathbb{P}(\mathcal{G}(\mathbf{n}, p, q) \in \mathtt{LC} \cap \mathtt{CNT} \cap \mathtt{D2}) \sim \mathbb{P}(\mathcal{G}(\mathbf{n}, p, q) \in \mathtt{D2})$, which then yields the asymptotic probability desired by Theorem 1. The proof of the lower bound of $\mathbb{P}(\mathcal{G}(\mathbf{n}, p, q) \in \mathtt{HAM})$ will be split into three cases: (1) $p, q = \omega(1/n)$; (2) $p = O(1/n)$; and (3) $q = O(1/n)$. In all three cases, we will use a multi-round exposure technique of $\mathcal{G}(\mathbf{n}, p, q)$. Roughly speaking, we will expose a subgraph $G_b \subseteq G$ where $G \sim \mathcal{G}(\mathbf{n}, p, q)$ and G_b contains most edges of G. Case 1 is the simplest case, in which we will define graph property $\mathtt{TPCL}$ which consists of a set of properties that hold a.a.s. for $\mathcal{G}(\mathbf{n}, p, q)$. Then we will define $\mathtt{COL}$ to be a set of properties that edges in $G \setminus G_b$ must satisfy. Then, we will prove that

$$\mathbb{P}(\{G \in \overline{\mathtt{LC}} \cap \mathtt{CNT} \cap \mathtt{D2} \cap \mathtt{TPCL}\} \cap \mathtt{COL}) \ll \mathbb{P}(\mathtt{COL} \mid G \in \overline{\mathtt{LC}} \cap \mathtt{CNT} \cap \mathtt{D2} \cap \mathtt{TPCL}).$$

This implies that $\mathbb{P}(G \in \overline{\mathtt{LC}} \cap \mathtt{CNT} \cap \mathtt{D2} \cap \mathtt{TPCL}) = o(1)$, which will lead us to derive the asymptotic probability for $\mathbb{P}(G \in \mathtt{LC} \cap \mathtt{CNT} \cap \mathtt{D2})$. While obtaining a lower bound for $\mathbb{P}(\mathtt{COL} \mid G \in \overline{\mathtt{LC}} \cap \mathtt{CNT} \cap \mathtt{D2} \cap \mathtt{TPCL})$ is rather straightforward, an upper bound for $\mathbb{P}(\{G \in \overline{\mathtt{LC}} \cap \mathtt{CNT} \cap \mathtt{D2} \cap \mathtt{TPCL}\} \cap \mathtt{COL})$ is obtained by using *Pósa rotations* and bounding the probability that the longest path does not get extended by the edges exposed in the second stage. This is a standard technique for proving Hamiltonicity in random graphs.

In Case 2, $\mathcal{G}(\mathbf{n}, p, q)$ is similar to the random k-partite graph $\mathcal{G}(\mathbf{n}, 0, q)$. Two complications arise in this case. Firstly, we cannot totally ignore block edges in $\mathcal{G}(\mathbf{n}, p, q)$ as they contribute to degree 2 vertices in $\mathcal{G}(\mathbf{n}, p, q)$ with a non-varnishing probability. More specifically, $\mathbb{P}(\mathcal{G}(\mathbf{n}, p, q) \in \mathtt{HAM}) > \mathbb{P}(\mathcal{G}(\mathbf{n}, 0, q) \in \mathtt{HAM})$ when $p = \Theta(1/n)$ and thus, the proof cannot be obtained by simply extending the proof for random bipartite graphs to random k-partite graphs. Secondly, due to the asymmetry between p and q, the edges exposed in later stages will not be uniformly distributed and we need to take care of the multipartition of the vertices. This is similar to the case of the random bipartite graphs.

In Case 3, $\mathcal{G}(\mathbf{n}, p, q)$ looks like a collection of $G_i \sim \mathcal{G}(n, p)$ plus a set of random edges between every pair of G_i, G_j, $1 \leq i < j \leq k$. A tempting approach would be to find a Hamilton cycle in each G_i and then somehow connect these cycles by using a few crossing edges to form a Hamilton cycle in $\mathcal{G}(\mathbf{n}, p, q)$. This approach would succeed if $q = o(1/n)$. However, when $q = \Theta(1/n)$, similar to Case 2, the crossing edges are contributing, with a non-negligible probability, to the degree 2 vertices in $\mathcal{G}(\mathbf{n}, p, q)$. Thus, we cannot purely focus on structures in G_i. Instead, inside each G_i, we will take particular care of the vertices with degree less than 2, and we will look for a small number of vertex disjoint paths covering all vertices in G_i. These paths have specified end vertices. Then we will stitch these paths together with some crossing edges to form a Hamilton cycle in $\mathcal{G}(\mathbf{n}, p, q)$.

As part of the overview of the proof, we define `TPCL` and `COL`. They will be used in the proof of the first case, and in the second case as well with some minor modifications. In Case 3, their definitions will be significantly modified.

5.1   TPCL

We say $G \in$ `TPCL` if G satisfies the following set of properties.

(T1) $G \in$ `SEXPN`.

(T2) There are at most $n^{0.4}$ small vertices.

(T3) If $p = o(\log n / n)$ then every vertex is incident with at most $\log n / 200$ block edges. If $q = o(\log n / n)$ then every vertex is incident with at most $\log n / 200$ crossing edges.

(T4) The maximum degree is at most $C \log n$ for some sufficiently large constant $C > 0$.

Lemma 6. *A.a.s. $G \in$ `TPCL`.*

Proof. (T1) follows by Lemma 5. (T2) follows by Lemma 3(b). (T3) and (T4) follow by a standard first moment argument similar to the proof of Lemma 3. We omit the details. ∎

5.2   COL

Let $L(G)$ denote the length of a longest path in G. Assume $G' \subseteq G$. Let $F = E(G \setminus G')$. We say $(G, G') \in$ `COL` if

(a) $|F \cap \mathcal{N}_G(v)|$ is 0 if v is small in G, and is at most $\log n / 100$ if v is large in G;

(b) $L(G) = L(G')$ if $L(G') < n - 1$, and $G \notin$ `HAM` if $L(G') = n - 1$.

In the proof of Theorem 1, we will use `COL` to denote the event that $(G, G_b) \in$ `COL`, although G and G_b are defined differently in the three cases. We will recall the definition of `COL` when we proceed to the proof in each case.

6   Proof of Theorem 1: when $p, q = \omega(1/n)$

We first define G_b.

6.1 G_b

Let $\bar{p} = a/n \log n$, where $a = 1$. We will choose similar parameters for multi-round exposures in Cases 2 and 3 with different values of a. We keep a in the definition of $\bar{p}$ for the ease of comparison. Define

$$p_1 = 1 - \frac{1-p}{1-\bar{p}}; \quad q_1 = 1 - \frac{1-q}{1-\bar{p}}.$$

In case 1, both p_1 and q_1 are real numbers between 0 and 1. We will run a two stage exposure of the edges in $\mathcal{G}(\mathbf{n}, p, q)$. First, generate $G_b \sim \mathcal{G}(\mathbf{n}, p_1, q_1)$, then independently for

6

every non-edge x in G_b, add x to the graph with probability $\bar{p}$. Call the resulting graph G. It is straightforward to verify that $G \sim \mathcal{G}(\mathbf{n}, p, q)$. For convenience, colour the edges in G_b blue and the edges in $E(G \setminus G_b)$ red.

By our definition of p_1 and q_1 it is easy to see that (A1) and (A2) are satisfied with p and q replaced by p_1 and q_1. Recall that COL denotes the event that $(G, G_b) \in$ COL.

The next two lemmas bound $\mathbb{P}(\text{COL} \mid G \in \overline{\text{LC}} \cap \text{D2} \cap \text{TPCL})$ and $\mathbb{P}(\{G \in \overline{\text{LC}} \cap \text{D2} \cap \text{TPCL}\} \cap \text{COL})$.

Lemma 7. *There exists a function $f = f(n) \to \infty$ as $n \to \infty$ such that*

$$\mathbb{P}(\text{COL} \mid G \in \overline{\text{LC}} \cap \text{D2} \cap \text{TPCL}) \geq \exp(-an/f \log n).$$

Lemma 8.
$$\mathbb{P}(\{G \in \overline{\text{LC}} \cap \text{D2} \cap \text{TPCL}\} \cap \text{COL}) \leq \exp(-\Omega(an/\log n)).$$

Now we are ready to prove Theorem 1 in case 1.
Proof of Theorem 1 (case 1). By Lemmas 7 and 8,

$$\mathbb{P}(G \in \overline{\text{LC}} \cap \text{D2} \cap \text{TPCL}) \leq \frac{\exp(-\Omega(an/\log n))}{\exp(-an/f \log n)}.$$

As $f \to \infty$ as $n \to \infty$, the above probability is $o(1)$. By Lemmas 5 and 6, $\mathbb{P}(G \in \text{CNT} \cap \text{TPCL}) = 1 - o(1)$. It follows immediately that

$$\mathbb{P}(G \in \text{LC} \cap \text{D2} \cap \text{CNT}) = \mathbb{P}(G \in \text{D2}) - \mathbb{P}(G \in \overline{\text{LC}} \cap \text{CNT} \cap \text{D2} \cap \text{TPCL}) + o(1) = \mathbb{P}(\text{D2}) + o(1).$$

By (3) and the fact that $G \in \text{HAM}$ implies $G \in \text{D2}$, we have $\mathbb{P}(G \in \text{HAM}) = \mathbb{P}(G \in \text{D2}) + o(1)$. Together with Lemma 3 this yields the asymptotic probability of $\mathbb{P}(G \in \text{HAM})$ as in Theorem 1. ∎

It remains to prove Lemmas 7 and 8.

