Hamilton cycles in 3-out

Tom Bohman*and Alan Frieze†
Department of Mathematical Sciences,
Carnegie Mellon University,
Pittsburgh PA 15213,
USA.

August 1, 2007

Abstract

Let $G_{3-\mathrm{out}}$ denote the random graph on vertex set $[n]$ in which each vertex chooses 3 neighbors uniformly at random. Note that $G_{3-\mathrm{out}}$ has minimum degree 3 and average degree 6. We prove that the probability that $G_{3-\mathrm{out}}$ is Hamiltonian goes to 1 as n tends to infinity.

1 Introduction

One of the natural questions for a sparse random graph model is whether or not a random instance is Hamiltonian **whp**. The Hamiltonicity threshold for the basic models $G_{n,m}$ and $G_{n,p}$ was established quite precisely by Komlós and Szemerédi [13]. See Bollobás [3], Ajtai, Komlós and Szemerédi [1] and Bollobás and Frieze [4] for refinements. If $m = \frac{n}{2}(\log n + \log\log n + c)$ then

$$\lim_{n\to\infty} \mathbf{Pr}(G_{n,m} \text{ is Hamiltonian}) = \lim_{n\to\infty} \mathbf{Pr}(\delta(G_{n,m}) \geq 2) = e^{-e^{-c}}$$

where δ denotes minimum degree. The lesson here is that the minimum degree seems to be the most important factor in determining the likelihood of Hamiltonicity in

*Supported in part by NSF grant DMS-0401147.
†Supported in part by NSF grant CCF-0502793.

a random graph. This naturally leads to the consideration of models where the minimum degree condition is automatically satisfied. If we simply condition on $G_{n,m}$ having minimum degree 2 then about $\frac{1}{6}n \log n$ random edges are needed (see Bollobás, Fenner and Frieze [5]). The barrier to Hamiltonicity in this model is the existence of three vertices of degree 2 that have a common neighbor. So, we expect even sparser random graph models with minimum degree at least three to have Hamilton cycles **whp** (note that if we have a linear number of vertices of degree 2 it may be very likely that there are three vertices of degree 2 that have a common neighbor). We focus our attention here on random graphs with only a linear number of edges and minimum degree 3. Bollobás, Fenner and Frieze [6] showed that if we condition on $G_{n,m}$ having minimum degree 3 then $m = 128n$ random edges suffice to give a Hamilton cycle **whp**. We conjecture that for any $c \geq 3/2$ the random graph $G_{n,cn}$ conditioned on minimum degree 3 has a Hamilton cycle with high probability.

Another well-studied sparse random graph is the random regular graph. Let G_r denote a graph chosen uniformly at random from the set of r-regular graphs with vertex set $[n]$. Robinson and Wormald [16] showed G_r is Hamiltonian **whp** for $r \geq 3$, r constant. Allowing r to grow with n presented some challenges, but they have now been resolved (see Cooper, Frieze and Reed [8] and Krivelevich, Sudakov, Vu and Wormald [14]).

We now come to the topic of this paper: $G_{m\text{-out}}$. We begin with vertex set $V = [n]$. Each $v \in V$ indepently chooses m random out-neighbors to create the random digraph $D_{m\text{-out}}$. We then obtain $G_{m\text{-out}}$ by ignoring orientation. Note that $G_{m\text{-out}}$ is a graph with minimum degree m and average degree $2m$. Note further that there is the potential for multiple edges, but the expected number is $O(m^2)$. So, we can either allow these multiple edges or condition on them not occurring. Since the probability that there are no multiple edges is bounded away from zero, any property that holds **whp** in the model that allows multiple edges also holds **whp** when we condition on no multiple edges. The Hamiltonicity of $G_{m\text{-out}}$ was first discussed by Fenner and Frieze [10]. They showed that $G_{23-\text{out}}$ is Hamiltonian **whp**. This was improved to $G_{10-\text{out}}$ by Frieze [11] and to $G_{5-\text{out}}$ by Frieze and Łuczak [12]. Cooper and Frieze [7] showed that the digraph $D_{2-\text{in},2-\text{out}}$ (here each vertex chooses 2 out-neighbors and 2 in-neighbors uniformly at random) has a directed Hamilton cycle **whp**. This implies that $G_{4-\text{out}}$ is Hamiltonian **whp**. So the main open question is whether $G_{3-\text{out}}$ is Hamiltonian **whp** or not (note that, **whp** $G_{2-\text{out}}$ contains 3 vertices of degree 2 that have a common neighbor and is therefore not Hamiltonian **whp**). We settle this question:

Theorem 1.

$$\lim_{n\to\infty} \mathbf{Pr}(G_{3-\text{out}} \textit{ is Hamiltonian}) = 1.$$

We give a proof of Theorem 1 in Section 2. The proof hinges on a few key Lemmas, the proofs of which are given in later sections.

2 Proof of Theorem 1

The proof has two main steps. Before the first step, we set aside a small set of edges that will not be used until the second step. In the first step we find a collection of vertex-disjoint paths and cycles that covers the vertex set. The set-aside edges are then used in the second step to transform this collection of cycles and paths into a Hamilton cycle.

Of course, we need some notation. For a digraph $D = (V, A)$ and $v \in V$ let $d^+_D(v), d^-_D(v)$ denote the out-degree and in-degree, respectively, of v in D. Let $V = [n]$ and let Ω denote the set of digraphs with vertex set V in which $d^+_D(v) = 3$ for $v \in V$. It will be convenient to let the out-arcs for each vertex be chosen with replacement and loops allowed. Thus $|\Omega| = n^{3n}$ and D_{3-out} is sampled uniformly from Ω. For a digraph D let Γ_D be the underlying graph obtained by ignoring orientation. Thus $G_{3-out} = \Gamma_{D_{3-out}}$.

We separate the edge set as follows. We begin by randomly selecting a set of vertices K and setting aside the third out-arc of each vertex in K. Formally, for $K \subseteq V$ define Ω_K to be the set of digraphs D for which $d^+_D(v) = 2$ for all $v \in K$ and $d^+_D(v) = 3$ for all $v \in V \setminus K$. Thus, $|\Omega_K| = n^{3(n-k)+2k} = n^{3n-k}$. Further, set

$$\Omega_k = \bigcup_{K \in \binom{[n]}{k}} \Omega_K.$$

We work with the uniform distribution on $\Omega_k \times [n]^k$. The first component gives all but k of the arcs in D_{3-out} and the second component determines the heads of the remaining k arcs. For $D \in \Omega_k$ we let $K(D) = \{v \in V : d^+_D(v) = 2\}$ and $L(D) = \{v \in K(D) : d^-_D(v) = 0\}$. We add the third out-arc of each vertex $v \in L(D)$ back into the graph (thereby arriving at graph with minimum degree 3). The resulting digraph is used for the first part of the proof. The remaining set-aside arcs (i.e. the third out-arc from each $v \in K \setminus L$) are used for the second part of the argument.

Set

$$k = \frac{n}{\log^{1/2} n}.$$

Let $D_1 = (V, A_1)$ denote the digraph drawn uniformly at random from Ω_k. Let A be the set of arcs then determined by the randomly chosen k-tuple in $[n]^k$. Let A_2 be the set of arcs in A with tails in $L(D_1)$ and set $A_3 = A \setminus A_2$. Let E_3 be the set of edges given by ignoring the orientations of the arcs in A_3. Set $D_2 = D_1 + A_2$ and $G_2 = \Gamma_{D_2}$. In the first part of the argument we find a collection of paths and cycles in G_2 that covers the vertex set. The arcs in A_3 are then used in the second part of the argument to convert this collection of paths and cycles into a Hamilton cycle. Since G_2 is a super-graph of G_{2-out} we have the following:

Lemma 2 (Fenner, Frieze [9])**.** *G_2 is connected* **whp***.*

Define a **simple 2-matching** in a graph H to be a subgraph of H with maximum degree 2, i.e. a collection of vertex-disjoint paths and cycles. We assume that the vertex set of a simple 2-matching in H is $V(H)$. So some paths in a 2-matching could be isolated vertices. If F is a 2-matching let $\kappa(F)$ be the number of components in F and let $\rho(F)$ be the number of path components in F. If F is a simple 2-matching in H and $x \in V(H)$ then we define F^x to the component in F that contains the vertex x. Finally, if F is a simple 2-matching, we let $\mathcal{P}(F)$ denote the set of vertices covered by the path components of F. The first step in the proof is establish that G_2 has a simple 2-matching with few components.

Lemma 3. Whp *G_2 contains a simple 2-matching F_2 with $\kappa(F_2) \leq \frac{6n}{\log n}$.*

Lemma 3 is proved in Section 3. The proof uses an extension of the Tutte 1-factor theorem.

The **Extension-Rotation Procedure** defined in Figure 1 converts F_2 into a Hamilton cycle. This procedure takes as input the graph $G_2 = ([n], E)$, the set of arcs A_3 and a 2-matching $F = F_2$ in G_2, and, if successful, it produces a Hamilton cycle. Appealing to Lemmas 2 and 3, we assume G_2 is connected and $\kappa(F_2) \leq 6n/\log n$. We reveal (some of) the random arcs in A_3 one at a time as the algorithm proceeds. In the course of the procedure we iteratively modify the 2-matching F and occasionally add edges from E_3 to the edge set E. In each iteration we attempt to reduce $\kappa(F)$, extend the longest path in F while keeping $\kappa(F)$ and $\rho(F)$ fixed, or replace a path component in F with a cycle component (thereby reducing $\rho(F)$). We use edges from E whenever possible. We resort to the arcs in A_3 only when we cannot perform any of the operations using only the edges in E and only to replace a path component with a cycle component. Since $\rho(F)$ only increases in situations where $\rho(F) = 0$, and in this case we increase $\rho(F)$ to 1 and simultaneously reduce $\kappa(F)$, it follows that the number of arcs from A_3 that we use is bounded by $\kappa(F) + \rho(F)$, which we assume is at most $12n/\log n$.

In order to analyze this procedure, we condition. For a graph G we set

$$\Omega_k^G = \{D \in \Omega_k : \Gamma_D = G\}.$$

Figure 1: **Extension Rotation Procedure**

While $\kappa(F) + \rho(F) > 1$:

If $\mathcal{P}(F) = \emptyset$:

Since G_2 is connected, there exists an edge $\{x, y\} \in E$ such that F^x and F^y are distinct cycle components.

