APPROXIMATE HAMILTON DECOMPOSITIONS OF RANDOM GRAPHS

FIACHRA KNOX, DANIELA KÜHN AND DERYK OSTHUS

Abstract. We show that if $pn \gg \log n$ the binomial random graph $G_{n,p}$ has an approximate Hamilton decomposition. More precisely, we show that in this range $G_{n,p}$ contains a set of edge-disjoint Hamilton cycles covering almost all of its edges. This is best possible in the sense that the condition that $pn \gg \log n$ is necessary.

1. Introduction

Consider the random graph model where one starts with an empty graph and successively adds edges which are chosen uniformly at random. One of the most striking results in the theory of random graphs is the fact that this random graph acquires a Hamilton cycle as soon as its minimum degree is at least 2. This was proved by Bollobás [1] and then soon afterwards generalized by Bollobás and Frieze [3] who showed that for any fixed k the above random graph contains k edge-disjoint Hamilton cycles as soon as its minimum degree is at least $2k$. (If the minimum degree is odd, one can guarantee an additional perfect matching, this is also the case for several of the conjectures and results discussed below.)

More recently, Frieze and Krivelevich [8] conjectured that a similar result even holds for arbitrary edge densities:

Conjecture 1 (Frieze and Krivelevich [8])**.** *For any p, whp the binomial random graph $G_{n,p}$ contains $\lfloor \delta(G_{n,p}) \rfloor /2$ edge-disjoint Hamilton cycles.*

Here we say that a property of a random graph on n vertices holds *whp* if the probability that it holds tends to 1 as n tends to infinity. The result from [3] implies that Conjecture 1 holds if $pn \leq \log n + O(\log \log n)$. Frieze and Krivelevich [9] extended this to all p with $pn = (1 + o(1)) \log n$. They proved also an approximate version of Conjecture 1 for constant edge probabilities.

Theorem 2 (Frieze and Krivelevich [8])**.** *Let $0 < p < 1$ be constant. Then whp $G_{n,p}$ contains $(1 - o(1))np/2$ edge-disjoint Hamilton cycles.*

As for constant p, whp all degrees are close to np, the result implies that $G_{n,p}$ can 'almost' be decomposed into edge-disjoint Hamilton cycles. As remarked in [10], the proof of [8] also works as long as p is a little larger than $n^{-1/8}$. In this paper, we extend this result to essentially the entire range of p.

Theorem 3. *For any $\eta > 0$, there exists a constant C such that if $p \geq \frac{C \log n}{n}$, then whp $G_{n,p}$ contains $(1 - \eta)np/2$ edge-disjoint Hamilton cycles.*

While finalizing this paper, we learned that this result was proven independently by Krivelevich (personal communication). Theorem 3 is best possible in the sense that if we relax the condition on p, then $G_{n,p}$ can no longer be 'almost' decomposed

Date: December 9, 2010.

D. Kühn was supported by the EPSRC, grant no. EP/F008406/1 and by the ERC, grant no. 258345. D. Osthus was supported by the EPSRC, grant no. EP/F008406/1.

into Hamilton cycles. Indeed, suppose that $pn = C \log n$ for some fixed C. Then there is an $\varepsilon > 0$ so that whp the minimum degree of $G_{n,p}$ is at most $(1-\varepsilon)np$ (see Exercise 3.4 in [2]).

A version of Conjecture 1 for random regular graphs of bounded degree was proved by Kim and Wormald [13]: if $r \geq 4$ is fixed, then whp a random r-regular graph contains $\lfloor r/2 \rfloor$ edge-disjoint Hamilton cycles.

Hypergraph versions of Theorem 2 for Hamilton ℓ-cycles were also recently considered by Frieze and Krivelevich [10] as well as Frieze, Krivelevich and Loh [11]. A Hamilton ℓ-cycle in this case consists of a cyclic ordering of the vertices together with a cyclic sequence of edges, each consisting of k consecutive vertices, such that consecutive edges intersect in exactly ℓ vertices.

A 'deterministic' version of these results was recently proved by Christofides, Kühn and Osthus [4]: Suppose that G is a d-regular graph on n vertices, where $d \geq (1+\varepsilon)n/2$ and n is sufficiently large. Then G contains $(1-\varepsilon)d/2$ edge-disjoint Hamilton cycles. This approximately confirms a conjecture of Nash-Williams [16] which states that any d-regular graph where $d \geq n/2$ has $\lfloor d/2 \rfloor$ edge-disjoint Hamilton cycles.

A related conjecture of Erdős (see [18]) states that almost all tournaments G contain at least $\min\{\delta^+(G), \delta^-(G)\}$ edge-disjoint Hamilton cycles. It follows from results in [8] that this is approximately true. Kühn, Osthus and Treglown [15] recently proved that we do not even require G to be random for this to hold: any almost regular tournament G contains a set of edge-disjoint Hamilton cycles which cover almost all edges of G.

The proof of Theorem 2 in [8] actually works for any quasi-random graph. (Here a graph G is defined to be quasi-random if it is almost regular and if the density of any large induced subgraph of G is close to that of G.) In contrast, our argument only seems to work for $G_{n,p}$ (for example, the proof of Lemma 17 breaks down for a quasi-random graph). It would be interesting to know whether our result can also be extended to some natural class of (sparse) quasi-random graphs.

2. Notation and organization of the paper

We consider the binomial random graph G_{n,p_0} and let

$$w_0 = \frac{np_0}{\log n}.$$

So using the notation and assumptions of Theorem 3, $w_0 \geq C$. Our results will hold provided that w_0 is sufficiently large depending on η, which we will assume throughout.

The graph G_1 is generated by including each edge of G_0 independently at random with probability $(1 - \frac{\eta}{4})$. We define $G_2 := G_0 \backslash G_1$. Note that $G_1 \sim G_{n,p_1}$ and $G_2 \sim G_{n,p_2}$, where $p_1 = (1-\frac{\eta}{4})p_0$ and $p_2 = \frac{\eta p_0}{4}$. (Of course, the distributions of these random graphs are not independent of each other.)

The outline of the proof of Theorem 3 is as follows: We begin in Section 3 by stating and applying some large deviation bounds on the number of edges in certain subgraphs of $G_{n,p}$, which we will use later on. Then in Section 4, we will use Tutte's r-factor theorem to show that one may find a regular subgraph of G_1 whose degree is close to np_1 (the average degree of G_1). In Section 5 we show that this subgraph can almost be decomposed into 2-factors in such a way that each 2-factor has relatively few cycles. Finally in Section 6 we convert each of these 2-factors into Hamilton cycles, using the edges of G_2 (along with any edges of G_1 which were not included in our collection of 2-factors). This is achieved using an appropriate variant of the well-known rotation-extension technique, first introduced by Pósa [17]. The fact that each of the 2-factors originally had few cycles will allow us to place an upper bound on the number of edges needed to perform the conversions, and thus to show that the process can be completed before all of the edges of G_2 have been used up.

Throughout the paper we use the following notation: for a graph G and sets A, B of vertices of G, $e_G(A, B)$ is the number of edges of G with one endpoint in A and the other in B. Let $e_G(A) = e_G(A, A)$. On the other hand if G is a graph then $e(G)$ denotes the number of edges, and for a graph G with a spanning subgraph H, $G \backslash H$ denotes the graph obtained by removing the edges of H from G. We omit floor and ceiling symbols in arguments where they do not have a significant effect. log denotes the natural logarithm.