6.2   Proof of Lemma 7

Equivalently we can define G_b as follows. Take $G \sim \mathcal{G}(\mathbf{n}, p, q)$. Define

$$p^* = \frac{\bar{p}(1-p)}{(1-\bar{p})p}, \quad q^* = \frac{\bar{p}(1-q)}{(1-\bar{p})q}.$$

Do the following independently for every edge $x \in G$: if x is a block edge, delete x with probability p^*; if x is a crossing edge, delete x with probability q^*. As $p(1-p^*) = p_1$ and $q(1-q^*) = q_1$ with our definition of (p^*, q^*), we immediately have

Claim 9. *The resulting graph is distributed as $\mathcal{G}(\mathbf{n}, p_1, q_1)$.*

We will prove that conditioning on $G = H$ for any $H \in \overline{\text{LC}} \cap \text{D2} \cap \text{TPCL}$, $\mathbb{P}(\text{COL} \mid G = H) \geq \exp(-an/f \log n)$, and Lemma 7 follows.

Consider the set of edges deleted in generating G_b from H. Colour these edges red.

Let P be a longest path in H. Note that COL is implied if

(B1) no large vertex in H is incident with more than $\log n/100$ red edges.

7

(B2) no small vertex in H is incident with a red edge;

(B3) no edge in P is red.

Let $\mathcal{X}$ be the union of the set of edges in P and the set of edges incident with small vertices. Let $X = |\mathcal{X}|$. By (T2), $|\mathcal{X}| \leq (n-1) + n^{0.4} \log n / 10 < 2n$. As $\max\{p^*, q^*\} = o(a/\log n)$ in case 1, the probability that none of the edges in $\mathcal{X}$ is deleted is at least $(1 - o(a/\log n))^{2n} \geq \exp(-an/f \log n)$ for some $f \to \infty$. Note that (B2) and (B3) are implied if no edges in $\mathcal{X}$ are red. Hence, $\mathbb{P}(B1 \cap B_2) \geq \exp(-an/f \log n)$.

Let $\overline{\mathcal{X}}$ be the set of edges in H that are not in $\mathcal{X}$. Condition on no edges in $\mathcal{X}$ were deleted (i.e. became red). We will prove that a.a.s. every vertex is incident with at most $\log n / 200$ red block edges in $\overline{\mathcal{X}}$, and at most $\log n / 200$ red crossing edges in $\overline{\mathcal{X}}$. By (T3) we may assume that $p, q = \Omega(\log n / n)$. By (T4), each vertex has degree $O(\log n)$. By the definition of p^* and q^*, every edge is deleted (i.e. becomes red) with probability $O(\bar{p} \cdot \max\{1/p, 1/q\}) = O(1/\log^2 n)$. By the tail bounds for the binomial distribution and the union bound, a.a.s. every vertex is incident with at most $o(\log n)$ red edges and thus $\mathbb{P}(B1 \mid B2 \cap B3) = 1 - o(1)$. Hence, $\mathbb{P}(\texttt{COL} \mid G = H) \geq \mathbb{P}(B1 \cap B2 \cap B3) \geq \exp(-an/f \log n)$ for some $f \to \infty$. Lemma 7 follows. ∎

6.3 Proof of Lemma 8

Recall the definition of $\texttt{EXPN}$ and $\texttt{SEXPN}$. Assuming (T1), there exists an absolute constant $\epsilon_0 > 0$ such that

$$\text{for every } S \subseteq V(G) \text{ where } |S| \leq \epsilon_0 n, \text{ we have } |N_G(S)| \geq 2|S| \cdot 1_{\{n_1(G)=0\}}. \qquad (4)$$

By the definition of $\texttt{COL}$ we immediately have the following claim.

Claim 10. $\{G \in \overline{\texttt{LC}} \cap \texttt{D2} \cap \texttt{TPCL}\} \cap \texttt{COL}$ *implies* $B \cap \{G_b \in \texttt{CNT} \cap \texttt{EXPN} \cap \texttt{D2}\}$, *where* $B = \{L(G) = L(G_b) < n-1\} \cup (\{L(G_b) = n-1\} \cap \{G \notin \texttt{HAM}\})$.

Proof. If $G \in \texttt{TPCL} \cap \texttt{D2}$ and $(G, G_b) \in \texttt{COL}$, then by (T1) we have $G_b \in \texttt{CNT} \cap \texttt{EXPN} \cap \texttt{D2}$. Moreover, $(G, G_b) \in \texttt{COL}$ implies B. This proves our claim. ∎

Hence, it is sufficient to prove

$$\mathbb{P}(B \mid G_b \in \texttt{CNT} \cap \texttt{EXPN} \cap \texttt{D2}) \leq \exp(-\Omega(an/\log n)),$$

as

$$\begin{aligned}
\mathbb{P}(\{G \in \overline{\texttt{LC}} \cap \texttt{D2} \cap \texttt{TPCL}\} \cap \texttt{COL}) &\leq \mathbb{P}(B \cap \{G_b \in \texttt{CNT} \cap \texttt{EXPN} \cap \texttt{D2}\}) \\
&\leq \mathbb{P}(B \mid G_b \in \texttt{CNT} \cap \texttt{EXPN} \cap \texttt{D2}),
\end{aligned}$$

by Claim 10. Note that G is obtained by adding every non-edge in G_b independently with probability $\bar{p}$. We will prove that conditioning on any graph $G_b \in \texttt{CNT} \cap \texttt{EXPN}$, adding approximately $\bar{p}\binom{n}{2} \sim an/2 \log n$ edges will either increase $L(G_b)$, or complete a Hamilton path in G_b to a Hamilton cycle, with sufficiently high probability. We will use the classical technique of Pósa rotations to bound the probability of B.

8

Pósa rotations. Let $P = v_0, v_1, \ldots, v_\ell$ be a longest path in G_b. Then v_0 is not adjacent to v_ℓ, and all the neighbours of v_ℓ in G_b must be in P, since otherwise we can extend P to a longer path. Assume $v_i v_\ell$ is an edge in G_b where $i < \ell - 1$, then another longest path $P' = v_0, \ldots, v_i v_\ell, v_{\ell-1}, \ldots, v_{i+1}$ can be obtained by using the edge $v_i v_\ell$ instead of $v_i v_{i+1}$. This operation from P to P' is called a Pósa rotation. Consider the set $\mathscr{P}$ of longest paths obtained by repeatedly rotating P. All of these paths start from v_0 and end at a vertex that is in P. Let $\mathrm{End}(v_0)$ denote the set of ends other than v_0 in the paths in $\mathscr{P}$. A key observation is the following. The reader may refer to [4] for a proof.

Lemma 11. $|N_{G_b}(\mathrm{End}(v_0))| < 2|\mathrm{End}(v_0)|$.

As $G_b \in \mathtt{EXPN} \cap \mathtt{D2}$ we immediately have that $|\mathrm{End}(v_0)| \geq \epsilon_0 n$ where ϵ_0 is specified in (4).

Now for every $v \in \mathrm{End}(v_0)$, there is a longest path P_v which is obtained from P by repeatedly applying Pósa rotations. Let $\mathrm{End}(v)$ denote the set of ends other than v in the longest paths obtained by rotating P_v. Again, we have $|N_{G_b}(\mathrm{End}(v))| < 2|\mathrm{End}(v)|$, which implies that $|\mathrm{End}(v)| \geq \epsilon_0 n$. Consider the set $\mathscr{E}$ of pairs of vertices (x, y) where $x \in \mathrm{End}(v_0)$, and $y \in \mathrm{End}(x)$. We have that $|\mathscr{E}| \geq \epsilon_0^2 n^2$. Moreover, adding any pair in $\mathscr{E}$ as an edge to G will either form a Hamilton cycle in G, if $\ell = n - 1$, or form a cycle with length $\ell + 1$, and then using the fact that G is connected, we can extend the cycle to a path of length $\ell + 1$, if $\ell < n - 1$. In either case, event B fails. For that reason, we call $\mathscr{E}$ a set of *boosters*. We have shown that B fails if $\mathscr{E} \cap E(G \setminus G_b) \neq \emptyset$.

