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 $V_1, V_2, \dots, V_k$ 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 probability 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

*Research supported in part by NSF grant DMS1363136
†Research supported by NSERC.

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. All asymptotics refer to $n \to \infty$. Given a graph property Γ, we say $\mathcal{G}(\mathbf{n}, p, q) \in \Gamma$ asymptotically almost surely (a.a.s.) if $\lim_{n\to\infty}(\mathcal{G}(\mathbf{n}, p, q) \in \Gamma) = 1$. 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 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 $\alpha + \alpha\ln(1/\alpha) < \gamma < 1$. A.a.s. $\mathcal{G}(\mathbf{n}, p, q)$ contains at most n^γ 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 $p^2n = o(1)$, $pj = o(1)$,

$j^2 = o(n_i)$, and $j^2 = o(n - n_i)$. Then,

$$\begin{aligned}
\mathbb{E}X_j(i) &= n_i \mathbb{P}(W_1 + W_2 = j) \\
&= n_i \sum_{s=0}^{j} \binom{n_i - 1}{s} p^s (1-p)^{n_i - 1 - s} \binom{n - n_i}{j - s} q^{j-s} (1-q)^{n - n_i - j + s} \\
&\sim n_i \exp(-pn_i - q(n - n_i)) \sum_{s=0}^{j} \frac{n_i^s (n - n_i)^{j-s}}{s!(j-s)!} p^s q^{j-s} \\
&= n_i e^{-\phi_i} \frac{\phi_i^j}{j!}, \qquad (1)
\end{aligned}$$

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. As an example of similar calculations, see Theorem 2.8 of [4]. Hence,

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

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

$$\begin{aligned}
\sum_{j \le \alpha \log n} \sum_{i \in [k]} \mathbb{E}X_j(i) &< (1 + o(1)) \sum_{j \le \alpha \log n} \sum_{i \in [k]} n_i \exp(-\log n_i) \frac{(\log n_i)^j}{j!} \\
&< k \sum_{j \le \alpha \log n} \left(\frac{e \log n}{j} \right)^j = (1 + o(1)) k n^{\rho(\alpha)},
\end{aligned}$$

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

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

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. *In the following and throughout the paper $\gamma \le 1$. A vertex in G is called γ-small, if its degree is less than $(\gamma \log n)/10$. A vertex with degree at least $(\gamma \log n)/10$ is called γ-large.*

We say that $V(G)$ naturally admits the block partition $V_1, V_2, ..., V_k$ if the block partition $V_1, V_2, ..., V_k$ is defined by the model used to generate G. (e.g. when G is generated via the stochastic block model). We say G has property `EXPN`, if there exists $\epsilon > 0$ such that

$$\text{for every } S \subseteq V(G) \text{ where } \ |S| \le \epsilon n,\ |N_G(S)| \ge 2|S| \cdot 1_{\{n_1(G)=0\}}.$$

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. A set $F \subseteq E(G)$ is said to be γ-*deletable* if (i) $|F \cap \mathcal{N}_G(v)| = 0$ if v is γ-small, and is at most $(\gamma \log n)/100$ otherwise and (ii) if $V(G)$ naturally admits the block partition $V_1, V_2, ..., V_k$ then F contains at most half of the edges between any pair of distinct blocks V_i, V_j, $i, j \in [k]$. We say G has property `SEXPN`(γ), if $G - F$ is connected and $G - F \in$ `EXPN` for all γ-deletable F.

5 Overview of the proof of Theorem 1

Since $G \notin$ `D2` implies that $G \notin$ `HAM`, the upper bound for the probability that $\mathcal{G}(\mathbf{n}, p, q) \in$ `HAM` is implied by Lemma 3(a). Next we prove the lower bound. Let `LC` denote the set of graphs where a longest path contains the same number of vertices as in a longest cycle, and let `CNT` denote the class of connected graphs. Note that if $G \in$ `LC`$\cap$`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 $G \in$ `LC`. It follows then that

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

Our goal is to prove that $\mathbb{P}(\mathcal{G}(\mathbf{n}, p, q) \in \texttt{LC} \cap \texttt{CNT} \cap \texttt{D2}) \sim \mathbb{P}(\mathcal{G}(\mathbf{n}, p, q) \in \texttt{D2})$, which then yields the asymptotic probability desired by Theorem 1. The proof of the lower bound of $\mathbb{P}(\mathcal{G}(\mathbf{n}, p, q) \in \texttt{HAM})$ 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 `TPCL` which consists of a set of properties that hold a.a.s. for $\mathcal{G}(\mathbf{n}, p, q)$. Then we will define `COL` to be a set of properties that edges in $G \setminus G_b$ must satisfy. Then, we will prove that

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

This implies that $\mathbb{P}(G \in \overline{\texttt{LC}} \cap \texttt{CNT} \cap \texttt{D2} \cap \texttt{TPCL}) = o(1)$, which will lead us to derive the asymptotic probability for $\mathbb{P}(G \in \texttt{LC} \cap \texttt{CNT} \cap \texttt{D2})$. While obtaining a lower bound for $\mathbb{P}(\texttt{COL} \mid$

$G \in \overline{\texttt{LC}} \cap \texttt{CNT} \cap \texttt{D2} \cap \texttt{TPCL}$) is rather straightforward, an upper bound for $\mathbb{P}(\{G \in \overline{\texttt{LC}} \cap \texttt{CNT} \cap \texttt{D2} \cap \texttt{TPCL}\} \cap \texttt{COL})$ 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, $\mathcal{G}(\mathbf{n}, p, q)$ 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, $1 \le i < j \le k$. 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 $\mathcal{G}(\mathbf{n}, p, q)$.

In Case 3, $\mathcal{G}(\mathbf{n}, p, q)$ 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, $\mathbb{P}(\mathcal{G}(\mathbf{n}, p, q) \in \texttt{HAM}) > \mathbb{P}(\mathcal{G}(\mathbf{n}, 0, q) \in \texttt{HAM})$ 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.

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 $G \in \texttt{TPCL}(\gamma)$ if G satisfies the following set of properties.

(T1) $G \in \texttt{SEXPN}(\gamma)$.

(T2) There are at most n^{ξ} γ-small vertices, where $0 < \xi = \xi(\gamma) < 1$.

(T3) If $p = o((\log n)/n)$ then every vertex is incident with at most $(\gamma \log n)/100$ block edges. If $q = o((\log n)/n)$ then every vertex is incident with at most $(\gamma \log n)/100$ crossing edges.

(T4) There are $\omega(n)$ edges between any pair of blocks .

(T5) The maximum degree is at most $K \log n$ for some sufficiently large constant $K > 0$.

Lemma 5. *There exists γ such that a.a.s. $\mathcal{G}(\mathbf{n}, p, q) \in \texttt{D2}$ implies $\mathcal{G}(\mathbf{n}, p, q) \in \texttt{TPCL}(\gamma)$.*

Proof. (T1) follows by Lemma 6. (T2) follows by Lemma 3(b). (T3) and (T5) follow by a standard first moment argument similar to the proof of Lemma 3. (T4) follows from a standard application of Chernoff bounds. 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 $(G, G') \in \texttt{COL}$ if

(a) F is γ-deletable;

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

In the proof of Theorem 1, we will use COL to denote the event that $(G, G_b) \in \texttt{COL}$, 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 $p, q = \omega(1/n)$

Lemma 6. *Assume (A1), (A2) and (A4). If $p, q = \omega(1/n)$ then a.a.s. $\mathcal{G}(\mathbf{n}, p, q) \in \texttt{D2}$ implies $\mathcal{G}(\mathbf{n}, p, q) \in \texttt{SEXPN}$.*

Its technical proof is postponed till Section 9.

In this section we take

$$\gamma = 1 \text{ and } \epsilon = b_{mim}^6/3 \text{ and } \xi = 2/5 \text{ in (T2)}.$$

We also denote by small, EXPN, SEXPN, TPCL the properties γ-small, EXPN, SEXPN(γ), TPCL(γ) with γ as given above. 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 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 $(G, G_b) \in \texttt{COL}$.

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

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

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

Lemma 8.

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

Now we are ready to prove Theorem 1 in Case 1.

Proof of Theorem 1 (case 1). By Lemmas 7 and 8,

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

As $f \to \infty$ as $n \to \infty$, the above probability is $o(1)$. By Lemmas 6 and 5, $\mathbb{P}(G \in \texttt{D2}) = \mathbb{P}(G \in \texttt{D2} \cap \texttt{TPCL} \cap \texttt{CNT}) + o(1)$. It follows immediately that

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

By (3) and the fact that $G \in \texttt{HAM}$ implies $G \in \texttt{D2}$, we have $\mathbb{P}(G \in \texttt{HAM}) = \mathbb{P}(G \in \texttt{D2}) + o(1)$. Together with Lemma 3 this yields the asymptotic probability of $\mathbb{P}(G \in \texttt{HAM})$ 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 $H \in \overline{\texttt{LC}} \cap \texttt{D2} \cap \texttt{TPCL}$, $\mathbb{P}(\texttt{COL} \mid G = H) \geq \exp(-an/f \log n)$, 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.