Replace F^x and F^y with a single path using the edge $\{x, y\}$.

If $\mathcal{P}(F) \neq \emptyset$:

Let P be a longest path in F.

Let X be the set of vertices spanned by P.

Let END be the set of end-points of paths that span X.

If there exists $\{x, y\} \in E$ such that $x \in$ END and $y \notin \mathcal{P}(F)$:

Replace P and F^y with a single path.

If there exists $\{x, y\} \in E$ such that $x \in$ END and $y \in \mathcal{P}(F) \setminus X$:

Replace P and F^y with two paths by extending a path on X ending with x along F^y to one of the endpoints of F^y.

If $N(x) \subseteq X$ for all $x \in$ END:

Choose an ordering of the elements of END uniformly at random.

Consider the elements of END one at a time. For each such $x \in$ END:

Let END(x) be the set of endpoints of paths spanning X such that x is the other endpoint.

Check whether or not $x \in K \setminus L$.

If $x \in K \setminus L$ determine the arc $(x, y_x) \in A_3$.

If $x \in K \setminus L$ and $y_x \in$ END(x) then use $\{x, y_x\}$ to replace P with a single cycle, add $\{x, y_x\}$ to E, and move onto the next iteration of the procedure.

If $x \notin K \setminus L$ or $y_x \notin$ END(x) then move onto the next element of END.

If $x \notin K \setminus L$ or $y_x \notin$ END(x) for all $x \in$ END then **Fail.**

We prove that the Extension Rotation Procedure succeeds **whp**, and thereby prove Theorem 1, by working with the partition $\mathcal{F}$ of $\Omega_k \times [n]^k$ in which (D_1, A) and (D_1', A') are in the same part if $\Gamma_{D_1} = \Gamma_{D_1'}$ (in other words, each part is of the form $\Omega_k^G \times [n]^k$ for some graph G). The principal virtue of this partition is that if we condition on $(D_1, A) \in \Omega_k^G$ then $K \setminus L$ is not determined, it can be viewed as a random set (L, on the other hand, is determined). Of course, the set $K \setminus L$ is not a uniform random set, but we will see below that it is approximately uniform.

We define a filtration $\mathcal{F} = \mathcal{F}_0, \mathcal{F}_1, \dots$ of $\Omega_k \times [n]^k$. The partition $\mathcal{F}_1$ is determined by the third out-arcs of the vertices in L: (D_1, A) and (D_1', A') are in the same part of $\mathcal{F}_1$ if $\Gamma_{D_1} = \Gamma_{D_1'}$ and the arcs in A with tails in $L(D_1) = L(D_1')$ are identical to the arcs in A' with tails in $L(D_1') = L(D_1)$. The remaining partitions are then determined by the execution of the Extension-Rotation Procedure. We refine the partition each time we consider an element of END. Thus, each time we arrive at a 2-factor F such that $\mathcal{P}(F) \neq \emptyset$ and $N(x) \subseteq X$ for all $x \in \text{END}$, where X is the vertex set of a longest path in F and END is the set of endpoints of paths spanning X, we will likely refine the partition many times as we will likely have to check many elements of END before we find $x \in \text{END}$ such that $x \in K \setminus L$ and $y_x \in \text{END}(x)$. Let C_i be the first i vertices checked in this step of the Extension Rotation Procedure. Further, define

$$I_i = C_i \cap (K \setminus L) \qquad \text{and} \qquad O_i = L \cup (C_i \setminus K).$$

In words, I_i are the vertices among the first i checked that are in $K \setminus L$ and O_i is L together with the vertices among the first i checked that are out of $K \setminus L$. If ordered pairs (D_1, A) and (D_1', A') are in the same part of $\mathcal{F}_i$ then $\Gamma_{D_1} = \Gamma_{D_1'}$, $I_i((D_1, A)) = I_i((D_1', A'))$, $O_i((D_1, A)) = O_i((D_1', A'))$ and the arcs in A with tails in I_i are identical to arcs in A' with tails in I_i. If $\mathcal{X}$ is a part in the partition $\mathcal{F}_i$ then we say that a digraph $D_1 \in \Omega_k$ is in $\mathcal{X}$ if there exists A determined by a vector in $[n]^k$ such that $(D_1, A) \in \mathcal{X}$.

We show that throughout the algorithm nearly every part $\mathcal{X}$ of the partition $\mathcal{F}_i$ has the property that nearly all the digraphs in $\mathcal{X}$ satisfy certain structural conditions (i.e.. small maximum degree, no small dense subgraphs, etc.). For $\mathcal{X}$ a part of the partition $\mathcal{F}_i$ we define $\Gamma_{\mathcal{X}}$ to be the underlying graph determined by $\mathcal{X}$ (i.e. $\Gamma_{D_1+A_2}$ together with the edges from E_3 that have been added by the extension-rotation procedure) and $A_{\mathcal{X}}$ to be the out-arcs from A_3 determined by $\mathcal{X}$. Furthermore, define $C_{\mathcal{X}}$, $I_{\mathcal{X}}$ and $O_{\mathcal{X}}$ to be the sets of vertices checked for membership in $K \setminus L$, in determined to be $K \setminus L$, and determined not to be in $K \setminus L$, respectively, en route to the part $\mathcal{X}$. We discard $\mathcal{X}$ if $\Gamma_{\mathcal{X}}$ has does not have the desired structural properties or if too many digraphs $D \in \mathcal{X}$ do not have the desired structural properties. In Sections 4 and 5, respectively, we formally define two events: $\mathcal{E}_G$ a collection of graphs and $\mathcal{E}_D$ a collection of digraphs. We will prove that if (D_1, A) is chosen uniformly at random from $\Omega_k \times [n]^k$ then

$$\mathbf{Pr}(\exists B \subseteq A_3 \text{ such that } \Gamma_{D_1+A_2+B} \in \mathcal{E}_G) = o(1) \tag{1}$$

$$\mathbf{Pr}(D_1 \in \mathcal{E}_D \ \wedge \ \Gamma_{D_1} \notin \mathcal{E}_G) \leq e^{-\log^2 n}. \tag{2}$$

At stage i of the algorithm we discard each part $\mathcal{X}$ of $\mathcal{F}_i$ such that $\Gamma_{\mathcal{X}} \in \mathcal{E}_G$. We also discard $\mathcal{X}$ if $|\{D_1 \in \mathcal{X} \cap \mathcal{E}_D\}| > n^2 e^{-\log^2 n}|\{D_1 \in \mathcal{X}\}|$. We bound the probability that digraph D is discarded at stage i by applying the following version of Markov's inequality. If $\mathcal{E}$ is an event and $\mathcal{G}$ is a partition of our probability space then

$$\mathbf{Pr}[\mathbf{Pr}[\mathcal{E}|\mathcal{G}] \geq \varepsilon] = \mathbf{Pr}[E[1_{\mathcal{E}}|\mathcal{G}] \geq \varepsilon] \leq \frac{1}{\varepsilon} E[E[1_{\mathcal{E}}|\mathcal{G}]] = \frac{E[1_{\mathcal{E}}]}{\varepsilon} = \frac{\mathbf{Pr}(\mathcal{E})}{\varepsilon}. \tag{3}$$

Applying (3) with $\varepsilon = n^2 e^{-\log^2 n}$ we see that the probability that there exists an index i such that our randomly chosen (D_1, A) is discarded at stage i due to (2) is bounded above by $1/n$.

So, we restrict our attention to parts $\mathcal{X}$ of the partition $\mathcal{F}_i$ in the filtration given by the Extension-Rotation algorithm such that $\Gamma_{\mathcal{X}} \notin \mathcal{E}_G$ and the probability D_1 is in $\mathcal{E}_D$ where D_1 is chosen uniformly at random from $D_1 \in \mathcal{X}$ is at most $n^2 e^{-\log^2 n}$. The remainder of our analysis hinges on the following two Lemmas. Of course, we must only show that with high probability the Extension Rotation Procedure does not arrive in a situation were the longest path in the current 2-factor cannot be completed to a cycle using the arcs in A_3. To this end, we first show that there are many end-points of paths that span the vertex set spanned by the longest path in the 2-matching.

Lemma 4. *Let G be a fixed graph on n vertices such that $\delta(G) \geq 3$. Suppose further that P is a maximal path in G, x is an endpoint of P, X is the set of vertices spanned by P and* $\mathrm{END}(x)$ *is the set of end-points (other than x) of paths that span X and have x as one end-point. Then we have*

$$G \notin \mathcal{E}_G \qquad \Rightarrow \qquad \mathrm{END}(x) > \frac{n}{100}.$$

Lemma 4 is proved in Section 4. One consequence of Lemma 4 is that each time we find $x \in \mathrm{END}$ such that $x \in K \setminus L$ then the probability that the third edge out of x completes the cycle is bounded away form zero. Thus, **whp** the number of cycles we close among the first i elements of $K \setminus L$ found is at least $i/200$ for, say, $i \geq \sqrt{n}$. Therefore, the number of elements of $K \setminus L$ found before the successful termination of the Extension-Rotation Procedure is **whp** at most $1200n/\log n$. It remains to show that each $x \in \mathrm{END}$ we consider (that is in neither $I_{\mathcal{X}}$ nor $O_{\mathcal{X}}$) has a good chance of being in $K \setminus L$.

Lemma 5. *Suppose $\mathcal{X}$ is a part of the partition $\mathcal{F}_i$ such that $\Gamma_{\mathcal{X}} \notin \mathcal{E}_G$ and $Pr(D \in \mathcal{E}_D | D \in \mathcal{X}) < n^2 e^{-\log^2 n}$. Suppose further that $|I_{\mathcal{X}}| \leq \frac{1200n}{\log n}$ and $|O_{\mathcal{X}}| = o(n)$. If D is chosen uniformly at random from $\mathcal{X}$ then*

$$\left| \left\{ x \in V : Pr(x \in K(D)) \leq \frac{k}{n} \cdot \frac{1}{4000(\log \log n)^3} \right\} \right| = o(n).$$

Lemma 5 is proved in Section 5. It follows from this Lemma that **whp** the number of elements of $K \setminus L$ found among the first i vertices queried is at least $i/(\log^{1/2} n \cdot 8000(\log \log n)^3))$ for, say, $i \geq \sqrt{n}$. Let N be the number of vertices queried for membership in $K \setminus L$ before the successful completion of the Extension-Rotation Procedure. We have

$$\frac{N}{\log^{1/2} n \cdot 8000(\log \log n)^3} \leq \frac{1200n}{\log n}.$$

and so

$$N = O\left(n \cdot \frac{(\log \log n)^3}{\log^{1/2} n} \right).$$

Therefore, **whp** we have $|O_{\mathcal{X}}| = o(n)$ throughout the execution of Extension-Rotation (inductively justifying our application of Lemma 5).

