A scaling limit for the length of the longest cycle in a sparse random digraph

Michael Anastos*and Alan Frieze†

July 4, 2021

Abstract

We discuss the length $\vec{L}_{c,n}$ of the longest directed cycle in the sparse random digraph $D_{n,p}, p = c/n$, c constant. We show that for large c there exists a function $\vec{f}(c)$ such that $\vec{L}_{c,n}/n \to \vec{f}(c)$ a.s. The function $\vec{f}(c) = 1 - \sum_{k=1}^{\infty} p_k(c)e^{-kc}$ where p_k is a polynomial in c. We are only able to explicitly give the values p_1, p_2, although we could in principle compute any p_k.

1 Introduction

In this paper we consider the length $\vec{L}_{c,n}$ of the longest cycle in the random digraph $D_{n,p}, p = c/n$ where we will assume that c is a sufficiently large constant. Here $D_{n,p}$ is the random subgraph of the complete digraph $\vec{K}_n$ obtained by including each of the $n(n-1)$ edges independently with probability p. Most of the literature on long cycles has been concerned with the length $L_{c,n}$ of the longest cycle in the random graph $G_{n,p}$. It was shown by Frieze [9] that w.h.p. $L_{c,n} \geq (1-(c+1+\varepsilon_c)e^{-c})n$ where $\varepsilon_c \to 0$ as $c \to \infty$. Using the elegant coupling argument of McDiarmid [14] we see that this implies that w.h.p. $\vec{L}_{c,n} \geq (1-(c+1+\varepsilon_c)e^{-c})n$. This was improved by Krivelevich, Lubetzky and Sudakov [13] who showed that w.h.p. $\vec{L}_{c,n} \geq (1-(2+\varepsilon_c)e^{-c})n$. Recently, Anastos and Frieze [1] have shown that if c is sufficiently large then w.h.p. $L_{c,n} \approx f(c)n$ as $n \to \infty$, for some function $f(c)$[1].

In this paper we use the ideas of [1] and show that w.h.p. $\vec{L}_{c,n} \approx \vec{f}(c)n$ and compute the first few terms of $\vec{f}(c) = 1 - \sum_{k=1}^{\infty} p_k(c)e^{-kc}$ where $p_k(c)$ is a polynomial in c for $k \geq 1$. I.e. we

*Department of Mathematics and Computer Science, Freie Universität Berlin, Berlin, Germany, email:**manastos@zedat.fu-berlin.de**

†Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh PA, U.S.A. email:**alan@random.math.cmu.edu**; the author is supported in part by NSF Grant DMS1363136

[1] Here we say $A_n \approx B_n$ if $A_n/B_n \to 1$ as $n \to \infty$.

prove a scaling limit for $\vec{L}_{c,n}$. The important point here is that we establish high probability errors that tend to zero with n, regardless of c.

Let K_1 denote the giant strong component of $D_{n,p}$, as discovered by Karp [12]. We consider a process that builds a large Hamiltonian subgraph of K_1. Our aim is to construct (something close) to a copy of the random graph $D_{5-in,5-out}$ as a large subgraph of K_1. In the random graph $D_{k-in,k-out}$ each $v \in [n]$ independently chooses k in-neighbors and k outneighbors to make a digraph with $\approx 2kn$ random edges. It has been shown by Cooper and Frieze [5], [6] that $D_{k-in,k-out}$ is Hamiltonian w.h.p. provided that $k \geq 2$. Taking $k = 5$ as opposed to $k = 2$ will greatly simplify the discussion. In order to do this, we will construct $D_{n,p}$ as the union of two independent copies D_{red}, D_{blue} of $D_{n,q}$ where $1 - p = (1 - q)^2$ so that $q = \frac{c}{2n} + O(n^{-2})$. One copy will have red edges and the other copy will have blue edges. A red edge (v, w) will be assoiciated with the vertex v and a blue edge (v, w) will be associated with the vertex w. In this way, the vertex v will be incident to a random number of red out-edges and to a random number of blue in-edges. These edge sets will be independent by construction. We say in the following that w is a blue in-neighbor of v if (w, v) is an edge of D_{blue} and that w is a red out-neighbor of v if (v, w) is an edge of D_{red}.

