

Cyclic Subsets in Regular Dirac Graphs

Nemanja Draganic^{1,*}, Peter Keevash² and Alp Müyesser³

¹Mathematical Institute, University of Oxford, UK

²Mathematical Institute, University of Oxford, UK

³New College, University of Oxford, UK

*Correspondence to be sent to: e-mail: nemanja.draganic@maths.ox.ac.uk

Communicated by Prof. Benny Sudakov

In 1996, in his last paper, Erdős asked the following question that he formulated together with Faudree: is there a positive c such that any $(n + 1)$ -regular graph G on $2n$ vertices contains at least $c2^{2n}$ distinct vertex-subsets S that are cyclic, meaning that there is a cycle in G using precisely the vertices in S . We answer this question in the affirmative in a strong form by proving the following exact result: if n is sufficiently large and G minimises the number of cyclic subsets then G is obtained from the complete bipartite graph $K_{n-1, n+1}$ by adding a 2-factor (a spanning collection of vertex-disjoint cycles) within the part of size $n + 1$. In particular, for n large, this implies that the optimal c in the problem is precisely $1/2$.

1 Introduction

An important recent theme in graph theory has been the study of when certain classical theorems hold in a “robust” or “resilient” way, according to various possible interpretations of these terms. While many classical results can be interpreted as part of this direction of research, it was first highlighted as a topic for systematic study by Sudakov and Vu [32]. The survey by Sudakov [31] discusses many types of robustness, primarily illustrated by the Hamiltonicity problem, which will also be our focus in this paper. Here the fundamental theorem is that of Dirac [9], stating that any graph G on n vertices with minimum degree $\delta(G) \geq n/2$ (commonly referred to as a *Dirac graph*) contains a cycle that is Hamiltonian (meaning that it uses all vertices of G).

There are many further results showing that Dirac graphs are Hamiltonian in a deeper and more robust sense, in strong contrast to graphs of lower minimum degree, which may not even be connected. These stronger theorems include showing that Dirac graphs have many Hamiltonian cycles [8, 30] and remain Hamiltonian even after passing to a p -random subgraph with $p \geq \Omega(\frac{\log n}{n})$; see [23]. Further results on resilience can be found in [3, 28] and on pancyclicity in [5, 10, 22, 27]. Another type of robustness considers decomposing G into Hamiltonian cycles [7, 25] or covering G by Hamiltonian cycles [11, 13, 16, 19, 24].

Here we consider another natural measure of robustness: are many induced subgraphs Hamiltonian? This problem was posed by Erdős and Faudree [12] (see also [4, Problem 622]), who asked whether in any regular Dirac graph, a constant fraction of vertex subsets are cyclic, meaning that they induce a Hamiltonian subgraph. We let $\text{Cyc}(G)$ count cyclic subsets in G and formulate the statement as a conjecture (they ask, “Is it true?”).

Conjecture 1.1. Any $(n + 1)$ -regular graph G on $2n$ vertices has $\text{Cyc}(G) \geq c2^{2n}$ for some absolute $c > 0$.

Received: March 19, 2025. Revised: June 23, 2025. Accepted: June 27, 2025

© The Author(s) 2025. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

We make the following remarks on this conjecture:

- (1) One cannot replace $n + 1$ by n , as if $G = K_{n,n}$ is a balanced complete bipartite graph then almost all subsets of $V(G)$ induce an unbalanced complete bipartite graph, which is clearly not Hamiltonian.
- (2) The regularity assumption cannot be weakened to minimum degree $n + 1$, as can be seen by considering $K_{n,n}$ with two spanning stars added within each part.
- (3) As Erdős notes in [12], one cannot take $\varepsilon > 1/2$, due to the examples $\mathcal{G}_n$ defined below.

Let $\mathcal{G}_n$ be the set of $(n + 1)$ -regular graphs G on $2n$ vertices obtained from the complete bipartite graph $K_{n-1, n+1}$ by adding a 2-factor (a spanning collection of vertex-disjoint cycles) within the part of size $n + 1$. Our main result is that for large n the extremal examples for Conjecture 1.1 belong to $\mathcal{G}_n$.

Theorem 1.2. For any $(n + 1)$ -regular graph G on $2n$ vertices with n sufficiently large we have $\text{Cyc}(G) \geq \min_{G' \in \mathcal{G}_n} \text{Cyc}(G')$.

In particular, this shows that Conjecture 1.1 holds with $c = 1/2$ for large n , as for any $G' \in \mathcal{G}_n$ the proportion of subsets that are cyclic is $\frac{1}{2} + \frac{3/2}{\sqrt{n\pi}} + O(n^{-3/2})$ (see Lemma 5.1 below).

Remark (2) above suggests the question of what minimum degree is needed in a graph on n vertices to ensure $\text{Cyc}(G) = \Omega(2^n)$. A similar construction with more stars shows that minimum degree $n/2 + o(\sqrt{n})$ is not sufficient. On the other hand, a well-known degree sequence generalisation of Dirac's theorem due to Chvátal [6] implies that a minimum degree of $n/2 + \Omega(\sqrt{n})$ is sufficient.

Proposition 1.3. Any graph G on n vertices with minimum degree $\delta(G) \geq n/2 + \Omega(\sqrt{n})$ has $\text{Cyc}(G) \geq \Omega(2^n)$.

To deduce Proposition 1.3 from [6] it suffices to observe via Chernoff bounds that for G as in the statement, a positive fraction of all induced subgraphs G' of G have minimum degree at least $0.49|V(G')|$ and at least $|V(G')|/2$ vertices of degree at least $|V(G')|/2$; we omit further details.

Notation. Let G be a graph. We denote its vertex set by $V(G)$, its edge set by $E(G)$, its minimum degree by $\delta(G)$ and its maximum degree by $\Delta(G)$. For $A \subseteq V(G)$ we write $\bar{A} = V(G) \setminus A$. For $v \in V(G)$ we let $d_G(v, A)$ count neighbours of v in A and $\bar{d}(v, A) = |A| - d_G(v, A)$ count non-neighbours of v in A . We denote by $e(G)$ the number of edges in G .

For $A, B \subseteq V(G)$ we write $G[A]$ for the subgraph of G induced by A . We let $e(A) = e(G[A])$ and $e(A, B) = |\{(a, b) \in E(G) \mid a \in A, b \in B\}|$, where edges in $A \cap B$ are counted twice. If A and B are disjoint, we denote by $G[A, B]$ the bipartite subgraph of G induced by (A, B) . We also use $\bar{e}$ for non-edges, for example, writing $\bar{e}(A, B) = |A||B| - e(A, B)$ for disjoint A, B .

Let G be a graph on n vertices. If $A \subseteq V(G)$ with $|A| \in \{\lfloor n/2 \rfloor, \lceil n/2 \rceil\}$ we call A a *half-set*. A cut of G is a partition of $V(G)$ into two parts. If (A, B) is a cut of G where A, B are half-sets, we call (A, B) a *balanced cut*.

For two constants $a, b > 0$ we write $a \ll b$ to indicate that a is small enough as a function of b so that our proofs go through; hierarchies with more parameters are defined similarly. We write $a = o_t(1)$ if $a \rightarrow 0$ as $t \rightarrow \infty$; if t is not specified then n is understood. Our asymptotic notation $O(\cdot)$, $\Omega(\cdot)$, etc., refers to large n . We say that an event E holds with high probability (whp) if $\mathbb{P}(E) = 1 - o(1)$.

Organisation. We start with an overview in Section 2, where we prove Conjecture 1.1 modulo three lemmas. These lemmas are proved in Section 3, which concerns various sufficient conditions for Hamiltonicity in near-Dirac graphs. Section 4 contains the proof of our asymptotic result, Theorem 4.1, which states that Conjecture 1.1 holds with $c = 1/2 - o(1)$. Extending the methods developed in previous sections, including a more delicate analysis, we prove the exact result Theorem 1.2 in Section 5. Section 6 contains some concluding remarks.

2 Overview

The first ingredient of our proof is a well-known classification of Dirac graphs into three cases, which we will call “bi-dense”, “almost two cliques” and “almost bipartite”. Here we use the following lemma of Krivelevich, Lee, and Sudakov [23], which has its origins in the work of Komlós, Sarközy, and Szemerédi [21] (see also [8, 29]). For the first case, we recall that a half-set in a graph is a set containing $\lfloor n/2 \rfloor$ or $\lceil n/2 \rceil$ vertices.

Lemma 2.1 ([23]). Let $\varepsilon \leq \frac{1}{320}$ and $\gamma \leq \frac{1}{10}$ be fixed positive reals such that $\gamma \geq 32\varepsilon$. If n is large enough, then for every graph G on n vertices with minimum degree at least $\frac{n}{2}$, one of the following cases holds:

- (i) **(Bi-dense)** $e(A, B) \geq \varepsilon n^2$ for all half-sets A and B (not necessarily disjoint),
- (ii) **(Almost two cliques)** There exists $A \subseteq V(G)$ of size $\frac{n}{2} \leq |A| \leq (\frac{1}{2} + 16\varepsilon)n$ such that $e(A, \bar{A}) \leq 6\varepsilon n^2$ and the induced subgraphs on both A and $\bar{A}$ have minimum degree at least $\frac{n}{5}$, or
- (iii) **(Almost bipartite)** There exists $A \subseteq V(G)$ of size $\frac{n}{2} \leq |A| \leq (\frac{1}{2} + 16\varepsilon)n$ such that $G[A, \bar{A}]$ has at least $(\frac{1}{4} - 14\varepsilon)n^2$ edges and minimum degree at least $\frac{\gamma}{2}n$. Moreover, either $|A| = \lceil \frac{n}{2} \rceil$, or the induced subgraph $G[A]$ has maximum degree at most γn .

Given this classification, the proof of Conjecture 1.1 will be completed by the following three lemmas, which give the required probability in each of the three cases that a random induced subgraph $G[\frac{1}{2}]$, obtained by including each vertex independently with probability $1/2$, is Hamiltonian, thus proving the following probabilistic reformulation of Conjecture 1.1.

Theorem 2.2. Any $(n+1)$ -regular graph G on $2n$ vertices has $\mathbb{P}(G[\frac{1}{2}] \text{ is Hamiltonian}) > \Omega(1)$.

In the first two cases, Hamiltonicity of $G[\frac{1}{2}]$ holds whp; it is the third case that determines the bound on the probability, as one would expect given the extremal example. The main difficulty of the proof will lie in the third case, and the accuracy of our probability estimates here determine whether we prove Theorem 2.2, or the stronger asymptotic result Theorem 4.1, or the even stronger exact result Theorem 1.2.

Lemma 2.3. Let $\varepsilon > 0$ and G be an $(n+1)$ -regular graph on $2n$ vertices. Suppose that $e_G(A, B) \geq \varepsilon n^2$ for all half-sets A and B . Then whp $G[\frac{1}{2}]$ is Hamiltonian.

Lemma 2.4. Let $n^{-1} \ll \varepsilon \ll 1$ and G be an $(n+1)$ -regular graph on $2n$ vertices. Suppose there is some $A \subseteq V(G)$ with $n \leq |A| \leq (1 + 32\varepsilon)n$ such that $e(A, \bar{A}) \leq 24\varepsilon n^2$ and $G[A], G[\bar{A}]$ both have minimum degree at least $2n/5$. Then whp $G[\frac{1}{2}]$ is Hamiltonian.