3 A simple 2-matching with few components

We apply the following consequence of the Tutte-Berge matching formula (see Theorem 30.7 in Shrijver [17]). If $G = (V, E)$ is a graph and A, B are disjoint subsets of V then we let $e(A, B)$ denote the number of edges in E that intersect both A and B and $e(A)$ denote the number of edges within A. The **size** of a simple 2-matching F is the number of edges in F.

Theorem 6. *Let $G = (V, E)$ be a graph. The maximum size of a simple 2-matching in G is equal to the minimum value of*

$$|V| + |U| - |S| + \sum_K \left\lfloor \frac{e(K, S)}{2} \right\rfloor, \tag{4}$$

where U and S are disjoint subsets of V, with S an independent set, and where K ranges over the components of $G - U - S$.

We also make use of the following two structural Lemmas. We define $\mathcal{E}_{\text{cycles}}$ to be the collection of graphs on vertex set $[n]$ for which there at least $n^{3/4}$ vertex sets C such that $|C| \leq \frac{\log n}{3.9}$, $G[C]$ is connected and $G[C]$ has at least $|C|$ edges. We make extensive use of the following observation:

$$S \subseteq \binom{[n]}{2} \qquad \Rightarrow \qquad \mathbf{Pr}(S \subseteq E(G_{3-out})) \leq \left(\frac{6}{n}\right)^{|S|}. \tag{5}$$

This is easily seen to be true for $|S| = 1$ and after this, knowing that $S' \subseteq E(G_{3-out})$ only reduces the probability that $e \notin S'$ is also in $E(G_{3-out})$.

Lemma 7. *With high probability $G_2 \notin \mathcal{E}_{\text{cycles}}$.*

Proof. We compute an upper bound on the expected number of small complex components in Γ_D where D is drawn uniformly at random from Ω:

$$\sum_{c=3}^{\frac{\log n}{3.9}} \binom{n}{c} c^{c-2} \binom{c}{2} \left(\frac{6}{n}\right)^c \leq \sum_{c=3}^{\frac{\log n}{3.9}} (6e)^c = O\left((6e)^{\frac{\log n}{3.9}}\right).$$

The Lemma then follows from Markov's inequality. $\square$

Lemma 8. *With high probability $\alpha(G_2) \leq 0.415n$.*

Proof. Let Z be the size of the largest independent set in $G_1 = \Gamma_{D_1}$. Applying the union bound (and noting that a maximum independent set is also a dominating set) we have

$$Pr(Z \geq \ell) \leq \sum_{i \geq \ell} \binom{n}{i} \left(1 - \frac{i}{n}\right)^{3i-k} \left[1 - \left(1 - \frac{1}{n-i}\right)^{3i} \left(1 - \frac{i}{n}\right)^3\right]^{n-i-k}$$

We checked numerically that the function

$$f(x) = \frac{(1-x)^{3x} \left(1 - (1-x)^3 e^{-3x/(1-x)}\right)^{1-x}}{x^x (1-x)^{1-x}}$$

is less than one for $x \geq 0.415$. We conclude that $Z \leq 0.415n$ with high probability. As G_1 is a subgraph of G_2, we have the desired result. $\square$

Rather than work with the event implicit in Theorem 6 directly, we consider a weaker event. Let $\mathcal{G}$ be the event that there exist disjoint vertex sets U, R such that

(i) $V \setminus (U \cup R)$ can be partitioned into sets S, T such that $G[T]$ is a forest and $G[S]$ is an independent set.

(ii) $|S| \geq \max\{|U|, 2n/(\log n)\}$,

(iii) $e(V \setminus (U \cup R)) + e(R, V \setminus (U \cup R)) = e(S \cup T) + e(R, S \cup T) \leq |V| + |S| - 3|U| - |R|$.

We now argue that $G \notin \mathcal{G} \vee \mathcal{E}_{\text{cycles}}$ implies that G has a simple 2-factor with at most $6n/\log n$ components. Suppose $G \notin \mathcal{G} \vee \mathcal{E}_{\text{cycles}}$. Let F be a simple 2-factor of maximum size. Since $G \notin \mathcal{E}_{\text{cycles}}$, F has at most $4n/\log n$ cycle components, and therefore the number of components in F is at most $4n/\log n$ plus n minus the size of F. So, it suffices to show that G has a simple 2-factor with at least $n - 2n/\log n$ edges. Let U, S be an arbitrary pair of disjoint subsets of V where S is an independent set. We may assume $|S| \geq 2n/\log n, |U|$ (otherwise the expression in (4) is clearly at least $n - 2n/\ln n$). For the given U and S, let R be the set of vertices in non-tree components in $G_2[V \setminus (U \cup S)]$. Set $T = V \setminus (U \cup S \cup R)$. Note that the graph induced on T is a forest. Suppose this is a forest with p components. Now, since $G \notin \mathcal{E}_{\text{cycles}}$ implies that there are at most $4n/\log n$ components in $G[R]$, we have

$$\sum_K \left\lfloor \frac{e(K,S)}{2} \right\rfloor \geq \frac{e(S,T) + e(S,R) - p - \frac{4n}{\log n}}{2}.$$

On the other hand, $G \notin \mathcal{G}$ implies that condition (iii) does not hold and we have

$$\begin{aligned} e(S,T) + e(S,R) &> |V| + |S| - 3|U| - |R| - e(T) \\ &= |V| + |S| - 3|U| - |R| - |T| + p \\ &= 2|S| - 2|U| + p. \end{aligned}$$

Therefore,

$$\sum_K \left\lfloor \frac{e(K,S)}{2} \right\rfloor > |S| - |U| - \frac{2n}{\log n}.$$

Applying Theorem 6, we see that there exists a simple 2-factor F in G with at at least $|V| - 2n/\log n$ edges.

It remains to show that $G_2 \notin \mathcal{G}$ with high probability. We apply the first moment method. Throughout these computations we take K to be a fixed set (the set L, however, is not determined). For ease of notation we set $u = |U|, r = |R|, s = |S|$, etc. We consider 3 cases based on the value of u/n.

Case 1. $u \leq \frac{n}{e^{18}}$.

Here we take special care with the vertices in $S \cup T$ that are in K. For disjoint sets U, Z, W such that $U, Z \subseteq V$ and $W \subseteq K$ let $\mathcal{G}_{U,Z,W}$ be the event that conditions (i), (ii) and (iii) are satisfied with

$$S \cup T = Z \cup W \qquad (S \cup T) \cap L = Z \cap K \qquad (S \cup T) \cap (K \setminus L) = W.$$

In words, Z is the set of vertices in $S \cup T$ of out-degree 3 in $D_1 + A_2$ and W is the set of vertices in $S \cup T$ of out degree 2 in $D_1 + A_2$. Note that this event includes conditions on arcs that point at the vertices in $Z \cap K$ and W; in particular, there is at least one arc directed at each vertex in W. Let α be the number of arcs directed from $Z \cup W = S \cup T$ to $S \cup T \cup R$ and let β be the number of edges that are directed from $Z \cup W = S \cup T$ to U. Finally, let γ be the number of arcs directed from $R = V \setminus (U \cup S \cup T)$ into vertices in W.

Before proceeding with the calculation, we establish an upper bound on α (and hence a lower bound on β). Note that, as we assume condition (iii) is satisfied, we have

$$\alpha + \gamma \ \le \ e(S \cup T) + e(R, S \cup T) \ \le \ |V| + |S| - 3|U| - |R| \ = \ 2s - 2u + t.$$

Therefore

$$\alpha \le 2s - 2u + t - \gamma \quad \text{ and } \quad \beta \ge s + 2t - w + 2u + \gamma.$$

For fixed U, Z, W we have

$$\begin{aligned} Pr(\mathcal{G}_{U,Z,W}) \ &\le \ \sum_{\beta,\gamma} \binom{3(s+t)-w}{\beta} \left(\frac{u}{n}\right)^{\beta} \left(\frac{3u+\alpha}{n}\right)^{w-\gamma} \qquad (6) \\ &\le \ 9n^2 \cdot 2^{3(s+t)-w} \left(\frac{u}{n}\right)^{s+2t+2u} \left(\frac{u+2s+t}{u}\right)^{w-\gamma}. \end{aligned}$$

Note that the last term in (6) accounts for the condition that there must be an arc directed into each vertex in W: There are at least $|W| - \gamma$ vertices in W that are not 'covered' by vertices in R, and each of these vertices must be covered by one of the arcs out of U or one of the α arcs out of $S \cup T$ that point to vertices in $S \cup T \cup R$.