By the construction of G, every edge in $\mathscr{E}$ is added to G in the second stage of edge exposure, independently with probability $\bar{p}$. The probability that none of these edges are added is $(1 - \bar{p})^{|\mathscr{E}|/2} \leq \exp(-\epsilon_0^2 an/2\log n)$. This completes the proof for Lemma 8. ∎

7 Proof of Theorem 1: when $p = O(1/n)$

In this case, we will define a 3-round edge exposure of $\mathcal{G}(\mathbf{n}, p, q)$.

7.1 G_b and G_y

Let $\bar{q} = a/n\log n$ where $a \to \infty$ and $a = o(\log n)$. Define

$$q_1 = 1 - \frac{1 - q}{(1 - \bar{q})^2}.$$

First, generate $\hat{G} \sim \mathcal{G}(\mathbf{n}, 0, q_1)$. Then, expose block edges that are incident to vertices with degree at most one in $\hat{G}$. Let G_b be the resulting graph. All edges in G_b are coloured blue. By bounding the expected number of vertices having degree at most 1 in $\hat{G}$ by $o(\log n)$, we immediately have the following claim.

Claim 12. *A.a.s. fewer than* $\log n$ *vertices in* G_b *are incident to a block edge.*

Then, for every crossing non-edge x in G_b, add x to the graph G_b with probability $\bar{q}$. All edges added in the second stage are coloured yellow. Call the resulting graph G_y. Finally, for each crossing non-edge in G_y, add it to the graph with probability $\bar{q}$ and colour it red.

The crossing edges of the resulting graph G have the same distribution as those in $\mathcal{G}(\mathbf{n}, p, q)$. Obviously we can couple G with $G' \sim \mathcal{G}(\mathbf{n}, p, q)$ such that $G \subseteq G'$. It is thus sufficient to prove that the lower bound for $\mathbb{P}(G \in \mathtt{HAM})$ matches the lower bound in Theorem 1.

Let now $\mathtt{TPCL}$ include properties (T1)–(T4) as well as the following property.

$(T5)$: Fewer than $\log n$ vertices of G are incident with fewer than 2 crossing edges.

Let $\mathtt{COL}$ denote $(G, G_b) \in \mathtt{COL}$ as defined in Section 5.2. Note that (T3), (T5) and $\mathtt{COL}$ implies the following.

$\mathtt{COL1}$: Fewer than $\log n$ vertices in G_b are incident to a block edge.

The next two lemmas bound $\mathbb{P}(\mathtt{COL} \mid G \in \overline{\mathtt{LC}} \cap \mathtt{D2} \cap \mathtt{TPCL})$ and $\mathbb{P}(\{G \in \overline{\mathtt{LC}} \cap \mathtt{D2} \cap \mathtt{TPCL}\} \cap \mathtt{COL})$ in case 2.

Lemma 13. *There exists a constant $K > 0$ such that*

$$\mathbb{P}(\mathtt{COL} \mid G \in \overline{\mathtt{LC}} \cap \mathtt{D2} \cap \mathtt{TPCL}) \geq \exp(-Kan/\log^2 n).$$

Lemma 14.
$$\mathbb{P}(\{G \in \overline{\mathtt{LC}} \cap \mathtt{D2} \cap \mathtt{TPCL}\} \cap \mathtt{COL}) \leq \exp(-\Omega(a^2 n/\log^2 n)).$$

Since $\mathbb{P}(\{G \in \overline{\mathtt{LC}} \cap \mathtt{D2} \cap \mathtt{TPCL}\} \cap \mathtt{COL}) \ll \mathbb{P}(\mathtt{COL} \mid G \in \overline{\mathtt{LC}} \cap \mathtt{D2} \cap \mathtt{TPCL})$ as $a \to \infty$, the proof of Theorem 1 for case 2 follows exactly as in case 1. It only remains to prove Lemmas 13 and 14.

7.2 Proof of Lemma 13

The proof is basically the same as that for Lemma 7. The probability bound is only different because of the different range for p and q. We only point out the differences in the proof. We give an equivalent definition of (G_b, G) as follows. Take $G' \sim \mathcal{G}(\mathbf{n}, p, q)$. Let q^* be such that $q(1 - q^*) = q_1$. Independently delete each crossing edge x in G' with probability q^*. Call the resulting graph G'_b. Next delete all block edges in G'_b except for the edges incident with a vertex which is only incident with at most one crossing edge in G'_b. The resulting graph has the same distribution as G_b. Let G be the graph obtained by taking all crossing edges in G' and all block edges in G_b. Note that G and G_b only differ on the crossing edges.

Now $\max\{p^*, q^*\}$ in the proof of Lemma 7 is replaced by q^*. By (A1), (A5) and the assumption $p = O(1/n)$ we have $q = \Theta(\log n/n)$ and thus $q^* = \Theta(a/\log^2 n)$. It follows that the probability that none of the edges in χ is deleted is at least $(1 - \Theta(a/\log^2 n))^{2n} \geq \exp(-\Theta(an/\log^2 n))$. The rest of the proof is the same as in Lemma 7. ∎

7.3 Proof of Lemma 14

Recall the definition of $\mathtt{TPCL}$, $\mathtt{EXPN}$ and $\mathtt{SEXPN}$. Assuming $G \in \mathtt{SEXPN}$, there exists an absolute constant $\epsilon_0 > 0$ such that (4) holds. Claim 10 continues to hold. Thus, it is sufficient to show
$$\mathbb{P}(B \mid \{G_b \in \mathtt{CNT} \cap \mathtt{EXPN} \cap \mathtt{D2}\}) \leq \exp(-\Omega(a^2 n/\log^2 n)).$$

As in the proof of Lemma 8, we will show that the probability that the additional yellow and red edges do not increase $L(G_b)$ is small. However, the graph induced by the yellow and red edges is k-partite. and so there is a subtle issue here with k-partite graphs. The yellow and red edges exposed in the second and third stage must respect to the vertex partition. If the set of boosters, the potential non-edges whose addition will allow an extension of the longest path, are all unichromatic, i.e. the ends of each booster have the same partition index, then the edges exposed in the second stage will not help to extend the paths. The purpose of using 3 rounds of edge exposure is to cope with this vertex partition issue. To explain how it works, we need a few definitions.

Let $P = (v_0, \ldots, v_\ell)$ be a longest path in the graph G_b. Let $\mathscr{P}$ be a list of longest paths obtained as follows. $\mathscr{P} = \{P\}$ initially and $\mathrm{End}(v_0) = \{v_\ell\}$. Take a path $P' \in \mathscr{P}$ and do a Pósa rotation, if it creates a path whose end other than v_0 is not yet in $\mathrm{End}(v_0)$. Then add this new path to $\mathscr{P}$ and add its end vertex other than v_0 to $\mathrm{End}(v_0)$. Since $|\mathrm{End}(v_0)| < n$, this process will terminate. As before, given that $G_b \in \mathtt{EXPN}$ we have $|\mathrm{End}(v_0)| \geq \epsilon_0 n$.

For each v, let $\sigma(v)$ denotes its partition index. For each $x \in \mathrm{End}(v_0)$, let P_x be the path that was added to $\mathscr{P}$ when x was added to $\mathrm{End}(v_0)$. Let $\phi(x)$ be the vertex adjacent to x on P_x. Let $\mathscr{P}(x)$ denote the set of longest paths obtained by rotating P_x with x being the fixed end, and let $\mathrm{End}(x)$ denote the set of the ends of the paths other than x in $\mathscr{P}(x)$. Then, all paths in $\mathscr{P}(x)$ must start with $x\phi(x)$. Moreover, $\sigma(x) \neq \sigma(\phi(x))$, unless x is incident with a block edge.

We call a vertex x in $\mathrm{End}(v_0)$ *good*, if at least $\frac{1}{2}|\mathrm{End}(x)|$ vertices in $\mathrm{End}(x)$ have distinct partition index from x.

Consider G_y. Let $\mathtt{Out}$ denote the set of vertices x for which there exists a longest path P in G_y such that $x \notin P$ and one end of P lies in distinct vertex part from x. We will show the following.

Claim 15. *Either there are $\frac{1}{2}|\mathrm{End}(v_0)|$ good vertices in $\mathrm{End}(v_0)$, or with probability $1 - \exp(-\Omega(an/\log n))$, either $L(G_y) > L(G_b)$ or $|\mathtt{Out}| = \Omega(an/\log n)$ in G_y.*

Note $\exp(-\Omega(an/\log n)) \leq \exp(-\Omega(a^2 n/\log^2 n))$ as $a \to \infty$ is chosen to be $o(\log n)$.