Lemma 2.5. Let $n^{-1} \ll \varepsilon \ll \gamma \ll 1$ and G be an $(n+1)$ -regular graph on $2n$ vertices. Suppose there is some $A \subseteq V(G)$ with $n \leq |A| \leq (1 + 32\varepsilon)n$ such that $G[A, \bar{A}]$ has at least $(1 - 56\varepsilon)n^2$ edges and minimum degree at least γn , where if $|A| > n$ then $G[A]$ has maximum degree at most γn . Then $\mathbb{P}[G[\frac{1}{2}] \text{ is Hamiltonian}] \geq 0.01$.

We prove these lemmas in the following section, thus completing the proof of Conjecture 1.1.

3 Conditions for Hamiltonicity

In this section we formulate conditions for Hamiltonicity in each of the three cases from Lemma 2.1. We use these to prove the lemmas stated in the previous section, thus completing the proof of Conjecture 1.1.

3.1 Bi-dense graphs

In this subsection we prove Lemma 2.3, which establishes Theorem 2.2 for bi-dense graphs. This will follow from the following stability result for Dirac's theorem, which is [20, Lemma 11].

Theorem 3.1 (Stability for Dirac's theorem). For any $\varepsilon > 0$, if n is large and G is a graph on n vertices with $\delta(G) \geq (\frac{1}{2} - \varepsilon)n$ then one of the following holds: (i) G is Hamiltonian; (ii) $e(G[A]) = 0$ for some $A \subseteq V(G)$ with $|A| = (1/2 - \varepsilon)n$; or (iii) $e_G(A, B) \leq n$ for some disjoint $A, B \subseteq V(G)$ with $|A| = |B| = (1/2 - \varepsilon)n$.

We remark that the source of this result appears not to be widely known (it is sometimes considered folklore) and that a nice short proof is given in the appendix of [20], using the following ideas. Supposing

that G is not Hamiltonian, one shows that the longest cycle C_1 has length $\sim n$ or $\sim n/2$, and in the latter case there is a disjoint cycle C_2 of length $\sim n/2$. In the first case, a set A as in (ii) is obtained by the standard Chvátal–Erdős/Pósa argument: consider any vertex not in C_1 and its shifted neighbourhood on C_1 . In the second case, by maximality of C_1 there can be at most n edges from C_1 to C_2 .

The idea of the proof of Lemma 2.3 is that whp $G[\frac{1}{2}]$ satisfies the hypotheses of Theorem 3.1 and does not satisfy conclusion (ii) or (iii), so satisfies conclusion (i), meaning it is Hamiltonian. The hypotheses are easy consequences of the following well-known Chernoff bound (see, e.g., [17, Theorem 2.1]), which will be used throughout the paper.

Theorem 3.2 (Chernoff bound). Let X be a binomial random variable and $\delta > 0$. Then

$$\mathbb{P}[|X - \mathbb{E}X| \geq \delta \mathbb{E}X] \leq 2 \exp(-\delta^2 \mathbb{E}X / (2 + \delta)).$$

To consider the conclusions of Theorem 3.1 for $G[\frac{1}{2}]$, we will need the standard fact that bidenseness is whp inherited by the random induced subgraph $G[\frac{1}{2}]$. For completeness, we will include a short argument for this in the proof of Lemma 2.3, using the following Frieze–Kannan Regularity Lemma [14] (or one could use the Szemerédi Regularity Lemma, but this would give much worse bounds for n).

Lemma 3.3. For any $\varepsilon > 0$ there are T, n_0 so that for any graph G on $n \geq n_0$ vertices there is a partition $(V_1, \dots, V_t)$ of $V(G)$ with $t \leq T$ and $|V_1| \leq \dots \leq |V_t| \leq |V_1| + 1$ such that for any $A, B \subseteq V(G)$ we have

$$\left| e_G(A, B) - \sum_{i,j} \frac{|A \cap V_i|}{|V_i|} \frac{|B \cap V_j|}{|V_j|} e(G[V_i, V_j]) \right| < \varepsilon n^2.$$

We conclude this subsection by proving Lemma 2.3.

Proof of Lemma 2.3. Let $\varepsilon > 0$ and G be an $(n+1)$ -regular graph on $2n$ vertices. Suppose that $e_G(A, B) \geq \varepsilon n^2$ for all half-sets A and B . Consider $G[\frac{1}{2}] = G[S]$ where S is a random subset of $V(G)$. By Chernoff bounds, whp $|S| = n \pm n^{0.6}$ and every vertex has degree $n/2 \pm n^{0.6}$ in $G[\frac{1}{2}]$.

We intend to apply Theorem 3.1 with $(|S|, 0.1\varepsilon)$ in place of (n, ε) . The assumptions of this theorem are satisfied, so to complete the proof, we need to show that whp $G[S]$ does not satisfy conclusion (ii) or (iii), and so it must satisfy conclusion (i), meaning it is Hamiltonian. To do so, it suffices to show that bidenseness is inherited, in that whp for any $A', B' \subseteq S$ with $|A'|, |B'| \sim n/2$ we have $e_G(A', B') \geq 0.1\varepsilon |S|^2$, say.

To see this, let $(V_1, \dots, V_t)$ be a partition of $V(G)$ obtained from Lemma 3.3 applied with 0.1ε in place of ε . By Chernoff bounds, whp each $|S \cap V_i| = \frac{n}{t} \pm n^{0.6}$. Consider any $A', B' \subseteq S$ with $|A'|, |B'| \sim n/2$. Construct $A, B \subseteq V(G)$ with $|A|, |B| \sim n$ so that each $|A \cap V_i| \sim 2|A' \cap V_i|$ and $|B \cap V_i| \sim 2|B' \cap V_i|$. Then $e_G(A, B) \geq \varepsilon n^2 - o(n^2)$ by bidenseness of G . Applying the conclusion of Lemma 3.3 to (A, B) and (A', B') , we see that $e_G(A, B)$ differs by at most $0.1\varepsilon n^2$ from $\sum_{i,j} \frac{|A \cap V_i|}{|V_i|} \frac{|B \cap V_j|}{|V_j|} e(G[V_i, V_j])$ and $e_G(A', B')$ differs by at most $0.1\varepsilon n^2$ from $\sum_{i,j} \frac{|A' \cap V_i|}{|V_i|} \frac{|B' \cap V_j|}{|V_j|} e(G[V_i, V_j]) \sim \sum /4$. Therefore, $e_G(A', B') \geq 0.1\varepsilon |S|^2$. ■

3.2 Almost two cliques

In this subsection we prove Lemma 2.4, which establishes Theorem 2.2 for graphs that are almost two cliques. We use the following sufficient condition for Hamiltonicity in such graphs.

Lemma 3.4. Let G be a graph on n vertices and (A, B) be a partition of $V(G)$ with $|A|, |B| \geq 0.49n$. Suppose that $G[A, B]$ contains two disjoint edges and $G[A], G[B]$ both have minimum degree at least $n/100$ and at most $10^{-4}n^2$ non-edges. Then G is Hamiltonian.

The proof of Lemma 3.4 uses the following folklore result on Hamilton-connectivity (we include the short proof for completeness).

Lemma 3.5. Let G be a graph on n vertices with minimum degree $\delta(G) \geq n/2 + 1$. Then G is Hamilton-connected, meaning for any distinct $a, b \in V(G)$, there exists a Hamilton path from a to b .

Proof. Consider any $a, b \in V(G)$ and obtain G' from G by deleting a and b . Then $|V(G')| = n - 2$ and $\delta(G') \geq |V(G')|/2$, so by Dirac's theorem G' has a Hamiltonian cycle C . As a and b each have at least $n/2$ neighbours in C , by the pigeonhole principle we can find x, y adjacent on C such that ax and by are edges, thus extending C to a Hamilton path from a to b . ■

Proof of Lemma 3.4. Fix disjoint edges a_1b_1 and a_2b_2 with $a_1, a_2 \in A$ and $b_1, b_2 \in B$. It suffices to find a Hamilton path from a_1 to a_2 in $G[A]$; then by symmetry we also have such a path from b_1 to b_2 in $G[B]$, so G is Hamiltonian. To do so, we consider the set $L_A := \{v \in V(G[A]) : d(v, A) \leq 0.3n\}$ of low degree vertices in A , find two short paths covering L_A , then connect them using Lemma 3.5.

Counting non-edges in A incident to L_A , we have $\frac{1}{2}|L_A|(|A| - 0.3n) \leq 10^{-4}n^2$, so $|L_A| < 2 \cdot 10^{-3}n$. As $G[A]$ has minimum degree at least $n/100$, we can greedily choose a “cherry matching” where for each $x \in L_A \setminus \{a_1, a_2\}$ we choose a path of length 2 centred at x , with all such paths being vertex-disjoint. If a_1 or a_2 is in L_A we also choose a disjoint edge connecting them to a vertex not in L_A .

By definition of L_A , any two vertices of $A \setminus L_A$ have at least $n/20$ common neighbours in $A \setminus L_A$, so we can greedily join all paths chosen so far to form two vertex disjoint paths P_1 from a_1 to some $a'_1 \in A \setminus L_A$ and P_2 from a_2 to some $a'_2 \in A \setminus L_A$ that cover L_A and have total length at most $n/100$.

The induced graph on $A \setminus (V(P_1) \cup V(P_2)) \cup \{a'_1, a'_2\}$ has minimum degree at least $0.29n$, so by Lemma 3.5 has a Hamiltonian path P from a'_1 to a'_2 . Then P_1PP_2 is a Hamiltonian path from a_1 to a_2 , as required. ■

The required two disjoint edges in $G[A, B]$ in Lemma 3.4 will be provided by the following lemma, for which regularity is essential. Given that Lemma 3.4 requires just two edges, a much weaker bound here would suffice, but we include the following nearly optimal bound as it serves as a good warm-up for the upcoming section of the paper.

Lemma 3.6. Let G be an $(n+1)$ -regular graph on $2n$ vertices and (A, B) be a partition of $V(G)$ with $|A|, |B| > \sqrt{n}/100$. Then $G[A, B]$ has a matching of size $\sqrt{n}/100$.

Proof. Let M be a maximal matching in $G[A, B]$. Suppose for contradiction that $|M| < \sqrt{n}/100$. Assume without loss of generality that $|A| \geq |B|$ and note that $|A| \leq n + \sqrt{n}/100$ as otherwise all vertices in B have degree at least $\sqrt{n}/100$ in A , so a matching of size $\sqrt{n}/100$ could be found greedily.

Let $(\tilde{A}, \tilde{B})$ be a balanced cut obtained from (A, B) by moving a set S of at most $\sqrt{n}/100$ vertices. Let C be a minimal vertex cover in the bipartite graph $G[\tilde{A}, \tilde{B}]$. If $|C| > \sqrt{n}/2$, by König's theorem this implies the existence of a matching of the same size, and thus we would have a matching of size $\sqrt{n}/2 - \sqrt{n}/100$ between A and B . Towards a contradiction, suppose that $|C| \leq \sqrt{n}/2$. To simplify notation, we write A, B in place of $\tilde{A}, \tilde{B}$ for the remainder of the proof.