We bound the probability that there exists a pair U, R satisfying the conditions of the event $\mathcal{G}$ with $|U| \le n/e^{18}$ by summing over all possible choices of U, Z and W:

$$
\begin{aligned}
Pr\left(\bigvee_{U,Z,W} \mathcal{G}_{U,Z,W}\right) &\le \sum_{u,z,w} \binom{n}{u}\binom{n}{z}\binom{k}{w} \\
&\qquad \cdot 9n^2 2^{3(s+t)-w} \left(\frac{u}{n}\right)^{s+2t+2u} \left(\frac{u+2s+t}{u}\right)^{w-\gamma} \\
&\le 9n^2 \sum_{u,z,w} \left(\frac{ne}{u}\right)^u \left(\frac{ne}{s+t-w}\right)^{s+t-w} \left(\frac{ke}{w}\right)^w \qquad (7) \\
&\qquad \cdot 8^{s+t} \left(\frac{1}{2}\right)^w \left(\frac{u}{n}\right)^{s+2t+2u} \left(\frac{u+2s+t}{u}\right)^{w-\gamma} \\
&\le 9n^2 \sum_{u,z,w} \left(\frac{3k}{2n}\right)^w \left(\frac{eu}{n}\right)^u \left(\frac{16eu}{s+t}\right)^{s+t} \left(\frac{s+t}{u}\right)^w .
\end{aligned}
$$

Note that we make some crude simplifications and use the bound

$$
\left(\frac{s+t}{s+t-w}\right)^{s+t-w} \left(\frac{s+t}{w}\right)^w \le 2^{s+t}.
$$

(This can be viewed as a special case of equation (16), which is used extensively in Section 4.) We bound the final summand in (7) by considering two subcases. Recall that we assume $s > 6n/\log n$. If $s + t \ge 32eu$ then we write

$$
\left(\frac{16eu}{s+t}\right)^{s+t} \left(\frac{s+t}{u}\right)^w \le \left(\frac{1}{2}\right)^{s+t-w} (16e)^w.
$$

If w is a positive proportion of $s + t$ then the $\frac{3k}{2n}$ term suffices to establish the desired upper bound, and otherwise the $(1/2)^{s+t-w}$ term suffices. If, on the other hand, $s + t < 32eu$ then we write

$$
\left(\frac{eu}{n}\right)^u \left(\frac{16eu}{s+t}\right)^{s+t} \le \left[e^{17}\frac{u}{n}\right]^u,
$$

and again we have the desired upper bound.

Case 2. $n/e^{18} \le u \le 0.015n$.

Suppose first that $s + t = n - u - r \ge 8eu$. As above, we let β be the number of arcs directed from $S \cup T$ to U. We have $\beta \ge s + 2t + 2u - o(n)$. The probability that there exist U, R that satisfy condition (iii) with $n - |U| - |R| \ge 8e|U|$ and

$n/e^{18} \le |U| \le 0.015n$ is bounded above by (we write $n-u-r = s+t$)

$$e^{o(n)} \sum_{u,r,\beta} \binom{n}{u}\binom{n}{s+t}\binom{3(s+t)}{\beta}\left(\frac{u}{n}\right)^{\beta} \le e^{o(n)} \sum_{u,r} \left(\frac{ne}{u}\right)^{u}\left(\frac{ne}{s+t}\right)^{s+t} 2^{3(s+t)}\left(\frac{u}{n}\right)^{s+2t+2u}$$
$$\le e^{o(n)} \sum_{u,r} \left(\frac{u}{n}\right)^{t}\left(\frac{eu}{n}\right)^{u}\left(\frac{8eu}{s+t}\right)^{s+t}.$$

Note that the $\left(\frac{eu}{n}\right)^u$ term is exponentially small and dominates the $e^{o(n)}$ term and the other terms are at most 1.

Suppose now that $s+t = n-u-r < 8eu$. Here we add the consideration of the edges from R to $S \cup T$ to the previous bound. Let γ be the number of arcs directed from R to $S \cup T$. Now we have $\beta \ge s+2t+2u+\gamma - o(n)$. The probability that there exist U, R that satisfy condition (iii) with $n - |U| - |R| < 8e|U|$ and $n/e^{18} \le |U| \le 0.015n$ is bounded above by

$$e^{o(n)} \sum_{u,r,\gamma,\beta} \binom{n}{u}\binom{n}{s+t}\binom{3(s+t)}{\beta}\left(\frac{u}{n}\right)^{\beta}\binom{3r}{\gamma}\left(\frac{s+t}{n}\right)^{\gamma}\left(1-\frac{s+t}{n}\right)^{3r-\gamma}$$
$$\le e^{o(n)} \sum_{u,r,\gamma} \left(\frac{ne}{u}\right)^{u}\left(\frac{ne}{s+t}\right)^{s+t} 2^{3(s+t)}\left(\frac{u}{n}\right)^{s+2t+2u+\gamma} \tag{8}$$
$$\cdot \left(\frac{3re}{\gamma}\right)^{\gamma}\left(\frac{s+t}{n}\right)^{\gamma}\left(1-\frac{s+t}{n}\right)^{3r-\gamma}.$$

We write

$$\left(\frac{3re \cdot \frac{s+t}{n} \cdot \frac{u}{n}}{\gamma}\right)^{\gamma}\left(1-\frac{s+t}{n}\right)^{3r-\gamma} \le e^{3r\frac{s+t}{n}\frac{u}{n}} e^{-\frac{s+t}{n}(3r-\gamma)} \le e^{-\frac{s+t}{n}(3n-6u-3t-3s-\gamma)}.$$

Since $0 \le \alpha \le 2s - 2u + t - \gamma$, we have $\gamma \le 2s - 2u + t$ and

$$6u + 3t + 3s + \gamma \le 4u + 4t + 5s \le (4 + 5 \cdot 8e)u \le 1.99n.$$

So, we can bound the expression in (8) by

$$e^{o(n)} \sum_{u,r} \left(\frac{ne}{u}\right)^{u}\left(\frac{ne}{s+t}\right)^{s+t}\left(\frac{8}{e}\right)^{s+t}\left(\frac{u}{n}\right)^{s+2t+2u} \le e^{o(n)} \sum_{u,r} \left(\frac{eu}{n}\right)^{u}\left(\frac{u}{n}\right)^{t}\left(\frac{8u}{s+t}\right)^{s+t}$$
$$\le e^{o(n)} \sum_{u,r} \left(\frac{eu}{n}\right)^{u}\left(\frac{u}{n}\right)^{t} e^{u\left(\frac{8}{e}\right)}$$

As $0.015 \cdot e \cdot e^{8/e} < 1$, we have the desired bound.

Case 3. $0.015n \leq u$.

Let U, R (and $S \cup T = V \setminus (U \cup R)$) be fixed disjoint sets such that $|U| \geq 0.015n$. Let α be the number of arcs directed from $S \cup T$ to $S \cup T \cup R$ and γ be the number of arcs directed from R to $S \cup T$. The probability that there exist such sets U, R with the specified number of arcs is bounded above by

$$e^{o(n)} \sum_{u,r,\alpha,\gamma} \binom{n}{u} \binom{n-u}{r} \binom{3(n-u-r)}{\alpha} \left(1 - \frac{u}{n}\right)^{\alpha} \left(\frac{u}{n}\right)^{3(n-u-r)-\alpha} \cdot$$
$$\binom{3r}{\gamma} \left(1 - \frac{u+r}{n}\right)^{\gamma} \left(\frac{u+r}{n}\right)^{3r-\gamma}. \quad (9)$$

As we are bounding the probability of the event $\mathcal{G}$, we assume (in particular) that condition (iii) holds; that is, we assume,

$$\alpha + \gamma = e(S \cup T) + e(S \cup T, R) \leq n + s - 3u - r,$$

which implies

$$\alpha \leq n + s - 3u - r - \gamma = 2s - 2u + t - \gamma \leq 2(s+t).$$

Given this inequality and $u \leq s \leq \alpha(G_2) \leq 0.5n$ (by Lemma 8), we see that for u, r, γ fixed the expression

$$\binom{3(s+t)}{\alpha} \left(1 - \frac{u}{n}\right)^{\alpha} \left(\frac{u}{n}\right)^{-\alpha}$$

is increasing in α. Therefore, the probability that there exist disjoint sets U, R with $|U| = u$ and $|R| = r$ that satisfy conditions (i), (ii) and (iii) is bounded above by

$$3n \sum_{u,r,\gamma} \binom{n}{u} \binom{n-u}{r} \binom{3(n-u-r)}{n+s-3u-r-\gamma} \left(1 - \frac{u}{n}\right)^{n+s-3u-r-\gamma} \left(\frac{u}{n}\right)^{2n-s-2r+\gamma-o(n)}$$
$$\binom{3r}{\gamma} \left(1 - \frac{u+r}{n}\right)^{\gamma} \left(\frac{u+r}{n}\right)^{3r-\gamma-o(n)}, \quad (10)$$

where we set $s = \max\{n - u - r, 0.415n\}$. Note that the only role s plays in (9) is in the upper bound on α. So, we take the maximum possible value of α by taking the maximum value for s (and this has no impact on the rest of the expression). Also note that we apply Lemma 8 to set an absolute upper bound on the s.

We turn now to calculations. We re-name $u \to x, r \to y, \gamma \to z$. Then for $a \geq b$, we let $\mathrm{Bin}(a,b) = \frac{b^b}{a^a (b-a)^{b-a}}$ be a replacement for a corresponding binomial.

We now define the function of three real variables

$$g(x,y,z) = \mathrm{Bin}(1,x)\cdot\mathrm{Bin}(1-x,y)\cdot\mathrm{Bin}(3(1-x-y),\alpha)\cdot\mathrm{Bin}(3y,z)\cdot \\ x^{\beta(x,y,z)}\cdot(1-x)^{\alpha(x,y,z)}\cdot(1-x-y)^{z}\cdot(x+y)^{3y-z}$$

where

$$\alpha(x,y,z) = 1+s(x,y,z)-3x-y-z \qquad \beta(x,y,z) = 2-s(x,y,z)-2y+z$$
$$s(x,y,z) = \max\{1-x-y, 0.415\}.$$

Claim 9. *If* $0.015 \le x \le 0.415$, $0 \le y \le 1-2x$ *and* $0 \le z \le 3y$ *then* $g(x,y,z) \le 0.995$.
(Note that $y \le 1-2x$ *is equivalent to* $u \le s+t$*.)*

Note that Claim 9 implies that the expression in equation (10) is exponentially small for all relevant choices of r, u, γ, and Lemma 3 follows.

It remains to prove Claim 9. Of course, this can be achieved by a direct assault using calculus (e.g. concavity is useful). However, the authors feel that a simple check by computer is more compelling. Posted at `www.math.cmu.edu/~tbohman` the reader will find a short program in C that divides the 3-dimensional region into small cubes (side length 0.00125) and bounds the function over each cube in a trivial manner, thereby bounding the function over the entire region.