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

(B3) no edge in P is red.

(B4) at most half of the edges between any pair of blocks are 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), $|\mathcal{X}| \le (n-1) + n^{0.4}(\log n)/10 < 2n$. 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 $(1 - o(a/\log n))^{2n} \ge \exp(-an/f \log n)$ for some $f \to \infty$. Note that (B2) and (B3) are implied if no edges in $\mathcal{X}$ are red. Hence, $\mathbb{P}(B_2 \cap B_3) \ge \exp(-an/f \log n)$. For (B4) we observe that $q^* = o(1)$, (T4) implies that there are $\omega(n)$ edges between every pair of blocks, and a standard application of Chernoff bounds shows that there exists a pair for which more than half of the edges are deleted with probability at most $e^{-\omega(n)}$.

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 (T5), 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, $\mathbb{P}(\texttt{COL} \mid G = H) \ge \mathbb{P}(B1 \cap B2 \cap B3) \ge \exp(-an/f \log n)$ for some $f \to \infty$. Lemma 7 follows.

6.3 Proof of Lemma 8

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

$$\text{for every } S \subseteq V(G) \text{ where } |S| \le \epsilon n, \text{ we have } |N_G(S)| \ge 2|S| \cdot 1_{\{n_1(G)=0\}}. \tag{4}$$

By the definition of COL we immediately have the following claim.

Claim 10. $\{G \in \overline{\texttt{LC}} \cap \texttt{D2} \cap \texttt{TPCL}\} \cap \texttt{COL}$ *implies* $\mathcal{B} \cap \{G_b \in \texttt{CNT} \cap \texttt{EXPN} \cap \texttt{D2}\}$, *where*

$$\mathcal{B} = \{L(G) = L(G_b) < n-1\} \cup (\{L(G_b) = n-1\} \cap \{G \notin \texttt{HAM}\}). \tag{5}$$

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 $\mathcal{B}$. This proves our claim. □

Hence, it is sufficient to prove

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

as

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

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 $\mathcal{B}$.

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. $|N_{G_b}(\mathit{End}(v_0))| < 2|\mathit{End}(v_0)|$.

Proof. As $G_b \in \texttt{EXPN} \cap \texttt{D2}$ we immediately have that $|\mathrm{End}(v_0)| \geq \epsilon n$ where ϵ 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 while keeping v fixed. Again, we have $|N_{G_b}(\mathrm{End}(v))| < 2|\mathrm{End}(v)|$, which implies that $|\mathrm{End}(v)| \geq \epsilon n$. 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 $|\mathscr{E}| \geq \epsilon^2 n^2$. 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 $\mathcal{B}$ fails. For that reason, we call $\mathscr{E}$ a set of *boosters*. We have shown that $\mathcal{B}$ 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 $(1 - \bar{p})^{|\mathscr{E}|/2} \leq \exp(-\epsilon^2 an/2\log n)$. This completes the proof for Lemma 8.

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

The parameters ϵ, γ, ξ as well as the relevant notation are unchanged from Section 6.

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 COL will be defined differently from the previous case. We will split COL into three parts. Let 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 12. *A.a.s.* $(G, G_b) \in \texttt{COL1}$.

Let $\texttt{COL} = \texttt{COL1} \cap \texttt{COL2} \cap \texttt{COL3}$ where COL2 and COL3 will be defined later. Note that $\{G \in \texttt{D2}\} \cap \texttt{COL1}$ implies $G_b \in \texttt{D2}$. If $\{G \in \texttt{D2}\} \cap \texttt{COL1}$ 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 $u_1 u_2 \dots u_\ell$, 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 $\{G \in \texttt{D2}\} \cap \texttt{COL1}$. Supplant every green 2-path in G by a new green edge. Call the resulting graph H. Note that H is not defined if $\{G \in \texttt{D2}\} \cap \texttt{COL1}$ fails. Note also that $H[V_i \setminus \mathcal{P}] = G[V_i \setminus \mathcal{P}]$ for every $1 \leq i \leq k$. 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 $1 \leq i < j \leq k$, 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 $i \neq j$. 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 13. *A.a.s. the following statements hold.*

(a) $E \neq \perp$.

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

(c) E *induces a matching.*

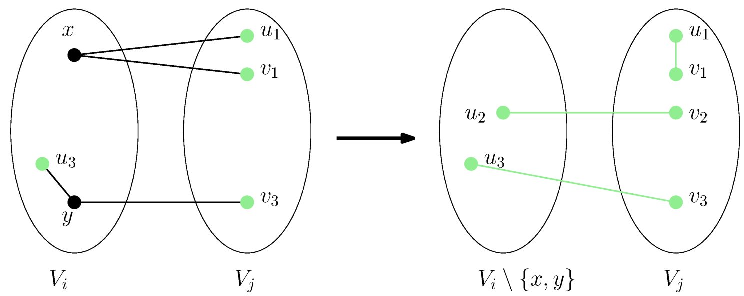

Figure 1: Construct H

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

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

Claim 14. *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 15. *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.

Proof of Theorem 1 in the case $q = O(1/n)$: Using Claim 13 it is easy to show that $B = E_g$ satisfies all assumptions of Lemma 15. 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 15 immediately implies HHAM, which gives

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

by Claim 12 and Lemma 15, which completes the proof. □

It only remains to prove Lemma 15. We will prove it by induction on k. The following key lemma will be used to complete the inductive argument.

Lemma 16. *Fix $1 \leq i \leq k$. Let $H_i = G[V_i \setminus \mathcal{P}]$ and 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 $\{G \in \texttt{D2}\} \cap \texttt{COL1}$ holds then $H_i + A_i$ has a Hamilton cycle containing all of the edges in A_i.

We will prove Lemma 15 in Section 7.1 and prove Lemma 16 in Section 7.2.

7.1 Proof of Lemma 15

We proceed with induction on k. The base case $k = 1$ follows by Lemma 16. 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 15 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 16, $H_k + A_k$ 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.

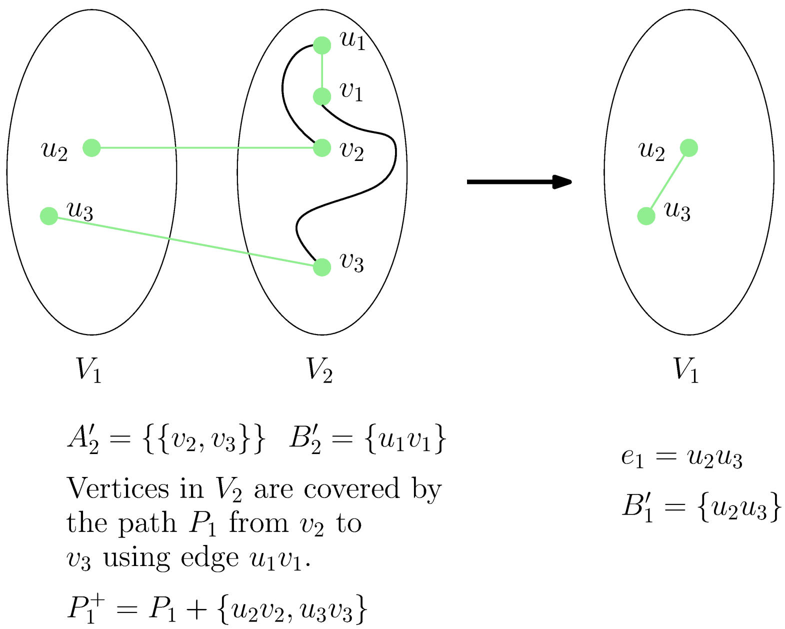

Figure 2: Construct P_j, P_j^+, e_j and $\widehat{B}$ with $k = 2$

Let $\widehat{B} = (B \cup \{e_1, \ldots, e_\ell\}) \setminus (B'_k \cup B''_k)$. Let $\widehat{H} = H[\cup_{i \le k-1} V_i]$ be the subgraph of H induced by $\cup_{i \le k-1} V_i$. 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 15 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 15 follows by induction. □

7.2 Proof of Lemma 16

We introduce/recall the following useful notations.

- $H_i = G[V_i \setminus \mathcal{P}]$;
- $H_i^+ = H_i + A_i$, the graph obtained by taking the union of the edges in H_i and 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

$$\mathbb{P}(\{G \in \mathtt{D2}\} \cap \mathtt{COL1} \cap \{H_i^+ \notin \mathtt{HAM}^+\}) = o(1). \tag{6}$$

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.

7.2.1 COL, SSEXPN, TPCL and LC

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

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

We say (G, V) has property SSEXPN, if the following holds: (i) $|V| \le \log n$, (ii) no two vertices in V share a common neighbour (iii) no vertex in V is adjacent to a vertex with degree at most 2, (iv) 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, we have that $G - F$ is connected, and $(G - F, V) \in \texttt{EXPN}^+$.

We have the following lemma whose proof is postponed until Section 9. For $1 \le i \le k$ we let $V(A_i) = \cup_{e \in A_i} e$.

Lemma 17. *A.a.s.* $G \in \texttt{D2}$ *implies* $(H_i, V(A_i)) \in \texttt{SSEXPN}$ *for every* $1 \le i \le k$.