Let $A' = A \cap C$ and $B' = B \cap C$. Assume without loss of generality that $e(A', B \setminus B') = t \geq e(B', A \setminus A')$. Since G is $(n+1)$ -regular, we have that the complement of G is $(n-1)$ -regular, so the number of non-edges

$$\bar{e}(A', A \setminus A') \geq |A'|(n-1) - \bar{e}(A', A' \cup B') - \bar{e}(A', B \setminus B') \geq (n-1)|A'| - n/4 - (|A'|n - t) > t - n/3,$$

since by assumption $|A' \cup B'| \leq \sqrt{n}/2$. This implies $e(A', A \setminus A') < |A'| |A \setminus A'| - t + n/3$, and hence again by regularity and recalling that $e(A \setminus A', B \setminus B') = 0$ (as C is a minimal vertex-cover of the bipartite graph) we have

$$\begin{aligned} e(A \setminus A', B') &\geq |A \setminus A'|(n+1) - e(A \setminus A', A \setminus A') - e(A \setminus A', A') \\ &> |A \setminus A'|(n+1) - (|A \setminus A'| - 1)|A \setminus A'| - (|A'| |A \setminus A'| - t + n/3) \\ &\geq 2|A \setminus A'| + t - n/3 > t, \end{aligned}$$

which contradicts the assumption on the maximality of $e(A', B \setminus B')$, and completes the proof. ■

We conclude this subsection by applying Lemma 3.4 to prove Lemma 2.4.

Proof of Lemma 2.4. Let $n^{-1} \ll \varepsilon \ll 1$ and G be an $(n+1)$ -regular graph on $2n$ vertices. Suppose there is some $A \subseteq V(G)$ with $n \leq |A| \leq (1 + 32\varepsilon)n$ such that $e(A, \bar{A}) \leq 24\varepsilon n^2$ and $G[A], G[\bar{A}]$ both have minimum degree at least $2n/5$. We note that the number of non-edges in $G[A]$ is at most $\binom{|A|}{2} - (|A|(n+1) - 24\varepsilon n^2)/2 < 40\varepsilon n^2 < 10^{-4}n^2$ for sufficiently small ε , and similarly for $G[\bar{A}]$.

Consider $G[\frac{1}{2}] = G[S]$ where S is a random subset of $V(G)$. By Chernoff bounds, whp $|S| = n \pm n^{0.6}$, $|S \cap A| = |A|/2 \pm n^{0.6}$, $|S \cap \bar{A}| = |\bar{A}|/2 \pm n^{0.6}$, and $G[S \cap A]$, $G[S \cap \bar{A}]$ both have minimum degree at least $0.1|S|$. By Lemma 3.6, there is a matching of size $\sqrt{n}/100$ in $G[A, \bar{A}]$, so whp there is a matching of size 2 in $G[S \cap A, S \cap \bar{A}]$ by Chernoff bounds. Finally, bounding the number of non-edges in $G[S \cap A]$, $G[S \cap \bar{A}]$ by those in $G[A]$, $G[\bar{A}]$, we can apply Lemma 3.4 to conclude that $G[S]$ is Hamiltonian. ■

3.3 Almost bipartite

In this subsection we prove Lemma 2.5, which establishes Theorem 2.2 for graphs that are almost bipartite. We use the following sufficient condition for Hamiltonicity in such graphs. For the statement, we require the following definitions, which will be important throughout the paper. A *linear forest* is a collection of vertex-disjoint paths. We say that a cut (X, Y) of a graph H is k -good if $|X| \geq |Y|$ and $H[X]$ has a linear forest with at least $k + |X| - |Y|$ edges, or if $|Y| \geq |X|$ and $H[Y]$ has a linear forest with at least $k + |Y| - |X|$ edges; we call a cut *good* if it is 0-good.

Lemma 3.7. Let $n^{-1} \ll \varepsilon \ll \gamma$ and G be a graph on n vertices in which all vertex degrees are $n/2 \pm n^{0.6}$. Suppose (A, B) is a good cut of G such that $|B| \leq |A| \leq |B| + \varepsilon n$ and $G[A, B]$ has at least $(1/4 - \varepsilon)n^2$ edges and minimum degree at least $\gamma n/3$. Then G has a Hamilton cycle.

The proof of Lemma 3.7 uses the following bipartite analogue of Lemma 3.5 on Hamilton-connectivity (again we include a short proof for completeness).

Lemma 3.8. Let $G = (A, B)$ be a bipartite graph with $|A| = |B| = n/2$ and $\delta(G) \geq n/4 + 1$. Then for any $a \in A$ and $b \in B$ there is a Hamilton path from a to b .

Proof. Consider any $a, b \in V(G)$ and obtain G' from G by deleting a and b . Then G' has a Hamiltonian cycle C by [6, Corollary 1.4]. Fix an order of C and for each $v \in V(G)$ let v^+ denote its successor on C . Then $N_{G'}(b) \setminus \{a\}$ and $\{v^+ : v \in N_{G'}(a) \setminus \{b\}\}$ must intersect, as they are both subsets of $A \setminus \{a\}$ of size at least $n/4$, so we can extend C to a Hamiltonian path from a to b . ■

Proof of Lemma 3.7. Similarly to the proof of Lemma 3.4, our plan will be to find a path P' from some $a' \in A$ to some $b' \in B$ so that $G' = G \setminus (V(P') \setminus \{a', b'\})$ satisfies the conditions of Lemma 3.8. Then, a Hamilton path from a' to b' will complete P' to the required Hamiltonian cycle. The path P' will thus need to cover the low degree vertices and balance the part sizes using the given linear forest in A .

We denote the sets of low degree vertices in each part by $L_A := \{v \in A : d(v, B) \leq 0.3n\}$ and $L_B := \{v \in B : d(v, A) \leq 0.3n\}$. To bound L_A , we note that $|L_A|(n/2 - n^{0.6} - 0.3n) \leq 2e(A) \leq (n/2 + n^{0.6})|A| - e(A, B) \leq (n/2 + n^{0.6})(n/2 + \varepsilon n) - (1/4 - \varepsilon)n^2 < 3\varepsilon n^2$, so $|L_A| < 20\varepsilon n$; similarly, $|L_B| < 20\varepsilon n$.

Since $|A| \geq |B|$ we can fix a linear forest F with exactly $|A| - |B| < \varepsilon n$ edges, which exists as (A, B) is a good cut. Next, as $G[A, B]$ has minimum degree at least $\gamma n/3$, we can greedily choose a cherry matching, where for each $x \in (L_A \cup L_B) \setminus V(F)$ we choose a path of length 2 in $G[A, B]$ centred at x , with all such paths vertex-disjoint from each other and from F . We can also choose a disjoint edge of $G[A, B]$ for each vertex in $(L_A \cup L_B) \cap V(F)$ connecting it to a vertex not in $L_A \cup L_B \cup V(F)$.

To complete the construction of the path P' , we greedily merge paths, where in each step we fix two endpoints x, y of two distinct existing paths and join them by a path of length 2 or 3 in $G[A, B]$, vertex-disjoint from the existing paths. To see that this is possible, note that if x, y are in the same part then they have at least $0.3n + 0.3n - (n/2 + \varepsilon n) > n/20$ common neighbours in the other part, so we can choose the required path of length 2, or if $x \in A, y \in B$ then we can connect x to some new $x' \in B$ then connect x', y by a disjoint path of length 2. This process ends with one path P' of length $< 10^3 \varepsilon n$; by possibly adding one more edge we can assume it has ends $a' \in A$ and $b' \in B$.

Finally, $G' = G \setminus (V(P') \setminus \{a', b'\})$ has balanced parts (A', B') by choice of F , and $G'[A', B']$ has minimum degree at least $0.3n - 10^3 \varepsilon n > |A'|/2 + 1$, so it satisfies the conditions of Lemma 3.8, which gives a Hamiltonian path in $G'[A', B']$ from a' to b' that completes P' to the required Hamiltonian cycle. ■

The required linear forest in Lemma 3.7 will be provided by the following lemma, which uses regularity, and plays an essential role here and later in the paper.

We recall that a balanced cut (A, B) is a partition of $V(G)$ with $|A| = |B| = n$.

Lemma 3.9. Let G be an $(n+1)$ -regular graph on $2n$ vertices and (A, B) be a balanced cut. Suppose A', B' are vertex covers of $G[A], G[B]$. Then, $(|A'| + 1)(|B'| + 1) \geq n + 1$.

Proof. Write $V := V(G)$. Suppose without loss of generality that $e(A \setminus A', A') \geq e(B \setminus B', B')$. Note that

$$e(A', B \setminus B') \leq e(A', B) = e(A', V) - e(A', A') - e(A', A \setminus A') \leq (n+1)|A'| - e(A', A \setminus A').$$

As $B \setminus B'$ is an independent set, we deduce

$$\begin{aligned} e(B \setminus B', B') &= e(B \setminus B', V) - e(B \setminus B', A \setminus A') - e(B \setminus B', A') - e(B \setminus B', B \setminus B') \\ &\geq (n+1)(n - |B'|) - (n - |A'|)(n - |B'|) - ((n+1)|A'| - e(A', A \setminus A')) - 0 \\ &= (n+1) - (|A'| + 1)(|B'| + 1) + e(A', A \setminus A'). \end{aligned}$$

As $e(A \setminus A', A') \geq e(B \setminus B', B')$, we conclude that $(n+1) - (|A'| + 1)(|B'| + 1) \leq 0$. ■

The following simple remarks will be used throughout the remainder of the paper.

Remark 3.10. Let M be a maximal matching in a graph H . Then $V(M)$ is a vertex cover in H . In particular, $|M| \geq |C|/2$ where C is a minimum vertex cover. Also, suppose H has maximum degree Δ and C is a vertex cover in H . Then $e(H) \leq \Delta|C|$.

As the size of a random subset of a set A has the binomial $B(|A|, 1/2)$ distribution, we will frequently need the following normal approximation, which follows from the Central Limit Theorem.

Lemma 3.11. If $X \sim B(n, 1/2)$ then $\mathbb{P}(a\sqrt{n}/2 \leq X - n/2 \leq b\sqrt{n}/2) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt + o(1)$ for any $a < b$.

Finally, we need the following well-known observation on the difference of independent binomials.

Lemma 3.12. If $X \sim B(n, 1/2)$, $Y \sim B(m, 1/2)$ are independent then $X + m - Y \sim B(n + m, 1/2)$.

Proof. Let $I_1, \dots, I_{n+m}$ be iid Bernoulli $B(1, 1/2)$ variables. We can write $X = \sum_{i=1}^n I_i$ and $Y = \sum_{i=n+1}^{n+m} (1 - I_i)$, as each $1 - I_i$ is also $B(1, 1/2)$. Then, $X + m - Y = \sum_{i=1}^{n+m} I_i \sim B(n + m, 1/2)$. ■

We conclude this subsection by applying Lemma 3.7 to prove Lemma 2.5, thus completing the proof of Theorem 2.2, and so Conjecture 1.1.