If there are $\frac{1}{2}|\mathrm{End}(v_0)|$ good vertices in $\mathrm{End}(v_0)$, then with the same proof as in case 1, we can bound $\mathbb{P}(B \mid G_b \in \mathtt{CNT} \cap \mathtt{EXPN} \cap \mathtt{D2})$ by $\exp(-\Omega(an/\log n)) \leq \exp(-\Omega(a^2 n/\log^2 n))$.

If instead $|\mathtt{Out}| = \Omega(an/\log n)$ in G_y, then we will prove the following claim which, together with Claim 15 and the argument above, completes the proof for Lemma 14.

Claim 16. *If $|\mathtt{Out}| = \Omega(an/\log n)$ then B holds with probability $\exp(-\Omega(a^2 n/\log^2 n))$.*

Proof. Let $x \in \mathtt{Out}$ and let P be the corresponding path, and y be the end on P where $\sigma(x) \neq \sigma(y)$. Consider Pósa rotations on P with y as the fixed end, and let $\mathrm{End}(y)$ denote the ends obtained. Either at least half of $\mathrm{End}(y)$ have partition indices distinct from $\sigma(x)$, or at least half with partition indices distinct from $\sigma(y)$. This implies that $|\mathscr{E}_x| = \Omega(n)$ where

$$\mathscr{E}_x = \{\{x, y\} : z \in \mathrm{End}(y), \sigma(x) \neq \sigma(z)\} \cup \{\{y, z\} : z \in \mathrm{End}(y), \sigma(y) \neq \sigma(z)\}.$$

Let $\mathscr{E} = \cup_{x \in \mathtt{Out}} \mathscr{E}_x$. Then $|\mathscr{E}| = \Omega(an^2/\log n)$. Moreover, B fails if $\mathscr{E} \cap E(G \setminus G_y) \neq \emptyset$. Since every booster in $\mathscr{E}$ appears in the second stage of edge exposure with probability $\bar{q}$. The probability that none of them appears is at most $(1 - \bar{q})^{|\mathscr{E}|} = \exp(-\Omega(a^2 n/\log^2 n))$. ∎

Proof of Claim 15. Let $\text{Bad}(v_0)$ denote the set of vertices in $\text{End}(v_0)$ that are not good and assume $|\text{Bad}(v_0)| > \frac{1}{2}|\text{End}(v_0)|$. Let $x \in \text{Bad}(v_0)$. Let $\texttt{Mono(x)}$ denote the set of vertices in $\text{End}(x)$ which have the same partition index as x. By the definition of $\text{Bad}(v_0)$ and because $x \in \text{Bad}(v_0)$, $|\texttt{Mono(x)}| \geq \frac{1}{2}|\text{End(x)}|$. Consequently, the vertices in $\texttt{Mono(x)}$ must have distinct partition indices from $\phi(x)$, unless x is incident with a block edge in G_b. Let $\texttt{A}$ denote the set of vertices that are not incident with any block edges in G_b. Consider all pairs of vertices $\mathscr{E}' = \{(\phi(x), y) : x \in \text{Bad}(v_0) \cap \texttt{A}, y \in \texttt{Mono(x)}\}$. As $\texttt{TPCL} \cap \texttt{COL}$ implies $\texttt{COL1}$, it follows that $|\mathscr{E}'| = \Omega(n^2)$. If any yellow edge exposed in the second round of edge exposure is in $\mathscr{E}'$, then we find a cycle C by deleting x from the path P', and then adding the edge $y\phi(x)$. G_b is connected and by (A3) there must exist a vertex z such that (a) z has distinct partition index from x; (b) z is not on P'; (c) there is a path from z to P'. Using the cycle C and the path from z to P' we obtain a path of length greater than that of P', which implies $L(G_y) > L(G_b)$, or we obtain a path of the same length as P' and one of its ends, namely z, has a distinct partition index from x. This implies that $x \in \texttt{Out}$. Assume it is always the latter case for each yellow edge in $\mathscr{E}'$. Since $|\mathscr{E}'| = \Theta(n^2)$, it follows that $\mathbb{E}|\texttt{Out}| = \bar{q} \cdot \Theta(n^2) = \Omega(an/\log n)$. The probability bound in the claim follows by the Chernoff bound. ∎

8 Proof of Theorem 1: when $q = O(1/n)$.

In this case we have $\min_{1 \leq i \leq k}\{pn_i - \log n_i\} = \log\log n + O(1)$. The subgraph induced by the vertices in block i is distributed as $\mathcal{G}(n_i, p)$.

Let $\bar{p} = a/n\log n$ with $a = 1$. Define

$$p_1 = 1 - \frac{1-p}{(1-\bar{p})}.$$

We give a quick overview of the proof in this case. First we generate $G_b \sim \mathcal{G}(\mathbf{n}, p_1, q)$. We call a vertex "problematic" if it is incident with fewer than 2 block edges in G_b. Let $\mathcal{P}$ denote the set of problematic vertices. All edges in G_b are coloured blue. In the second round of edge exposure, we expose block edges that are not present in G_b with probability $\bar{p}$. The resulting graph is G. All edges in $G \setminus G_b$ are coloured red. It is easy to see that $G \sim \mathcal{G}(\mathbf{n}, p, q)$. We obtain our Hamilton cycle by (i) finding a collection $\mathcal{A}_1$ of paths of length 1 or 2 that cover the problematic vertices, (ii) finding a collection $\mathcal{A}_2$ of crossing edges so that the number of crossing edges in $\mathcal{A}_1 \cup \mathcal{A}_2$ between any pairs V_i and V_j is even and positive, and (iii) finding a collection of vertex disjoint paths that connect the ends of the paths and edges in $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2$ into a Hamilton cycle. These longer paths found in (iii) only use block edges.

The property $\texttt{COL}$ will be defined differently from previous cases. We will split $\texttt{COL}$ into three parts. Let $\texttt{COL1}$ denote the property that

$$|F \cap \mathcal{N}_G(v)| \leq d_G(v) - 2 \text{ for all } v \in G \text{ such that } d_G(v) \geq 2.$$

where $F = E(G) \setminus E(G_b)$. With a simple first moment argument we can prove the following.

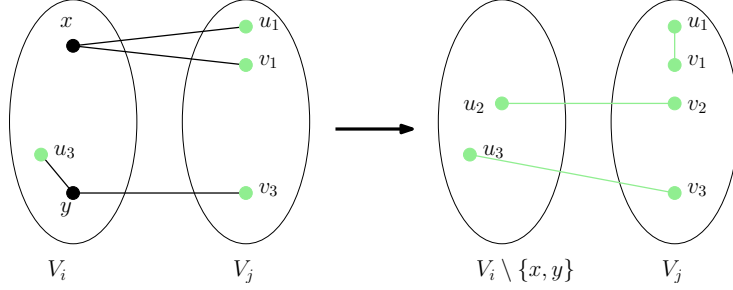**Claim 17.** *A.a.s.* $(G, G_b) \in \texttt{COL1}$.

Figure 1: Construct M

Let $\mathtt{COL} = \mathtt{COL1} \cap \mathtt{COL2} \cap \mathtt{COL3}$ where $\mathtt{COL2}$ and $\mathtt{COL3}$ will be defined later. Note that $\{G \in \mathtt{D2}\} \cap \mathtt{COL1}$ implies $G_b \in \mathtt{D2}$. If $\{G \in \mathtt{D2}\} \cap \mathtt{COL1}$ holds then every problematic vertex is incident with at least 2 edges. For every $u \in \mathcal{P}$, randomly choose 2 edges incident with u, colour them green. For each green edge, colour the end other than the problematic vertex green. See the left side of Figure 1 for an example.

Given a path $u_1 u_2 \ldots u_\ell$, we say that we *supplant* the path by an edge $e = u_1 u_\ell$ if we delete all the internal vertices on the path and their incident edges, and add edge e. Assume $\{G \in \mathtt{D2}\} \cap \mathtt{COL1}$. Supplant every green 2-path in G by a new green edge. Call the resulting graph H. Note that H is not defined if $\{G \in \mathtt{D2}\} \cap \mathtt{COL1}$ fails. Note also that $H[V_i \setminus \mathcal{P}] = G[V_i \setminus \mathcal{P}]$ for every $1 \le i \le k$. Let E_0 denote the set of green edges and let U_0 denote the set of green vertices obtained so far in H.

Next, we will choose a set of blue crossing edges and recolour them green, and colour the ends of these edges green. For every $1 \le i < j \le k$, if there are an odd number of green edges between V_i and V_j in H, then randomly choose a blue crossing edge x between V_i and V_j in G_b and recolour it green. Colour the end vertices of x green. If there is no green edge between V_i and V_j in H, then randomly choose two blue crossing edges x, y between V_i and V_j in G_b and recolour them green. Colour the end vertices of x and y green.