3. Large Deviation Bounds

We will need the following Chernoff bounds, which are proved e.g., in Janson, Łuczak and Ruciński [12]:

Lemma 4. *Let $X \sim Bin(n, p)$. Then the following properties hold:*

(i) *If $\varepsilon < \frac{1}{2}$, then $\mathbb{P}(|X - np| \geq \varepsilon np) \leq e^{-\frac{\varepsilon^2 np}{3}}$.*
(ii) *If $t \geq 7np$, then $\mathbb{P}(X \geq t) \leq e^{-t}$.*

We can use these bounds to deduce some facts about the number of edges between subsets of vertices of a random graph, as follows:

Lemma 5. *Let $G \sim G(n, p)$. Then whp, for any disjoint $A, B \subseteq [n]$, the following properties hold: Let $a = |A|$ and $b = |B|$. Then*

(i) *If $\left(\frac{1}{a} + \frac{1}{b}\right) \frac{\log n}{p} \geq \frac{7}{2}$, then $e_G(A, B) \leq 2(a + b) \log n$, and*
(ii) *If $\left(\frac{1}{a} + \frac{1}{b}\right) \frac{\log n}{p} \leq \frac{7}{2}$, then $e_G(A, B) \leq 7abp$.*

Proof. (i) Let $X = e_G(A, B)$ and let $t = 2(a + b) \log n$. Since $X \sim Bin(ab, p)$, we have that $t \geq 7abp = 7\mathbb{E}X$. If $a + b < 3$ then the result is trivial; otherwise, by Lemma 4 we have that $\mathbb{P}(X \geq t) \leq e^{-t} = \left(\frac{1}{n^a n^b}\right)^2 \leq \frac{1}{n^3}\left(\frac{1}{n^a n^b}\right)$, and a union bound immediately gives the result.

(ii) Similarly, we have $\mathbb{P}(X \geq 7abp) \leq e^{-7abp} \leq e^{-t}$ and the result follows. ☐

In an exactly similar way, we can show that

Lemma 6. *Let $G \sim G_{n,p}$. Then whp, for every $A \subseteq [n]$ the following properties hold: Let $a = |A|$. Then*

(i) *If $\frac{\log n}{ap} \geq \frac{7}{4}$, then $e_G(A) \leq 2a \log n$, and*
(ii) *If $\frac{\log n}{ap} \leq \frac{7}{4}$, then $e_G(A) \leq \frac{7a^2 p}{2}$.*

For larger sets, Lemma 7 gives a more precise result. Note that we allow $\alpha, \beta \to 0$ in the statement.

Lemma 7. *Let $G \sim G_{n,p}$. Then whp, for all pairs $A, B \subseteq [n]$ of disjoint sets the following property holds: Let $\alpha = |A|/n$ and $\beta = |B|/n$, and suppose that $\alpha\beta np \geq 700$. Then*

$$\frac{13}{14}\alpha\beta n^2 p \leq e_G(A, B) \leq \frac{15}{14}\alpha\beta n^2 p.$$

Proof. $e_G(A, B) \sim Bin(\alpha\beta n^2, p)$, so by Lemma 4,

$$\mathbb{P}\left(e_G(A, B) < \frac{13}{14}\alpha\beta n^2 p\right) \leq e^{-\frac{\alpha\beta n^2 p}{3 \cdot 14^2}} \leq e^{-\frac{700n}{588}} \leq e^{-n \log(3.1)} = \frac{1}{(3.1)^n}.$$

A union bound over all 3^n possibilities now gives the result. The right-hand inequality follows in an exactly similar manner using the opposite tail estimate. $\square$

4. Regular subgraphs of a random graph

We first show that G_1 contains a regular subgraph of degree at least

$$r_1 = np_0\left(1 - \frac{3\eta}{4}\right),$$

where r_1 is taken to be even.

We do this by using a theorem of Tutte: Let G be an arbitrary graph, r a positive integer and suppose that S, T, U is a partition of $V(G)$. Then define

$$R_r(S,T) = \sum_{v\in T} d(v) - e_G(S,T) + r(|S| - |T|)$$

and let $Q_r(S,T)$ be the number of *odd* components of $G[U]$, where a component C is odd if and only if $r|C| + e_G(C,T)$ is odd. (In our case it will often suffice to bound $Q_r(S,T)$ simply by the total number of components of $G[U]$.)

Theorem 8 (Tutte [19])**.** *Let r be a positive integer. A graph G contains an r-factor if and only if $R_r(S,T) \geq Q_r(S,T)$ for every partition S, T, U of $V(G)$.*

In order to apply Theorem 8 we will need an upper bound on $Q_r(S,T)$. We do this by observing that if $G[U]$ has many components, then it must contain a large *isolated set* of vertices; that is, a set $A \subseteq U$ such that $e_G(A, U\backslash A) = 0$. This becomes useful when looking at a random graph, since (as we will prove in Lemma 11) it follows that whp A has many neighbours in $S\cup T$. This gives a lower bound on $|S\cup T|$ in terms of $Q_r(S,T)$, and thus gives an upper bound on $Q_r(S,T)$ in terms of $|S\cup T| = |S| + |T|$.

Lemma 9. *Let G be a graph with v components. Then for any $v' \leq \frac{v}{2}$, there exists a set $W \subseteq V(G)$ which is isolated in G, such that $v' \leq |W| \leq \max\{2v', \frac{2|G|}{v}\}$.*

Proof. Call a component C of G *small* if its order is at most v', and *large* otherwise. Suppose first that the union of all small components of G also has order at most v'. Then the number of small components is at most v', and hence there are at least $v - v' \geq \frac{v}{2}$ large components. So one of these components must have order at most $\frac{2|G|}{v}$, and we can set W to be this component.

On the other hand, if the sum of the orders of small components is greater than v' then we can form W by starting with $\emptyset$ and adding small components one by one until $|W| \geq v'$. Now since the last component added has size at most v', we have that $|W| \leq 2v'$. $\square$

Given a graph G, define the *boundary* $B_G(A)$ of a set $A \subseteq V(G)$ to be the set of vertices which are adjacent (in G) to some vertex of A, but are not themselves elements of A. We will use the following two lemmas to give a lower bound on $|B_{G_1}(A)|$:

Lemma 10. *Whp,*

(i) $\delta(G_1) \geq (1 - \frac{\eta}{2})np_0$,
(ii) $\Delta(G_1) \leq np_0$, *and*
(iii) $\delta(G_2) \geq \frac{\eta np_0}{5}$.

Proof. Note that for a vertex x of G_1, $d(x) \sim Bin(n-1, p_1)$ and $\mathbb{E}(d(x)) = (1-\frac{\eta}{4})(n-1)p_0$. By Lemma 4, we have

$$\mathbb{P}\left(d(x) \le (1-\frac{\eta}{2})np_0\right) \le e^{-(\frac{\eta}{5})^2\frac{np_0}{3}} = n^{-\frac{\eta^2 w_0}{75}} \le \frac{1}{n^2},$$

and a union bound gives the result. The bound on the maximum degree follows similarly, as does that on the minimum degree of G_2. □

Lemma 11. *The following holds whp: Let H be a spanning subgraph of G_0, and let $A \subseteq [n]$ be nonempty. Let $\delta_A = \min_{x\in A} d_H(x)$. Then setting $a = |A|$ and $b = |B_H(A)|$, the following properties hold:*

(i) *If $\frac{\log n}{ap_0} \ge \frac{7}{2}$, then $b \ge \frac{a(\delta_A - 6\log n)}{2\log n}$. In particular, if $H = G_1$, then $b \ge a$.*
(ii) *If $\frac{\log n}{ap_0} \le \frac{7}{2}$, then $3a + b \ge \frac{\delta_A}{7p_0}$. In particular, if $H = G_1$, then $3a + b \ge \frac{n}{14}$.*

Proof. Let $B = B_H(A)$.