Proof of Lemma 2.5. Let $n^{-1} \ll \varepsilon \ll \gamma \ll \delta \ll 1$ and G be an $(n+1)$ -regular graph on $2n$ vertices. Suppose there is some $A \subseteq V(G)$ with $n \leq |A| \leq (1 + 32\varepsilon)n$ such that $G[A, \bar{A}]$ has at least $(1 - 56\varepsilon)n^2$ edges and minimum degree at least γn , where if $|A| > n$ then $G[A]$ has maximum degree at most γn .

Consider $G[\frac{1}{2}] = G[S]$ where S is a random subset of $V(G)$. By Chernoff bounds, whp $|S| = n \pm n^{0.6}$, all vertex degrees are $|S|/2 \pm |S|^{0.6}$, $|S \cap A| = |A|/2 \pm n^{0.6}$, $|S \cap \bar{A}| = |\bar{A}|/2 \pm n^{0.6}$ and $G' := G[S \cap A, S \cap \bar{A}]$ has minimum degree at least $\gamma n/3$.

Also, G' has at most as many non-edges as $G[A, \bar{A}]$, so $e(G') \geq |A \cap S||\bar{A} \cap S| - 56\varepsilon n^2 \geq (1/4 - 150\varepsilon)n^2$.

To complete the proof via Lemma 3.7 (with ε replaced by 150ε), it remains to show that the event E that $(S \cap A, S \cap \bar{A})$ is a good cut of $G[S]$ has $\mathbb{P}(E) \geq 0.01$.

To do so, we consider two cases according to the size of $k := |A| - n$.

Case 1: $k > \delta\sqrt{n}$.

Here $|A| > n$, so $G[A]$ has maximum degree at most γn , by assumption. Also $G[A]$ has minimum degree at least $\delta(G) - |B| \geq k + 1$, so by Remark 3.10 every vertex cover has size at least $e(G[A])/(\gamma n) \geq nk/(2\gamma n) = k/2\gamma$, and hence $G[A]$ has a matching of size at least $k/4\gamma$. As $k > \delta\sqrt{n}$ and $\gamma \ll \delta$, by Chernoff bounds $\mathbb{P}[|M[S]| > k/20\gamma] > 1 - \delta$, say. Also, by Lemma 3.12 and Chernoff we have $\mathbb{P}[0 \leq |S \cap A| - |S \cap \bar{A}| \leq k/20\gamma] = \mathbb{P}[n - k \leq B(2n, 1/2) \leq n - k + k/20\gamma] > 1/2 - \delta$, say, so $\mathbb{P}(E) \geq 1/2 - 2\delta$.

Case 2: $k \leq \delta\sqrt{n}$.

We consider any balanced cut (A^*, B^*) obtained from $(A, \bar{A})$ by moving k vertices from A to $\bar{A}$ and minimum vertex covers A', B' of $G[A^*], G[B^*]$. By Lemma 3.9, we have $\max\{|A'|, |B'|\} \geq \sqrt{n} - 1$.

Suppose first that $|B'| \geq \sqrt{n} - 1$. Then $G[B^*]$ has a matching M of size $(\sqrt{n} - 1)/2$ by Remark 3.10. By Chernoff we have $\mathbb{P}[|M[S]| > \sqrt{n}/9] > 0.99$, say. At most $k \leq \delta\sqrt{n}$ edges in $M(S)$ contain a moved vertex, so this event gives a matching of size $0.1\sqrt{n}$ in $S \cap \bar{A}$. Also, by Lemma 3.12 and Lemma 3.11,

$$\mathbb{P}[0 \leq |S \cap \bar{A}| - |S \cap A| \leq 0.1\sqrt{n}] = \mathbb{P}[n + k \leq B(2n, 1/2) \leq n + k + 0.1\sqrt{n}],$$

which by Lemma 3.11 differs by at most δ from $\int := \int_0^{0.1/\sqrt{2}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt > 0.02$. Thus, $\mathbb{P}(E) \geq 0.01$.

Similarly, if $|A'| \geq \sqrt{n} - 1$, we have a matching of size $0.1\sqrt{n}$ in $S \cap A$ with probability > 0.99 and

$$\mathbb{P}[0 \leq |S \cap A| - |S \cap \bar{A}| \leq 0.1\sqrt{n}] = \mathbb{P}[n - k \leq B(2n, 1/2) \leq n - k + 0.1\sqrt{n}]$$

again differs by at most δ from $\int > 0.02$. In all cases, $\mathbb{P}(E) \geq 0.01$, as required. $\blacksquare$

4 Asymptotic Result

In this section we strengthen our solution of Conjecture 1.1 to the following asymptotically tight result.

Theorem 4.1. Any $(n+1)$ -regular graph G on $2n$ vertices has $\mathbb{P}(G[\frac{1}{2}])$ is Hamiltonian $> 1/2 - o(1)$.

In the first subsection we give the proof of Theorem 4.1, assuming a key lemma (Lemma 4.2), which will also be used in the proof of our exact result Theorem 1.2 in the next section. In the second subsection we prove the key lemma, assuming a technical lemma that we prove by elementary calculus in the final subsection.

4.1 The key lemma

Here we prove Theorem 4.1 assuming the following key lemma. For the statement, recall that a cut (X, Y) of a graph H is k -good if $|X| \geq |Y|$ and $H[X]$ has a linear forest with at least $k + |X| - |Y|$ edges, or if $|Y| \geq |X|$ and $H[Y]$ has a linear forest with at least $k + |Y| - |X|$ edges.

Lemma 4.2. Let $n^{-1} \ll \delta \ll \lambda \ll \eta \ll 1$. Let G be an $(n+1)$ -regular graph on $2n$ vertices. Let (A, B) be a balanced cut of G . Let A', B' be minimum vertex-covers of $G[A], G[B]$ of sizes $\alpha\sqrt{n}, \beta\sqrt{n}$. Suppose that $\min\{\alpha, \beta\} > \eta$. Let $G[S]$ be a random induced subgraph of G , where each vertex is included in S independently and with probability $1/2$. Then $(A \cap S, B \cap S)$ is a $\delta\sqrt{n}$ -good cut of $G[S]$ with probability at least $1/2 + \lambda$.

Proof of Theorem 4.1. Let $n^{-1} \ll \varepsilon \ll \gamma \ll \delta \ll \eta \ll \theta \ll 1$ and G be an $(n+1)$ -regular graph on $2n$ vertices. Let $G[\frac{1}{2}] = G[S]$ be a random induced subgraph of G . We will show $\mathbb{P}(G[S] \text{ is Hamiltonian}) > 1/2 - \theta$. As in the proof of Theorem 2.2, we consider three cases according to Lemma 2.1, where in the first two cases $\mathbb{P}(G[S] \text{ is Hamiltonian}) = 1 - o(1)$ by Lemma 2.3 and Lemma 2.4. Thus, it suffices to consider the third near-bipartite case: there is some $A \subseteq V(G)$ with $n \leq |A| \leq (1 + 32\varepsilon)n$ such that $G[A, \bar{A}]$ has at least $(1 - 56\varepsilon)n^2$ edges and minimum degree at least γn , where if $|A| > n$ then $G[A]$ has maximum degree at most γn .

As in the proof of Lemma 2.5, whp $G[S]$ satisfies all conditions of Lemma 3.7, with the possible exception of the event E that $(S \cap A, S \cap \bar{A})$ is a good cut of $G[S]$. To complete the proof, it suffices to show $\mathbb{P}[E] > 1/2 - \theta/2$. Again, we consider two cases according to $k := |A| - n$. If $k > \delta\sqrt{n}$ then $\mathbb{P}(E) \geq 1/2 - 2\delta$ as in Case 1 of the proof of Lemma 2.5. Thus, we can assume $k \leq \delta\sqrt{n}$. As in Case 2 of the proof of Lemma 2.5, we consider any balanced cut (A^*, B^*) obtained from $(A, \bar{A})$ by moving k vertices from A to $\bar{A}$ and minimum vertex covers A', B' of $G[A^*], G[B^*]$ of sizes $\alpha\sqrt{n}, \beta\sqrt{n}$. By Lemma 3.9, we have $\max\{|A'|, |B'|\} \geq \sqrt{n} - 1$.

Let E' be the event that $(S \cap A^*, S \cap B^*)$ is a k -good cut of $G[S]$. Then E' implies E , so it suffices to show $\mathbb{P}[E'] > 1/2 - \theta/2$. This holds by Lemma 4.2 if $\min\{\alpha, \beta\} \geq \eta$, so we can assume otherwise, say $\alpha < \eta$ and $\beta > \frac{1}{2}\eta^{-1}$. Then $G[B^*]$ contains a matching of size $\frac{1}{4}\eta^{-1}\sqrt{n}$ by Remark 3.10. By Chernoff bounds, whp $G[S \cap B^*]$ contains a matching of size $\frac{1}{20}\eta^{-1}\sqrt{n}$, and as $\eta \ll \theta$, with probability at least $1/2 - \theta/2$ we have $0 \leq |B \cap S| - |A \cap S| \leq \frac{1}{20}\eta^{-1}\sqrt{n}$. The result follows. $\blacksquare$

4.2 Proof of the key lemma

In this subsection we prove Lemma 4.2, assuming Lemma 4.7 below, which will be proved in the following subsection. To achieve the asymptotically optimal probability of finding a good cut in $G[\frac{1}{2}]$ we will require a much tighter argument for finding linear forests, based on the following lemma.

Lemma 4.3. Let G be an $(n+1)$ -regular graph on $2n$ vertices. Let (A, B) be a balanced cut of G . Let A', B' be minimal vertex-covers of $G[A], G[B]$ of sizes $\alpha\sqrt{n}, \beta\sqrt{n}$. Suppose $\alpha, \beta = \Theta(1)$ and $\bar{e}(A', B \setminus B') \geq \bar{e}(B', A \setminus A')$. Then there exist (bipartite) subgraphs G_A of $G[A', A \setminus A']$ and G_B of $G[B', B \setminus B']$ with

$$e(G_B) \geq n - O(\sqrt{n}), \quad \Delta(G_B) \leq \alpha\sqrt{n} + 1, \quad e(G_A) \geq (2 - \alpha\beta)n - O(\sqrt{n}), \quad \Delta(G_A) \leq \beta\sqrt{n} + 1.$$

Proof. Recall for any $X, Y \subseteq V(G)$ and $v \in V(G)$ that $\bar{e}(X, Y)$ counts non-edges in $X \times Y$ and $\bar{d}(v, X)$ counts non-neighbours of v in X . Note that as G is $(n+1)$ -regular and $|A| = |B| = n$, for $v \in B$ we have $d(v, B) = t$ if and only if $\bar{d}(v, A) = t - 1$, and similarly interchanging A and B ; we will use this observation repeatedly below.

Write $cn := \bar{e}(A', B \setminus B')$ and $c'n = \bar{e}(B', A \setminus A')$. Then $c' \leq c$ by assumption.

We also write $sn := \bar{e}(A \setminus A', B \setminus B')$. As $B \setminus B'$ is independent we have

$$e(B \setminus B', B') = \sum_{v \in B \setminus B'} (\bar{d}(v, A) + 1) \geq (c + s)n + |B \setminus B'| \geq (1 + c + s)n - \beta\sqrt{n}. \quad (1)$$

Let $B_1 = \{v \in B' : d(v, B \setminus B') \geq \alpha\sqrt{n}\}$ and $B_2 = \{v \in B \setminus B' : d(v, B') \geq \alpha\sqrt{n}\}$.