Let COL2(i) denote the event that $L(H_i^+) = L(H_i^{b+})$ if $L(H_i^{b+}) < n - 1$, and $H_i^+ \notin \texttt{HAM}^+$ if $L(H_i^{b+}) = n - 1$. We may define $\texttt{COL2} = \cup_{i \in [k]} \texttt{COL2}(i)$, although in the proof of Lemma 16 we only need to consider $\texttt{COL2}(i)$. Let 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

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

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^+.

7.2.2 Completing the proof of Lemma 16

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

$$\begin{aligned}
\mathbb{P}(G \in \texttt{D2} \cap \texttt{COL1} \cap H_i^+ \notin \texttt{HAM}^+) &\le \mathbb{P}(G \in \texttt{D2} \cap \texttt{COL1} \cap \{H_i^+ \notin \texttt{LC} \cap \texttt{CNT}\}) \\
&\le \mathbb{P}(G \in \texttt{D2} \cap \texttt{COL1} \cap H_i^+ \notin \texttt{LC}) + \mathbb{P}(G \in \texttt{D2} \cap \texttt{COL1} \cap H_i^+ \notin \texttt{CNT}) \\
&= \mathbb{P}(G \in \texttt{D2} \cap \texttt{COL1} \cap H_i^+ \notin \texttt{LC}) + o(1) \quad \text{(by Lemma 17).}
\end{aligned}$$

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

Lemma 18. $\mathbb{P}(\mathtt{COL2}(i) \cap \mathtt{COL3} \mid H_i^+ \in \overline{\mathtt{LC}} \cap \{G \in \mathtt{D2} \cap \mathtt{COL1} \cap \mathtt{TPCL}\}) \geq \exp(-O(an/\log^2 n))$.

The proof is almost identical to the proof of Lemma 7, with a few trivial modifications as in Lemma 23. We omit the details.

Lemma 19. $\mathbb{P}(H_i^+ \in \overline{\mathtt{LC}} \cap \{G \in \mathtt{D2} \cap \mathtt{COL1} \cap \mathtt{TPCL}\} \cap \mathtt{COL2}(i) \cap \mathtt{COL3}) \leq \exp(-\Omega(an/\log n))$.

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

Hence,

$$\begin{aligned}
&\mathbb{P}(H_i^+ \in \overline{\mathtt{LC}} \cap (G \in \mathtt{D2} \cap \mathtt{COL1} \cap \mathtt{TPCL}) \cap \mathtt{COL2}(i) \cap \mathtt{COL3}) \\
&\quad \leq \mathbb{P}(H_i^+ \in \overline{\mathtt{LC}} \cap \{H_i^{b+} \in \mathtt{CNT} \cap \mathtt{D2}\} \cap \{(H_i^{b+}, V(A_i)) \in \mathtt{EXPN}^+\} \cap \mathtt{COL2}(i) \cap \mathtt{COL3}) \\
&\quad \leq \mathbb{P}(\mathtt{COL2}(i) \cap \mathtt{COL3} \mid \{H_i^{b+} \in \mathtt{CNT} \cap \mathtt{D2}\} \cap (H_i^{b+}, V(A_i)) \in \mathtt{EXPN}^+)
\end{aligned}$$

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 $\mathrm{End}(v_0)$ be the set of admissible paths obtained by doing admissible Pósa rotations on P. We first show that

Claim 21. $|N_{H_i^{b+}}(End(v_0)) \setminus V(A_i)| < 2|End(v_0)|$.

Proof. The proof is similar to the standard Pósa rotation argument. Consider any $y \in \mathrm{End}(v_0)$ and the path P' obtained via a Pósa rotation when y is added to $\mathrm{End}(v_0)$. 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 $\mathrm{End}(v_0)$ 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 $\mathrm{End}(v_0)$ before y. Hence, either $\{v_{i-1}, v_{i+1}\} \cap \mathrm{End}(v_0) \neq \emptyset$ or $v_i \in V(A_i)$. Our claim follows immediately. □

By (T1') and Claim 21, we have that $|\mathrm{End}(v_0)| = \Omega(n)$. Take any $x \in \mathrm{End}(v_0)$, consider $\mathrm{End}(x)$, the set of longest admissible paths starting from x by performing admissible Pósa rotations. Then we also have $|\mathrm{End}(x)| = \Omega(n)$ for every $x \in \mathrm{End}(v_0)$. If any of the edges in $E(H_i^+ \setminus H_i^{b+})$ belongs to the set $\mathscr{E} := \{xy : x \in \mathrm{End}(v_0),\ y \in \mathrm{End}(x)\}$, 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

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

completing the proof. □

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

Our strategy in this case is to produce a random bipartite graph and apply the methodology of [2]. We will define a 3-round edge exposure of $\mathcal{G}(\mathbf{n}, p, q)$. Let $\bar{q} = a/n \log n$ where $a \to \infty$ and $a = o(\log n)$. Define

$$q_1 = 1 - \frac{1-q}{(1-\bar{q})^2} \geq \frac{(1+\eta)\log n}{n},$$

where η is a positive constant. The inequality above follows by (A1), (A4) and the assumption that $p = O(1/n)$.

Let $\widehat{G} \sim \mathcal{G}(\mathbf{n}, p, q_1)$ and let $E_1 = E(\widehat{G})$. For every crossing $uv \notin E_1$, add uv with probability $\bar{q}$ and colour uv yellow. Let E_y be the set of yellow edges. Finally, for every crossing $uv \notin E_1 \cup E_y$, add uv with probability $\bar{q}$ and colour uv red. Let E_r be the set of red edges. The graph with edge set $E_1 \cup E_y \cup E_r$ has the distribution $\mathcal{G}(n, p, q)$.

8.1 G_b

Assume that $\widehat{G} \in \mathtt{D2}$ (note that $\mathbb{P}(\mathcal{G}(\mathbf{n}, p, q_1) \in \mathtt{D2}) = \mathbb{P}(\mathcal{G}(\mathbf{n}, p, q) \in \mathtt{D2}) + o(1)$). We will create a bipartite graph G_b which is a subgraph of $\widehat{G}$ except for a few additional *golden* edges which we will denote by $\mathfrak{G}$. We will also denote by $\mathfrak{G}_V$ the set of endpoints of edges in $\mathfrak{G}$. Given a graph G and two disjoint subsets of vertices U and V let $G[U, V]$ denote the subgraph of G on $V(G)$ with the set of edges in G that join vertices in U to vertices in V.

Property $\mathtt{EXPN}$ is modified to state the following: in equation (7) below, A, B is a partition of $V(G)$.

$$\text{For all } S \subseteq A \text{ (resp. } B) \text{ where } |S| \leq 0.24n,\ |N_G(S) \setminus \mathfrak{G}_V| \geq 2|S| 1_{n_1(G)=0}. \tag{7}$$

Property $\mathtt{SEXPN}$ is defined as before.

We will define a bipartite graph G_b that has the following property a.a.s.

Property $\mathcal{P}$: (a) $V(G_b) \subseteq [n]$, $|V(G_b)| = n - O(n^{1/2})$ and $|V(G_b)|$ is even; (b) G_b has a set $\mathfrak{G}$ of golden edges and $|\mathfrak{G}| = O(n^{1/2})$; (c) a Hamilton cycle in $G_b \cup E_y \cup E_r$ covering $\mathfrak{G}$ can be modified to create a Hamilton cycle in $\widehat{G} \cup E_y \cup E_r$; (d) there is a bipartition (A,B) of $V(G_b)$ and a block X satisfying the following conditions: (d1) $|A| = |B|$; (d2) $\min\{V_j \cap A, V_j \cap B\} < n/\log^2 n$ for each block $V_j \neq X$; (d3) $|X| \leq 0.22n$ if $k \geq 3$, and $X = \emptyset$ if $k = 2$; (d4) $G_b = G_b[A, B]$, and if $\widehat{G} \in \mathtt{D2}$ then $G_b \in \mathtt{D2} \cap \mathtt{EXPN} \cap \mathtt{CNT}$.

A partition $(V_1, \ldots, V_k)$ of V is said to be *special* if the smallest part has size greater than $0.22n$. We will discuss the special case in Section 10. So assume for now that we have Property $\mathcal{P}$.

Claim 22. *Assume that $(V_1, \ldots, V_k)$ is not special. Then, a.a.s. there exists a graph G_b with Property $\mathcal{P}$ such that $(G_b \setminus \mathfrak{G}) \subseteq \widehat{G}$.*

We postpone the proof of Claim 22 until later. Let G_b be the graph from Claim 22 and let X be the block that satisfies condition (d) of Property $\mathcal{P}$. Let $G = G_b \cup E_y \cup E_r$. We will show the existence of a Hamilton cycle of G covering $\mathfrak{G}$. Then $\mathcal{G}(n,p,q) \in \mathtt{HAM}$ follows by Property $\mathcal{P}$ as $\widehat{G} \cup E_y \cup E_r$ has the same distribution as $\mathcal{G}(n,p,q)$.