See Figure 1 for an illustration of the construction of H and E_g. Let E denote the set of crossing edges recoloured from blue to green. Let U denote the set of end vertices of edges in E. Write $E = \perp$ if the above construction cannot be completed. This happens only if $|E_{G_b}(V_i, V_j)| < 2$ for some $i \ne j$. However (A2) and (A4) ensure that $\mathbb{P}(|E_{G_b}(V_i, V_j)| < 2) = o(1)$. The following a.a.s. properties are straightforward and we omit their proofs.

Claim 18. *A.a.s. the following statements hold.*

(a) $E \ne \perp$.

(b) $E \cap E_0 = \emptyset$.

(c) E *induces a matching.*

(d) $U \cap U_0 = \emptyset$.

Let $E_g = E_0 \cup E$ denote all the green edges obtained so far in H.

Claim 19. *If H has a Hamiltonian cycle containing all edges in E_g then G contains a Hamilton cycle.*

Proof. Let C be such a Hamilton cycle in H. Replacing each edge $x \in E_0 \cap C$ by the green 2-path whose supplantation yielded x gives a Hamilton cycle in G. ∎

We will prove that H has a Hamiltonian cycle containing all edges in E_g by using the following lemma. Given a set of edges B, let $V(B)$ denote the set of vertices spanned by B. Let $B_{i,j}$, $i \leq j$, denote the set of pairs of vertices of B with one end in i and the other end in j. Given a graph G and a set E of edges on $V(G)$, let $G + E$ denote the graph on $V(G)$ obtained by taking the union of the edges from G and E.

Lemma 20. *Let B be a set of pairs of vertices such that*

- *the pairs in B are pairwise disjoint;*

- $|B_{i,j}| > 0$ *for every* $1 \leq i < j \leq k$ *and* $\sum_{j \neq i} |B_{i,j}|$ *is even for every* $1 \leq i \leq k$;

- $V(B) = O(\log n)$;

- $E_g \subseteq B$;

- *no two vertices in $V(B)$ share a common neighbour in H;*

- *no vertex in $V(B)$ is adjacent to a vertex with degree at most 2 in H.*

Then a.a.s. if $\{G \in \mathtt{D2}\} \cap \mathtt{COL1}$ holds then $H + B$ has a Hamilton cycle that contains all edges in B.

Using Claim 18 it is easy to show that $B = E_g$ satisfies all assumptions of Lemma 20. Let $\mathtt{HHAM}$ denote the event that H has a Hamilton cycle containing all edges in E_g. If $\{G \in \mathtt{D2}\} \cap \mathtt{COL1}$ holds then taking $B = E_g$ in Lemma 20 immediately implies $\mathtt{HHAM}$, which gives

$$
\begin{aligned}
\mathbb{P}(G \in \mathtt{HAM}) &\geq \mathbb{P}(\{G \in \mathtt{D2}\} \cap \mathtt{COL1} \cap \mathtt{HHAM}) \\
&\geq \mathbb{P}(G \in \mathtt{D2}) - \mathbb{P}(\{G \in \mathtt{D2}\} \cap \mathtt{COL1} \cap \overline{\mathtt{HHAM}}) - \mathbb{P}(\overline{\mathtt{COL1}}) \\
&= \mathbb{P}(G \in \mathtt{D2}) - o(1),
\end{aligned}
$$

by Claim 17 and Lemma 20, which completes the proof of Theorem 1 in the case $q = O(1/n)$. ∎

It only remains to prove Lemma 20. We will prove it by induction on k. The following key lemma will be used to complete the inductive argument.

Lemma 21. *Fix $1 \leq i \leq k$. Let A_i be a set of pairs of vertices of H_i such that*

- *the pairs in A_i are pairwise disjoint;*

- $V(A_i) \leq \log n$;

- *no two vertices in $V(A_i)$ share a common neighbour in H_i;*

- *no vertex in $V(A_i)$ is adjacent to a vertex with degree at most 2 in H_i.*

Then a.a.s. if $\{G \in \mathtt{D2}\} \cap \mathtt{COL1}$ holds then $H_i + A_i$ has a Hamilton cycle containing all of the edges in A_i.

We will prove Lemma 20 in Section 8.1 and prove Lemma 21 in Section 8.2.
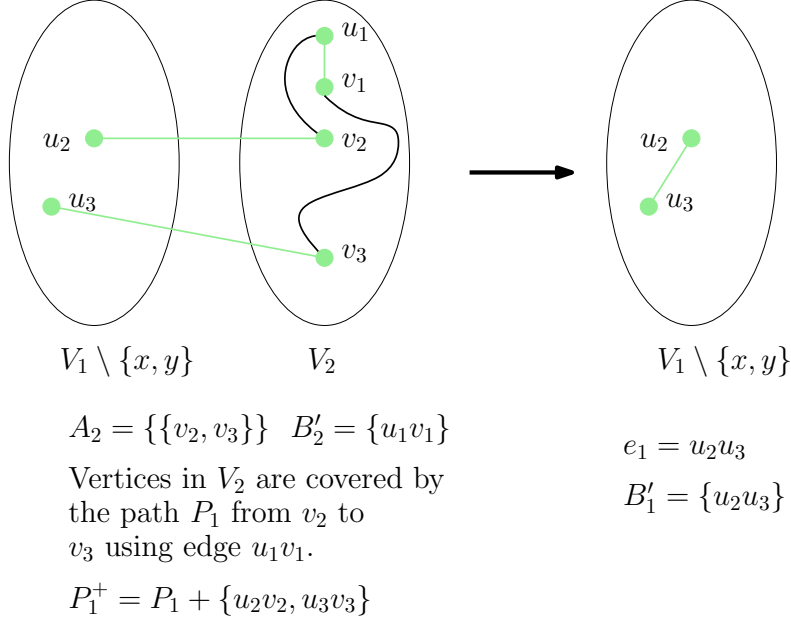
14

$$A_2 = \{\{v_2, v_3\}\} \quad B_2' = \{u_1 v_1\}$$

Vertices in V_2 are covered by the path P_1 from v_2 to v_3 using edge $u_1 v_1$.

$$P_1^+ = P_1 + \{u_2 v_2, u_3 v_3\}$$

$$e_1 = u_2 u_3$$

$$B_1' = \{u_2 u_3\}$$

Figure 2: Construct P_j, P_j^+ and e_j with $k = 2$

8.1 Proof of Lemma 20

We proceed with induction on k. The base case $k = 1$ follows by Lemma 21. Assume $k \geq 2$ and that the assertion holds for $k - 1$.

With a slight abuse of notation we call the pairs in B edges, even though they are not necessarily edges present in H. Let B_k' denote the set of edges in B with both ends in V_k and let B_k'' denote the set of edges of B with exactly one end in V_k. Let $V_k(B_k'')$ denote the ends of the edges in B_k'' that are in V_k. The second assumption of Lemma 20 implies that $|B_k''|$ is even. Take an arbitrary pairing A_k' of the vertices in $V_k(B_k'')$ and let $A_k = B_k' \cup A_k'$. By Lemma 21, $H_k + A_k$ has a Hamilton cycle C which uses all edges in A_k. Delete all edges in A_k' from C. This results in a collection of vertex disjoint paths $P_1, \ldots, P_\ell$ such that

- the ℓ paths cover all vertices in H_k and use all the edges in B_k';

- $\ell = |A_k'| = \frac{1}{2}|V_k(B_k'')|$;

- the ends of the ℓ paths are the set of vertices in $V_k(B_k'')$;

For every P_j above, the ends of P_j are each incident with an edge in $E_g \subseteq B_k''$. Let P_j^+ denote the path obtained by adding these two edges to P_j. Supplant P_j^+ by a new edge e_j. Now both ends of e_j are in $\cup_{i \leq k-1} V_i$. See Figure 2 for an example of the construction of P_j, P_j^+ and e_j.

Let $\widehat{B} = (B \cup \{e_1, \ldots, e_\ell\}) \setminus (B_k' \cup B_k'')$. Let $\widehat{H} = H[\cup_{i \leq k-1} V_i]$ be the subgraph of H induced by $\cup_{i \leq k-1} V_i$. Now $\widehat{H}$ has $\widehat{k} = k - 1$ blocks of vertices, and $\widehat{B}$ is a set of pairs of vertices with both ends in $\widehat{H}$. Moreover, all assumptions of Lemma 20 are satisfied by $\widehat{B}$ and $\widehat{H}$ with k replaced by $\widehat{k}$. By the induction hypothesis, there is a Hamilton cycle C' in $\widehat{H} + \widehat{B}$

which uses all the edges in $\widehat{B}$. Now replacing every edge e_j in C' by P_j^+. The resulting is a Hamilton cycle in $H + B$ which uses all of the edges in B.