For $v \in B_1$ we have $\bar{d}(v, A \setminus A') \geq \bar{d}(v, A) - |A'| \geq d(v, B \setminus B') - 1 - \alpha\sqrt{n}$.

Similarly, for $v \in B_2$ we have $\bar{d}(v, A \setminus A') \geq d(v, B') - 1 - \alpha\sqrt{n}$, so

$$\sum_{v \in B_1} (d(v, B \setminus B') - 1 - \alpha\sqrt{n}) \leq c'n \leq cn, \quad \sum_{v \in B_2} (d(v, B') - 1 - \alpha\sqrt{n}) \leq sn.$$

Thus, we can delete at most $sn + cn$ edges from $G[B', B \setminus B']$ to obtain G_B with maximum degree at most $\alpha\sqrt{n} + 1$ and $e(G_B) \geq n - \beta\sqrt{n}$ by (1).

From (1) we also obtain $\bar{e}(B', A) = \sum_{v \in B'} \bar{d}(v, A) = \sum_{v \in B'} (d(v, B) - 1) \geq (1 + c + s)n - 2\beta\sqrt{n}$, so

$$c'n = \bar{e}(B', A \setminus A') \geq \bar{e}(B', A) - |A'| |B'| \geq (1 + c + s - \alpha\beta)n - 2\beta\sqrt{n}.$$

Similarly to (1), as $A \setminus A'$ is independent we deduce

$$e(A \setminus A', A') \geq \sum_{v \in A \setminus A'} (\bar{d}(v, B) + 1) \geq \bar{e}(B', A \setminus A') + |A \setminus A'| \geq (2 + c + s - \alpha\beta)n - (\alpha + 2\beta)\sqrt{n}.$$

Now by the same argument as for G_B we can delete at most $sn + cn$ edges from $G[A', A \setminus A']$ to obtain G_A with maximum degree at most $\beta\sqrt{n} + 1$ and $e(G_A) \geq (2 - \alpha\beta)n - O(\sqrt{n})$. $\blacksquare$

We also need to sharpen our argument for passing from minimum vertex covers in G to matchings in $G[\frac{1}{2}]$. Previously, we only used minimality and lost a factor of 8, using the naive estimate that one may lose a factor of 2 in finding a matching in G and then each edge survives with probability $1/4$ in $G[\frac{1}{2}]$. In the following lemma we consider a *minimum vertex cover*, meaning that it is not only minimal but also of minimum possible size, and improves the factor 8 to 4 under mild assumptions on the size.

Lemma 4.4. Let G be a graph on vertices with minimum vertex cover C , where $\log^2 n \leq |C| \leq n/4$. Then whp $G[\frac{1}{2}]$ has a matching of size $(1 - o(1))|C|/4$.

Proof. Write $G[\frac{1}{2}] = G[S]$, where S is a uniformly random subset of $V(G)$. We condition on $C' := S \cap C$ and fix a maximal matching M in $G[C']$. By Chernoff whp $|C'| \geq (1 - o(1))|C|/2$. We can assume $|M| < (1 - o(1))|C'|/2$, otherwise we are done. Then $C'' := C' \setminus V(M)$ is an independent set, by maximality of M .

Next we note that the bipartite graph $G' = G[C'', V(G) \setminus C]$ satisfies the Hall condition that $|N_{G'}(X)| \geq |X|$ for every $X \subseteq C''$. Indeed, if this failed for some X then $(C \setminus X) \cup N_{G'}(X)$ would be a vertex cover of G that contradicts C having minimum possible size. Thus, there is a matching M' in G' which covers C'' .

Now we reveal the rest of $S \setminus C$. Each edge of M' survives in $M'[S]$ with probability $1/2$, so by Chernoff whp $(1 - o(1))|C''|/2$ edges survive. Combining M and $M'[S]$ gives the required matching as

$$|M \cup M'[S]| = |M| + (1 - o(1))(|C'| - |M|)/2 \geq (1 - o(1))|C'|/2 \geq (1 - o(1))|C|/4. \quad \blacksquare$$

We will obtain linear forests via a well-known result of Alon [1] (see also [26]) giving an approximate form of the linear arboricity conjecture.

Theorem 4.5 (Asymptotic linear arboricity). Let $1/\Delta \ll \varepsilon$ and G be a graph of maximum degree at most Δ . Then G can be edge-decomposed into at most $(1 + \varepsilon)\Delta/2$ linear forests.

So far we have only relied on the Chernoff concentration inequality, but for the asymptotically correct count on edges in random induced subgraphs we also need the following consequence of Azuma's inequality [2].

Lemma 4.6. Let G be a graph on n vertices with $\Delta(G) = o(e(G))$. Then whp $e(G[\frac{1}{2}]) = (1 + o(1))e(G)/4$.

Proof. Let $X = e(S)$, where S is a uniformly random subset of $V(G)$. Then $\mathbb{E}X = e(G)/4$, and changing whether some vertex v is in S can affect $e(G[\frac{1}{2}])$ by at most $d_G(v)$. By Azuma's inequality $\mathbb{P}(|X - \mathbb{E}X| > t) \leq e^{-t^2/2Q}$, where $Q = \sum_{v \in V(G)} d(v)^2 \leq \Delta(G) \sum_{v \in V(G)} d(v) = 2\Delta(G)e(G) = o(e(G)^2)$. The lemma follows. $\blacksquare$

We conclude this subsection with the proof of the key lemma, assuming the following calculus lemma to be proved in the next subsection.

Lemma 4.7. Let $n^{-1} \ll \delta \ll \lambda \ll \eta \ll 1$ and $\alpha, \beta > \eta$. Define $m_1 = \max\{\alpha/4, 2/\beta - \alpha\}$ and $m_2 = \max\{\beta/4, 1/\alpha\}$. Let $X, Y \sim B(n, 1/2)$ be independent binomials. Consider the event E that $-(m_1 - \delta)\sqrt{n} \leq X - Y \leq (m_2 - \delta)\sqrt{n}$. Then $\mathbb{P}[E] \geq 1/2 + \lambda$.

Proof of Lemma 4.2. Let $n^{-1} \ll \delta \ll \lambda \ll \eta \ll 1$. Let G be an $(n+1)$ -regular graph on $2n$ vertices and (A, B) be a balanced cut of G . Let A', B' be minimum vertex-covers of $G[A], G[B]$ of sizes $\alpha\sqrt{n}, \beta\sqrt{n}$, where $\min\{\alpha, \beta\} > \eta$. Let $G[\frac{1}{2}] = G[S]$ be a random induced subgraph of G . Let E' be the event that $(A \cap S, B \cap S)$ is a $\delta\sqrt{n}$ -good cut of $G[\frac{1}{2}]$. We will show $\mathbb{P}[E'] \geq \frac{1}{2} + \lambda$.

We assume without loss of generality that $\bar{e}(A', B \setminus B') \geq \bar{e}(B', A \setminus A')$.

As $\min\{\alpha, \beta\} > \eta$, we can apply Lemma 4.3, obtaining subgraphs G_B of $G[B', B \setminus B']$ and G_A of $G[A', A \setminus A']$ with $e(G_B) \geq n - O(\sqrt{n})$, $\Delta(G_B) \leq \alpha\sqrt{n} + 1$, $e(G_A) \geq (2 - \alpha\beta)n - O(\sqrt{n})$ and $\Delta(G_A) \leq \beta\sqrt{n} + 1$.

By Lemma 4.6 and Chernoff whp $G'_A := G_A[A \cap S]$ has $e(G'_A) \geq (2 - \alpha\beta - o(1))n/4$ and $\Delta(G_A) \leq (1 + o(1))\beta\sqrt{n}/2$, and $G'_B := G_B[B \cap S]$ has $e(G'_B) \geq (1 - o(1))n/4$ and $\Delta(G_B) \leq (1 + o(1))\alpha\sqrt{n}/2$.

By Theorem 4.5, we decompose G'_A into $(1 + o(1))\beta\sqrt{n}/4$ linear forests, so by averaging we find a linear forest with $(2/\beta - \alpha - o(1))\sqrt{n}$ edges. Similarly, in G'_B we find a linear forest with $(1/\alpha - o(1))\sqrt{n}$ edges.

Also, as $\min\{\alpha, \beta\} > \eta$, by Lemma 4.4 whp we have matchings of size $(\alpha/4 - \delta)\sqrt{n}$ in $G[A \cap S]$ and $(\beta/4 - \delta)\sqrt{n}$ in $G[B \cap S]$. Thus, whp $G[A \cap S]$ contains a linear forest with $(m_1 - \delta)\sqrt{n}$ edges and $G[B \cap S]$ contains a linear forest with $(m_2 - \delta)\sqrt{n}$ edges, where $m_1 = \max\{\alpha/4, 2/\beta - \alpha\}$ and $m_2 = \max\{\beta/4, 1/\alpha\}$.

Finally, applying Lemma 4.7 to $|A \cap S|, |B \cap S| \sim B(n, 1/2)$, replacing (δ, λ) by $(2\delta, 2\lambda)$, with probability at least $1/2 + 2\lambda$ we have $-(m_1 - 2\delta)\sqrt{n} \leq |B \cap S| - |A \cap S| \leq (m_2 - 2\delta)\sqrt{n}$, so E' holds. $\blacksquare$

4.3 The technical lemma

We conclude this section by proving Lemma 4.7, thus completing the proof of the key lemma, and so of the asymptotic result Theorem 4.1.

Proof of Lemma 4.7. Let $n^{-1} \ll \delta \ll \lambda \ll \eta \ll 1$ and $\alpha, \beta > \eta$. Let $X, Y \sim B(n, 1/2)$ be independent binomials and E be the event that $-(m_1 - \delta)\sqrt{n} \leq X - Y \leq (m_2 - \delta)\sqrt{n}$, where $m_1 = \max\{\alpha/4, 2/\beta - \alpha\}$ and

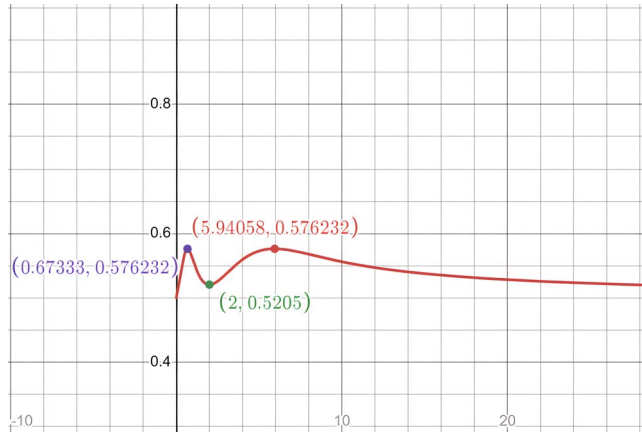


Fig. 1. An illustration of the function $f(\alpha)$ and its local extrema.

$m_2 = \max\{\beta/4, 1/\alpha\}$. We will show $\mathbb{P}[E] \geq 1/2 + \lambda$. As $X + n - Y \sim B(2n, 1/2)$ by Lemma 3.12, by Lemma 3.11 for any $a < b$ we have $\mathbb{P}(a\sqrt{n} \leq X - Y \leq b\sqrt{n}) = I[a, b] + o(1)$, where $I[a, b] := \int_{a\sqrt{2}}^{b\sqrt{2}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$. It suffices to show $I[-m_1, m_2] \geq 1/2 + 2\lambda$, as then by continuity $\mathbb{P}[E] \geq 1/2 + \lambda$ for $\delta \ll \lambda$.