4 Rotations: Proof of Lemma 4

We begin by defining the event $\mathcal{E}_G$. This event is comprised of three properties: large maximum degree, small dense sets and a more complicated non-monotone condition for larger sets. Set

$$\Delta_0 = \frac{2\log n}{\log\log n}.$$

The event $\mathcal{E}_G$ is the union of events $\Delta(G) \ge \Delta_0$, $\mathcal{E}_{\mathrm{small}}$ and $\mathcal{E}_{\mathrm{pieces}}$ (the latter two are defined below). As the events $\Delta(G) \ge \Delta_0$ and $\mathcal{E}_{\mathrm{small}}$ are increasing, we can establish that the probability that there exists $B \subseteq A_3$ such that $\Gamma_{G_1+A_2+B}$ is in one of these events by working with Ω itself. The event $\mathcal{E}_{\mathrm{pieces}}$ requires a more delicate argument.

Define $\mathcal{E}_{\mathrm{small}}$ to be the event that there is a vertex set S and an integer i such that $1 \le i \le 10$, $|S| < \frac{i}{4}\log n$ and the graph induced by S is connected and spans at least $|S|+i$ edges.

Lemma 10. *If D is chosen uniformly at random from Ω then*

$$Pr(\Delta(\Gamma_D) \geq \Delta_0 \ \vee \ \Gamma_D \in \mathcal{E}_{small}) = o(1).$$

Proof. The degree of a vertex in Γ_D is distributed as $3 + Bi(3n, 1/n)$. Therefore

$$Pr(\Delta(\Gamma_D) \geq \Delta_0) \leq n \binom{3n}{\Delta_0 - 3} \frac{1}{n^{\Delta_0 - 3}} = o(n^{-1/2}).$$

Another application of the union bound gives

$$\begin{aligned} Pr(\Gamma_D \in \mathcal{E}_{\text{small}}) &\leq \sum_{s=4}^{\frac{10}{4}\log n} \binom{n}{s} s^{s-2} \binom{s}{2}^{1+\left\lceil \frac{s}{\frac{1}{4}\log n} \right\rceil} \left(\frac{6}{n}\right)^{s+\left\lceil \frac{s}{\frac{1}{4}\log n} \right\rceil} \\ &\leq \sum_{s=4}^{\frac{10}{4}\log n} \left(\frac{ne}{s} \cdot s \cdot \frac{6}{n}\right)^s \cdot \left(\frac{3s^2}{n}\right)^{\left\lceil \frac{s}{\frac{1}{4}\log n} \right\rceil} \\ &= o(n^{-1/2}). \end{aligned}$$

□

Let $N(S) = \{y \notin S : \ \exists x \in S$ such that y is adjacent to $x\}$ be the set of neighbors of a vertex set S. Define $\mathcal{E}_{\text{pieces}}$ to be the event that there exist sets S_1, S_2, T_1, T_2 such that

(i) S_1 and S_2 are disjoint with $n^{1/30} \leq |S_1| + |S_2| \leq \frac{n}{100}$.

(ii) For each $x \in S_1$ there exists $y_x \in S_1 \cup T_1$ such that $\{x, y_x\}$ is an edge, with the additional property that if $u, x \in S_1$ and $u = y_x$ then $x \neq y_u$.

(iii) For each $x \in S_2$ there exist distinct $y_x, z_x \in T_2$ such that $\{x, y_x\}, \{x, z_x\}$ are edges.

(iv) $T_1 = \{y_x : x \in S_1$ and $y_x \notin S_1\}$ and $T_2 = \{y_x, z_x : x \in S_2\}$.

(v) $N(S_1), N(S_2) \subseteq T_1 \cup T_2$ and S_2 is an independent set.

Lemma 11. *Let (D_1, A) be chosen uniformly at random from $\Omega_k \times [n]^k$.*

$$Pr\left(\exists B \subseteq A_3 : \Gamma_{D_1 + A_2 + B} \in \mathcal{E}_{\text{pieces}}\right) \leq e^{-\Omega(n^{1/30})}.$$

The proof of Lemma 11 is given in Section 4.1 below. Note that Lemmas 10 and 11 imply (1). We now proceed with the proof of Lemma 4.

We show $|S|$ is large in two stages.

Stage 1: $|S| \geq n^{1/30}$.

Let T_i be the set of end-points (other than x_0) of paths that can be obtained from P by i rotations with x_0 fixed. We view the T_i's as the level sets in a tree rooted at x_0. Note that the children of a vertex $y \in T_i$ in this tree are second-neighbors of y in D. Since each end-point $y \in T_i$ yields $d(y) - 1$ end-points in $T(y_{i+1})$ and $\delta(G) \geq 3$ we have

$$|T_{i+1}| \geq 2|T_i| - \delta_i, \tag{11}$$

where δ_i accounts for vertices in T_i that have second-neighbors in common.

If $\sum_{i=1}^{\ell} \delta_i = k$ then there exists a connected subgraph H with v vertices such that $v \leq 4\ell k$ and at least $v + k - 1$ edges. Since $G \notin \mathcal{E}_{\text{small}}$, it follows that $k \leq 10$ implies

$$4\ell k \geq v \geq \frac{k-1}{4} \cdot \log n \quad \Rightarrow \quad \ell \geq \frac{k-1}{k} \cdot \frac{1}{16} \cdot \log n.$$

Therefore,

$$\sum_{i=1}^{(\log n)/18} \delta_i \leq 9$$

and

$$|S| \geq \left|T_{(\log n)/18}\right| \geq 2^{\frac{\log n}{18} - 9} \geq n^{1/30}$$

for n sufficiently large.

Stage 2: $|S| \geq \frac{n}{100}$.

Let P be $x = x_0, x_1, x_2, \ldots$ and $S = \mathrm{END}(x)$. Suppose that the vertices in S induce a collection of sub-paths $P_1, P_2, \ldots, P_m$ of P and that P_i contains s_i vertices. Let Y be the collection of paths P_i that consist of a single vertex. (So, if $s_i = 1$ then $P_i \in Y$. Note that for these paths we abuse notation by referring to a single vertex as a path.) Pósa [15] proves that

$$x_i \in N(S) \quad \Rightarrow \quad x_{i-1} \in S \text{ or } x_{i+1} \in S. \tag{12}$$

Note that (12) implies

$$u \in Y \text{ and } v \in S \quad \Rightarrow \quad \{u, v\} \notin E(G). \tag{13}$$

Set

$$\begin{aligned} S_2 &= Y, & T_2 &= \{x_i : x_{i-1} \in S_2 \text{ or } x_{i+1} \in S_2\} \\ S_1 &= S \setminus Y, & T_1 &= \{x_i : x_i \notin S \text{ and } (x_{i-1} \in S_1 \text{ or } x_{i+1} \in S_1)\}. \end{aligned}$$

In words, S_2 is the set of vertices that are singleton paths in the collection $P_1, \ldots, P_m$ and S_1 is the set of vertices in longer paths. The sets T_2 and T_1 are the vertices adjacent along P to the end-vertices of the paths in $P_1, \ldots, P_m$. Since S_1, S_2, T_1 and T_2 satisfy conditions (ii), (iii), (iv) and (v) in the definition of the event $\mathcal{E}_{\text{pieces}}$ and $G \notin \mathcal{E}_{\text{pieces}}$, it follows that $|S| = |S_1| + |S_2| \geq \frac{n}{100}$.

4.1 Proof of Lemma 11

We apply the first moment method.

Let $s = |S_1| + |S_2|$ and $\nu = |S_2|$.

We view K as a fixed set. The set L, on the other hand, is random. We then let $K' \subseteq K$ be the tails of the arcs in $A \setminus (A_2 \cup B)$. So K' is the set of vertices (for a given choice of B) for which the third out-arc is still not in the graph. For $X \subseteq K$ and sets $S_1, S_2, T_1, T_2 \subseteq [n]$ let $\mathcal{E}_{X,S_1,S_2,T_1,T_2}$ be the event that $K' = X$ and S_1, S_2, T_1, T_2 satisfy conditions (i)–(v). We bound the probability of the union of all events of this form using the first moment method. A key property of this event is that, since we set $X = K'$, each vertex in X has in-degree at least 1.

We do not work with the events $\mathcal{E}_{X,S_1,S_2,T_1,T_2}$ directly. Rather we consider the events $\mathcal{F}_{Y,S_1,S_2,T_1,T_2}$ where $Y \subseteq (S_1 \cup S_2) \cap K$ defined by

$$\mathcal{F}_{Y,S_1,S_2,T_1,T_2} = \bigvee_{\substack{X \subseteq K: \\ X \cap (S_1 \cup S_2) = Y}} \mathcal{E}_{X,S_1,S_2,T_1,T_2}.$$

In words, we consider the event where we specify membership in K' only for those elements of K that are in $S_1 \cup S_2$. We choose Y, S_1, S_2 using the following method (in fact, we break S_2 into more parts in the actual calculation below, we indicate here how one can recover the set Y from the expression given below). We choose $Q, R \subseteq [n]$ and $\hat{Q}, \hat{R} \subseteq K$. Then we set

$$S_1 = Q \cup \hat{Q} \qquad S_2 = R \cup \hat{R} \qquad Y = \hat{Q} \cup \hat{R}.$$

The important point to note here is that if $x \in (R \cup Q) \cap K$ then we are considering the event where $x \notin K'$. So the the third arc out of x is present in the graph. Furthermore, as we specify $\hat{Q} \cup \hat{R} \subseteq K'$, there is at least one arc pointing into each vertex in $\hat{Q} \cup \hat{R}$.