Now Lemma 20 follows by induction. ∎

8.2 Proof of Lemma 21

We introduce/recall the following useful notations.

- $H_i = H[V_i \setminus \mathcal{P}] = G[V_i \setminus \mathcal{P}]$;

- $H_i^+ = H_i + A_i$;

- $H_i^b = G_b[V_i \setminus \mathcal{P}]$;

- $H_i^{b+} = H_i^b + A_i$.

We say a Hamilton cycle in H_i^+ is admissible if it uses all edges in A_i. Let $H_i^+ \in \mathtt{HAM}^+$ denote the property that H_i^+ contains an admissible Hamilton cycle. It is sufficient to prove that

$$\mathbb{P}(\{G \in \mathtt{D2}\} \cap \mathtt{COL1} \cap \{H_i^+ \notin \mathtt{HAM}^+\}) = o(1). \tag{5}$$

We will use Pósa rotations to bound the above probability. A path in H_i^+ (H_i^{b+}) is said admissible if for each edge in A_i it uses either that edge or none of its vertices. Let $L(H_i^+)$ and $L(H_i^{b+})$ denote the length of a longest admissible path in H_i^+ and H_i^{b+} respectively. We will adapt the previous Pósa rotation arguments to cope with admissible paths. This requires modifications of several previous definitions.

8.2.1 COL, SSEXPN, TPCL and LC

First, we define $\mathtt{SSEXPN}$ which is a stronger condition than $\mathtt{SEXPN}$. For $V \subseteq V(G)$, we say that (G, V) has property $\mathtt{EXPN}^+$, if there exists an absolute constant $\epsilon_0 > 0$ such that

$$\text{for every } S \subseteq V(G) \text{ where } |S| \leq \epsilon_0 n, \ |N_G(S) \setminus V| \geq 2|S| \cdot 1_{\{n_1(G)=0\}}.$$

We say G has property $\mathtt{SSEXPN}$, if the following holds: For any $F \subseteq E(G)$ such that $|F \cap \mathcal{N}_G(v)|$ is 0 if v is small, and is at most $\log n / 100$ if v is large, and for any $V \subseteq V(G)$ such that $|V| \leq \log n$, no two vertices in V share a common neighbour, and no vertex in V is adjacent to a vertex with degree at most 2, we have that $G - F$ is connected, and $(G - F, V) \in \mathtt{EXPN}^+$. We have the following lemma whose proof is postponed until Section 9.

Lemma 22. *A.a.s. $H_i \in \mathtt{SSEXPN}$ for every $1 \leq i \leq k$.*

Let $\mathtt{COL2}(i)$ denote the event that $L(H_i^+) = L(H_i^{b+})$ if $L(H_i^{b+}) < n - 1$, and $H_i^+ \notin \mathtt{HAM}^+$ if $L(H_i^{b+}) = n - 1$. We may define $\mathtt{COL2} = \cup_{i \in [k]} \mathtt{COL2}(i)$, although in the proof of Lemma 21 we only need to consider $\mathtt{COL2}(i)$. Let $\mathtt{COL3}$ denote the event that

$$|F \cap \mathcal{N}_G(v)| \text{ is 0 if } v \text{ is small in } G, \text{ and is at most } \log n / 100 \text{ if } v \text{ is large in } G,$$

where $F = E(G \setminus G_b)$.

We redefine TPCL so that (T1) is replaced by

$$(\text{T1'}): \quad H_i \in \texttt{SSEXPN} \text{ for every } 1 \le i \le k.$$

Let $H_i^+ \in \texttt{LC}$ denote the event that the longest admissible path has the same number of vertices as the longest admissible cycle in H_i^+.

8.2.2 Completing the proof of Lemma 21

Because A_i is a set of vertex-disjoint edges, no two edges in A_i appear next to each other in the longest admissible cycle. Then, if H_i^+ is connected, one can always extend a longest admissible cycle to a longer admissible path, unless $H_i^+ \in \texttt{HAM}^+$. Hence, $H_i^+ \in \texttt{LC} \cap \texttt{CNT}$ implies that $H_i^+ \in \texttt{HAM}^+$. Recall again that H_i and H_i^+ are defined only if $\{G \in \texttt{D2}\} \cap \texttt{COL1}$ holds. Hence

$$\mathbb{P}(G \in \texttt{D2} \cap \texttt{COL1} \cap H_i^+ \notin \texttt{HAM}^+) \le \mathbb{P}(G \in \texttt{D2} \cap \texttt{COL1} \cap \{H_i^+ \notin \texttt{LC} \cap \texttt{CNT}\})$$
$$\le \mathbb{P}(G \in \texttt{D2} \cap \texttt{COL1} \cap H_i^+ \notin \texttt{LC}) + \mathbb{P}(G \in \texttt{D2} \cap \texttt{COL1} \cap H_i^+ \notin \texttt{CNT})$$
$$= \mathbb{P}(G \in \texttt{D2} \cap \texttt{COL1} \cap H_i^+ \notin \texttt{LC}) + o(1) \quad \text{(by Lemma 22(a))}.$$

It is sufficient to prove that $\mathbb{P}(G \in \texttt{D2} \cap \texttt{COL1} \cap H_i^+ \notin \texttt{LC}) = o(1)$, which follows from the following two lemmas and the fact that a.a.s. $G \in \texttt{TPCL}$.

Lemma 23. $\mathbb{P}(\texttt{COL2}(i) \cap \texttt{COL3} \mid H_i^+ \in \overline{\texttt{LC}} \cap \{G \in \texttt{D2} \cap \texttt{COL1} \cap \texttt{TPCL}\}) \ge \exp(-O(an/\log^2 n))$.

The proof is almost identical to the proof of Lemma 7, with a few trivial modifications as in Lemma 13. We omit the details.

Lemma 24. $\mathbb{P}(H_i^+ \in \overline{\texttt{LC}} \cap \{G \in \texttt{D2} \cap \texttt{COL1} \cap \texttt{TPCL}\} \cap \texttt{COL2}(i) \cap \texttt{COL3}) \le \exp(-\Omega(an/\log n))$.

Proof. Recall that $V(A_i)$ denotes the set of vertices spanned by the edges in A_i. We have the following claim similar to Claim 10.

Claim 25. $\{G \in \texttt{D2} \cap \texttt{COL3} \cap \texttt{TPCL}\}$ *implies that* $\{H_i^{b+} \in \texttt{CNT} \cap \texttt{D2}\} \cap \{(H_i^{b+}, V(A_i)) \in \texttt{EXPN}^+\}$.

Hence,

$$\mathbb{P}(H_i^+ \in \overline{\texttt{LC}} \cap (G \in \texttt{D2} \cap \texttt{COL1} \cap \texttt{TPCL}) \cap \texttt{COL2}(i) \cap \texttt{COL3})$$
$$\le \mathbb{P}(H_i^+ \in \overline{\texttt{LC}} \cap \{H_i^{b+} \in \texttt{CNT} \cap \texttt{D2}\} \cap \{(H_i^{b+}, V(A_i)) \in \texttt{EXPN}^+\} \cap \texttt{COL2}(i) \cap \texttt{COL3})$$
$$\le \mathbb{P}(\texttt{COL2}(i) \cap \texttt{COL3} \mid \{H_i^{b+} \in \texttt{CNT} \cap \texttt{D2}\} \cap (H_i^{b+}, V(A_i)) \in \texttt{EXPN}^+)$$

Let $P = v_0 v_1, \dots v_\ell$ be a longest admissible path in H_i^{b+}. A Pósa rotation which adds edge $v_h v_\ell$ and deletes edge $v_h v_{h+1}$ is said to be admissible if $v_h v_{h+1} \notin A_i$. Let $\text{End}(v_0)$ be the set of admissible paths obtained by doing admissible Pósa rotations on P. We first show that

Claim 26. $|N_{H_i^{b+}}(\text{End}(v_0)) \setminus V(A_i)| < 2|\text{End}(v_0)|$.