(i) Then

$$a\delta_A \le \sum_{x\in A} d_H(x) = e_H(A,B) + 2e_H(A) \le e_{G_0}(A,B) + 2e_{G_0}(A) \tag{1}$$

which by Lemmas 5(i) and 6(i) is at most $2(a+b)\log n + 4a\log n$. The final part follows from Lemma 10(i).

(ii) We claim that $e_{G_0}(A,B) \le 7ap_0(a+b)$. Indeed, if $\left(\frac{1}{a}+\frac{1}{b}\right)\frac{\log n}{p_0} \le \frac{7}{2}$, then by Lemma 5(ii), $e_{G_0}(A,B) \le 7abp_0 \le 7ap_0(a+b)$. On the other hand, if $\left(\frac{1}{a}+\frac{1}{b}\right)\frac{\log n}{p_0} \ge \frac{7}{2}$, then $e_{G_0}(A,B) \le 2(a+b)\log n \le 7ap_0(a+b)$. Similarly, Lemma 6 implies that $e_{G_0}(A) \le 7a^2p_0$. Now

$$a\delta_A \overset{(1)}{\le} e_{G_0}(A,B) + 2e_{G_0}(A) \le 7ap_0(a+b) + 14a^2p_0$$

and the result follows immediately. Again the final part follows by Lemma 10(i). □

We will later use the above lemmas to show that taking successive neighbourhoods of a set will give us a set of size linear in n in a reasonably short time. For now, they allow us to give a bound on the size of $Q_r(S,T)$ in terms of $|S|$ and $|T|$.

Lemma 12. *In the graph G_1, whp, for any partition S, T, U of $[n]$, $Q_{r_1}(S,T) \le 150(|S|+|T|)$.*

Proof. Let v be the number of components of $G_1[U]$. Consider first the case when $150 \le v \le \frac{n}{75}$. Then by applying Lemma 9 to the graph $G_1[U]$, we have that there exists a set $W \subseteq [n]$, isolated in $G_1[U]$, such that $\frac{v}{2} \le |W| \le \frac{n}{75}$. Now by Lemma 11 with $A = W$,

$$|B_{G_1}(W)| \ge \min\{|W|, \frac{n}{14} - 3|W|\} \ge |W| \ge \frac{v}{2}.$$

But if W is isolated in $G_1[U]$, then its boundary in G_1 lies entirely in $S \cup T$. So $\frac{v}{2} \le |B_{G_1}(W)| \le |S| + |T|$, and hence $Q_{r_1}(S,T) \le v \le 2(|S|+|T|)$.

Now consider the case $v \ge \frac{n}{75}$. Setting $v' = \frac{n}{150}$ in Lemma 9, we have a set W, isolated in $G_1[U]$, such that $\frac{n}{150} \le |W| \le \frac{n}{75}$. Again the boundary of W in G_1 has size at least $|W| \ge \frac{n}{150}$, and hence $\frac{n}{150} \le |S| + |T|$. So $150(|S|+|T|) \ge n$ and the result holds trivially, since $Q_{r_1}(S,T)$ cannot be greater than n.

Finally if $v \le 150$, then $Q_{r_1}(S,T) \le 150 \le 150(|S|+|T|)$ unless we are in the trivial case $|S| = |T| = 0$. But if S, T are both empty then $U = [n]$ and $G_1[U] = G_1$, which whp has only one component (this follows from the fact that a random graph $G_{n,p}$ with $pn \geq 2\log n$ is whp connected, as proved in [6]). Since r_1 is even this component cannot be odd, whence $Q_{r_1}(S,T) = 0$. $\square$

Lemma 13. *In the graph G_1, whp, we have that $R_{r_1}(S,T) \geq Q_{r_1}(S,T)$ for any partition S, T, U of $[n]$.*

Proof. Let d_S, d_T be the average degrees of the vertices in S, T respectively. Let $\rho = \frac{|T|}{|S|}$, and $s = |S|$. We consider the following cases:

Case 1: $\rho \leq \frac{1}{2}$. Then since $e_{G_1}(S,T) \leq d_T|T|$ and $|S| \geq 2|T|$, we have

$$R_{r_1}(S,T) \geq r_1(|S| - |T|) \geq \frac{r_1}{3}(|S| + |T|)$$

and for sufficiently large n, $\frac{r_1}{3}(|S| + |T|) \geq 150(|S| + |T|) \geq Q_{r_1}(S,T)$.

Case 2: $\rho \geq 4$. Observe that by the definition of r_1 and by Lemma 10, we have

$$d_T - r_1 \geq \left(1 - \frac{\eta}{2}\right) np_0 - \left(1 - \frac{3\eta}{4}\right) np_0 = \frac{\eta np_0}{4} \tag{2}$$

and

$$d_S - r_1 \leq np_0 - \left(1 - \frac{3\eta}{4}\right) np_0 = \frac{3\eta np_0}{4}.$$

Now since $e_{G_1}(S,T) \leq d_S|S|$ and $|T| \geq 4|S|$,

$$\begin{aligned} R_{r_1}(S,T) &\geq d_T|T| - d_S|S| + r_1(|S| - |T|) = (d_T - r_1)|T| - (d_S - r_1)|S| \\ &\geq \frac{\eta np_0}{4}(|T| - 3|S|) \geq \frac{\eta np_0}{20}(|S| + |T|) \end{aligned}$$

which again is at least $Q_{r_1}(S,T)$ for sufficiently large n.

Case 3: $\frac{1}{2} \leq \rho \leq 4$ and $\left(\frac{1}{s} + \frac{1}{\rho s}\right) \frac{\log n}{p_1} \geq \frac{7}{2}$. In this case by Lemma 5 we have that $e_{G_1}(S,T) \leq 2(\rho+1)s\log n$, and so it suffices to prove that

$$\rho s(d_T - 2\log n - r_1 - 150) + s(r_1 - 2\log n - 150) \geq 0. \tag{3}$$

(3) holds if $d_T - 2\log n - r_1 - 150 \geq 0$ and $r_1 - 2\log n - 150 \geq 0$. But the latter inequality holds since $r_1 = \left(1 - \frac{3\eta}{4}\right) np_0 = \left(1 - \frac{3\eta}{4}\right) w_0 \log n$, and the former since

$$d_T - r_1 \overset{(2)}{\geq} \frac{\eta np_0}{4} = \frac{\eta w_0 \log n}{4} \geq 3\log n,$$

as $w_0 \geq \frac{12}{\eta}$.