Denote $f(\alpha) := I[-\frac{\alpha}{4}, \frac{1}{\alpha}]$. As $\alpha/4 \leq 2/\beta - \alpha$ if and only if $\beta \leq \frac{8}{5\alpha}$, while $\beta/4 \leq 1/\alpha$ if and only if $\beta \leq \frac{4}{\alpha}$, we have

$$I[-m_1, m_2] = \begin{cases} I\left[\alpha - \frac{2}{\beta}, \frac{1}{\alpha}\right] \geq f(\alpha) & \text{if } 0 < \beta \leq \frac{8}{5\alpha}, \\ I\left[-\frac{\alpha}{4}, \frac{1}{\alpha}\right] = f(\alpha) & \text{if } \frac{8}{5\alpha} < \beta < \frac{4}{\alpha}, \\ I\left[-\frac{\alpha}{4}, \frac{\beta}{4}\right] \geq f(\alpha) & \text{if } \beta \geq \frac{4}{\alpha}. \end{cases}$$

Thus, it remains to show $f(\alpha) \geq 1/2 + 2\lambda$. Figure 1 shows a computational illustration of this inequality.

The reader who is satisfied with this computational evidence can skip the remainder of the proof, in which we verify the inequality using calculus. As f is continuous and $\lambda \ll \eta$, it suffices to show that $f(\alpha) > 1/2$ for all $\alpha > 0$. As $\alpha \rightarrow 0$ or $\alpha \rightarrow \infty$ we have $f(\alpha) \rightarrow I[0, \infty] = 1/2$. We will show that f is at first increasing, until it reaches a local maximum, then it is decreasing until it reaches a local minimum at $\alpha = 2$, then it is increasing again until it reaches another local maximum, after which it is again decreasing. One can verify $f(2) = I[-1/2, 1/2] > 1/2$ via tables of the normal distribution, so this will complete the proof.

We set $x := \alpha^2$ and calculate the derivative of f via the Leibniz rule:

$$f'(\alpha) = \frac{1}{\sqrt{2\pi}} e^{-(\sqrt{2}\alpha/4)^2/2} \cdot \frac{\sqrt{2}}{4} - \frac{1}{\sqrt{2\pi}} e^{-(\sqrt{2}/\alpha)^2/2} \cdot \frac{\sqrt{2}}{\alpha^2} = \frac{1}{\sqrt{2\pi}} \left(\frac{1}{4} e^{-x/16} - \frac{1}{x} e^{-1/x} \right).$$

Considering $x = 4$, we observe that $f'(2) = 0$, as illustrated by the green dot in Figure 1.

To analyse the sign of $f'(\alpha)$, it will be more convenient to consider

$$g(x) := -\frac{x}{16} + \frac{1}{x} + \ln \frac{x}{4}, \quad g'(x) = -\frac{1}{16} - \frac{1}{x^2} + \frac{1}{x},$$

noting that $g(x)$ has the same sign as $f'(\alpha)$ and is zero iff $f'(\alpha)$ is zero.

We also note that $g'(x) > 0$ iff $-\frac{x^2}{16} + x - 1 > 0$, where the quadratic roots are $8 \pm 4\sqrt{3}$, so $g(x)$ is decreasing in $R_1 := (0, 8 - 4\sqrt{3})$, increasing in $R_2 := (8 - 4\sqrt{3}, 8 + 4\sqrt{3})$ and decreasing in $R_3 := (8 + 4\sqrt{3}, \infty)$.

Thus, $g(x)$ has at most one root in each R_i , one of which we have seen is $g(4) = 0$ with $r_2 = 4 \in R_2$.

As $g'(4) > 0$, if $g(x)$ had no root in R_1 then by the Intermediate Value Theorem it would be negative in R_1 . However, $g(x) \rightarrow \infty$ as $x \rightarrow 0$, so $g(x)$ must have a root $r_1 \in R_1$, so $g(x) > 0$ in $(0, r_1)$ and $g(x) < 0$ in (r_1, r_2) .

Similarly, if g had no root in R_3 it would be positive in R_3 , but $g(x) \rightarrow -\infty$ as $x \rightarrow \infty$, so g must have a root $r_3 \in R_3$. Then $g(x) > 0$ in (r_2, r_3) and $g(x) < 0$ for $x > r_3$.

This verifies the properties of $f(\alpha)$ that we showed above suffice to complete the proof. ■

5 Exact Result

In this section we prove our exact result, Theorem 1.2 in the following probabilistic form. For any graph G , let $p(G)$ be the probability that the random induced subgraph $G[\frac{1}{2}]$ is Hamiltonian. We need to show for any $(n+1)$ -regular graph G on $2n$ vertices with n sufficiently large that $p(G) \geq p_n := \min_{G' \in \mathcal{G}_n} p(G')$.

The proof requires a more careful implementation of the strategy for the asymptotic result in the previous section, and also more accurate estimates for $p(G)$. Our first lemma estimates the second order term in p_n .

Lemma 5.1. We have $p_n = \frac{1}{2} + \frac{3/2}{\sqrt{n\pi}} + O(n^{-3/2})$.

Proof. Fix $G \in \mathcal{G}_n$, with parts (A, B) , where $|A| = n+1$ and $|B| = n-1$. Recall that G is obtained from the complete bipartite graph $A \times B$ by adding some 2-factor inside A . Then $G[\frac{1}{2}] = G[S]$ is Hamiltonian if and only if $|A \cap S| \geq |B \cap S|$ and $G[A \cap S]$ contains a linear forest of size $|A \cap S| - |B \cap S|$. By Chernoff bounds, with probability $1 - e^{-\Theta(n)}$ we have $||A \cap S| - |B \cap S|| < n/100$ and $G[A \cap S]$ contains a linear forest of size $n/100$.

Thus, $p_n = \mathbb{P}(|A \cap S| \geq |B \cap S|) \pm e^{-\Theta(n)}$. By Lemma 3.12 we have $\mathbb{P}(|A \cap S| \geq |B \cap S|) = \mathbb{P}(B(2n, 1/2) \geq n-1)$. As $\mathbb{P}(B(2n, 1/2) \geq n+1) = \mathbb{P}(B(2n, 1/2) \leq n-1) = (1 - \mathbb{P}(B(2n, 1/2) = n))/2$, we deduce the required estimate $\mathbb{P}(B(2n, 1/2) \geq n-1) = \frac{1}{2} + \frac{3/2}{\sqrt{n\pi}} \pm n^{-3/2}$, using $4^{-n} \binom{2n}{n} = (n\pi)^{-1/2} \pm n^{-3/2}$ by Stirling's formula. ■

The following lemma gives an estimate for binomial probabilities that improves on the Chernoff bound for small or moderate deviations.

Lemma 5.2. Let $n \in \mathbb{N}$ be large, and let $k, k_1, k_2 \in \mathbb{Z}$ be such that $k \in [k_1, k_2] \subseteq [-n, n]$. Write $f_n(t) := \mathbb{P}(B(2n, 1/2) \geq n+t)$. Then

$$\mathbb{P}(B(n+k, 1/2) - B(n-k, 1/2)) \in [k_1, k_2] = 1 - f_n(k - k_1 + 1) - f_n(k_2 - k + 1),$$

where $f_n(t) \leq e^{-t^2/(3n+t)}$, and for $|t| \leq \sqrt{n}/100$ we have $f_n(t) = 1/2 - \frac{t-1/2}{\sqrt{n\pi}} \pm (2t^3 + 1)/n^{3/2}$.

Proof. We have $\mathbb{P}(B(n+k, 1/2) - B(n-k, 1/2)) \in [k_1, k_2] = \mathbb{P}(B(2n, 1/2) \in [n-k+k_1, n-k+k_2])$ by Lemma 3.12, so the stated equality holds. The first estimate on $f_n(t)$ follows from the Chernoff bound. For the second estimate, we may assume $t \geq 0$ by symmetry. Note that

$$\mathbb{P}[B(2n, 1/2) = n+t] = \frac{\binom{2n}{n+t}}{\binom{2n}{n}} \mathbb{P}[B(2n, 1/2) = n] = (1 \pm 2t^2/n) \left(\frac{1}{\sqrt{n\pi}} \pm n^{-3/2} \right).$$

Hence, using $\mathbb{P}(B(2n, 1/2) \geq n-1) = \frac{1}{2} + \frac{3/2}{\sqrt{n\pi}} \pm n^{-3/2}$, we conclude

$$f_n(t) = \mathbb{P}[B(2n, 1/2) \geq n-1] - \sum_{i=n-1}^{n+t-1} \mathbb{P}[B(2n, 1/2) = i] = 1/2 - \frac{t-1/2}{\sqrt{n\pi}} \pm (2t^3 + 1)/n^{3/2}. \quad \blacksquare$$

We conclude by proving our main theorem.

Proof of Theorem 1.2. Let $n^{-1} \ll \varepsilon \ll \gamma \ll \delta \ll \eta \ll \theta \ll 1$ and G be an $(n+1)$ -regular graph on $2n$ vertices. Let $G[\frac{1}{2}] = G[S]$ be a random induced subgraph of G and $p(G)$ be the probability that $G[S]$ is Hamiltonian. Suppose for contradiction that $p(G) < p_n$. We will show that there is a cut $(\tilde{A}, \tilde{B})$ of G with $|\tilde{A}| = n+1$, $|\tilde{B}| = n-1$ and $|G[\tilde{B}]| = 0$; call this an *extremal cut*. Then regularity will imply that $G[\tilde{A}, \tilde{B}]$ is complete bipartite and $G[\tilde{A}]$ is a 2-factor, that is, $G \in \mathcal{G}_n$, contradicting the definition of p_n .

As in the proof of Theorem 2.2 and Theorem 4.1, it suffices to consider the near-bipartite case of Lemma 2.1: there is some $A \subseteq V(G)$ with $n \leq |A| \leq (1 + 32\epsilon)n$ such that $G[A, \bar{A}]$ has at least $(1 - 56\epsilon)n^2$ edges and minimum degree at least γn , where if $|A| > n$ then $G[A]$ has maximum degree at most γn . Set $B := \bar{A}$.

As before, whp $G[S]$ satisfies all conditions of Lemma 3.7, with the possible exception of the event E that $(S \cap A, S \cap B)$ is a good cut of $G[S]$. Again we consider two cases according to the value of $k := |A| - n$. We start by ruling out the case $k > \delta\sqrt{n}$. To see that this is impossible, note that $G[A]$ has a matching of size $\frac{k}{4\gamma}$ by Remark 3.10, and that $p(G)$ is at least the probability that $G[A \cap S]$ has a matching of size $\frac{k}{20\gamma}$ while at the same time $0 \leq |A \cap S| - |B \cap S| \leq \frac{k}{20\gamma}$. Thus, we have $p(G) \geq \mathbb{P}(0 \leq B(|A|, 1/2) - B(|B|, 1/2) \leq \frac{k}{20\gamma}) - e^{-\Theta(\sqrt{n})}$ by Chernoff. Then by Lemma 5.2 and since $f_n(t)$ is decreasing with t , we have

$$\begin{aligned} \mathbb{P}(0 \leq B(|A|, 1/2) - B(|B|, 1/2) \leq \frac{k}{20\gamma}) &\leq \frac{k}{20\gamma} = 1 - f_n(k+1) - f_n(\frac{k}{20\gamma} - k + 1) \\ &\geq 1/2 + \frac{\delta}{\sqrt{\pi}} - O(\delta^3) - e^{-\Theta(\delta/\gamma)^2} \geq 1/2 + \delta/2, \end{aligned}$$

as $\gamma \ll \delta$, so $p(G) > p_n$, contradiction. Thus, we can assume $k \leq \delta\sqrt{n}$.