As in Section 7, we will restrict our attention to admissible paths and admissible Pósa rotations with respect to golden edges. That is, a path is admissible if for all $e \in \mathfrak{G}$ it uses e or contains neither of the ends of e. A Pósa rotation is admissible if it does not delete an edge of $\mathfrak{G}$. We say that $G \in \mathtt{HAM}^+$ if G has a Hamilton cycle containing all golden edges in G. A matching lower bound as in Theorem 1 on $\mathbb{P}(G \in \mathtt{HAM}^+)$ follows by the following two lemmas, as in the previous cases.

Let

$$b_{min} = \min\{|V_i|/n : i \in [k]\} \text{ and let } \gamma = b_{min}^{20}/k. \tag{8}$$

Every γ-large vertex $v \in A$ (and $\in B$ resp.) in G has at least $\gamma \log n$ neighbors in B (and A resp).

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

Lemma 23. *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 24.

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

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 3 where $(V_1, \ldots, V_k)$ is not special follows exactly as in Case 1.

It only remains to prove Lemmas 23 and 24.

8.2 Proof of Lemma 23

The proof is basically the same as that of Lemma 7, except that the probability bounds are different. G is obtained from $\mathcal{G}(\mathbf{n}, p.q)$ by deleting edges with probability $q^* = \Theta(a/\log^2 n)$ where $q(1-q^*) = q_1$. Let $\mathcal{X}$ be as in Lemma 7, for some longest path P of G'. The probability that none of the edges in $\mathcal{X}$ are 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.

8.3 Proof of Lemma 24

Recall the definition of $\mathtt{TPCL}$, $\mathtt{EXPN}$ and $\mathtt{SEXPN}$. Claim 10 continues to hold. Thus, it is sufficient to show

$$\mathbb{P}(\mathcal{B} \mid \{G_b \in \mathtt{CNT} \cap \mathtt{EXPN} \cap \mathtt{D2}\}) \leq \exp(-\Omega(a^2 n/\log^2 n)),$$

where event $\mathcal{B}$ is as defined in (5).

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 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 are in the same block, then the edges exposed in the second stage will not help to extend the paths. The purpose of using several rounds of edge exposure and the *bipartitasion* is to cope with this vertex partition issue. To explain how it works, we need a few definitions.

The proof of finding a Hamilton cycle covering $\mathfrak{G}$ will be an adaptation of [2]. The differences are that we will restrict to admissible paths, admissible Hamilton cycles, and admissible Pósa rotations. A pair $x \in A, y \in B$ is said to be *crossing*. A crossing pair (x, y) is said to be *valid* if x and y are contained in distinct blocks. The other difference from [2] is that we need to take care of invalid crossing pairs. Let $P = (v_0, \dots, v_\ell)$ be a longest admissible path in the graph G_b. For $v \in V(G_b)$ we let $\sigma(v) = 1_{v\in A}$. The main effect of restricting to admissible paths and admissible rotations is to make $\sigma(x) = \sigma(y)$ for all $x, y \in \text{End}(v_0)$.

We replace Lemma 11 by the following:

Lemma 25. *$|N_{G_b}(End(v_0)) \setminus \mathfrak{G}_V| < 2|End(v_0)$, where $\mathfrak{G}_V$ denotes the end points of the edges of $\mathfrak{G}$.*

The proof of the above Lemma is identical to the proof of Claim 21.

Let $\mathscr{P}$ be a list of longest paths obtained as follows. $\mathscr{P} = \mathscr{P}' = \{P\}$ initially and $\text{End}(v_0) = \{v_\ell\}$. Take a path $P' \in \mathscr{P}$ and do an admissible Pósa rotation (that has not been performed yet). If it creates a path whose end other than v_0 is not yet in $\text{End}(v_0)$ then add this new path to both $\mathscr{P}$ and $\mathscr{P}'$, and add its end vertex other than v_0 to $\text{End}(v_0)$. Otherwise add it to $\mathscr{P}'$ only. Since at the end $|\mathscr{P}'| \le n!$ and each path in $\mathscr{P}'$ may yield up to $n - 1$ Pósa rotations, this process will terminate.

Given that $G_b \in \texttt{EXPN}$ we have $|\text{End}(v_0)| \ge 0.24n$.

For each $x \in \text{End}(v_0)$, let P_x be the path that was added to $\mathscr{P}$ when x was added to $\text{End}(v_0)$. Let $\phi(x)$ be the vertex adjacent to x on P_x. We show next that

$$|\phi^{-1}(z)| \le 2 \text{ for all vertices on } P. \tag{9}$$

Proof of (9)**:** Assume $x \in \text{End}(v_0)$, u and v are the two neighbours of x on P. Then we know u and v cannot be in $\text{End}(v_0)$, since $\sigma(u), \sigma(v) \neq \sigma(x)$, the two edges xu and xv were not deleted until the rotation step where x is added to $\text{End}(v_0)$. Suppose xu was deleted in that step, then $\phi(x) = v$ in the path P_x. If xv was deleted, then $\phi(x) = u$ in P_x. Hence for $z \in V(P)$ we have that $x \in \phi^{-1}(z)$ only if $zx \in E(P)$. This completes the proof of (9). □

Let $\mathscr{P}(x)$ denote the set of longest paths obtained by rotating P_x with x being the fixed end, and let $\text{End}(x)$ 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 $x\phi(x)$. Remove from $\text{End}(v_0)$ those $o(n)$ vertices x for which $\phi(x)$ is incident with a golden edge. Let a path P be EVEN if its endpoints x, y satisfy

$\sigma(x) = \sigma(y)$. Otherwise P is ODD. Since G_b is bipartite, σ is constant over $\mathrm{End}(x)$ and so if P is EVEN (resp. ODD) then so are all the paths in $\mathscr{P}(x), x \in \mathrm{END}(v_0)$ and furthermore for all EVEN paths, $\sigma(x) \neq \sigma(\phi(x))$.

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 y of P satisfies $\sigma(y) \neq \sigma(x)$. We will show the following.

Claim 26. *Either all longest paths are ODD or with probability* $1 - \exp(-\Omega(an/\log n))$, *either* $L(G_y) > L(G_b)$ *or* $|\mathtt{Out}| = \Omega(\mathtt{an}/\log \mathtt{n})$ *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 all the paths in $\mathscr{P}$ are ODD 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 $\exp(-\Omega(an/\log n)) \leq \exp(-\Omega(a^2 n/\log^2 n))$ by using the set of red edges. There is one small point. Rotations produced $\Omega(n^2)$ paths with endpoints x, y where $x \in B$ and $y \in \mathrm{End}(x)$. Such a pair $x, y \in \mathrm{End}(x)$ forms a booster unless $x \in X \cap B$ and $y \in X \cap A$. Because of Property $\mathcal{P}$ (d3), $|X \cap A|, |X \cap B| \leq 0.22n$. Thus the number of boosters is now at least $[(0.24 - 0.22)n]^2 = \Omega(n^2)$.

If instead $|\mathtt{Out}| = \Omega(an/\log n)$ in G_y, then we will prove the following claim which, together with Claim 26 and the argument above, completes the proof for Lemma 24.

Claim 27. *If* $|\mathtt{Out}| = \Omega(\mathtt{an}/\log \mathtt{n})$ *then* $\mathcal{B}$ *holds with probability* $\exp(-\Omega(a^2 n/\log^2 n))$.

Proof. Let $x \in \mathtt{Out}$, let P be the corresponding path and we may assume that P is EVEN, and let w be an end on P where $\sigma(x) \neq \sigma(w)$ and let y be the other end of P. Consider admissible Pósa rotations on P with y as the fixed end, and let $\mathrm{End}(y)$ denote the ends obtained. Since P is EVEN, for every $z \in \mathrm{End}(y)$ we have $\sigma(z) \neq \sigma(x)$. This implies that $|\mathscr{E}_x| = \Omega(n)$ where

$$\mathscr{E}_x = \{\{x, z\} : \ z \in \mathrm{End}(y) \setminus X, z \in A\}.$$

Let $\mathscr{E} = \cup_{x \in \mathtt{Out}} \mathscr{E}_x$. Then $|\mathscr{E}| = \Omega(an^2/\log n)$. Moreover, $\mathcal{B}$ fails if $\mathscr{E} \cap E(G \setminus G_y) \neq \emptyset$. Since every booster in $\mathscr{E}$ appears in the final 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 26. Assume that all longest paths are EVEN. Consider all pairs of vertices $\mathscr{E}' = \{(\phi(x), w) : x \in \mathrm{End}(v_0), w \in \mathtt{End}(x) \setminus X\}$. As $\mathtt{TPCL} \cap \mathtt{COL}$ implies $\mathtt{COL}$, it follows, using (9), that $|\mathscr{E}'| \geq |End(v_0)|/2 \times (0.24 - 0.22)n = \Omega(n^2)$. If any yellow edge is a valid pair in $\mathscr{E}'$, then we find a cycle $C_x, |C_x| = \Omega(n)$ by deleting x from the corresponding path P', and then adding the edge $w\phi(x)$. Using the Chernoff bounds we see that we get $\Omega(an/\log n)$ distinct cycles, with the required probability. G_b is connected and because G_b is bipartite there must exist a vertex z such that (a) $\sigma(z) \neq \sigma(x)$; (b) z is not on C_x; (c) there is a path from z to C_x. Using the cycle C_x and the path from z to C_x we distinguish two cases: (i) we obtain a path of length greater than that of P, which implies $L(G_y) > L(G_b)$, or (ii) we obtain a path of the same length as P, with $x \notin P$, and one of the ends of P, namely z, satisfies $\sigma(z) \neq \sigma(x)$. That gives us the required number of vertices in $\mathtt{Out}$. □

8.4 Proof of Claim 22

We first assume $k \geq 3$. The case $k = 2$ is much simpler which will be discussed at the end of this section. We start with a tripartition A, B, C with the following conditions:

(i) C is the smallest block and $|C| \leq 0.22n$, and $|A| \geq |B|$; (ii) if S is a block other than C then $S \subseteq A$ or $S \subseteq B$; (iii) $|C| \geq |A| - |B|$.