We construct a sequence of sets $S_0 = \emptyset, S_1, S_2, \ldots, S_L \subseteq K_1$ as follows: suppose now that we have constructed S_ℓ, $\ell \geq 0$. We construct $S_{\ell+1}$ from S_ℓ via one of two cases:

Construction of S_L
Case a: If there is a vertex $v \in S_\ell$ that has at most four blue in-neighbors outside S_ℓ then we add the blue in-neighbors of v outside S_ℓ to S_ℓ to make $S_{\ell+1}$. Similarly, if there is a vertex $v \in S_\ell$ that has at most four red out-neighbors outside S_ℓ then we add the red out-neighbors of v outside S_ℓ to S_ℓ to make $S_{\ell+1}$.
Case b: If there is a vertex $v \in K_1 \setminus S_\ell$ that has at most four blue in-neighbors in $K_1 \setminus S_\ell$ then we add v and the blue in-neighbors of v to S_ℓ to make $S_{\ell+1}$. Similarly, if there is a vertex $v \in K_1 \setminus S_\ell$ that has at most four red out-neighbors in $K_1 \setminus S_\ell$ then we then we add v and the red out-neighbors of v to S_ℓ to make $S_{\ell+1}$.

S_L is the set we end up with when there are no more vertices to add. We note that S_L is well-defined and does not depend on the order of adding vertices. Indeed, suppose we have two distinct outcomes $O_1 = v_1, v_2, \ldots, v_r$ and $O_2 = w_1, w_2., \ldots, w_s$. Assume without loss of generality that there exists i which is the smallest index such that $w_i \notin O_1$. Then, $X = \{w_1, w_2, \ldots, w_{i-1}\} \subseteq O_1 = \{v_1, v_2, \ldots, v_r\}$. If w_i invoked Case a or Case b then w_i has at most 4 blue in-neighbors or at most 4 red out- neighbors in $K_1 \setminus X$ hence in $K_1 \setminus O_1 \subseteq K_1 \setminus X$. This contradicts the fact that $w_i \notin O_1$. Otherwise w_i was added to X because there exists a vertex $u \in X$ such that w_i is a blue in-neighbor (or a red out-neighbor respectively) of u and u has at most 4 blue in-neigbors (red out-neighbors resp.) in $K_1 \setminus X$. Thus $u \in O_1$ has at most 4 blue in-neigbors (red out-neighbors resp.) in $K_1 \setminus X \subseteq K_1 \setminus X$. Once again, this contradicts the fact that $w_i \notin O_1$.

We will argue below in Section 1.1 that w.h.p. the graph Γ_L *underlying* the digraph D_L induced by S_L is a forest plus a few small components (the graph underlying a digraph is obtained by ignoring orientation). Each tree in Γ_L will w.h.p. have at most $\log n$ vertices and w.h.p. Γ_L will have $o(n)$ vertices lying on non-tree components. From now on, when we refer to trees, they are either trees of Γ_L or digraphs whose underlying graphs are trees of Γ_L.

Notation 1: Let $\vec{\mathcal{T}}$ denote the set of trees in Γ_L. Each tree T of Γ_L will appear as a digraph $\vec{T}$ in D_L when we take account of orientation. For $\vec{T} \in \vec{\mathcal{T}}$ let $\vec{\mathcal{P}}_T$ be the set of vertex disjoint packings of *properly oriented* paths in $\vec{T}$ where we allow only paths whose start vertex has in-neighbors in $K_1 \setminus V(\vec{T})$ and whose end vertex has red out-neighbors in $K_1 \setminus V(\vec{T})$. Here we allow paths of length 0, so that a single vertex with neighbors in $K_1 \setminus V(\vec{T})$ counts as a path. For $P \in \vec{\mathcal{P}}_T$ let $n(\vec{T}, P)$ be the number of vertices in $\vec{T}$ that are not covered by P. Let $\phi(\vec{T}) = \min_{P \in \vec{\mathcal{P}}_T} n(\vec{T}, P)$ and $\vec{\mathcal{Q}}(\vec{T}) \in \vec{\mathcal{P}}$ denote a set of paths that leaves $\phi(\vec{T})$ vertices of $\vec{T}$ uncovered i.e. satisfies $n(\vec{T}, \vec{\mathcal{Q}}(\vec{T})) = \phi(\vec{T})$.