Proof. The proof is similar to the standard Pósa rotation argument. Consider any $y \in \text{End}(v_0)$ and the path P' obtained via a Pósa rotation when y is added to $\text{End}(v_0)$. Assume xy is an edge where x in on P. Assume $x = v_i$ and assume $x \notin V(A_i)$. Then, either the two neightbours of x on P' are exactly v_{i-1} and v_{i+1}, in which case one of them can be added to $\text{End}(v_0)$ by a Pósa rotation; or the two neighbours of x on P' are not v_{i-1} and v_{i+1}, which implies that one of them must have been added to $\text{End}(v_0)$ before y. Hence, either $\{v_{i-1}, v_{i+1}\} \cap \text{End}(v_0) \neq \emptyset$ or $v_i \in V(A_i)$. Our claim follows immediately. ∎

By (T1') and Claim 26, we have that $|\text{End}(v_0)| = \Omega(n)$. Take any $x \in \text{End}(v_0)$, consider $\text{End}(x)$, the set of longest admissible paths starting from x by performing admissible Pósa rotations. Then we also have $|\text{End}(x)| = \Omega(n)$ for every $x \in \text{End}(v_0)$. If any of the edges in $E(H_i^+ \setminus H_i^{b+})$ belongs to the set $\mathscr{E} := \{xy : x \in \text{End}(v_0), \ y \in \text{End}(x)\}$, then the event COL2(i) fails. As $|\mathscr{E}| = \Omega(n^2)$, the probability that $E(H_i^+ \setminus H_i^{b+}) \cap \mathscr{E} = \emptyset$ is at most $(1 - \bar{p})^{\Omega(n^2)} = \exp(-\Omega(an/\log n))$. Hence

$$\mathbb{P}(\text{COL2}(i) \mid H_i^{b+} \in \text{CNT} \cap \text{D2} \cap ((H_i^{b+}, V(A_i)) \in \text{EXPN}^+)) \leq \exp(-\Omega(an/\log n)),$$

completing the proof. ∎

9 Proof of Lemmas 5 and 22

We first state a lemma, whose proof is sketched below.

Lemma 27. *Assume (A1). A.a.s. $\mathcal{G}(\mathbf{n}, p, q)$ satisfies the following graph properties.*

(C1) *For every $i \in [k]$, at most $n^{0.9}$ vertices in V_i have degree less than $\dfrac{1}{2}\log n$.*

(C2) *No two vertices with degree less than 100 are within distance 5.*

(C3) *No vertices with degree less than 100 are contained in cycles of length at most 5.*

(C4) *Every set S with $|S| < n/\log^2 n$ induces at most $3|S|$ edges.*

(C5) *For all $\epsilon > 0$ there exists $\delta > 0$ such that for all S where $\Omega(n/\log^2 n) = |S| \leq \delta n$, $|E(S)| < \epsilon|S|\log n$.*

Proof of Lemma 5. Let G be a graph satisfying properties (C1)–(C5). We also assume that the minimum degree of G is at least 2, as otherwise the lemma is trivially true. In the proof of the lemma we consider various ranges for $|S|$. Colour the edges in F red and let $G' = G - F$. Our assumption on F implies that

$$\text{every vertex is incident with at most } \log n/100 \text{ red edges.} \tag{6}$$

Let $\epsilon = 1/24$, and let $\delta > 0$ be the constant in (C5).

Case a: $n/(\log n)^2 \leq |S| \leq (\delta/3)n$. Let $E_1 = E_G(S, \overline{S})$ and $E_2 = E_{G'}(S, \overline{S})$ and let $U = \mathcal{N}_{G'}(S)$. Suppose that $|U| < 2|S|$. Then, $|S \cup U| < 3|S| \leq \delta n$. By (C5), $S \cup U$ induces at most $\epsilon|S \cup U|\log n \leq 3\epsilon|S|\log n$ edges in G. This implies that $|E_2| \leq 3\epsilon|S|\log n$. By (C1), the total degree of vertices in S is at least $(|S| - n^{0.9}) \cdot (1/2)\log n \geq |S|\log n/3$. On the other hand, by (C5), S induces at most $\epsilon|S|\log n$ edges. Thus, $|E_1| \geq |S|\log n/3 - 2\epsilon|S|\log n =$

18

$|S| \log n/4$. Consequently, $|F \cap E_G(S, \overline{S})| = |E_1| - |E_2| \geq (5/24)|S| \log n$, contradicting condition (6).

Case b $|S| \leq n/\log^2 n$. A vertex in G is called *extremely small* if its degree is less than 100. Let $\mathcal{X}$ denote the set of extremely small vertices in S, and $\mathcal{Y}$ denote $S \setminus \mathcal{X}$.

Case b1: $|\mathcal{Y}| = 0$. Then F is not incident with any vertex in S. By (C2), S must induce an independent set. By our assumption, all vertices in S has degree at least 2. By (C2), $\mathcal{N}_G(a) \cap \mathcal{N}_G(b) = \emptyset$ for every distinct $a, b \in S$. It follows immediately then that $|\mathcal{N}_{G'}(S)| = |\mathcal{N}_G(S)| \geq 2|S|$.

Case b2: $|\mathcal{Y}| \geq 1$. Now F is not incident with any vertex in $\mathcal{X}$. Let $\mathcal{Z}_1 = \mathcal{N}_{G'}(\mathcal{X}) \setminus \mathcal{Y} = \mathcal{N}_G(\mathcal{X}) \setminus \mathcal{Y}$ be the set of neighbours of $\mathcal{X}$ that are not in $\mathcal{Y}$, $\mathcal{Z}_2 = N_{G'}(S) \setminus \mathcal{Z}_1$ be the neighbours of S in G' that are not in $\mathcal{Z}_1$. Then, $|\mathcal{N}_{G'}(S)| = |\mathcal{Z}_1| + |\mathcal{Z}_2|$. Let $\mathcal{Y}_1 = \mathcal{N}_{G'}(\mathcal{X}) \cap \mathcal{Y}$ be the set of neighbours of $\mathcal{X}$ in $\mathcal{Y}$. By (C2) and our assumption that the minimum degree of G is at least 2,

$$|\mathcal{Y}_1| = |E_{G'}(\mathcal{X}, \mathcal{Y})| = |E_G(\mathcal{X}, \mathcal{Y})| \tag{7}$$
$$|\mathcal{N}_{G'}(\mathcal{X})| = |\mathcal{Z}_1| + |\mathcal{Y}_1| \geq 2|\mathcal{X}|. \tag{8}$$

We prove next that every vertex in $\mathcal{Y}$ can be incident to at most one vertex in $\mathcal{Z}_1$ in G (and G'). Assume $a \in \mathcal{Y}$ is adjacent to two vertices b and c in $\mathcal{Z}_1$. If b and c have a common neighbour $z \in \mathcal{X}$, then $abzc$ forms a 4-cycle in G, violating (C3). Assume b and c each adjacent to $b' \in \mathcal{X}$ and $c' \in \mathcal{X}$ respectively. Then $b'bacc'$ is a 4-path in G connecting two light vertices, violating (C2). Hence, $|\mathcal{N}_{G'}(z) \cap \mathcal{Z}_1| \leq 1$ for every $z \in \mathcal{Y}$. Consequently,

$$|E_{G'}(\mathcal{Y}, \mathcal{Z}_1)| \leq |\mathcal{Y}|. \tag{9}$$

Assume to the contrary that $|\mathcal{Z}_1| + |\mathcal{Z}_2| = |\mathcal{N}_{G'}(S)| < 2|S| = 2(|\mathcal{X}| + |\mathcal{Y}|)$. Then, by (8) we have

$$|\mathcal{Z}_2| < 2|\mathcal{Y}| + |\mathcal{Y}_1|. \tag{10}$$

Every vertex in $\mathcal{Y}$ has degree at least 100 in G. Also, a vertex is incident with red edges only if its degree is at least $\log n/10$ and then at most $\log n/100$ red edges. It follows that every vertex in $\mathcal{Y}$ has degree at least 100 in G' as well. Thus,

$$|E_{G'}(\mathcal{X}, \mathcal{Y})| + 2|E_{G'}(\mathcal{Y})| + |E_{G'}(\mathcal{Y}, \mathcal{Z}_1)| + |E_{G'}(\mathcal{Y}, \mathcal{Z}_2)| \geq 100|\mathcal{Y}|. \tag{11}$$

By (C4), $|E_{G'}(\mathcal{Y})| \leq 3|\mathcal{Y}|$, $|E_{G'}(\mathcal{Y}, \mathcal{Z}_2)| \leq |E_{G'}(\mathcal{Y} \cup \mathcal{Z}_2)| \leq 3(|\mathcal{Y}| + |\mathcal{Z}_2|)$. We have shown that $|E_{G'}(\mathcal{X}, \mathcal{Y})| \leq |\mathcal{Y}_1|$ and $|E_{G'}(\mathcal{Y}, \mathcal{Z}_1)| \leq |\mathcal{Y}|$. Hence, by (10), the left hand side of (11) is at most $16|\mathcal{Y}| + 4|\mathcal{Y}_1| \leq 20|\mathcal{Y}|$, whereas the right hand side is $100|\mathcal{Y}|$, contradiction. This confirms that $|N_{G'}(S)| \geq 2|S|$. ∎

Proof of Lemma 27. The proof is standard and straightforward. We give a sketch only and omit the somewhat tedious calculations. For (C1), following the same argument as in Lemma 3, we can show that the expected number of vertices with degree less than $\frac{1}{2} \log n$ is $o(n^{0.9})$.