Claim 28. *A partition (A, B, C) satisfying the above conditions exists.*

Let (A, B, C) be a partition as in Claim 28. Next, we will construct G_b and partition the vertices in C and assign them to A and B according to $\widehat{G}$. We will then move a small number of vertices from A to B and vice versa.

Recall that $b_{min} = \min\{|V_i|/n : i \in [k]\}$. By (A4) $b_{\min} = \Omega(1)$. Consider $\widehat{G}$ and assume $\widehat{G} \in \mathtt{D2}$. Let SAFE be the set of vertices v in $\widehat{G}$ which have at least $b_{min} \log n/20$ neighbours in every block other than the block containing v. Let VERYSMALL be the set of vertices with degree less than $6/\delta$, where δ is a constant to be specified in the lemma below. Let TWO be the set of vertices with degree two.

Lemma 29. *A.a.s. there exists $0 < \delta = \delta(c) < 1/10$ such that $\widehat{G}$ has the following properties.*

(a) $|[n] \setminus SAFE| < n^{1-\delta}$, $|VERYSMALL| < n^{1/2-\delta}$.

(b) The graph distance between any pair of distinct vertices in VERYSMALL is at least 10.

(c) No vertex in VERYSMALL is in a cycle of length at most 5.

(d) No vertex in $\widehat{G}$ has more than $2/\delta$ neighbors in $[n] \setminus SAFE$.

We will sketch the proof in Section 9.

Now, arbitrarily assign the vertices in C to A and B so that $\big||A| - |B|\big| = 1_{\{\text{n is odd}\}}$. Since initially $|C| \geq |A| - |B|$ this can be done. Next, we reallocate vertices in $[n] \setminus (\text{SAFE} \setminus \text{VERYSMALL})$ so that they all have degree at least $3/\delta - 2/\delta = 1/\delta$ by Lemma 29(d) and the definition of VERYSMALL. Finally we sequentially reallocate vertices in $v \in$ VERYSMALL so that v is put into A if it has more neighbours in B than in A, and put it into B otherwise. After this we have $\big||A| - |B|\big| < n^{1-\delta} + 1$ since we only moved at most $n^{1-\delta}$ vertices by Lemma 29(a). The moves above do not guarantee that the minimum degree of $\widehat{G}[A, B]$ is at least 2. A vertex v will be of degree less than 2 in $\widehat{G}[A, B]$ only when $v \in$ TWO had two neighbours, one in A and the other in B at the time that v was processed. Let $\mathcal{B}_A$ and $\mathcal{B}_B$ denote the sets of these vertices in A and B respectively. By Lemma 29(a), $|\mathcal{B}_A| + |\mathcal{B}_B| < n^{1/2-\delta}$. For each vertex $u \in \mathcal{B}_A \cup \mathcal{B}_B$, let u_x and u_y be the two neighbours of u. Delete u from $\widehat{G}$ and from $A \cup B$, and add a golden edge between u_x and u_y. By this point we have $|A| + |B| = n - \mathcal{B}_A - \mathcal{B}_B$ and $\big||A| - |B|\big| < n^{1-\delta} + 1$ since we only moved at most $n^{1-\delta}$ vertices by Lemma 29(a).

If $n - \mathcal{B}_A - \mathcal{B}_B$ is odd, choose an arbitrary vertex $u \in$ SAFE and delete all edges incident with u except two, one joining u to a vertex u_x in A and the other joining u to a vertex u_y in B. Delete u from the graph and from $A \cup B$ and add a golden edge between u_x and

u_y. By our construction so far, we have a bipartition of the vertices of G_b into A and B; the absolute difference between $|A|$ and $|B|$ is at most $n^{1-\delta}+1+n^{1/2-\delta}$. There are at most $n^{1/2-\delta}$ golden edges, all of which join vertices in A to vertices in B. Now, we move vertices in SAFE that are not incident with any golden edge to make $|A| = |B|$, without making any vertex having less than 2 neighbours on their (possibly new) opposite side. Assume that $|A| \geq |B|$ so that we are looking for vertices to move from A to B. We will avoid VERYSMALL and its neighborhood and we will choose a set of vertices in A such that each vertex has at least 2 neighbors in A and each pair of vertices are at least distance two in G. The latter condition ensures that we do not move any vertices in the neighborhood of a moved vertex. Let $A_{\delta\geq 2}$ be the set of vertices in A that have at least 2 neighbors in A. Also let $VALID = (A_{\delta\geq 2}) \cap \text{SAFE} \setminus (\text{VERYSMALL} \cup N(\text{VERYSMALL}))$. In the case that $|VALID| = \omega(n^{1-\delta}\log^2 n)$ the greedy algorithm will ensure that one can find the corresponding set because the maximum degree in G is $O(\log n)$ a.a.s. The existence of that many vertices is justified as follows:

In the case where $p = \Omega(n^{-1-0.49\delta})$, partition V_1 into two equal sets V_1^1, V_1^2. Then the probability that there do not exist $n^{1-0.985\delta}$ vertices in V_1^1 that are adjacent to 2 vertices in V_1^2 is bounded above by $\mathbb{P}(Bin(\Omega(n), \Omega(n^{-0.98\delta})) \leq n^{1-0.985\delta}) = o(1)$. Thus a.a.s $|VALID| = \Omega(n^{1-0.985\delta}) - O(n^{1-\delta}) = \omega(n^{1-\delta}\log^2 n)$. In the case where $n/2 - |V_1| = \Omega(n^{1-0.99\delta})$, then by construction, there is a block V_j with $j \neq 1$ such that $|V_j \cap A| = \Omega(n^{1-0.99\delta})$. Any vertex in $V_j \cup \text{SAFE}$ is adjacent to $c\log n/20$ vertices in V_1, hence to at least 2 vertices in A and can be moved to A. Lemma 29 (a) implies that $|V_j \cap VALID| = \Omega(n^{1-0.99\delta}) = \omega(n^{1-\delta}\log n)$.

Finally in the case where $p = O(n^{-1-0.49\delta})$ and $n/2 - |V_1| = O(n^{1-0.99\delta})$, (A3) and (A4) imply that there exists $\beta > 0$ such that every block other than V_1 consists of at most $(1/2-\beta)n$ vertices. Hence the minimum in (A1) is due to block V_1 and $q \geq 2(\log n + \log\log n + O(1))/n$ and the expected number of vertices in block V_i , $i \neq 1$ of degree at most 2 is $O(n\binom{n}{2}q^2(1-q)^{n-(1/2-\beta)n}) = O(n(\log^2 n)e^{-(1+2\beta)(\log n+\log\log n+O(1))}) = O((\log^{1-\beta} n)n^{-2\beta})$. Also, the expected number of vertices in V_1 of degree 2 in G that are adjacent to 2 or 1 or 0 vertices in V_1 is $O(n^{-(0.98+o(1))\delta})$, $O(n^{-(0.49+o(1))\delta})$ and $O(\log n)$ respectively (here we have used the fact that such a vertex has at most 0,1,2 neighbors in $V \setminus V_1$ each event occurring with probability $o(n^{-1}), O(n^{-1})$ and $O(\log n/n)$ respectively). Hence in the case $p = O(n^{-1-0.49\delta})$ and $n/2 - |V_1| = O(n^{1-0.99\delta})$ the Markov inequality implies that a.a.s TWO consists of at most $\log^2 n$ vertices and TWO $\subset V_1$. In addition a.a.s. the neighborhood of TWO consists of $2|\text{TWO}| \leq 2\log^2 n$ vertices, is a subset of SAFE and will be assigned to B. Hence a.a.s no golden edges will be created, $|A| = |B|$ and there is no need to move any vertices between the blocks.