We will prove

Theorem 1.1. *Let $p = c/n$ where $c > 1$ is a sufficiently large constant. Then w.h.p.*

$$\vec{L}_{c,n} \approx |V(K_1)| - \sum_{\vec{T} \in \vec{\mathcal{T}}} \phi(\vec{T}). \tag{1}$$

The RHS of (1), modulo the $o(n)$ vertices that are spanned by non-tree components in Γ_L, is clearly an upper bound on the largest directed cycle in K_1. Any cycle must omit at least $\phi(\vec{T})$ vertices from each $\vec{T} \in \vec{\mathcal{T}}$. On the other hand, as we show below, w.h.p. there is cycle H that spans $V^* = (K_1 \setminus S_L) \cup \bigcup_{T \in \mathcal{T}} V(\mathcal{Q}(T))$. The length of H is equal to the RHS of (1).

The size of K_1 is well-known. Let x be the unique solution of $xe^{-x} = ce^{-c}$ in $(0, 1)$. Then w.h.p. (see e.g. [10], Theorem 13.2),

$$|K_1| \approx \left(1 - \frac{x}{c}\right)^2 n. \tag{2}$$

Equation (4.5) of Erdős and Rényi [8] tells us that

$$x = \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} (ce^{-c})^k = ce^{-c} + c^2 e^{-2c} + O(c^3 e^{-3c}). \tag{3}$$

We will argue below that w.h.p., as c grows, that

$$\sum_{\vec{T} \in \vec{\mathcal{T}}} \phi(\vec{T}) = (c^2 e^{-2c} + O(c^3 e^{-3c}))n. \tag{4}$$

The term $c^2 e^{-2c} n$ arises from vertices of out-degree one sharing a common out-neighbor or vertices of in-degree one sharing a common in-neighbor.

We therefore have the following improvement to the estimate in [13].

Corollary 1.2. *W.h.p., as c grows,*

$$\vec{L}_{c,n} \approx \left(1 - 2e^{-c} - (c^2 + 2c - 1)e^{-2c} - O(c^3 e^{-3c})\right) n. \tag{5}$$

Note the term $2e^{-c} - e^{-2c}$ accounts for vertices of in- or out-degree 0. In principle we can compute more terms than what is given in (5). We claim next that there exists some function $\vec{f}(c)$ such that the sum in (1) is concentrated around $\vec{f}(c)n$. In other words, the sum in (1) has the form $\approx \vec{f}(c)n$ w.h.p.

Theorem 1.3. *(a) There exists a function $\vec{f}(c)$ such that for any fixed $\epsilon > 0$, there exists n_ε such that for $n \geq n_\varepsilon$,*

$$\left| \frac{\mathbb{E}[\vec{L}_{c,n}]}{n} - \vec{f}(c) \right| \leq \epsilon. \tag{6}$$

(b)

$$\frac{\vec{L}_{c,n}}{n} \to \vec{f}(c) \text{ a.s.}$$

We will show that taking $c \geq 200$ in Theorems 1.1 and 1.3 suffices.

We will prove Theorem 1.3 in Section 3. We are grateful to a reviewer for pointing out that $\vec{L}_{c,n}/n \to f(c)$ in $L^r, r \geq 1$ because $\vec{L}_{c,n}/n$ is an a.s. bounded random variable.

1.1 Structure of D_L:

We first bound the size of S_L. We need the following lemma on the density of small sets.

Lemma 1.4. *W.h.p., every set $S \subseteq [n]$ of size at most $n_0 = n/100c^3$ contains less than $3|S|/2$ edges in $D_{n,p}$.*

Proof. The expected number of sets invalidating the claim can be bounded by

$$\sum_{s=3}^{n_0} \binom{n}{s} \binom{s(s-1)}{3s/2} \left(\frac{c}{n}\right)^{3s/2} \leq \sum_{s=3}^{n_0} \left(\frac{ne}{s} \cdot \left(\frac{2se}{3}\right)^{3/2} \cdot \left(\frac{c}{n}\right)^{3/2} \right)^s$$
$$= \sum_{s=3}^{n_0} \left(\frac{e^{5/2}(2c)^{3/2}s^{1/2}}{3^{3/2}n^{1/2}} \right)^s = o(1).$$

□

Now consider the construction of S_L. Let $A \subseteq K_1$ be the set of the vertices with blue in-degree less than $D = 30$ or red out-degree less than D in K_1. Let $S_0' = (A \cup N_b(A) \cup N_r(A)) \cap S_L \subseteq S_L$, where $N_b(A)$ is the set of blue in-neighbors of vertices in A and $N_r(A)$ is the set of red out-neighbors of vertices in A. If we start with $S_0 = S_0'$ and run the process for constructing Γ_L then we will produce the same S_L as if we had started with $S_0 = \emptyset$. This is because, as we have shown, the order of adding vertices does not matter. Now w.h.p. there are at most $n_D = \frac{2c^D e^{-c}}{D!} n$ vertices of blue in-degree at most D or red out-degree at most D, (see for example Theorem 3.3 of [10] that deals with the same question as it relates to degrees in $G_{n,p}$).