By (A1), there is $i \in [k]$ such that $pn_i + (n - n_i)q - \log n_i = \log \log n + O(1)$. Together with (A4), this implies that $p, q = O(\log n/n)$. Using this, the expected number of S with $|S| = s$ which induce more than $3s$ edges is at most

$$\binom{n}{s}\binom{s^2}{3s}(C \log n/n)^{3s}, \quad \text{for some constant } C > 0.$$

It is straightforward to see that summing the above over $s < n/\log^2 n$ yields $o(1)$. This proves (C4). The proof for (C5) is almost the same, and we can bound the expected number of sets S inducing more than $\epsilon|S|\log n$ by $o(1)$ by choosing a sufficiently small δ.

For (C2), we bound the expected number of ℓ-paths, $\ell \leq 5$, where the ends are vertices of degree less than 100. There are at most $n^{\ell+1}$ ways to choose the $\ell+1$ vertices. Using the probability bounds as in Lemma 3, the probability of both of the chosen end-vertices having degree less than 100 is at most $\log^{200} n/n^2$. The probability that the chosen $\ell+1$ vertices form a path is bounded by $(C\log n/n)^\ell$. Multiplying all together we have that the expected number of such paths is at most $n^{\ell+1} \cdot (\log^{200} n/n^2) \cdot (C\log n/n)^\ell = o(1)$. This proves (C2). The proof of (C3) is similar. ∎

Proof of Lemma 22. We only need to consider $\mathcal{G}(n_i, p)$ where $pn_i \geq \log n + \log\log n + O(1)$. The proof is almost the same as the proof of Lemma 5 with only small modifications which we point out below. Again we assume that H_i is a graph satisfying (C1)–(C5) (for (C1), we only need to consider a fixed i). Let V be an arbitrary set of vertices in H_i such that $|V| \leq \log n$, no two vertices in V share a common neighbour, and no vertex in V is adjacent to a vertex of degree at most 2. Colour the vertices in V red. Let F be an arbitrary set of edges such that $|F \cap \mathcal{N}_{H_i}(v)|$ is 0 if v is small, and is at most $\log n/100$ if v is large. Let $H_i' = H_i - F$.

Let $\epsilon = 1/24$, and let $\delta > 0$ be the constant in (C5). For S where $n/(\log n)^2 \leq |S| \leq (\delta/4)n$, let $E_1 = E_{H_i}(S, \overline{S})$ and let $E_2 = E_{H_i'}(S, \overline{S})$. Let $U = \mathcal{N}_{H_i'}(S)$. Assume $|U \setminus V| < 2|S|$. Then, $|S \cup U| < 3|S| + |V| \leq 4|S| \leq \delta n$. Now with the same proof as in Lemma 5, we can lead to a contradiction with condition (6). Thus, we must have $|U \setminus V| \geq 2|S|$ in this case.

For S where $|S| \leq n/\log^2 n$, again, call a vertex in H_i *extremely small* if its degree is less than 100. Let $\mathcal{X}$ denote the set of extremely small vertices in S, and $\mathcal{Y}$ denote $S \setminus \mathcal{X}$.

Case 1: $|\mathcal{Y}| = 0$. Then F is not incident with any vertex in S. By (C2), S must induce an independent set. Recall that $H_i = G_b[V_i \setminus \mathcal{P}]$. It follows immediately that every vertex in H_i has degree at least 2, as all problematic vertices are in $\mathcal{P}$. Moreover, by the assumptions on V, the vertices with degree 2 are not adjacent to any vertex in V, and every vertex in H_i can be adjacent to at most one vertex in V. Hence, every vertex has at least 2 non-red neighbours. By (C2), $\mathcal{N}_{H_i}(a) \cap \mathcal{N}_{H_i}(b) = \emptyset$ for every distinct $a, b \in S$. It follows immediately then that $|\mathcal{N}_{H_i'}(S) \setminus V| = |\mathcal{N}_{H_i}(S) \setminus V| \geq 2|S|$.

Case 2: $|\mathcal{Y}| \geq 1$. Now F is not incident with any vertex in $\mathcal{X}$. Let $\mathcal{Z}_1 = \mathcal{N}_{G'}(\mathcal{X}) \setminus \mathcal{Y} = \mathcal{N}_G(\mathcal{X}) \setminus \mathcal{Y}$ be the set of neighbours of $\mathcal{X}$ that are not in $\mathcal{Y}$, $\mathcal{Z}_2 = N_{G'}(S) \setminus \mathcal{Z}_1$ be the neighbours of S in G' that are not counted in $\mathcal{Z}_1$. Let $\mathcal{Z}_i' = \mathcal{Z}_i \setminus V$ for $i \in \{1, 2\}$. Then, $|\mathcal{N}_{H_i'}(S) \setminus V| = |\mathcal{Z}_1'| + |\mathcal{Z}_2'|$. Let $\mathcal{Y}_1 = \mathcal{N}_{G'}(\mathcal{X}) \cap \mathcal{Y}$ be the set of neighbours of $\mathcal{X}$ in $\mathcal{Y}$. With the same argument for (7)–(9), together with the fact that every vertex in $\mathcal{X}$ has at least 2 non-red neighbours, we have

$$|\mathcal{Y}_1| = |E_{H_i'}(\mathcal{X}, \mathcal{Y})| = |E_{H_i}(\mathcal{X}, \mathcal{Y})| \tag{12}$$

$$|\mathcal{N}_{H_i}(\mathcal{X}) \setminus V| = |\mathcal{Z}_1'| + |\mathcal{Y}_1| \geq 2|\mathcal{X}| \tag{13}$$

$$|E_{H_i'}(\mathcal{Y}, \mathcal{Z}_1)| \leq |\mathcal{Y}|. \tag{14}$$

Assume to the contrary that $|\mathcal{Z}_1'| + |\mathcal{Z}_2'| = |\mathcal{N}_{H_i'}(S) \setminus V| < 2|S| = 2(|\mathcal{X}| + |\mathcal{Y}|)$. Then, by (13) we have

$$|\mathcal{Z}_2'| < 2|\mathcal{Y}| + |\mathcal{Y}_1|. \tag{15}$$

Every vertex is adjacent to at most one vertex in V. It follows immediately that

$$|\mathcal{Z}_2| \leq |\mathcal{Z}_2'| + |\mathcal{Y}|. \tag{16}$$

Every vertex in $\mathcal{Y}$ has degree at least 100 in H_i. Also, a vertex is incident with red edges only if its degree is at least $\log n/10$ and then at most $\log n/100$ red edges. It follows that every vertex in $\mathcal{Y}$ has degree at least 100 in G' as well. Thus,

$$|E_{H_i'}(\mathcal{X}, \mathcal{Y})| + 2|E_{H_i'}(\mathcal{Y})| + |E_{H_i'}(\mathcal{Y}, \mathcal{Z}_1)| + |E_{H_i'}(\mathcal{Y}, \mathcal{Z}_2)| \geq 99|\mathcal{Y}|. \tag{17}$$

By (C4), $|E_{H_i'}(\mathcal{Y})| \leq 3|\mathcal{Y}|$, $|E_{H_i'}(\mathcal{Y}, \mathcal{Z}_2)| \leq |E_{H_i'}(\mathcal{Y} \cup \mathcal{Z}_2)| \leq 3(|\mathcal{Y}| + |\mathcal{Z}_2|) \leq 3(4|\mathcal{Y}| + |\mathcal{Y}_1|)$ by (15) and (16). We have shown that $|E_{H_i'}(\mathcal{X}, \mathcal{Y})| = |\mathcal{Y}_1|$ and $|E_{H_i'}(\mathcal{Y}, \mathcal{Z}_1)| \leq |\mathcal{Y}|$. Hence, the left hand side of (11) is at most $19|\mathcal{Y}| + 4|\mathcal{Y}_1| \leq 23|\mathcal{Y}|$, whereas the right hand side is $99|\mathcal{Y}|$, contradicting with $|\mathcal{Y}| \geq 1$. This shows that $|N_{H_i'}(S) \setminus V| \geq 2|S|$. ∎

10   Conclusion

We have analysed the Hamiltonicity of a particular stochastic block model and given tight estimates for the threshold. The most natural extension of this work will be to the case where P is an arbitrary symmmetric stochastic matrix. This will be the subject of further research.

References

[1] E. Abbe, Community Detection and Stochastic Block Models: Recent Developments, *Journal of Machine Learning Research* 18 (2018) 1-86.

[2] A.M. Frieze, Limit distribution for the existence of hamiltonian cycles in random bipartite graphs, *European Journal of Combinatorics* 6 (1985) 327-334.

[3] A.M. Frieze, Hamilton Cycles in Random Graphs: a bibliography.

[4] A.M. Frieze and M. Karoński, Introduction to Random Graphs, Cambridge University Press, 2015.

[5] J. Komlós and E. Szemerédi, Limit distributions for the existence of Hamilton circuits in a random graph, *Discrete Mathematics* 43 (1983) 55-63.