Finally, let G_b be the graph obtained by deleting all edges inside A or inside B. Thus, G_b is bipartite, i.e. $G_b = G_b[A, B]$. Our construction guarantees conditions (a),(b),(d1) and (d2) required by property $\mathcal{P}$. Obviously a Hamilton cycle in $G_b \cup E_y \cup E_r$ can be modified to create a Hamilton cycle in $\widehat{G} \cup E_y \cup E_r$ by replacing each golden edge by a 2-path. This verifies condition (c). Condition (d3) is guaranteed by Claim 28. For condition (d4), our construction guarantees $G_b = G_b[A, B]$ and if $\widehat{G} \in \texttt{D2}$ then $G_b \in \texttt{D2}$ and therefore $G \in \texttt{D2}$. The second part of (d4) is verified by the following claim whose proof is given in Section 9.
Claim 30. *If $G \in \texttt{D2}$ then $G \in \texttt{SEXPN}$ a.a.s.*

Proof of Claim 28 Order the $k \geq 3$ blocks so the block sizes are non-increasing. If $|V_1| = |V_2|$ then we will let $C = V_3$. Then $|C| \leq n/3$. To construct A and B, initially let $A = V_1$ and $B = V_2$. Sequentially adding V_j, $j \geq 4$ to the smaller between A and B. It is easy to see that (A, B, C) satisfies conditions (i)–(iii).

If $|V_1| > |V_2|$, let t be the minimum integer such that $\sum_{i=2}^{t} |V_i| \geq |V_1|$. Such a t exists by (A3), and $t \geq 3$. Let $C = V_t$. Then $|C| \leq n/3$. To construct A and B, initially let $A = V_1$ and $B = \cup_{j=2}^{t-1} V_j$. Sequentially adding V_j, $j \geq 4$ to the smaller between A and B. It is easy to see that (A, B, C) satisfies conditions (i)–(iii). □

Finally if $k = 2$ then $|V_1| = |V_2|$ by (A3). Let $A = V_1$ and $B = V_2$. Similar to the case $k \geq 3$ we consider two cases. The first one is $p = O(n^{-1-0.49\delta})$. In this case we can construct G_b and a bipartition of the vertices in G_b by moving $o(n/\log^2 n)$ vertices and deleting $o(n^{1/2})$ vertices as done earlier. The second case is $p = \Omega(n^{-1-0.49\delta})$ (and $|V_1| = n/2 = n/2 - O(n^{1-0.985\delta})$). As shown earlier in this case no further golden edges will be created and no further modifications are needed. The argument of finding an admissible Hamilton cycle in G_b will be exactly the same.

9 Proof of Technical Lemmas

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

Lemma 31. *Assume (A1). A.a.s. $\mathcal{G}(\mathbf{n}, p, q)$ satisfies the following graph properties.*

- $(C0)$ *The maximum degree is $O(\log n)$.*
- $(C1)$ *For some constant $0 < \zeta < 1$, and every $i \in [k]$, at most n^ζ vertices in V_i have degree less than $0.9 \log n$.*
- $(C2)$ *No two vertices with degree less than $\log \log n$ are within distance 10.*
- $(C3)$ *No vertex with degree less than $\log \log n$ is contained in cycles of length at most 5.*
- $(C4)$ $|E(S)| \leq 3|S|$ *for all* $|S| < 4n/\log^2 n$.
- $(C5)$ $|E(S)| \leq (4s^2/b_{min}n) \log n$ *for all* $n/\log^3 n \leq |S| \leq b_{min}^6 n$.

In addition, in the case of Section 8 where $p = O(1/n)$ we have the following:

(C6) For all $S \subseteq A$ (resp. B) satisfying $b_{min}^8 \leq |S| \leq 0.24n$ and all γ-deletable edge sets F $|N_{G-F}(S)| \geq (2 + 10^{-4})|S|$.

9.1 Proof of Lemma 6

Let G be a graph with minimum degree at least 2 that satisfies properties (C0)–(C5). 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

$$\text{every vertex is incident with at most } \log n/100 \text{ red edges.} \tag{10}$$

We first show that $G - F \in \texttt{EXPN}$ a.a.s.

Case a: $n/(\log n)^3 \le |S| \le b_{min}^6 n/3$. 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, $|S \cup U| < 3|S| \le b_{min}^6 n$. By (C5), $S \cup U$ induces at most $(4|S \cup U|^2/b_{min}n)\log n \le (36|S|^2/b_{min}n)\log n$ edges in G. This implies that $|E_2| \le (36|S|^2/b_{min}n)\log n$. By (C1), the total degree of vertices in S is at least $(|S| - n^{\zeta}) \times 0.9\log n \ge 0.85|S|\log n$. On the other hand, by (C5), S induces at most $4(|S|^2/b_{min}n)\log n$ edges. Thus, $|E_1| \ge 0.85|S|\log n - 4(|S|^2/b_{min}n)\log n$. Consequently, $|F \cap E_G(S, \overline{S})| = |E_1| - |E_2| \ge 0.85|S|\log n - 4(|S|^2/b_{min}n)\log n - (36|S|^2/b_{min}n)\log n \ge (0.85 - 120b_{min}^{11})|S|\log n \ge (0.85 - 120 \times 2^{-11})|S|\log n > (1/10)|S|\log n$, contradicting condition (10).

Case b $|S| < n/\log^3 n$. A vertex in G is called *moderately small* if its degree is less than $\log\log n$. Let $\mathcal{X}$ denote the set of moderately 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 have 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 $|\mathcal{N}_{G'}(S)| = |\mathcal{N}_G(S)| \ge 2|S|$.

Case b2: $|\mathcal{Y}| \ge 1$. 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 $\mathcal{Y}$, $\mathcal{Z}_2 = \mathcal{N}_{G'}(S) \setminus \mathcal{Z}_1$ 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,

$$|\mathcal{Y}_1| = |E_{G'}(\mathcal{X}, \mathcal{Y})| = |E_G(\mathcal{X}, \mathcal{Y})| \tag{11}$$

$$|\mathcal{N}_{G'}(\mathcal{X})| = |\mathcal{Z}_1| + |\mathcal{Y}_1| \ge 2|\mathcal{X}|. \tag{12}$$

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, $|\mathcal{N}_{G'}(z) \cap \mathcal{Z}_1| \le 1$ for every $z \in \mathcal{Y}$. Consequently,

$$|E_{G'}(\mathcal{Y}, \mathcal{Z}_1)| \le |\mathcal{Y}|. \tag{13}$$

Assume to the contrary that $|\mathcal{Z}_1| + |\mathcal{Z}_2| = |\mathcal{N}_{G'}(S)| < 2|S| = 2(|\mathcal{X}| + |\mathcal{Y}|)$. Then, by (12) we have

$$|\mathcal{Z}_2| < 2|\mathcal{Y}| + |\mathcal{Y}_1|. \tag{14}$$

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

$$|E_{G'}(\mathcal{X}, \mathcal{Y})| + 2|E_{G'}(\mathcal{Y})| + |E_{G'}(\mathcal{Y}, \mathcal{Z}_1)| + |E_{G'}(\mathcal{Y}, \mathcal{Z}_2)| \ge |\mathcal{Y}|\log\log n. \tag{15}$$

By (C0), (C4) and $|S| \le n/\log^3 n$, $|E_{G'}(\mathcal{Y})| \le 3|\mathcal{Y}|$, $|E_{G'}(\mathcal{Y}, \mathcal{Z}_2)| \le |E_{G'}(\mathcal{Y} \cup \mathcal{Z}_2)| \le 3(|\mathcal{Y}| + |\mathcal{Z}_2|)$. We have shown that $|E_{G'}(\mathcal{X}, \mathcal{Y})| \le |\mathcal{Y}_1|$ and $|E_{G'}(\mathcal{Y}, \mathcal{Z}_1)| \le |\mathcal{Y}|$. Hence, by (14), the left hand side of (15) is at most $16|\mathcal{Y}| + 4|\mathcal{Y}_1| \leq 20|\mathcal{Y}|$, whereas the right hand side is $|\mathcal{Y}| \log\log n$, contradiction. This confirms that $|\mathcal{N}_{G'}(S)| \geq 2|S|$.

We will now show that if $p, q = \omega(1/n)$ then a.a.s. $G - F \in \texttt{CNT}$. $G - F \in \texttt{EXPN}$ implies that that every connected component of $G - F$ has size at least ϵn, where $\epsilon = b_{min}^6/3$. Let C be the smallest connected component in $G - F$. If $G - F \notin \texttt{CNT}$ then $|C| \in [\epsilon n, 0.5n]$.

Case 1: Assume first that $b_{min} np \leq 0.5 \log n$. The number of edges spanned by C and not contained in a block is equal to a binomial random variable with mean at least $\frac{1}{2}\sum_{i=1}^k |C \cap V_i|(n - n_i)q = \Omega(n \log n)$. The Chernoff bounds imply that a.a.s C spans

$$\frac{1}{2}\sum_{i=1}^k |C \cap V_i|(n - n_i) \times 0.99q - (\gamma|C| \log n)/100 \geq$$

$$0.49 \sum_{i=1}^k |C \cap V_i|(\log n - 0.5 \log n) - (\gamma|C| \log n)/100 \geq 0.49|C|(\log n - np) - (\gamma|C| \log n)/100$$

$$\geq 0.24|C| \log n - (\gamma|C| \log n/100) \geq 0.23|C| \log n$$

edges in $G - F$ that are not contained in any block. (Recall that $\gamma \leq 1$, see (8).) A calculation similar to the one done to prove Lemma 31 (C5) shows that every set S of size $\epsilon n \leq |S| \leq n/10$ spans less than $\binom{|S|}{2} \times 1.01q < (1.05|S|^2 \log n)/n \leq 0.105|S| \log n$ edges that are not contained in any block. (Note that (A1) implies that $q \leq (2.1 \log n)/n$.)