Case 4: $\frac{1}{2} \leq \rho \leq 4$ and $(\frac{1}{s} + \frac{1}{\rho s})\frac{\log n}{p_1} \leq \frac{7}{2}$ and $\rho s \leq \frac{n}{30}$. In this case by Lemma 5 we have that $e_{G_1}(S,T) \leq 7\rho s^2 \left(1 - \frac{\eta}{4}\right) p_0$, and so it suffices to prove that

$$\rho s(d_T - r_1 - 150) + s(r_1 - 7\rho s\left(1 - \frac{\eta}{4}\right) p_0 - 150) \geq 0,$$

and hence it suffices that $d_T - r_1 - 150 \geq 0$ and $r_1 - 7\rho s(1 - \frac{\eta}{4})p_0 - 150 \geq 0$. But the former inequality holds as before, and the latter since

$$r_1 - 150 \geq \frac{14}{15} r_1 = \frac{14}{15}\left(1 - \frac{3\eta}{4}\right) np_0 \geq 28\rho s \left(1 - \frac{3\eta}{4}\right) p_0 \geq 7\rho s \left(1 - \frac{\eta}{4}\right) p_0.$$

Case 5: $\frac{1}{2} \leq \rho \leq 4$ and $\rho s \geq \frac{n}{30}$. Note that we still have $\frac{s}{n} \leq \frac{1}{\rho+1}$, since S, T are disjoint. Now by Lemma 7,

$$e_{G_1}(S,T) \leq \frac{15}{14}\rho s^2 \left(1 - \frac{\eta}{4}\right) p_0 \leq \frac{15}{14}\frac{\rho}{\rho+1} sn \left(1 - \frac{\eta}{4}\right) p_0 \leq \frac{6}{7} sn \left(1 - \frac{\eta}{4}\right) p_0.$$

So it suffices to prove that

$$\rho sn\left(1-\frac{\eta}{2}\right)p_0-\frac{6}{7}sn\left(1-\frac{\eta}{4}\right)p_0+(1-\rho)sn\left(1-\frac{3\eta}{4}\right)p_0-150(\rho+1)s\geq 0,$$

i.e., that $\frac{\eta\rho}{4}+(1-\frac{3\eta}{4})-\frac{6}{7}(1-\frac{\eta}{4})-\frac{150(\rho+1)}{np_0}\geq 0$, which is true if η is not too large (which we can assume without loss of generality). □

Corollary 14. *Whp, G_1 contains an even-regular subgraph of degree $r_1=(1-\frac{3\eta}{4})np_0$.*

Proof. This follows immediately from Theorem 8 and Lemma 13. □

5. 2-factors of regular subgraphs of a random graph

In this section we will show that any even-regular subgraph of G_0 of sufficiently large degree contains a 2-factor with fewer than $\frac{\kappa n}{\log n}$ cycles, where

$$\kappa = 2\log\left(\frac{16}{\eta}\right). \tag{4}$$

It will follow immediately that we can decompose almost all of our regular subgraph into 2-factors with at most this many cycles. Roughly, our strategy will be to show that the number of 2-factors with many cycles in the original graph is rather small; smaller, in fact, than the minimum number of 2-factors which an even-regular graph of degree r_1 must contain.

Lemma 15. *Let $k_0=\frac{\kappa n}{\log n}$. Then whp, for any r-regular subgraph $H\subseteq G_0$ with $r\geq 2np_0e^{-\frac{\kappa}{2}}$, H contains a 2-factor with at most k_0 cycles.*

To prove Lemma 15 we will need a number of further lemmas. We use Lemmas 16 and 17 to bound the number of 2-factors in $G_{n,p}$ with many cycles, while Lemma 18 gives a bound on the total number of 2-factors in H.

Lemma 16. *For any k and for $n\geq 3k$, we have*

$$\sum\prod_{i=1}^{k}\frac{1}{a_i}\leq\frac{k}{n}(\log n)^{k-1},$$

where the sum is taken over all ordered k-tuples $(a_1,a_2,\dots,a_k)$ such that $a_1+\ldots+a_k=n$ and $a_i\geq 3$ for each $i\in[n]$.

Proof. We proceed by induction on k. The case $k=1$ is trivial since both sides equal $\frac{1}{n}$. Supposing that the result holds for $k-1$, we have

$$\sum\prod_{i=1}^{k}\frac{1}{a_i}=\sum_{a_k=3}^{n-3(k-1)}\frac{1}{a_k}\sum\prod_{i=1}^{k-1}\frac{1}{a_i},$$

where again the second sum on the right-hand side is taken over all ordered $(k-1)$-tuples $(a_1,\dots,a_{k-1})$ such that $a_1+\ldots+a_{k-1}=n-a_k$ and $a_i\geq 3$ for all $i\in[k-1]$.

By induction, this is bounded above by

$$
\begin{aligned}
&\sum_{a_k=3}^{n-3(k-1)} \frac{1}{a_k}\frac{k-1}{n-a_k}(\log(n-a_k))^{k-2}\\
&= \frac{k-1}{n}\sum_{a_k=3}^{n-3(k-1)}\left(\frac{1}{a_k}+\frac{1}{n-a_k}\right)(\log(n-a_k))^{k-2}\\
&\le \frac{k-1}{n}\left((\log n)^{k-2}\left(\sum_{a_k=3}^{n-3}\frac{1}{a_k}\right)+\sum_{a_k=3}^{n-3}\frac{1}{n-a_k}(\log(n-a_k))^{k-2}\right)\\
&\le \frac{k-1}{n}\left((\log n)^{k-1}+\frac{1}{k-1}(\log n)^{k-1}\right)=\frac{k}{n}(\log n)^{k-1},
\end{aligned}
$$

where the last inequality follows from the fact that $\log n = \int_1^n \frac{1}{x}\,dx$ and $\frac{1}{k-1}(\log n)^{k-1} = \int_1^n \frac{1}{x}(\log x)^{k-2}\,dx$. □

Lemma 17. *Let $G \sim G_{n,p}$. Then whp, for any $k \ge \log n$ the number A_k of 2-factors in G with at least k cycles satisfies*

$$A_{k+1} < \frac{n!(\log n)^{2k}p^n}{k!2^k}.$$

Proof. Note that it suffices to show that if A'_k is the number of 2-factors in G with *exactly* k cycles, then

$$\mathbb{E}(A'_k) \le \frac{(n-1)!(\log n)^{k-1}p^n}{(k-1)!2^k}. \tag{5}$$

Indeed, we then have

$$
\begin{aligned}
\mathbb{E}(A_{k+1}) &= \sum_{i=k+1}^{\frac{n}{3}} \mathbb{E}(A'_i) \le \sum_{i=k+1}^{n} \frac{(n-1)!(\log n)^{i-1}p^n}{(i-1)!2^i}\\
&\le \sum_{i=k+1}^{n} \frac{(n-1)!(\log n)^{k}p^n}{k!2^i} \le \frac{(n-1)!(\log n)^{k}p^n}{k!2^k}
\end{aligned}
$$

and Markov's inequality implies that

$$\mathbb{P}\left(A_{k+1} \ge \frac{n!(\log n)^{2k}p^n}{k!2^k}\right) \le \frac{1}{n(\log n)^k} \le \frac{1}{n^2}.$$

A union bound now gives that whp the result holds for all $\log n \le k \le \frac{n}{3}$.

To prove (5), it suffices to show that K_n contains at most $\frac{(n-1)!(\log n)^{k-1}}{(k-1)!2^k}$ 2-factors with exactly k cycles. We can count these as follows: Define an *ordered 2-factor* to be a 2-factor together with an ordering of its cycles. We can count the number of ordered 2-factors by first choosing some k-tuple $(a_1, a_2, \ldots, a_k)$ and counting those ordered 2-factors whose cycles have lengths $a_1, a_2, \ldots, a_k$ in that order. This can be done by simply ordering $V(G)$ and placing vertices 1 to a_1 in the first cycle, vertices $a_1 + 1$ to $a_1 + a_2$ in the second, etc. This procedure will count each ordered 2-factor of the appropriate type $(2a_1)(2a_2)\cdots(2a_k)$ times, and hence the number of these ordered

2-factors is $\frac{n!}{2^k a_1 a_2 \cdots a_k}$. Summing over all valid k-tuples, we have that the total number of ordered 2-factors with k cycles is

$$\sum \frac{n!}{2^k} \prod_{i=1}^{k} \frac{1}{a_i} \leq \frac{n!}{2^k} \frac{k}{n} (\log n)^{k-1}$$

by Lemma 16. But the number of ordered 2-factors (with k cycles) is simply $k!$ times the total number of such 2-factors, and the result follows immediately. □

We now need a lower bound on the total number of 2-factors. To do this we use the following well known result (see e.g. the proof of Lemma 2 in [8]).