Now we explain the further division S_2. We write $\nu = \nu_0 + \nu_1 + \nu_2 + \hat{\nu}_0 + \hat{\nu}_1 + \hat{\nu}_2$ where ν_i is the number of vertices $x \in S_2 \setminus K'$ for which i of the arcs between x and $\{y_x, z_x\}$ are directed into x and $\hat{\nu}_i$ is the number of vertices $x \in S_2 \cap K'$ for which i of the arcs between x and $\{y_x, z_x\}$ are directed into x. Let $\rho = |S_1 \setminus K'|$ and $\hat\rho = |S_1 \cap K'|$. The probability that Γ_D is in $\mathcal{E}_{\text{pieces}}$ is at most

$$\begin{aligned}\sum_{s=n^{1/30}}^{n/100} \sum_{\substack{\nu_0,\nu_1,\nu_2\\ \hat\nu_0,\hat\nu_1,\hat\nu_2\\ \rho,\hat\rho}} & \binom{n}{\rho}\binom{k}{\hat\rho}\binom{n}{\nu_0,\nu_1,\nu_2}\binom{k}{\hat\nu_0,\hat\nu_1,\hat\nu_2} \times n^{\rho+\hat\rho}\binom{n}{2}^{\nu} \\ & \times \left(6\cdot\frac{1}{n^2}\cdot\frac{2s}{n}\right)^{\nu_0}\left(18\cdot\frac{1}{n^2}\cdot\left(\frac{2s}{n}\right)^2\right)^{\nu_1}\left(9\cdot\frac{1}{n^2}\cdot\left(\frac{2s}{n}\right)^3\right)^{\nu_2} \\ & \quad \left(2\cdot\frac{1}{n^2}\cdot\frac{6s}{n}\right)^{\hat\nu_0}\left(12\cdot\frac{1}{n^2}\cdot\frac{2s}{n}\right)^{\hat\nu_1}\left(9\cdot\frac{1}{n^2}\cdot\left(\frac{2s}{n}\right)^2\right)^{\hat\nu_2} \\ & \times \left(\frac{6}{n}\cdot\left(\frac{2s}{n}\right)^2\right)^{\rho}\left(\frac{6}{n}\cdot\frac{2s}{n}\right)^{\hat\rho} \times \left(1-\frac{s}{n}\right)^{3(n-3s-k)}. \end{aligned} \tag{14}$$

Explanation of (14) View the summand as five terms separated by the symbol $\times$. The first is an upper bound on the number of choices for S_1, S_2 and Y. For such a fixed choice of S_1, S_2 and Y the rest of the expression in (14) is an upper bound on the probability of the union of all events of the form $\mathcal{F}_{Y,S_1,S_2,T_1,T_2}$. The second term is an upper bound on the number of choices for T_1, T_2 and the explicit edges from $S_1 \cup S_2$ to $S_1 \cup T_1 \cup T_2$ given by conditions (ii) and (iii). Then we consider the probability that the vertices in S_2 have the prescribed neighbors. Note that if v is among the $\hat\nu_0$ vertices in $S_2 \cap K'$ that has out-arcs directed to y_v and z_v then, since $v \in K'$ and therefore v has in-degree at least 1, there is a vertex in $T_1 \cup T_2$ that directs one of its out-arcs to v. The accounts for the 'extra' $6s/n$ factor in this term of the product. The fourth term is a bound on the probability that the arcs out of S_1 are in the prescribed locations. Finally we multiply by the probability that no vertex outside of $S_1 \cup S_2 \cup T_1 \cup T_2$ sends an arc into $S_1 \cup S_2$.

The bound in (14) can be approximated from above by

$$\begin{aligned}\sum_{s=n^{1/10}}^{n/100} \sum_{\substack{\nu_0,\nu_1,\nu_2\\ \hat\nu_0,\hat\nu_1,\hat\nu_2\\ \rho,\hat\rho}} & \left(\frac{24es^2}{\rho n}\right)^{\rho}\left(\frac{12eks}{\hat\rho n}\right)^{\hat\rho} \times \left(\frac{6es}{\nu_0}\right)^{\nu_0}\left(\frac{36es^2}{\nu_1 n}\right)^{\nu_1}\left(\frac{36es^3}{\nu_2 n^2}\right)^{\nu_2} \\ & \left(\frac{6esk}{\hat\nu_0 n}\right)^{\hat\nu_0}\left(\frac{12esk}{\hat\nu_1 n}\right)^{\hat\nu_1}\left(\frac{18es^2k}{\hat\nu_2 n^2}\right)^{\hat\nu_2} \times e^{s(-3+9s/n+3k/n)}. \end{aligned} \tag{15}$$

We repeatedly apply the following observation (the proof of which is given below):

$$x, y, \alpha, \beta > 0 \quad \Rightarrow \quad \left(\frac{\alpha}{x}\right)^x \left(\frac{\beta}{y}\right)^y \le \left(\frac{\alpha+\beta}{x+y}\right)^{x+y}. \tag{16}$$

The expression in (15) is bounded above by

$$\sum_{s=n^{1/10}}^{n/100} \sum_{\substack{\nu_0,\nu_1,\nu_2\\ \hat{\nu}_0,\hat{\nu}_1,\hat{\nu}_2\\ \rho,\hat{\rho}}} e^{s(-3+9s/n+3k/n)} \times e^s \times \left(\frac{\ell(s,k,n)}{s}\right)^s.$$

where

$$\begin{aligned}\ell(s,k,n) &= 24s(s/n) + 12s(k/n) + 6s + 36s(s/n) + 36s(s/n)^2 + 6s(k/n)\\ &\qquad + 12s(k/n) + 18s(ks/n^2)\\ &= s\left(6 + 48s/n + 36s^2/n^2 + o(1)\right).\end{aligned}$$

Since we assume $s \le n/100$, we have $\ell(s,k,n) \le 6.5s$ and the summand in (15) is at most

$$\left[e^{-1.9}6.5\right]^s \le (0.98)^s.$$

Thus the sum in (14) can be bounded by $n^8(.98)^{n^{1/30}}$, and this proves Lemma 11.

Proof of (16). Let $f(x) = \left(\frac{\alpha}{x}\right)^x \left(\frac{\beta}{z-x}\right)^{z-x}$. Then we have

$$\begin{aligned}\frac{f'(x)}{f(x)} &= \log\alpha - 1 - \log x - \log\beta + 1 + \log(z-x)\\ &= \log\left(\frac{(z-x)\alpha}{\beta x}\right).\end{aligned}$$

So, $f'(x) = 0$ iff $x = x^* = \frac{z\alpha}{\alpha+\beta}$.
Differentiating once more, we get

$$\frac{f''(x)}{f(x)} - \frac{f'(x)^2}{f(x)^2} = -\frac{1}{x} - \frac{1}{z-x} < 0$$

and so $f''(x^*) < 0$ and x^* is the maximum over $x \in [0, z]$. $\square$

5 Conditioning: Proof of Lemma 5

We begin with some definitions and initial observations, including the definition of the event $\mathcal{E}_D$. Set

$$\tau = 3 - \frac{1}{\log^{1/2} n} = 3 - \alpha \qquad \ell = 10 \log_\tau \log n.$$

Recall that $\mathcal{X}$ is a fixed part is some partition in the filtration of $\Omega_k \times [n]^k$ given by the execution of the Extension Rotation Procedure. Here we work with the space of digraphs Ω_k; the edges that have been added to raise the minimum degree to 2 or in the course of the Extension Rotation Procedure play no role.

For $x \in [n] \setminus (I_{\mathcal{X}} \cup O_{\mathcal{X}})$ let $\{D \in \mathcal{X}\}$ be partitioned into $\mathcal{X}_x, \overline{\mathcal{X}_x}$ where $D \in \mathcal{X}_x$ if $x \in K(D)$. Note that for fixed x we have $Pr(x \in K(D)) = |\mathcal{X}_x|/|\mathcal{X}|$. Thus, to prove Lemma 5 it suffices to show

$$\left| \left\{ x \in [n] \setminus (I_{\mathcal{X}} \cup O_{\mathcal{X}}) : \frac{|\mathcal{X}_x|}{|\mathcal{X}|} \leq \frac{1}{4000(\log\log n)^3} \cdot \frac{k}{n} \right\} \right| = o(n). \tag{17}$$

Define a bipartite graph $\Sigma = \Sigma_x$ on vertex set $\{\mathcal{X} \in D\}$ by including an edge joining $D \in \mathcal{X}_x$ and $D' \in \overline{\mathcal{X}_x}$ if D' is obtained by reversing the edges along a shortest path of length ℓ in D that starts at a vertex in $V \setminus (K(D) \cup O_{\mathcal{X}})$ and ends at x. We have, where d_Σ denotes degree in Σ,

$$\sum_{D \in \mathcal{X}_x} d_\Sigma(D) = \sum_{D \in \overline{\mathcal{X}_x}} d_\Sigma(D). \tag{18}$$

We get a lower bound on $|\mathcal{X}_x|/|\mathcal{X}|$ by bounding the degrees in Σ: We establish upper bounds on $d_\Sigma(D)$ for $D \in \mathcal{X}_x$ and lower bounds on $d_\Sigma(D)$ for $D \in \overline{\mathcal{X}_x}$.

For a digraph $D \in \Omega_k$ let $\Pi_{D,x \leftarrow V}$ be the set of vertices $y \in V$ such that there exists a shortest path of length ℓ from y to x. Let $\Pi_{D,x \to K}$ be the set of vertices $z \in K$ such that there is a shortest path from x to z of length ℓ. We bound $|\Pi_{D,x \leftarrow V}|$ and $|\Pi_{D,x \to K}|$ for most $D \in \mathcal{X}$ and $x \in [n]$. To this end, we define the following cut-offs:

$$\pi_{\text{in}} = (1000 \log\log n) \log^{10} n \qquad \pi_{\text{out}} = \frac{\log^{10} n}{(\log\log n)^2} \cdot \frac{k}{n}$$
$$\rho_{\text{in}} = (\log n)^{20}.$$

Let $\Psi_{\leftarrow V} \subseteq \Omega_k$ be the event

$$|\{x \in [n] : |\Pi_{D,x \leftarrow V}| \geq \pi_{\text{in}}\}| \geq \frac{n}{(\log n)^{100}}, \text{ or}$$
$$\exists x \in [n] \text{ such that } |\Pi_{D,x \leftarrow V}| \geq \rho_{\text{in}}.$$

Similarly, define $\Psi_{\to K} \subseteq \Omega_k$ to be the event

$$|\{x \in [n] : |\Pi_{D,x\to K}| \leq \pi_{\text{out}}\}| \geq \frac{n}{(\log n)^{100}}.$$

We now define the event $\mathcal{E}_D$: Set $\mathcal{E}_D = \Psi_{\to K} \cup \Psi_{\leftarrow V}$.

Lemma 12. *If D is chosen uniformly at random from Ω_k then*

$$Pr(D \in \Psi_{\leftarrow V} \cup \Psi_{\to K} \ \wedge \ \Gamma_D \notin \mathcal{E}_G) \leq e^{-\frac{1}{4}\log^3 n}.$$

The proof of Lemma 12 is given in Section 5.1. Note that condition (2) follows from Lemma 12. We now proceed with the proof of Lemma 5.