We will consider below a balanced cut (A^*, B^*) obtained by moving k vertices from A to B and minimum vertex covers A', B' of $G[A^*], G[B^*]$. As before, we consider the event E' that $(S \cap A^*, S \cap B^*)$ is a k -good cut of $G[S]$, which implies E , and to conclude the proof it suffices to show $\mathbb{P}[E'] \geq p_n$. Again we can assume $\min\{\alpha, \beta\} \leq \eta$, otherwise this holds by Lemma 4.2.

We consider the two cases, depending on whether $|A| = |B|$ or not.

Case A: $|A| > |B|$

As mentioned above, we consider any balanced cut (A^*, B^*) obtained by moving $k \geq 1$ vertices from A to B . Let A', B' be minimum vertex covers of $G[A^*], G[B^*]$ of sizes $a = \alpha\sqrt{n}, b = \beta\sqrt{n}$. By Lemma 3.9, we have $(\alpha\sqrt{n} + 1)(\beta\sqrt{n} + 1) \geq n + 1$.

Let M_A^*, M_B^* be maximal matchings in $G[A^*], G[B^*]$. Then, $|M_A^*| \geq a/2$ and $|M_B^*| \geq \lceil b/2 \rceil$.

Let M_A, M_B be maximal matchings in $G[A], G[B]$. Then, $|M_A| \geq |M_A^*| \geq a/2$ and $|M_B| \geq |M_B^*| - k \geq \lceil b/2 \rceil - k$. As noted above, we can assume $\min\{\alpha, \beta\} \leq \eta$.

Case A1: Suppose $\beta \leq \eta$.

Then $a \geq \frac{1}{2}\eta^{-1}\sqrt{n}$. By Chernoff bounds, with probability $1 - e^{-\Theta(\sqrt{n})}$ there is a matching of size $a/9$ in $G[A \cap S]$. Next we consider the choice of the random set $S' := V(M_B) \cap S \subseteq B \cap S$. The expected number of surviving edges of M_B is $\mathbb{E}[|M_B[S']|] = |M_B|/4$, so by averaging $\mathbb{P}(|M_B[S']| \geq b') \geq 1/8$, where $b' := \max\{\lceil (b/2) - k \rceil / 8, 0\}$.

Writing $B^\circ = B \setminus V(M_B)$, to estimate $p(G)$ we consider two events depending on which of $|A \cap S|$ or $|B \cap S|$ is larger, giving

$$\begin{aligned} p(G) &\geq \mathbb{P}[0 \leq |A \cap S| - |B \cap S| \leq a/9] - e^{-\Theta(\sqrt{n})} \\ &\quad + \frac{1}{8} \mathbb{E}_S \mathbb{P}[-b' \leq |A \cap S| - |S'| - |B^\circ \cap S| \leq -1], \end{aligned}$$

where $|A \cap S| \sim B(n + k, 1/2)$, $|B \cap S| \sim B(n - k, 1/2)$, and $|B^\circ \cap S| \sim B(n - k - 2|M_B|, 1/2)$.

We estimate the first term in the bound on $p(G)$ by Lemma 5.2 with parameters $(n, k, k_1, k_2) = (n, k, 0, a/9)$, to get that it is larger than $1 - (1/2 - \frac{k+1/2}{\sqrt{n\pi}}) - (2k^3 + 1)/n^{3/2} - e^{-\alpha^2/200}$.

For the second, as $|B^\circ \cap S| - |A \cap S| + n + k \sim B(2n - 2|M_B|, 1/2)$, writing $\mu = n - |M_B|$, we have

$$\begin{aligned} \mathbb{P}(-b' \leq |A \cap S| - |S'| - |B^\circ \cap S| \leq -1) \\ &= \mathbb{P}[n + k - |S'| + 1 \leq B(2n - 2|M_B|, 1/2) \leq n + k - |S'| + b'] \\ &= f_\mu(k + |M_B| - |S'| + 1) - f_\mu(k + |M_B| - |S'| + b' + 1). \end{aligned}$$

Using $b', |S'| \leq |M_B| \leq \eta\sqrt{n}$ where $\eta \ll 1$, and $b' \geq |M_B|/8$, and also $k \leq \delta\sqrt{n}$, so that $(k + |M_B|)^2/n = o((k + b')/\sqrt{n})$ for η sufficiently small, by Lemma 5.1 we deduce

$$\begin{aligned} p(G) - p_n &= 1 - (1/2 - \frac{k+1/2}{\sqrt{n\pi}}) - e^{-\alpha^2/200} + 1_{b' > 0} \frac{1}{8} \frac{b'}{\sqrt{n\pi}} - (\frac{1}{2} + \frac{3/2}{\sqrt{n\pi}}) - O\left(\frac{(k + |M_B|)^3}{n^{3/2}}\right) \\ &\geq \frac{k-1+1_{b' > 0}b'/8}{\sqrt{n\pi}} - e^{-\alpha^2/200} - o\left(\frac{k + b'}{\sqrt{n}}\right). \end{aligned} \tag{2}$$

Recalling that $(\alpha\sqrt{n}+1)(\beta\sqrt{n}+1) \geq n+1$, where $\beta \leq \eta \ll 1$, and also that $b' \geq (\lceil b/2 \rceil - k)/8 \geq \sqrt{n}/(40\alpha) - k/8$, we deduce $p(G) - p_n \geq \Omega(n^{-1/2}) > 0$ unless $k = 1$ and $b' = 0$.

To conclude this subcase, it suffices to show that $b = 1$. Indeed, as $|A| = n + 1$ and $|B| = n - 1$, this would imply that $G[B^*]$ is a star, whose root must be the vertex v moved from A to B when creating (A^*, B^*) , as v has at least γn neighbours in B . Thus, $G[B]$ is an independent set, so (A, B) is an extremal cut, which is a contradiction.

As $b' = 0$ we have $b \leq 2$, so it remains to rule out the possibility $b = 2$, because $b \neq 0$ as there is at least one edge in $G[B^*]$. Now, $b = 2$ implies that $G[B^*]$ is either a union of a triangle and isolated vertices or has a matching of size 2. The former case violates the regularity of G , whereas in the latter case $\mathbb{P}(|M_B[S']| \geq 1) \geq 1/16$, so a similar calculation to above (with $b' = 1$ defined in the beginning of the proof, instead of $b' = 0$) gives the contradiction $p(G) - p_n \geq \Omega(n^{-1/2}) > 0$.

Case A2: Suppose $\alpha \leq \eta$.

Then $b \geq \frac{1}{2}\eta^{-1}\sqrt{n}$. By Chernoff bounds, with probability $1 - e^{-\Theta(\sqrt{n})}$ there is a matching of size $b/9$ in $G[B \cap S]$. We obtain (A^*, B^*) by moving k vertices from A to B to maximise $e(A^*)$. Then by averaging $e(A^*) \geq (1 - 2k/n)e(A) \geq (1 - 2k/n)(k+1)(n+k)/2 \geq (k+1)n/2 - k^2$, as $e(A)$ has minimum degree at least $k+1$. On the other hand, $e(A^*)$ has maximum degree at most γn , so its minimum vertex cover has size $a \geq e(A^*)/2\gamma n \geq (k+1)/5\gamma$.

The remainder of this case is similar to Case A1; we give the details for completeness.

We consider the choice of the random set $S' := V(M_A) \cap S \subseteq A \cap S$. We have $\mathbb{E}[|M_A[S']|] = |M_A|/4 \geq a/8$, so $\mathbb{P}(|M_B[A]| \geq a') \geq 1/8$, where $a' = \lceil a/16 \rceil \geq (k+1)/90\gamma$. We write $A^\circ = A \setminus V(A_B)$ and estimate

$$p(G) \geq \mathbb{P}[0 \leq |B \cap S| - |A \cap S| \leq b/9] - e^{-\Theta(\sqrt{n})} + \frac{1}{8}\mathbb{E}_S\mathbb{P}[1 \leq |A^\circ \cap S| + |S'| - |B \cap S| \leq a'],$$

where $|A \cap S| \sim B(n+k, 1/2)$, $|B \cap S| \sim B(n-k, 1/2)$, and $|A^\circ \cap S| \sim B(n+k-2|M_A|, 1/2)$.

As $|A^\circ \cap S| - |B \cap S| - (n-k) \sim B(2n-2|M_A|, 1/2)$, writing $\mu = n - |M_A|$, as in Case A1 we have

$$\mathbb{P}(1 \leq |A^\circ \cap S| + |S'| - |B \cap S| \leq a') = f_\mu(k + |M_A| - |S'| + 1) - f_\mu(k + |M_A| - |S'| + a' + 1).$$

Since $k \leq a' \leq |M_A| \leq 100\eta\sqrt{n}$ where $\eta \ll 1$, as before we deduce

$$\begin{aligned} p(G) - p_n &= 1 - (1/2 - \frac{k+1/2}{\sqrt{n\pi}}) - e^{-\beta^2/200} + \frac{1}{8}\frac{a'}{\sqrt{n\pi}} - (\frac{1}{2} + \frac{3/2}{\sqrt{n\pi}}) - O\left(\frac{|M_A|^3}{n^{3/2}}\right) \\ &\geq \frac{k-1+a'/8}{\sqrt{n\pi}} - e^{-\beta^2/200} + o\left(\frac{a'}{\sqrt{n}}\right). \end{aligned}$$

Recalling that $(\alpha\sqrt{n}+1)(\beta\sqrt{n}+1) \geq n+1$, we deduce $p(G) - p_n \geq \Omega(n^{-1/2}) > 0$ for any $k \geq 0$, using $a' \geq (k+1)/90\gamma$ and $\gamma \ll 1$. Thus, we obtain a contradiction in this case.

Case B: $|A| = |B|$

By symmetry we can suppose that $\beta \leq \eta$. If $G[B]$ contains a vertex with degree at least γn , then the proof is identical to Case A1 with $A := A \cup \{v\}$ and $B := B \setminus \{v\}$. Otherwise, since $G[B]$ spans at least $n/2$ edges because of regularity, it contains a matching of size $1/5\gamma$. Hence, we can again follow Case A1 until Equation (2), where we will have $k = 0$ and $b' \geq 1/50\gamma$, so clearly $p(G) - p_n > 0$ also in this case. ■

6 Concluding Remarks

One question left unanswered by our paper is to determine which graph(s) $G \in \mathcal{G}_n$ achieve $p(G) = p_n$. The proof of Lemma 5.1 shows that $|p(G) - p(G')| < e^{-\Theta(n)}$ for all $G, G' \in \mathcal{G}_n$, so this seems to be a delicate question about large deviation rate functions. A natural guess would be that $p(G)$ is minimised when the number of independent sets in A is maximised, that is, the optimal 2-factor should be a C_4 -factor (see [33]).