Lemma 18. *Let r be even, and let H be an r-regular graph on n vertices. Then H contains at least $(\frac{r}{2n})^n n!$ 2-factors.*

Proof. It is easy to see that the number of perfect matchings of a d-regular bipartite graph B with vertex classes of size n equals the permanent of the incidence matrix of B. Egorychev [5] and Falikman [7] proved that the value of this permanent is at least $\left(\frac{d}{n}\right)^n n!$ (this confirmed a conjecture of van der Waerden). So we take an orientation of H in which every vertex has in- and out-degree $\frac{r}{2}$. Form a bipartite graph B whose vertex classes X, Y are each copies of $V(H)$, with an edge xy for each $x \in X$, $y \in Y$ such that $\vec{xy}$ is an edge of the orientation of H. Now B is $\frac{r}{2}$-regular and hence has at least $\left(\frac{r}{2n}\right)^n n!$ perfect matchings. But any perfect matching in B yields a 2-factor in H, and distinct matchings yield distinct 2-factors. □

Proof of Lemma 15. It suffices to show that whp $A_{k_0+1} < (\frac{r}{2n})^n n!$, and hence by Lemma 17 it suffices that

$$\left(\frac{r}{2n}\right)^n n! \geq \frac{n!(\log n)^{2k_0} p_0^n}{k_0! 2^{k_0}},$$

which holds as long as

$$\left(\frac{2np_0}{r}\right)^n \leq \frac{2^{k_0} k_0!}{(\log n)^{2k_0}}.$$

Noting that $k_0! \geq (\frac{k_0}{e})^{k_0}$, it suffices that

$$\begin{aligned} n \log \frac{2np_0}{r} &\leq k_0(\log k_0 + \log 2 - 2 \log\log n - 1) \\ &= \frac{\kappa n}{\log n}(\log n - 3\log\log n + \log\kappa + \log 2 - 1). \end{aligned}$$

Since $\log\kappa + \log 2 - 1 > 0$, this follows immediately from

$$\log \frac{2np_0}{r} \leq \frac{\kappa}{2} \leq \kappa\left(1 - \frac{3\log\log n}{\log n}\right).$$

□

Corollary 19. *Let $m = \frac{1}{2}(1-\eta)np_0$. Then G_1 contains a collection of at least m edge-disjoint 2-factors, each with at most k_0 cycles.*

Proof. By Corollary 14, G_1 contains a regular subgraph H of degree $r_1 = \left(1 - \frac{3\eta}{4}\right) np_0$. By Lemma 15 (noting that H is also a regular subgraph of G_0), we can remove 2-factors with at most k_0 cycles from H one by one as long as the degree of the resulting graph remains above $2np_0 e^{-\frac{\kappa}{2}}$. Recalling by (4) that $e^{-\frac{\kappa}{2}} = \frac{\eta}{16}$, this gives us a collection of $\frac{1}{2}\left(r_1 - \frac{\eta np_0}{8}\right) \geq \frac{1}{2}(1-\eta)np_0 = m$ 2-factors, each with at most k_0 cycles. □

6. Converting 2-factors into Hamilton cycles

Applying Corollary 19 yields a collection $F_1, F_2, \dots, F_m$ of edge-disjoint 2-factors in G_1, each with at most k_0 cycles, where

$$m = \frac{1}{2}(1-\eta)np_0 \quad \text{and} \quad k_0 = \frac{\kappa n}{\log n}. \tag{6}$$

Now we wish to convert these 2-factors into Hamilton cycles. Our proof develops ideas from Krivelevich and Sudakov [14]. Our strategy will be to show that for each F_i, we can connect the cycles of F_i into a Hamilton cycle using edges of $G_0 \backslash (F_1 \cup F_2 \cup \dots \cup F_m)$. We do this by incorporating the cycles of F_i one by one into a long path, and then finally closing this path to a Hamilton cycle.

Let

$$E_0 = \frac{\log n}{\log(\frac{\eta w_0}{20})} \quad \text{and} \quad E_1 = \frac{\log n}{\log(\frac{\eta^2 w_0}{10^5})}.$$

Definition 20. *Let $P = v_1 v_2 \dots v_\ell$ be a path in G_0 with endpoints $v_1 = x, v_\ell = y$, and let Γ be a spanning subgraph of G_0, whose edges are disjoint from those of P. Let v_i be a vertex of P such that $v_i y$ is an edge of Γ. A rotation of P about y with pivot v_i is the operation of deleting the edge $v_i v_{i+1}$ from P and adding the edge $v_i y$ to form a new path $v_1 v_2 \dots v_i y v_{\ell-1} v_{\ell-2} \dots v_{i+1}$ with endpoints x and $y' = v_{i+1}$. Call the edge $v_i v_{i+1}$ the broken edge of the rotation.*

Call a spanning subgraph F of G_0 a *broken 2-factor* if F consists of a collection of vertex-disjoint cycles together with a vertex-disjoint path, which we call the *long path* of F. The key to our proof of Theorem 3 is the following lemma:

Lemma 21. *Let P be a path in G_0. Let Γ be a spanning subgraph of G_0 whose edges are disjoint from those of P, such that*

$$e(G_2 \backslash \Gamma) \le \frac{\eta^6 n^2 p_0}{10^{17}} \tag{7}$$

and

$$\delta(\Gamma) \ge \frac{n\eta p_0}{4}. \tag{8}$$

Then there exists a sequence of at most $2E_0 + 2E_1$ rotations which can be performed on P, using edges of Γ, to produce a new path P', such that at least one of the following holds:

(i) *$|P| \ge \frac{\eta n}{200}$ and the endpoints x, y of P' are joined by an edge of Γ.*
(ii) *One of the endpoints x, y of P' is joined to a vertex outside P' by an edge of Γ.*

Before we prove Lemma 21, we will first show how it is used to prove Theorem 3. Our aim will be to convert each 2-factor F_i into a Hamilton cycle H_i in turn.