For each $D \in \mathcal{X}$ we define three sets of vertices:

$$\begin{aligned} V_1 = V_1(D) &= \{x \in [n] : |\Pi_{D,x\leftarrow V}| \geq \pi_{\text{in}}\} \\ V_2 = V_2(D) &= \{x \in [n] : |\Pi_{D,x\to K}| \leq \pi_{\text{out}}\} \\ V_3 = V_3(D) &= \{x \in [n] : |\Pi_{D,x\to K} \cap I_{\mathcal{X}}| \geq \pi_{\text{out}}/2\} \end{aligned}$$

The key observation is that most $x \in [n]$ have the property that $x \in V_i(D)$ for very few $D \in \mathcal{X}$ for $i = 1, 2, 3$. Indeed, the number of ordered pairs (x, D) where $x \in [n]$, $D \in \mathcal{X} \setminus \Psi_{\leftarrow V}$ and $x \in V_1(D)$ is at most $|\mathcal{X}|n/(\log^{100} n)$. It follows that

$$\left|\left\{x \in [n] : |\{D \in \mathcal{X} \setminus \Psi_{\leftarrow V} : x \in V_1(D)\}| \geq \frac{|\mathcal{X}|}{\log^{99} n}\right\}\right| \leq \frac{n}{\log n}. \tag{19}$$

Similarly, we have

$$\left|\left\{x \in [n] : |\{D \in \mathcal{X} \setminus \Psi_{\to K} : x \in V_2(D)\}| \geq \frac{|\mathcal{X}|}{\log^{99} n}\right\}\right| \leq \frac{n}{\log n}. \tag{20}$$

Now we turn to the set of vertices $x \in [n]$ with the property that $x \in V_3(D)$ for many digraphs D. Consider a fixed digraph $D \in \mathcal{X} \setminus \Psi_{\leftarrow V}$. The number of ordered pairs (a, b) where $a \in V$, $b \in I_{\mathcal{X}}$ and there is a shortest path of length ℓ from a to b is at most

$$\frac{n}{\log^{100} n} \cdot \rho_{\text{in}} + |I_{\mathcal{X}}| \cdot \pi_{\text{in}} \leq \frac{n \log^{10} n}{\log n} \cdot (\log\log n)^2$$

for n sufficiently large. It follows that $|V_3(D)| \leq 2n(\log\log n)^4/(\log^{1/2} n)$. Therefore

$$\left|\left\{x \in [n] : |\{D \in \mathcal{X} \setminus \Psi_{\leftarrow V} : x \in V_3(D)\}| \geq \frac{|\mathcal{X}|}{\log^{1/5} n}\right\}\right| \leq \frac{n}{\log^{1/5} n}. \tag{21}$$

As it suffices to establish (17), we may restrict our attention to vertices x in the complements of the sets in equations (19), (20) and (21).

Now apply (18). We have

$$|\mathcal{X}_x|\pi_{\text{in}} + \frac{|\mathcal{X}|}{\log^{99} n}\rho_{\text{in}} + \frac{|\mathcal{X}|}{e^{\log^2 n}}\Delta_0^\ell \geq \sum_{D \in \mathcal{X}_x} d_\Sigma(D)$$
$$= \sum_{D \in \overline{\mathcal{X}_x}} d_\Sigma(D) \geq \left(|\mathcal{X}| - |\mathcal{X}_x| - \frac{|\mathcal{X}|}{\log^{99} n} - \frac{|\mathcal{X}|}{\log^{1/5} n}\right)\frac{\pi_{\text{out}}}{2}$$

Therefore

$$|\mathcal{X}_x|\left(\pi_{\text{in}} + \frac{\pi_{\text{out}}}{2}\right) \geq |\mathcal{X}|\frac{\pi_{\text{out}}}{2}\left(1 - \frac{1}{\log^{99} n} - \frac{1}{\log^{1/5} n} - \frac{2\rho_{\text{in}}}{\pi_{\text{out}}\log^{99} n} - \frac{2\Delta_0^\ell}{\pi_{\text{out}} e^{\log^2 n}}\right)$$

and

$$\frac{|\mathcal{X}_x|}{|\mathcal{X}|} \geq \frac{\pi_{\text{out}}}{4\pi_{\text{in}}} = \frac{1}{4000(\log\log n)^3} \cdot \frac{k}{n}.$$

5.1 Proof of Lemma 12

We begin with the event $\Psi_{\leftarrow V}$. Let x be a fixed vertex. For $i = 1, \ldots, \ell$ let $S_i = S_i(x)$ be the set of vertices that can reach x in a path of length i but not via a shorter path. We reveal the sets S_i iteratively (i.e. using breadth first search).

Claim 13. *If $f(n)$ is an arbitrary positive real function then*

$$Pr\left(|S_\ell| \geq f(n)\tau^\ell \wedge \Gamma_D \notin \mathcal{E}_G\right) \leq e^{\frac{1-f(n)}{2}}$$

for n sufficiently large.

Proof. Note that $\Gamma_D \notin \mathcal{E}_G$ implies

$$\sum_{j=0}^{\ell} |S_i| \leq \sum_{j=0}^{\ell} \Delta_0^j \leq (\ell+1)\Delta_0^\ell = o(n^{1/3}).$$

Thus, we may discard the set of digraphs for which there exists an index $1 \leq i \leq \ell$ such that $|S_i| > \sqrt{n}/\ell$. In other words, we iteratively condition on the event $|S_i| \leq \sqrt{n}/\ell$.

Now an observation. If $0 < \lambda < 1$ and $\sum_{j=0}^{i} |S_j| \leq \sqrt{n}$ then we have

$$
\begin{aligned}
E[e^{\lambda |S_{i+1}|} \mid |S_i|] &\leq \sum_{k=0}^{\tau n} \binom{\tau n}{k} \left(\frac{|S_i|}{n-\sqrt{n}} \right)^k \left(1 - \frac{|S_i|}{n-\sqrt{n}} \right)^{\tau n - k} e^{\lambda k} \\
&= \left(1 - \frac{|S_i|}{n-\sqrt{n}} + \frac{|S_i| e^{\lambda}}{n-\sqrt{n}} \right)^{\tau n} \\
&\leq \left(1 + \lambda(1+\lambda) \frac{|S_i|}{n-\sqrt{n}} \right)^{\tau n} \\
&\leq e^{\tau |S_i| \lambda (1+\lambda)(1+2/\sqrt{n})}.
\end{aligned}
$$

Now set $\lambda_0 = 1/2$. For $i = 1, \ldots, \ell$ we define λ_i by setting

$$
\lambda_{i-1} = \tau \lambda_i (1 - \lambda_i)(1 + 2/\sqrt{n})
$$

and $\lambda_i < 1/2$. Note that so long as $\lambda_i \geq 2/\sqrt{n}$, we inductively have $\lambda_i \geq \lambda_{i-1}/\tau$. Therefore, after verifying that $1/2\tau^{\ell} \geq 2/\sqrt{n}$, we see that

$$
\lambda_{\ell} \geq \frac{1}{2\tau^{\ell}}. \tag{22}
$$

We now define a random variable Y_i. If $|S_1|, |S_2|, \ldots, |S_{i-1}| \leq \sqrt{n}/\ell$ then set $Y_i = |S_i|$. If, on the other hand, there exists $j < i$ such that $|S_j| > \sqrt{n}/\ell$ then set $Y_i = \lambda_{i-1} Y_{i-1}/\lambda_i$. Let $\mathcal{B}_i$ be the event that there exists $j < i$ such that $|S_j| > \sqrt{n}/\ell$. For $i = 1, \ldots, \ell$ we have

$$
\begin{aligned}
E\left[e^{\lambda_i Y_i} \mid \overline{\mathcal{B}_i}\right] &= E\left[E\left[e^{\lambda_i |S_i|} \mid |S_{i-1}|\right] \mid \overline{\mathcal{B}_i}\right] \\
&\leq E\left[e^{\tau |S_{i-1}| \lambda_i (1-\lambda_i)(1+2/\sqrt{n})} \mid \overline{\mathcal{B}_i}\right] = E\left[e^{\lambda_{i-1} |S_{i-1}|} \mid \overline{\mathcal{B}_i}\right] = E\left[e^{\lambda_{i-1} Y_{i-1}} \mid \overline{\mathcal{B}_i}\right].
\end{aligned}
$$

Of course, the same inequality holds if we condition instead on $\mathcal{B}_i$. It follows that $E\left[e^{\lambda_\ell Y_\ell}\right] \leq E\left[e^{\lambda_0 Y_0}\right] = e^{1/2}$. Applying Markov's inequality and (22) we have

$$
Pr\left(Y_\ell \geq f(n)\tau^{\ell}\right) \leq E\left[e^{\lambda_\ell Y_\ell}\right] e^{-\lambda_\ell f(n) \tau^{\ell}} \leq e^{\frac{1}{2} - \frac{1}{2} f(n)}.
$$

As the event $|S_\ell| \neq Y_\ell$ is a subset of $\Gamma_D \in \mathcal{E}_G$, the Lemma follows. □

Let X be the set of vertices x such that $|\Pi_{D,x\leftarrow V}| = |S_\ell(x)| \geq 1000 \log\log n \cdot \log^{10} n$. Applying Claim 13 we have

$$
Pr(x \in X \wedge \Gamma_D \notin \mathcal{E}_G) \leq \frac{e^{1/2}}{\log^{500} n}.
$$

Therefore, $E[|X| \mid \Gamma_D \notin \mathcal{E}_G] \leq 2n/\log^{500} n$ and $E[|X|] \leq 3n/\log^{500} n$ (as $Pr(\Gamma_D \in \mathcal{E}_G) = o(1/\sqrt{n})$ and $|X| \leq n$) . We bound the probability that X is significantly larger than its expected value by applying the Azuma-Hoeffding inequality to the arc-exposure martingale. We observe the out-arcs one at a time. Let a_i be the i^{th} arc observed in this process. Set $X_i = E[|X| \mid a_1, a_2, \ldots, a_k]$. Note that if D and D' differ by a single out-arc then we have

$$||X(D)| - |X(D')|| < \sum_{i=0}^{\ell} 3^i \leq \log^{11} n.$$

It follows that $|X_{k+1} - X_k| \leq \log^{11} n$ for $k \geq 0$. The Azuma-Hoeffding inequality then implies

$$Pr\left(X \geq \frac{n}{\log^{100} n}\right) \leq Pr\left(X \geq E[X] + \frac{n}{\log^{200} n}\right) \leq e^{-n^{1-o(1)}}.$$

To bound the probability that there exists a vertex x such that $|\Pi_{D,x\leftarrow V}| \geq \rho_{\text{in}}$ we apply Claim 13 with $f(n) = \log^{10} n$. It follows that $Pr(|S_\ell| \geq \rho_{\text{in}} \wedge \Gamma_D \notin \mathcal{E}_G) \leq e^{-(\log^{10} n)/2}$.