Another natural direction would be to generalise the parameters of our problem, namely (a) the assumed degree of regularity in the graph G and (b) the distribution on its induced subgraphs. Concretely, one may consider a d -regular graph on m vertices and ask about the probability $h(G, p)$ that $G[p]$ is Hamiltonian, where $G[p]$ denotes a random induced subgraph with each vertex included independently with probability p . If we stick to $(m, d) = (2n, n+1)$ and decrease p then this puts the

spotlight on a competing construction which we saw lurking in the background throughout the paper: suppose (for convenience) that $n = k^2$ is a perfect square, fix a copy of $K_{n,n}$, add k vertex-disjoint k -vertex stars spanning each part, delete crossing edges between the star centres to make an $(n + 1)$ -regular graph. When G is this construction we saw that $h(G, 1/2) \geq 0.52$ (see Figure 1), so it is not a very serious competitor with $\mathcal{G}_n$ when $p = 1/2$, but for smaller p there may be a phase transition at which the competing construction takes over as the optimum. Furthermore, $h(G, p) \rightarrow 0$ as $p \rightarrow 0$, as the largest linear forest in either part of $G[p]$ has $O(p\sqrt{n})$ edges, whereas $G[p]$ almost surely picks $\Omega(\sqrt{pn})$ more elements from one of the two parts of G .

On the other hand, if we stick to $p = 1/2$ and decrease d , then to obtain a sensible question one should add an additional assumption, such as k -connectivity for some k , otherwise G may be disconnected, in which case $h(G, 1/2)$ decays exponentially in $m = |V(G)|$. Alternatively, one could simply assume that G is Hamiltonian, as suggested to us by Alex Scott (personal communication). We conjecture that the extremal examples G for such questions are essentially disjoint unions of bipartite graphs, with a few edges added to ensure the connectivity or Hamiltonicity assumption. A more explicit and weaker form of this conjecture, which still seems interesting, would be to show that if $d = cm$ for fixed $c \in (0, 1/2)$ and m large then $p(G) = \Omega(m^{-k/2})$ where $k = \lfloor (2c)^{-1} \rfloor$.

One may also consider other classical combinatorial theorems and ask for robust analogues in the sense of this paper. For example, consider the Hajnal-Szemerédi Theorem [15] (see also [18]) that any n -vertex graph with $\delta(G) \geq (r - 1)n/r$ (where r divides n) contains a K_r -factor, that is, a partition of $V(G)$ into sets inducing copies of K_r in G . By analogy with Conjecture 1.1, we pose the following conjecture.

Conjecture 6.1. For any $r \geq 2$ there is some $\varepsilon > 0$ so that if G is an $(\frac{(r-1)n}{r} + 1)$ -regular graph on n vertices, where $r \mid n$, then at least $\varepsilon 2^n$ subsets of $V(G)$ induce a K_r -factor.

A plausible class of extremal constructions for Conjecture 6.1 may be the r -partite analogue of $\mathcal{G}_n$, that is, slightly unbalanced complete r -partite graphs with a suitable factor in the largest part A . This suggests that the optimal ε should be $1/r^2$, where in the random induced subgraph $G[S]$ we have one probability factor of $1/r$ for $r \mid |S|$ and another for $A \cap S$ being the largest part.

Funding

This work was supported by SNSF project 217926 [to N.D.] and by ERC Advanced Grant 883810 [to P.K.].

Acknowledgments

We would like to thank Zach Hunter, Teng Liu, and Aleksa Milojević for carefully reading the first version of our manuscript and helping us eliminate some minor errors. We also thank the anonymous referee for their helpful remarks.

References

- Alon, N. "The linear arboricity of graphs." *Israel J. Math.* **62** (1988): 311–25. <https://doi.org/10.1007/BF02783300>.
- Azuma, K. "Weighted sums of certain dependent random variables." *Tohoku Math. J.* **19**, no. 2 (1967): 357–67. <https://doi.org/10.2748/tmj/1178243286>.
- Ben-Shimon, S., M. Krivelevich, and B. Sudakov. "On the resilience of Hamiltonicity and optimal packing of Hamilton cycles in random graphs." *SIAM J. Discr. Math.* **25** (2011): 1176–93. <https://doi.org/10.1137/110821299>.
- Bloom, T. *Erdős problems*. <https://www.erdosproblems.com/622>. Accessed: 20 October 2024.
- Bondy, J. A. "Pancyclic graphs I." *J. Combin. Theory Ser. B* **11** (1971): 80–4. [https://doi.org/10.1016/0095-8956\(71\)90016-5](https://doi.org/10.1016/0095-8956(71)90016-5).
- Chvátal, V. "On Hamilton's ideals." *J. Combin. Theory Ser. B* **12** (1972): 163–8. [https://doi.org/10.1016/0095-8956\(72\)90020-2](https://doi.org/10.1016/0095-8956(72)90020-2).
- Csaba, B., D. Kühn, A. Lo, D. Osthus, and A. Treglown. "Proof of the 1-factorization and Hamilton decomposition conjectures." *Mem. Amer. Math. Soc.* **244** (2016): 1. <https://doi.org/10.1090/memo/1154>.

8. Cuckler, B. and J. Kahn. "Hamiltonian cycles in Dirac graphs." *Combinatorica* **29** (2009): 299–326. <https://doi.org/10.1007/s00493-009-2360-2>.
9. Dirac, G. A. "Some theorems on abstract graphs." *Proc. Lond. Math. Soc. (3)* **2**, no. 3 (1952): 69–81.
10. Draganić, N., D. M. Correia, and B. Sudakov. "Pancyclicity of Hamiltonian graphs." *J. Eur. Math. Soc.* (to appear).
11. Draganić, N., S. Glock, D. M. Correia, and B. Sudakov. "Optimal Hamilton covers and linear arboricity for random graphs." *Proc. Amer. Math. Soc.* **153** (2025): 921–35. <https://doi.org/10.1090/proc/16924>.
12. Erdős, P. "A selection of problems and results in combinatorics." *Combin. Probab. Comput.* **8** (1999): 1–6. <https://doi.org/10.1017/S0963548398003496>.
13. Ferber, A. and V. Jain. "1-factorizations of pseudorandom graphs." 2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS), 698–708. IEEE, 2018.
14. Frieze, A. and R. Kannan. "Quick approximation to matrices and applications." *Combinatorica* **19** (1999): 175–220. <https://doi.org/10.1007/s004930050052>.
15. Hajnal, A. and E. Szemerédi. "Proof of a conjecture of P. Erdős." *Combinatorial Theory and Its Applications Proc. Colloq. Balatonfüred 1969*, edited by Erdős P., Rényi A., and Sós V. 601–23. Amsterdam: North-Holland, 1970.
16. Hefetz, D., D. Kühn, J. Lapinskas, and D. Osthus. "Optimal covers with Hamilton cycles in random graphs." *Combinatorica* **34** (2014): 573–96. <https://doi.org/10.1007/s00493-014-2956-z>.
17. Janson, S., A. Ruciński, and T. Łuczak. *Random Graphs*. John Wiley & Sons, 2011.
18. Kierstead, H. and A. Kostochka. "A short proof of the Hajnal–Szemerédi theorem on equitable colouring." *Combin. Probab. Comput.* **17** (2008): 265–70. <https://doi.org/10.1017/S0963548307008619>.
19. Knox, F., D. Kühn, and D. Osthus. "Edge-disjoint Hamilton cycles in random graphs." *Random Struct. Algorithms* **46** (2015): 397–445. <https://doi.org/10.1002/rsa.20510>.
20. Komlós, J., G. N. Sárközy, and E. Szemerédi. "On the square of a Hamiltonian cycle in dense graphs." *Random Struct. Algorithms* **9** (1996): 193–211. [https://doi.org/10.1002/\(SICI\)1098-2418\(199608/09\)9:1/2<\(\(\193::AID-RSA12\\(\\)\)3.0.CO;2-P](https://doi.org/10.1002/(SICI)1098-2418(199608/09)9:1/2<((\193::AID-RSA12\(\))3.0.CO;2-P).
21. Komlós, J., G. N. Sárközy, and E. Szemerédi. "Proof of the Seymour conjecture for large graphs." *Ann. Combin.* **2** (1998): 43–60. <https://doi.org/10.1007/BF01626028>.
22. Krivelevich, M., C. Lee, and B. Sudakov. "Resilient pancyclicity of random and pseudorandom graphs." *SIAM J. Discr. Math.* **24** (2010): 1–16. <https://doi.org/10.1137/090761148>.
23. Krivelevich, M., C. Lee, and B. Sudakov. "Robust Hamiltonicity of Dirac graphs." *Trans. Amer. Math. Soc.* **366** (2014): 3095–130. <https://doi.org/10.1090/S0002-9947-2014-05963-1>.
24. Krivelevich, M. and W. Samotij. "Optimal packings of Hamilton cycles in sparse random graphs." *SIAM J. Discr. Math.* **26** (2012): 964–82. <https://doi.org/10.1137/110849171>.
25. Kühn, D. and D. Osthus. "Hamilton decompositions of regular expanders: a proof of Kelly's conjecture for large tournaments." *Adv. Math.* **237** (2013): 62–146. <https://doi.org/10.1016/j.aim.2013.01.005>.
26. Lang, R. and L. Postle. "An improved bound for the linear arboricity conjecture." *Combinatorica* **43** (2023): 547–69. <https://doi.org/10.1007/s00493-023-00024-9>.
27. Letzter, S. "Pancyclicity of highly connected graphs." (2023): arXiv preprint arXiv:2306.12579.
28. Montgomery, R. "Hamiltonicity in random graphs is born resilient." *J. Combin. Theory Ser. B* **139** (2019): 316–41. <https://doi.org/10.1016/j.jctb.2019.04.002>.
29. Sárközy, G. and S. Selkow. "Distributing vertices along a Hamiltonian cycle in Dirac graphs." *Discr. Math.* **308** (2008): 5757–70. <https://doi.org/10.1016/j.disc.2007.10.042>.
30. Sárközy, G. N., S. M. Selkow, and E. Szemerédi. "On the number of Hamiltonian cycles in Dirac graphs." *Discr. Math.* **265** (2003): 237–50. [https://doi.org/10.1016/S0012-365X\(02\)00582-4](https://doi.org/10.1016/S0012-365X(02)00582-4).
31. Sudakov, B. Robustness of graph properties. In *Surveys in Combinatorics*, 2017. London Math. Soc. Lecture Note Ser., vol. 440, 372–408. Cambridge: Cambridge Univ. Press, 2017, <https://doi.org/10.1017/9781108332699.009>.
32. Sudakov, B. and V. H. Vu. "Local resilience of graphs." *Random Struct. Algorithms* **33** (2008): 409–33. <https://doi.org/10.1002/rsa.20235>.
33. Zhao, Y. "Extremal regular graphs: independent sets and graph homomorphisms." *Amer. Math. Monthly* **124** (2017): 827–43. <https://doi.org/10.4169/amer.math.monthly.124.9.827>.