For this, we using the following algorithm: Let F^* be a broken 2-factor formed by removing an edge of F_i arbitrarily. Let j be the number of *steps* performed so far during the conversion process (both on F^* and on the 2-factors which have already been converted into Hamilton cycles). Here a step is taken to mean a single application of Lemma 21 to either obtain a broken 2-factor with fewer cycles or to close a Hamilton path to a cycle. Let

$$\Gamma_j = G_0 \backslash (H_1 \cup \dots \cup H_{i-1} \cup F^* \cup F_{i+1} \cup \dots \cup F_m).$$

Then since $\delta(G_0) \ge \delta(G_1) \ge (1 - \frac{\eta}{2})np_0$ by Lemma 10, and

$$\Delta(H_1 \cup \dots \cup H_{i-1} \cup F^* \cup F_{i+1} \cup \dots \cup F_m) \le 2m = (1-\eta)np_0,$$

we have that $\delta(\Gamma_j) \geq \frac{\eta n p_0}{4}$. Assume also that $e(G_2 \backslash \Gamma_j) \leq 4jE_1$, and that

$$j \leq k_0 m \overset{(6)}{\leq} \frac{\kappa n^2 p_0}{2 \log n}. \tag{9}$$

Then

$$e(G_2 \backslash \Gamma_j) \overset{(9)}{\leq} \frac{2\kappa n^2 p_0}{\log(\frac{\eta^2 w_0}{10^5})} \overset{(4)}{\leq} \frac{\eta^6 n^2 p_0}{10^{17}}. \tag{10}$$

Let P^* be the long path of F^*. If P^* is a Hamilton path, then Lemma 21 applied with $\Gamma = \Gamma_j$ and $P = P^*$ shows that after at most $2E_0 + 2E_1$ rotations we can close P^* to a Hamilton cycle H_i. We then move on to the next 2-factor F_{i+1}. If there are no 2-factors remaining (i.e., if $i = m$), then we have constructed the required set of m edge-disjoint Hamilton cycles.

Otherwise by Lemma 21, after at most $2E_0 + 2E_1$ rotations we can either join an endpoint of P^* to a vertex x outside P^*, or we can close P^* to form a cycle C^*. In the first case x will be a vertex of some cycle C_x of F^*, and we can delete one of the edges of C_x incident to x to form a new path P^{**} which incorporates C_x. We then redefine F^* to be the union of P^{**} with the remaining cycles of F_i; this is a broken 2-factor with long path P^{**}, which has one cycle fewer than before. The algorithm then proceeds to the next step.

In the second case we have that $|C^*| = |P^*| \geq \frac{\eta n}{200}$. Now if $|C^*| \leq n - \frac{\eta n}{200}$, then by Lemma 7 we have $e_{G_2}(V(C^*), [n] \backslash V(C^*)) \geq \frac{\eta^2 n^2 p_0}{50000}$. Since

$$e(G_2 \backslash \Gamma_j) \overset{(10)}{\leq} \frac{\eta^6 n^2 p_0}{10^{17}} < \frac{\eta^2 n^2 p_0}{50000},$$

there must exist an edge in Γ_j from some vertex y of C^* to a vertex outside C^*. On the other hand, if $|C^*| \geq n - \frac{\eta n}{200}$ then applying Lemma 11 with $H = \Gamma_j$ and $A = [n] \backslash V(C^*)$ implies the same. We then delete one of the edges of C^* incident to y and extend the resulting path as in the first case.

We run this algorithm until the last 2-factor F_m has been converted into a Hamilton cycle. Now since each step either reduces the number of cycles in a broken 2-factor or closes a Hamilton path to a Hamilton cycle, the algorithm will terminate after at most $k_0 m$ steps. It remains to justify our assumption that $e(G_2 \backslash \Gamma_j) \leq 4jE_1$, for each j (i.e., at each step). We can prove this by induction: $G_2 \subseteq \Gamma_0$, and since at most $2E_0 + 2E_1$ rotations are performed at each step, it follows that $e(\Gamma_j \backslash \Gamma_{j+1}) \leq 2E_0 + 2E_1 + 2 \leq 4E_1$. So $e(G_2 \backslash \Gamma_{j+1}) \leq 4jE_1 + 4E_1 = 4(j+1)E_1$, as required. □

It remains to prove Lemma 21. Our strategy will be as follows: We can assume that whenever we have an endpoint x of a path P' obtainable by fewer than $2E_0 + 2E_1$ rotations of P, then all of its neighbours lie on P' (otherwise (ii) holds). So assuming this, we try to form some large sets A, B, such that for any $a \in A, b \in B$, we can obtain a path P' with endpoints a, b. Then Lemma 7 together with (7) will allow us to close P' to a cycle.

We will (eventually) obtain the sets A, B by dividing the path P into two segments, and showing that we can perform a large number of rotations using only those pivots which lie all in one half or all in the other. This will ensure that the rotations involving the first endpoint do not interfere with those involving the second endpoint, and vice versa. In order to do this we need to show two things: Firstly, there exists a subset C_1 of the first segment of the path and a subset C_2 of the second segment of the path, such that for $i = 1, 2$, each vertex in C_i has many neighbours which also lie in C_i. In fact since we are concerned with the successors or predecessors of the neighbours rather than the neighbours themselves, we will require the neighbours to lie in the interior (taken along P) of C_i. Secondly we will show that we can force the endpoints of the path to actually lie in these subsets.

We can accomplish the latter property by showing that the subsets are sufficiently large and by performing rotations until each endpoint lies in its corresponding subset. The obvious problem with this is that as we perform these rotations, C_1 and C_2 will cease to lie in their respective segments. So instead of defining C_1, C_2 immediately, we construct a subset C of $V(P)$ with certain properties; then after rotating so that a, b lie in C, we will define C_1, C_2 to be subsets of C, and the properties of C will ensure that the vertices of each C_i have many neighbours in $int(C_i)$. Here the *interior* $int(C_i)$ of C_i is the set of elements x of C_i such that both of the vertices adjacent to x along P also lie in C_i.

We start with the following lemma, where $k = \log n$.

Lemma 22. *Let $\varepsilon = \frac{\eta}{600}$, and $P \subseteq G_0$ be a path, $n' := |P| \geq \frac{\eta n}{200}$. Let Γ be a spanning subgraph of G_0, edge-disjoint from P, which satisfies (7). Let $W_1, W_2, \dots, W_k$ be a partition of P into segments whose lengths are as equal as possible. Then there exists $S \subseteq [k]$ with $|S| \geq (1-\varepsilon)k$, and subsets $W_i' \subseteq W_i$ for each $i \in S$ with $|int(W_i')| \geq (1-\varepsilon)\frac{n'}{k}$, such that for any $x \in W_i'$, and for at least $|S| - \varepsilon k$ of the sets W_j', $|N_\Gamma(x) \cap int(W_j')| \geq \frac{\eta p_0 n'}{20k}$.*

Proof. We start with $S = [k]$ and $W_i' = W_i$, and as long as there exists $i \in S$ and a vertex $x \in W_i'$, such that $|N_\Gamma(x) \cap int(W_j')| \leq \frac{\eta p_0 n'}{20k}$ for at least εk values of $j \in S$, we remove x. (In this case, call x *weakly connected* to W_j'.) Further, if at any stage there exists $i \in S$ such that $|int(W_i')| \leq (1-\varepsilon)\frac{n'}{k}$, then we remove i from S.

We claim that this process must terminate before $\frac{\varepsilon^2 n'}{4}$ vertices are removed. Indeed, suppose we have removed $\frac{\varepsilon^2 n'}{4}$ vertices and let R be the set of removed vertices. Now $|R| = \frac{\varepsilon^2 n'}{4}$, and so $\sum_{i=1}^k |int(W_i')| \geq (1 - \frac{3\varepsilon^2}{4})n'$. Hence we have $|int(W_i')| \geq (1-\varepsilon)\frac{n'}{k}$ for at least $1 - \frac{3\varepsilon k}{4}$ values of i, i.e., at most $\frac{3\varepsilon k}{4}$ indices have been removed from our original set S. So each $x \in R$ is still weakly connected to at least $\frac{\varepsilon k}{4}$ sets W_i' with $i \in S$. For each $i \in S$, let $WC(i)$ be the set of vertices $x \in R$ which are weakly connected to W_i'.