Observe from (A1) that $q \geq (0.5 + o(1))(\log n)/n$. Now let $n/10 \leq |C| \leq n/2$. Let $c = |C|$ and $n_j = |V_j|$ and $c_j = |C \cap V_j|$ for $j = 1, 2, \ldots, k$. Now $v \in V_j \setminus C$ has $Bin(c - c_j, q)$ G-neighbors in C for $j = 1, 2, \ldots, k$. Let $J_{0.75} = \{j : c_j \leq 0.75 \min\{c, n_j\}\}$ and suppose that $J_{0.75} \neq \emptyset$. Let $j \in J_{0.75}$. Then $v \in V_j \setminus C$ has at least $Bin(c - c_j, q)$ neighbors in C, where $c - c_j \geq 0.25c \geq n/40$. And so the probability it has fewer than $\log n/100$ G-neighbors in C is at most $n^{-\xi}$ for some $\xi > 0$. Thus the probability that all vertices in $V_j \setminus C$ have fewer than $\log n/100$ G-neighbors in C is at most $n^{-(v_j - c_j)\xi} \leq n^{-0.25\xi v_j} = e^{-\Omega(n \log n)}$. There are at most 2^{kn} choices for the blocks and so, if $J_{0.75} \neq \emptyset$, then a.a.s. there is a vertex not in C with at least $\log n/100$ G-neighbors in C, and so C is not a component in $G - F$ (recall that $\gamma \leq 1$ and every vertex is incident to at most $\gamma/100 \log n$ neighbors in F).

To complete this case, we must show that $J_{0.75} \neq \emptyset$. Now $J_{0.75} = \emptyset$ means that $c_j \leq 0.75c$ implies that $c_j \geq 0.75n_j$. So, if $c_j \leq 0.75c$ for all j then $c = c_1 + \cdots + c_k \geq 0.75(n_1 + \cdots + n_k) = 0.75n$, which is a contradiction. So, assume that $c_1 > 0.75c > 0.25c > c_2 + \cdots + c_k$ and $c_j \geq 0.75n_j$ for $j \geq 2$. It follows that $0.25c \geq 0.75\sum_{i=2}^k n_i \geq 0.75(n - n_1) \geq 0.75n/2 > 0.25c$, which is a contradiction.

Case 2: Now assume that $b_{min} np \geq 0.5 \log n$. Fix a block V_j. Let $large_j$ be the set of vertices in V_j that have at least $0.4 \log n$ G-neighbors in V_j. A first moment argument implies that a.a.s $large_j = (1 + o(1))|V_j|$. An argument similar to the one used to show that $G - F \in \texttt{EXPN}$ implies that in $G - F$ no vertex in $large_j$ lies in a component of size smaller than ϵn. Fix disjoint sets $S, T \subseteq V_j$ let $\mathcal{E}_j(S, T)$ denote the event that there are at most $|S|\,|T|p/2$ edges between S and T. Then, where $m = b_{min} n/3$,

$$\Pr(\exists |S| = \Omega(n), |T| = m : \mathcal{E}_j(S, T)) \leq 2^{2n} e^{-|S|mp/8} = o(1).$$

Now F contains at most $(|S|\log n)/100 \leq |S||T|p/10$ edges between such a pair S, T. It follows that after deletion of F, each block contains a component C_j of size $(1+o(1))n_j$. A.a.s. there are at most $5n_in_jq/4$ crossing edges between V_i and V_j and at least $3n_in_jq/4$ crossing edges between $large_i$ and $large_j$. Because F deletes at most half of the edges between a pair of blocks we see that $G-F$ contains a component containing $\cup_{i\in k} large_i$, hence of size $n-o(n)$. Since $G-F$ dose not have a component of size smaller than ϵn we have that $G-F \in \texttt{CNT}$. □

9.2 Proof of Lemma 31

The proof is standard and straightforward. We give a sketch only and omit the somewhat tedious calculations. For (C0) we use the first moment method. 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 $p, q \lesssim 2(\log n)/b_{min}n$. Using this, the expected number of S with $|S| = s$ which induce more than $3s$ edges is at most

$$\binom{n}{s}\binom{\binom{s}{2}}{3s}\left(\frac{(1+o(1))\log n}{b_{min}n}\right)^{3s} \leq \left(\frac{(2+o(1))^3 e^4 (\log^3 n)s^2}{27 b_{min} n^2}\right)^{s}.$$

It is straightforward to see that summing the above over $s < 4n/\log^2 n$ yields $o(1)$. This proves (C4). The proof for (C5) is the same.

For (C2), we bound the expected number of ℓ-paths, $\ell \leq 10$, where the ends are vertices of degree less than $\log\log n$. 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 $\log\log n$ is at most $n^{o(1)-2}$. 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 $n^{\ell+1}\cdot n^{o(1)-2}\cdot((C\log n)/n)^{\ell} = o(1)$. This proves (C2). The proof of (C3) is similar.

For (C6), fix $S \subseteq A$ such that $b_{min}^8 n \leq |S| \leq 0.24n$. Either there exists some i such that $|V_i \cap S| \geq b_{min}^8 n/10^3 k$ and $V_i \neq X$ or $|S \cap X| \geq (1-10^{-3})|S|$. We thus consider 2 cases.

Case 1: $|V_i \cap S| \geq b_{min}^8 n/10^3 k$ for some $V_i \neq X$.

Let S' be a subset of S of size $b_{min}^8 n/10^3 k$ and S^* be the set of vertices in $\widehat{G}$ that have at least $\gamma \log n/100 + 2$ (recall $\gamma = b_{min}^{20}/k$) neighbors in S'.

$$\mathbb{P}(|S^*| \leq 0.999(n-|V_i|)) \leq \binom{n}{b_{min}^8 n/10^3 k}\binom{n}{0.001n}(e^{-\Omega(\log n)})^{cn} \leq 2^{2n} e^{-\Omega(n\log n)} = o(1).$$

Thus a.a.s. $|S^*| \geq 0.999(n-|V_i|)$. Now observe that $o(n)$ of the vertices in $B \cap S^*$ may have been moved to A and each of the rest of the vertices may have a single neighbor from A moved to B (see construction of G_b). After removing F from G each of the $(0.499-o(1))n$

remaining vertices in $S^* \cap B$ has at least $\gamma \log n/100+2-(\gamma \log n/100+1)$ neighbors in $S' \cap A$. Hence $|N_{G-F}(S')| \geq (1+o(1))|S^* \cap B| \geq (0.499-o(1))n \geq (2+10^{-4}) \times 0.24n \geq (2+10^{-4})|S|$.

Case 2: $(1-10^{-3})|S| \leq |S \cap X|$. The probability that $b \in B \setminus X$ has at least $(\gamma \log n)/50$ neighbors in $S \cap X$ is $1-O(n^{-c})$ for some constant $c > 0$. Hence, we have that a.a.s., after removing F, $|N(S \cap X)| \geq 0.999|B \setminus X|$. $S \subset A$ implies that

$$|N(S)| \geq 0.999|B \setminus X| = 0.999(|B| - |X \cap B|) = 0.999(|B| - |X| + |X \cap A|) \geq \\ 0.999 \times n/2 - 0.22n + (1-10^{-3})|S| \geq (2+10^{-4})|S|.$$

In the last inequality we used the fact that $|S| \leq 0.22n$. □

9.3 Proof of Lemma 29

(b) and (c) follow from (C2), (C3) and the remaining properties are also first moment calculations.

9.4 Proof of Lemma 17

There are three types of vertices in $V(A_i)$: (a) they are the neighbours of some vertex in V_i whose degree in $G[V_i]$ equals 1; (b) they are the neighbours of some vertex V_j (for some $j \neq i$) whose degree in $G[V_j]$ is at most 1; (c) they are ends of some crossing edge recoloured from blue to green. Vertices of the first two types are neighbours of vertices with degree at most 1, and vertices of the last type are random vertices in V_i. It is standard argument to show $|V(A_i)| \leq \log n$, and no two vertices in $V(A_i)$ share a common neighbour, and no vertex in $V(A_i)$ is adjacent to a vertex of degree at most 2. This confirms that $V(A_i)$ satisfies the required properties. 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 6 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 $\psi > 0$ be the constant in (C5). For S where $n/(\log n)^2 \leq |S| \leq (\psi/4)n$, 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, $|S \cup U| < 3|S| + |V| \leq 4|S| \leq \psi n$. Now with the same proof as in Lemma 6, we can lead to a contradiction with condition (10). Thus, we must have $|U \setminus V| \geq 2|S|$ in this case.