Now we turn to the event $\Psi_{x\to K}$. Let y be a fixed vertex. Let $T_i = T_i(y)$ be the sum of the out-degrees of the set of vertices that can be reached from y via paths of length i but not via shorter paths (e.g $T_0 = 3$). As above, we reveal T_i iteratively. Recall that $\ell = 10 \log_\tau \log n$. Set $\ell_1 = \ell - \log_2(1000 \log\log n)$.

Claim 14.

$$Pr\left(T_\ell \leq \tau^{\ell_1}\right) \leq \frac{1}{\log^{500} n}.$$

Proof. We begin with an observation. Define $\gamma_i = 3^{i+2}/(2n)$. Note that the probability that an out-arc from T_i hits a vertex that has already been seen is bounded above by γ_i. Then

$$E\left[e^{-\lambda|T_{i+1}|} \mid |T_i|\right] \leq \left(\gamma_i + \alpha e^{-2\lambda} + (1-\alpha-\gamma_i)e^{-3\lambda}\right)^{|T_i|}. \tag{23}$$

We apply (23) in two ways. First, if $\lambda > 0$ and $\lambda e^{-3\lambda} \geq 2\gamma_i$ then we have

$$\gamma_i + \alpha e^{-2\lambda} + (1-\alpha-\gamma_i)e^{-3\lambda} \leq e^{-2\lambda}$$

and

$$E\left[e^{-\lambda|T_{i+1}|} \mid |T_i|\right] \leq e^{-2\lambda|T_i|}.$$

Furthermore, if $3\gamma_i \leq \lambda$ then we have

$$
\begin{aligned}
E\left[e^{-\lambda|T_{i+1}|} \mid |T_i|\right] &\leq \left(\gamma_i + \alpha(1-2\lambda+3\lambda^2) + (1-\alpha-\gamma_i)(1-3\lambda+5\lambda^2)\right)^{|T_i|} \\
&\leq (1-\lambda(\tau-6\lambda))^{|T_i|} \\
&\leq e^{-\lambda(\tau-6\lambda)|T_i|}.
\end{aligned}
$$

Set $\lambda_0 = 1000 \log\log n$ and $i_0 = \ell - \ell_1 + 4$. For $i = 1, \ldots, \ell$ define λ_i by

$$
\lambda_{i-1} = \begin{cases} 2\lambda_i & \text{if } i \leq i_0 \\ \lambda_i(\tau - 6\lambda_i) & \text{if } i > i_0 \end{cases}
$$

(with the additional condition $\lambda_i < \tau/12$ in the latter case). Of course, we make these choices in order to establish $E[e^{-\lambda_{i+1}|T_{i+1}|}] \leq E[e^{-\lambda_i|T_i|}]$ for $i = 0, \ldots, \ell-1$ and therefore

$$
E[e^{-\lambda_\ell|T_\ell|}] \leq E[e^{-\lambda_0|T_0|}] = \frac{1}{\log^{3000} n}.
$$

Applying Markov's inequality we have

$$
Pr[|T_\ell| \leq s\tau^\ell] \leq e^{\lambda_\ell s\tau^\ell} E[e^{-\lambda_\ell|T_\ell|}] \leq \frac{e^{\lambda_\ell s\tau^\ell}}{\log^{3000} n}.
$$

It remains to establish an upper bound on λ_ℓ. Note first that $\lambda_{i_0} = \frac{1}{16}$ and $\lambda_{i_0+1} \leq \frac{1}{32}$. Then observe that $i > i_0$ and $\lambda_i \leq \frac{1}{24}$ implies

$$
\lambda_i \leq \frac{\lambda_{i-1}}{\tau}\left(1 + \frac{8\lambda_{i-1}}{\tau}\right). \tag{24}
$$

Furthermore, we have $\lambda_i \leq \lambda_{i-1}/2$ for all i and the bound

$$
\prod_j \left(1 + \frac{8\lambda_j}{\tau^2}\right) \leq \exp\left\{\sum_j \frac{8\lambda_j}{\tau}\right\} \leq e^{16/\tau} \leq e^6,
$$

then implies

$$
\lambda_\ell \leq 2^{-4}\tau^{-\ell_1+4}e^6.
$$

Taking $s = \tau^{-\ell+\ell_1}$, we have

$$
Pr\left(|T_\ell| \leq \tau^{\ell_1}\right) \leq \frac{e^{e^6(\tau/2)^4}}{\log^{3000} n}.
$$

□

Now let Y be the set of vertices y such that $T_\ell(y) \leq \tau^{\ell_1}$. It follows from Claim 14 that we have $E[|Y|] \leq n/\log^{2000} n$. We again apply the Azuma-Hoeffding inequality to the arc exposure martingale to show that the Y is concentrated around its expected value. Recall that a_i denotes the i^{th} arc observed in this process. Set

$$Y_k = \begin{cases} E[|Y| \mid a_1, a_2, \dots, a_k] & \text{if } Pr\left(\Delta(\Gamma_D) \geq \log^3 n \mid a_1, \dots a_{k-1}\right) < \frac{1}{n^2} \\ Y_{k-1} & \text{if } Pr\left(\Delta(\Gamma_D) \geq \log^3 n \mid a_1, \dots a_{k-1}\right) \geq \frac{1}{n^2}. \end{cases}$$

Note that if D and D' differ by a single out-arc and $\Delta(\Gamma_D), \Delta(\Gamma_{D'}) < \log^3 n$ and $D, D' \notin \mathcal{E}_G$ then we have

$$||X(D)| - |X(D')|| < (\log^3 n)^\ell < n^{1/10}.$$

Since

$$Pr\left(\Delta(\Gamma_D) \geq \log^3 n \mid a_1, \dots a_k\right) < nPr\left(\Delta(\Gamma_D) \geq \log^3 n \mid a_1, \dots a_{k-1}\right)$$

and $|Y| \leq n$ we have $|Y_{i+1}| - |Y_i| \leq 2n^{1/10}$. The Azuma Hoeffding inequality then implies

$$Pr\left(Y_{3n} \geq E[Y] + \frac{n}{\log^{200} n}\right) \leq e^{-n^{8/10-o(1)}}.$$

Noting that

$$\begin{aligned} Pr(Y_{3n} \neq Y) &\leq n^2 Pr(\Delta(\Gamma_D) > \log^3 n) \\ &< n^2 \cdot n \binom{3n}{\log^3 n - 3} \left(\frac{1}{n}\right)^{\log^3 n - 3} \\ &\leq e^{-\log^3 n}, \end{aligned}$$

we conclude that

$$Pr\left(Y \geq n/(\log^{100} n)\right) \leq e^{-(\log^3 n)/2}.$$

□

Final Remark. We have established that $G_{3-\text{out}}$ is Hamiltonian with high probability. In general, it may be the case if $m \geq 3$ then with high probability $G_{m-\text{out}}$ contains $\lfloor m/2 \rfloor$ disjoint Hamilton cycles plus, in the case that m is odd, a disjoint matching that saturates all but at most one vertex. We leave this as an open problem.

References

[1] M. Ajtai, J. Komlós and E. Szemerédi, The first occurrence of Hamilton cycles in random graphs. *Annals of Discrete Mathematics* **27** (1985) 173-178.

[2] N. Alon and J. Spencer, The probabilistic method, Wiley, 1992.

[3] B. Bollobás, The evolution of sparse graphs, in *Graph Theory and Combinatorics* (Proceedings of the Cambridge Combinatorics Conference in Honour of Paul Erdős (B. Bollobás; Ed)), Academic Press (1984) 35-57.

[4] B. Bollobás and A.M. Frieze, On matchings and hamiltonian cycles in random graphs. *Annals of Discrete Mathematics* **28** (1985) 23-46.

[5] B. Bollobás, T.I. Fenner and A.M. Frieze, An algorithm for finding hamilton paths and cycles in random graphs. *Combinatorica* **7** (1987) 327-341

[6] B. Bollobás, T.I. Fenner and A.M. Frieze, Hamilton cycles in random graphs with minimal degree at least k, in *A tribute to Paul Erdős*, edited by A. Baker, B. Bollobás and A. Hajnal (1990) 59 - 96.

[7] C. Cooper and A.M. Frieze, Hamilton cycles in random graphs and directed graphs. *Random Structures and Algorithms* **16** (2000) 369-401.

[8] C. Cooper, A.M. Frieze and B. Reed, Random regular graphs of non-constant degree: connectivity and Hamilton cycles. *Combinatorics, Probability and Computing* **11** (2002) 249-262.

[9] T.I. Fenner and A.M. Frieze, On the connectivity of random m-orientable graphs and digraphs. *Combinatorica* **2** (1982) 347-359.

[10] T.I. Fenner and A.M. Frieze, On the existence of hamiltonian cycles in a class of random graphs. *Discrete Mathematics* **45** (1983) 301-305.

[11] A.M. Frieze, Finding Hamilton cycles in sparse random graphs. *Journal of Combinatorial Theory B* **44** (1988) 230-250.

[12] A. Frieze and T. Łuczak, Hamiltonian cycles in a class of random graphs: one step further, in *Proceedings of Random Graphs '87*, Edited by M.Karonski, J.Jaworski and A.Rucinski, John Wiley and Sons, 53-59.

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

[14] M. Krivelevich, B. Sudakov, V. Vu and N.C. Wormald, Random regular graphs of high degree. *Random Structures and Algorithms* **18** (2001) 346-363.

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

[16] R.W. Robinson and N.C. Wormald, Almost all regular graphs are Hamiltonian. *Random Structures and Algorithms* **5** (1994) 363-374.

[17] A. Schrijver, Combinatorial optimization: polyhedra and efficiency, Springer 2003.