Now consider the set $S_0 = \{i \in S \mid |WC(i)| \geq \frac{\varepsilon^3 n'}{32}\}$. Note that if $i \in S_0$, then

$$\frac{|int(W_i')|}{n}\frac{|WC(i)|}{n}np_2 \geq \frac{(1-\varepsilon)\varepsilon^3(n')^2\eta w_0 \log n}{128kn^2} \geq \frac{\eta^6 w_0}{10^{16}} \geq 700.$$

So Lemma 7 implies that the number of edges of G_2 between $int(W_i')$ and $WC(i)$ is at least

$$\frac{13}{14}\frac{\eta p_0 |WC(i)|(1-\varepsilon)n'}{4k} \geq \frac{\eta p_0 |WC(i)| n'}{10k}.$$

But by the definition of $WC(i)$, Γ contains at most $\frac{\eta p_0 |WC(i)| n'}{20k}$ edges between $int(W_i')$ and $WC(i)$, and hence $G_2 \backslash \Gamma$ contains at least this many edges between $int(W_i')$ and $WC(i)$.

Observe that $\sum_{i \in S} |WC(i)| \geq |R|\frac{\varepsilon k}{4} = \frac{\varepsilon^3 n' k}{16}$, since each $x \in R$ is weakly connected to W_i' for at least $\frac{\varepsilon k}{4}$ values of $i \in S$. But since $\sum_{i \in S \backslash S_0} |WC(i)| \leq \frac{\varepsilon^3 n' k}{32}$, we have that $\sum_{i \in S_0} |WC(i)| \geq \frac{\varepsilon^3 n' k}{32}$. Hence $G_2 \backslash \Gamma$ contains at least

$$\frac{\eta n' p_0}{20k}\sum_{i \in S_0} |WC(i)| \geq \frac{\eta \varepsilon^3 (n')^2 p_0}{640} \geq \frac{\eta^3 \varepsilon^3 n^2 p_0}{64 \cdot 4 \cdot 10^5} \geq \frac{\eta^6 n^2 p_0}{10^{16}}$$

edges, which would contradict (7). This proves the claim, and now we consider the sets W_i' as they are at the point at which the process terminates. It is immediate that for each $i \in S$, W_i' satisfies the requirements of the lemma. But since we have removed at most $\frac{3\varepsilon k}{4}$ indices from our original set S, we also have $|S| \geq (1-\varepsilon)k$. $\square$

Let $C = \bigcup_{i \in S} W_i'$ and note that $|C| \geq (1-\varepsilon)^2 n'$. We now need to show that the set of vertices which we can make into endpoints of P with relatively few rotations is of size at least $2\varepsilon n$. Doing this gives immediately that one of these endpoints must be an element of C.

Lemma 23. *Let P be a path in G_0 with endpoints a, b, and Γ be a spanning subgraph of G_0, edge-disjoint from P, which satisfies (8). Let S_t be the set of vertices $x \in P \backslash \{b\}$ such that a path P' with endpoints x, b can be obtained from P by at most t rotations. Then $|S_{t+1}| \geq \frac{1}{2}|B_\Gamma(S_t)| - |S_t|$.*

Proof. For a vertex $x \in P$, let x^-, x^+ be the predecessor and successor of x along P, respectively. Let $T = \{x \in B_\Gamma(S_t) \mid x^-, x^+ \notin S_t\}$. If $x \in T$, then since neither x nor any of its neighbours on P are in S_t, the neighbours of x are preserved by every sequence of at most t rotations of P; i.e., x^+ and x^- are adjacent to x along any path obtained from P by at most t rotations. It follows that one of x^-, x^+ must be in S_{t+1}. Indeed, starting from P, we can perform t rotations to obtain a path with endpoints z, b, such that zx is an edge of Γ. Now by one further rotation with pivot x and broken edge either xx^+ or xx^-, we obtain a path whose endpoints are either x^+, b or x^-, b.

Now let $T^+ = \{x^+ \mid x \in T, x^+ \in S_{t+1}\}$ and $T^- = \{x^- \mid x \in T, x^- \in S_{t+1}\}$. It follows from the above that either $|T^+| \geq \frac{|T|}{2}$ or $|T^-| \geq \frac{|T|}{2}$, and both of these are subsets of S_{t+1}. Hence $|S_{t+1}| \geq \frac{|T|}{2} \geq \frac{1}{2}(|B_\Gamma(S_t)| - 2|S_t|)$. $\square$

Corollary 24. *Either $|S_{E_0}| \geq \frac{\eta n}{200}$, or some element of S_{E_0} has a neighbour in Γ lying outside P (or both).*

Proof. It suffices to show that as long as $|S_t| \leq \frac{\eta n}{200}$, and assuming no element of S_t has a neighbour outside P, we have that $|S_{t+1}| \geq \min\{\frac{\eta w_0}{20}|S_t|, \frac{\eta n}{200}\}$. We apply Lemma 11, setting $H = \Gamma$ and $A = S_t$. Now in the notation of Lemma 11, $\delta_A \geq \delta(H[V(P)]) \geq \frac{\eta n p_0}{4}$ by (8), and so we have that either (i) $|B_\Gamma(S_t)| \geq (\frac{\eta w_0}{8} - 3)|S_t|$, or (ii) $3|S_t| + |B_\Gamma(S_t)| \geq \frac{\eta n}{28}$. If (i) holds then

$$|S_{t+1}| \geq \frac{1}{2}|B_\Gamma(S_t)| - |S_t| \geq (\frac{\eta w_0}{20} + 1)|S_t| - |S_t| = \frac{\eta w_0}{20}|S_t|.$$

On the other hand if (ii) holds then

$$|S_{t+1}| \geq \frac{1}{2}|B_\Gamma(S_t)| - |S_t| \geq \frac{\eta n}{56} - \frac{5}{2}|S_t| \geq \frac{\eta n}{200}.$$

$\square$

Corollary 24 implies that if our path P in Lemma 21 satisfies $|P| < \frac{\eta n}{200}$, then alternative (ii) of Lemma 21 holds. So suppose that $|P| \geq \frac{\eta n}{200}$. Then we can apply Lemma 22 to obtain a set $C = \bigcup_{i \in S} W_i'$. Now since $\frac{\eta n}{200} + |C| > n' = |P|$, we have that either alternative (ii) of Lemma 21 holds, or we can obtain in at most E_0 rotations a path with endpoints a', b such that $a' \in C$. Suppose we are in the latter case. Repeating the argument for b gives us a path P''' with endpoints $a', b' \in C$ which is obtained from P by at most $2E_0$ rotations.

Call a segment W_i of P *unbroken* if none of the rotations by which P''' is obtained had their pivot in W_i. Note that each unbroken segment is still a segment of P''' in the sense that the vertices are consecutive and their adjacencies along the path are preserved. Since we have arrived at the path P''' by at most $2E_0$ rotations, there are at least $k-2E_0$ unbroken segments W_i, and for at least $k-2E_0-\varepsilon k$ of these we have that $i \in S$. Noting that $E_0 \le \frac{k}{10}$, we are still left with at least $\frac{3k}{5}$ unbroken segments W_i for which $i \in S$. Let us relabel these segments W_i according to their order along P''', and take $C_1 = \bigcup_{i \le \frac{3k}{10}} W_i'$ and $C_2 = \bigcup_{i > \frac{3k}{10}} W_i'$. Note that for any $x \in C$ (and in particular for $x \in C_1$ and for a'),

$$|N_\Gamma(x) \cap int(C_1)| \ge \frac{\eta p_0 n'}{20k}\left(\frac{3k}{10} - \varepsilon k\right) \ge \frac{\eta p_0 n'}{70} \ge \frac{\eta^2 n p_0}{14000}. \tag{11}$$

Now let x_0 be a vertex separating C_1, C_2 along P''', and let x_0 divide P''' into paths $P_{a'}, P_{b'}$. Let U_t be the set of endpoints of paths obtainable by at most t rotations about a' with pivots lying only in $int(C_1)$. So these rotations affect only $P_{a'}$, and $P_{b'}$ is left intact in each of the resulting paths.