For S where $|S| \leq n/\log^2 n$, 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 (11)–(13), together with the fact that every vertex in $\mathcal{X}$ has at least 2 non-red neighbours, we have

$$|\mathcal{Y}_1| = |E_{H'_i}(\mathcal{X}, \mathcal{Y})| = |E_{H_i}(\mathcal{X}, \mathcal{Y})| \tag{16}$$

$$|\mathcal{N}_{H_i}(\mathcal{X}) \setminus V| = |\mathcal{Z}'_1| + |\mathcal{Y}_1| \geq 2|\mathcal{X}| \tag{17}$$

$$|E_{H'_i}(\mathcal{Y}, \mathcal{Z}_1)| \leq |\mathcal{Y}|. \tag{18}$$

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 (17) we have

$$|\mathcal{Z}'_2| < 2|\mathcal{Y}| + |\mathcal{Y}_1|. \tag{19}$$

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

$$|\mathcal{Z}_2| \leq |\mathcal{Z}'_2| + |\mathcal{Y}|. \tag{20}$$

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,

$$|E_{H'_i}(\mathcal{X}, \mathcal{Y})| + 2|E_{H'_i}(\mathcal{Y})| + |E_{H'_i}(\mathcal{Y}, \mathcal{Z}_1)| + |E_{H'_i}(\mathcal{Y}, \mathcal{Z}_2)| \geq 99|\mathcal{Y}|. \tag{21}$$

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 (19) and (20). 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 (15) 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 $|N_{H'_i}(S) \setminus V(A_i)| \geq 2|S|$.

It remains to show that $H'_i \in$ `CNT`. As done in Lemma 6 one may show that each component of H_i has size $\Omega(n)$ and that the largest component of H_i spans $(1 + o(1))|H_i|$ vertices a.a.s. Hence $H'_i \in$ `CNT` a.a.s.

9.5 Proof of Claim 30

Assume that $G \in$ `D2`. Recall $\gamma = b_{min}^{20}/k$. Let $F \subset E(G)$ be a γ-deletable set and let $G' = G - F$. We will first show that $G' \in$ `EXPN` a.a.s. Let $S \subset A$, $|S| \leq 0.24n$. We consider two cases:

Case 1: $|S \setminus SAFE| \geq |S|/\log\log n$. Lemma 29 (a) implies that $|S| \leq \frac{n}{\log^4 n}$. Let $S_1 = S \cap \text{VERYSMALL}$ and $S_2 = S \setminus S_1$. Lemma 29 implies that (i) $|N(S_1) \setminus \mathfrak{G}_V| \geq 2|S_1|$, (ii) each vertex in S_2 has at most a single neighbor in $N(S_1)\cup\mathfrak{G}_V$ hence $|N(S_2)\cap(N(S_1)\cup\mathfrak{G}_V)| \leq |S_2|$. In addition $S_2 \cup N(S_2)$ spans at least $(6/\delta)|S_2|/2 \leq 30|S_2|$ edges (recall $\delta < 1/10$). Since $|S_2| \leq |S| \leq n/\log^4 n$, Lemma 31 (C4) implies that no set of size $10|S_2|$ spans $30|S_2|$ edges and therefore $|N(S_2)| \geq 10|S_2|$. Thus $|N(S)\setminus\mathfrak{G}_V| = |N(S_1)|+10|N(S_2)|-|N(S_2)\cap(N(S_1)\cup \mathfrak{G}_V)| \geq 2|S_1|+10|S_2|-|S_2| = 2|S|+8|S_2| \geq 2|S|$.

Case 2: $|S \setminus SAFE| \leq |S|/\log\log n$. If $|S| \leq \frac{n}{\log^4 n}$ then the above argument implies that $|N(S)\setminus\mathfrak{G}_V| \geq 2|S|$. Hence we may assume that $|S| \geq \frac{n}{\log^4 n}$ and $|S\cap SAFE| = (1-o(1))|S|$. Let $S' = S \setminus SAFE$. We will show that if $|S'| \leq 0.24n$ then $|N(S')| \geq (2+10^{-4})|S'|$. Hence, $|N(S) \setminus \mathfrak{G}_V| \geq |N(S')| - |\mathfrak{G}_V| = |N(S')| - o(S') \geq (2+10^{-4}-o(1))|S'| \geq 2|S|$. At the first equality we used that Lemma 29 (a) implies that $|\mathfrak{G}_V| = o(|S|) = o(|S'|)$. As before Lemma 29 implies each vertex in S' has at most one neighbor in $\mathfrak{G}_V$. Hence $S' \cup N(S')$ span at least $(b_{min}\log n/30)|S'|$ edges in G'. Lemma 31 (C4) implies that if $|S'| < n/\log^2 n$ then $|N(S')| \geq 3|S'|$. Thereafter if $n/\log^2 n \leq |S'| \leq b_{min}^8 n$ Lemma 31 (C5) implies that any set of size at most $4|S|$ spans at most $(4|4S|^2/n)\log n \leq 64b_{min}^{16}|S|\log n < 64\times 2^{-15}\times b_{min}|S|\log n/30 < b_{min}|S|\log n/30$ edges, hence $N(S') \geq 3|S'|$. Finally 31 (C5) implies that if $|S| \geq b_{min}^8 n$ then $|N(S')| \geq (2+10^{-4})|S'|$.

Now we will show that if $G' \in \texttt{EXPN}$ then $G' \in \texttt{CNT}$. Assume that $G' \in \texttt{EXPN}$. In the event that $G' \notin \texttt{CNT}$ there exists a bipartition of A (and B respectively) into A_1, A_2 (and B_1, B_2 resp.) such that there are no edges from A_1 to B_2 and from A_2 to B_1. However by considering a subset of A_i (and $N(A_i)$ resp.) of size $\max\{|A_i|, 0.24\}$ (and $\max\{|N(A_i)|, 0.24\}$ resp.) $G' \in \texttt{EXPN}$ and $N(N(A_i)) = A_i$ for $i = 1,2$ imply that $|A_1|, |A_2| \geq 2\times 0.24n$ which gives a contradiction. □

10 Special case

In the special case we have $k = 3$ or 4 and $0.22n \leq |V_i|$ for $i \in [k]$.

First assume that $k = 3$ and $0.22n \leq |V_3| \leq |V_2| \leq |V_1| \leq 0.5$. Let V_1' (and V_2' respectively) be random subsets of V_1 (V_2 resp.) of size $|V_1| - |V_2| + \frac{|V_3|-(|V_1|-|V_2|)}{2}$ (and $\frac{|V_3|-(|V_1|-|V_2|)}{2}$ resp.).

Consider the graphs $G_1 = \widehat{G}[V_1 \setminus V_1', V_2 \setminus V_2']$ and $G_2 = \widehat{G}[V_3, V_1' \cup V_2']$. In this case in addition to E_y, E_r we will need a fourth set of random edges E_f in which every edge not in $E_1\cup E_y\cup E_r$ belongs with probability $\Theta(\alpha/n\log n)$, where $\alpha \to \infty$ slowly. Call these the fuchsia edges. We let SAFE be the set of vertices that have at least $\log n/100$ neighbors in each of the 4 sets of vertices that are defined by the bipartitions of the 2 graphs except from the block that they belong to. As done earlier we can move $o(n)$ vertices around in order to ensure that (i) each vertex has at least 2 neighbors in its own graph, (ii) no vertex of degree larger than 2 in $\widehat{G}$ will contribute to the creation of the golden edges and (iii) after considering the golden edges the partition defined by each of the two bipartite graphs is an equi-bipartition. Then using the same tools as earlier we can show that after revealing $E_y \cup E_r$, both of the two resulting graphs are Hamilonian. Notice that we have avoided the special case in both graphs.

Let H_1, H_2 be Hamilton cycles of G_1 and G_2 respectively. Now observe that for every $a \in V_1 \cap V(G_1)$, $b \in V_2 \cap V(G_1)$, $c \in V_3 \cap V(G_2)$ and $d \in V(G_2)$ such that $ab \in E(H_1), cd \in E(H_2)$ if $ad \in E(\widehat{G})$ ($bd \in E(\widehat{G})$ resp.) and ac (bc resp) $\in E_f$ then $G_b \cup E_y \cup E_r \cup E_f$ is Hamiltonian. Now a.a.s there will be $\Omega(n \log n)$ edges ad such that (i) $a \in$ SAFE and (ii) a is at distance at least two from any vertex that was "reshuffled" in the process of creating H_1, H_2. The latest condition implies that the neighbors of a and d respectively on H_1 and H_2 resp. lie in V_2 and V_3 resp. Thus a.a.s there exists a set of $\Omega(n \log n)$ edges which if they belong to $E_1 \cup E_y \cup E_r \cup E_f$ then our we get the requested Hamiltonicity. Since $\alpha \to \infty$, the latest event occurs with probability $\mathbb{P}(Bin(\Omega(n \log n), \Theta(\alpha/n \log n)) \geq 1) = 1 - o(1)$.

In the case $k = 4$ we have $0.22n \leq |V_4| \leq |V_3| \leq |V_2| \leq |V_1| \leq 0.36n$. We then let $U_1 = U_3 \cup U_4$ (hence $|U_1| \geq 0.44n \geq |V_1|$), $U_2 = V_1$ and $U_3 = V_2$. We then repeat the above construction with U_1, U_2, U_3 in place of V_1, V_2 and V_3 resp.

11 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 symmetric 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.