Lemma 25. *Suppose that* $|U_t| \le \frac{\eta^2 n}{10^6}$. *Then*

$$|B_\Gamma(U_t) \cap int(C_1)| \ge \min\left\{\frac{\eta^2 n}{150000}, \frac{\eta^2 w_0}{40000}|U_t|\right\}.$$

Proof. Let $u = |U_t|$ and $u' = |B_\Gamma(U_t) \cap int(C_1)|$. Consider the case $\frac{\log n}{u p_0} \ge \frac{7}{2}$. Then similarly to the proof of Lemma 11,

$$\begin{aligned} \frac{u\eta^2 n p_0}{14000} &\overset{(11)}{\le} \sum_{x \in U_t} |N_\Gamma(x) \cap int(C_1)| \le e_\Gamma(U_t, B_\Gamma(U_t) \cap int(C_1)) + 2e_\Gamma(U_t) \\ &\le 2(u+u')\log n + 4u\log n, \end{aligned}$$

whence

$$u' \ge \frac{(\eta^2 n p_0 - 84000\log n)u}{28000\log n} = \frac{(\eta^2 w_0 - 84000)u}{28000} \ge \frac{\eta^2 w_0}{40000}|U_t|.$$

On the other hand, if $\frac{\log n}{u p_0} \le \frac{7}{2}$ then

$$\begin{aligned} \frac{u\eta^2 n p_0}{14000} &\le \sum_{x \in U_t} |N_\Gamma(x) \cap int(C_1)| \le e_\Gamma(U_t, B_\Gamma(U_t) \cap int(C_1)) + 2e_\Gamma(U_t) \\ &\le 7up_0(u+u') + 14u^2 p_0 \end{aligned}$$

and so $3u + u' \ge \frac{\eta^2 n}{98000}$. Hence $u' \ge \frac{\eta^2 n}{150000}$. □

Corollary 26. $|U_{E_1}| \ge \frac{\eta^2 n}{10^6}$.

Proof. It suffices to prove that for each t such that $|U_t| \le \frac{\eta^2 n}{10^6}$, either $|U_{t+1}| \ge \frac{\eta^2 n}{10^6}$ or $|U_{t+1}| \ge \frac{\eta^2 w_0}{10^5}|U_t|$ (or both). Similarly to Lemma 23, we have that $|U_{t+1}| \ge \frac{1}{2}|B_\Gamma(U_t) \cap int(C_1)| - |U_t|$, and now Lemma 25 immediately gives the result. □

Proof of Lemma 21. Suppose first that $|P| \leq \frac{\eta n}{200}$. Then Corollary 24 immediately implies that we can obtain a path, one of whose endpoints has a neighbour in Γ lying outside P, in at most E_0 rotations. So we may assume that $|P| \geq \frac{\eta n}{200}$, and hence the conditions of Lemma 22 are satisfied. Now we proceed as above to obtain a path $P''' = P_{a'} \cup P_{b'}$, with endpoints a', b' and with sets C_1, C_2 satisfying (11), such that $a' \in C_1 \subseteq P_{a'}$ and $b' \in C_2 \subseteq P_{b'}$.

Let $U = U_{E_1}$. Now similarly, we can rotate about b' using only pivots in C_2, to obtain another set U' of endpoints in another E_1 rotations, such that $|U'| \geq \frac{\eta^2 n}{10^6}$. Now by Lemma 7, there are at least $\frac{13}{14}\frac{\eta^4 n^2 p_0}{10^{12}}$ edges of G_2 between U and U', and since by (7) $G_2 \backslash \Gamma$ contains fewer edges than this, it follows that there exists an edge xy of Γ between U and U'. Now by the definition of U and U', we can obtain a path P'' with endpoints x, b' from P''' by a sequence of at most E_1 rotations, none of which affect the second half $P_{b'}$ of the path P'''. From P'' we can obtain a path P' with endpoints x, y by at most E_1 rotations. □

7. Acknowledgements

We thank the referees for their detailed comments.

References

[1] B. Bollobás, The evolution of sparse graphs, *Graph theory and Combinatorics*, Academic Press, London (1984), 35–57.

[2] B. Bollobás, *Random Graphs*, Academic Press, London, 1985.

[3] B. Bollobás and A. Frieze, On matchings and Hamiltonian cycles in random graphs. *Random graphs '83* (Poznan, 1983), North-Holland Math. Stud., 118, North-Holland, Amsterdam (1985), 23–46.

[4] D. Christofides, D. Kühn and D. Osthus, Edge-disjoint Hamilton cycles in graphs, preprint.

[5] G. Egorychev, The solution of the Van der Waerden problem for permanents, *Dokl. Akad. Nauk SSSR* **258** (1981), 1041–1044.

[6] P. Erdős and A. Rényi, On Random Graphs. I., *Publicationes Mathematicae* **6** (1959), 290–297.

[7] D. Falikman, A proof of the Van der Waerden problem for permanents of a doubly stochastic matrix, *Mat. Zametki*, **29** (1981) 931–938.

[8] A. Frieze and M. Krivelevich, On packing Hamilton cycles in ε-regular graphs, *J. Combin. Theory B* **94** (2005), 159–172.

[9] A. Frieze and M. Krivelevich, On two Hamilton cycle problems in random graphs, *Israel Journal of Mathematics* **166** (2008), 221–234.

[10] A. Frieze and M. Krivelevich, Packing Hamilton cycles in random and pseudo-random hypergraphs, preprint.

[11] A. Frieze, M. Krivelevich and P.-S. Loh, Packing tight Hamilton cycles in 3-uniform hypergraphs, preprint.

[12] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, Wiley, 2000.

[13] J.H. Kim and N.C. Wormald, Random matchings which induce Hamilton cycles, and Hamiltonian decompositions of random regular graphs, *J. Combin. Theory B* **81** (2001), 20–44.

[14] M. Krivelevich and B. Sudakov, Sparse pseudo-random graphs are Hamiltonian, *J. Graph Theory* **42** (2003), 17–33.

[15] D. Kühn, D. Osthus and A. Treglown, Hamilton decompositions of regular tournaments, *Proc. London Math. Soc.* **101** (2010), 303-335.

[16] C. St. J. A. Nash-Williams, Hamiltonian arcs and circuits, in *Recent Trends in Graph Theory*, Springer 1971, 197–210.

[17] L. Pósa, Hamiltonian circuits in random graphs, *Discrete Mathematics* **14** (1976), 359–364.

[18] C. Thomassen, Edge-disjoint Hamiltonian paths and cycles in tournaments, *Proc. London Math. Soc.* **45** (1982), 151–168.

[19] W. T. Tutte, The factors of graphs, *Canad. J. Math.* **4** (1952), 314–328.

Fiachra Knox, Daniela Kühn & Deryk Osthus
School of Mathematics
University of Birmingham
Edgbaston
Birmingham
B15 2TT
UK

E-mail addresses:
{knoxf,kuehn,osthus}@maths.bham.ac.uk