Now suppose that the process runs for another k rounds. Then S_k contains at least kD edges and at most $Dn_D + 5k$ vertices. This is because round k adds at most five *new* vertices to S_k and the k vertices that take the role of v have either (i) blue in-degree at least D with all blue in-neighbors in S_k or (i) red out-degree at least D with all red out-neighbors in S_k. If k reaches $2n_D$ then

$$\frac{e(S_k)}{|S_k|} \geq \frac{2Dn_D}{(D+10)n_D} = \frac{3}{2}.$$

So, by Lemma 1.4, we can assert that w.h.p. the process runs for less than $2n_D$ rounds and,

$$|V(\Gamma_L)| \leq (D+10)n_D = (D+10)\frac{2c^D e^{-c}}{D!}n \leq 2(D+10)\left(\frac{ec}{D}\right)^D ne^{-c} \leq ne^{-c/2}. \tag{7}$$

The last inequality holds for $c \geq 200$ and $D = 30$.

We note the following properties of S_L. Let

$$V_1 = K_1 \setminus S_L \text{ and } V_2 = \{v \in S_L : \ v \text{ has at least one blue in-neighbor and at least one red out-neighbor in } V_1\}.$$

Then,

G1 Each vertex $v \in S_L \setminus V_2$ has no blue in-neighbors or no red out-neighbors in V_1.

G2 Each $v \in V_1 \cup V_2$ has at least five blue in-neighbors and five red out-neighbors in V_1.

Now consider a component K of Γ_L. Let $C_0 = C_0(K) = \{v_1, v_2, \ldots, v_L\}$ denote the set of vertices in K that are v in some step in the construction of D_L, indexed by the round in which they are added. Since a vertex may invoke some step in the construction of D_L at most twice we have,

$$|C_0(K)| \geq L/2. \tag{8}$$

At the same time, at each step the set $|K \setminus C_0(K)|$ may grow by at most 4 and so

$$|K \setminus C_0(K)| \leq 4L \leq 8|C_0(K)|. \tag{9}$$

Hence

$$|C_0(K)| \geq \frac{|K|}{9}. \tag{10}$$

We next show that w.h.p., only a small component K can satisfy (10). K will have at least $|K|/9$ vertices for which either there are no blue in-neghbors outside K or no red out-neighbors outside of K. It will also contain a spanning tree in the graph undelying $D_{n,p}$. So, the expected number of components of size $k \leq ne^{-c/2}$ that satisfy this condition is at most

$$\binom{n}{k} k^{k-2} \left(\frac{c}{n}\right)^{k-1} \binom{k}{k/9} \times \left(2\left(1-\frac{c}{2n}\right)^{(n-k)}\right)^{k/9} \leq \left(\frac{ne}{k}\right)^k k^{k-2} \left(\frac{c}{n}\right)^{k-1} 2^{10k/9} e^{-ck/20}$$

$$\leq \frac{n}{ck^2}\left(2^{10/9}ce^{1-c/20}\right)^k = o(n^{-2}), \quad (11)$$

if $c \geq 200$ and $k \geq \log n$.

So, we can assume that all components are of size at most $\log n$. Then the expected number of vertices on components that are not trees is bounded by

$$\sum_{k=2}^{\log n}\binom{n}{k}k^{k+1}\left(\frac{c}{n}\right)^k\binom{k}{k/9}\times\left(2\left(1-\frac{c}{2n}\right)^{(n-k)}\right)^{k/9} \leq \sum_{k=2}^{\log n}\left(\frac{ne}{k}\right)^k k^{k+1}\left(\frac{c}{n}\right)^k 2^{10k/9}e^{-ck/20}$$
$$\leq \sum_{k=2}^{\log n} k\left(2^{10/9}ce^{1-c/20}\right)^k = O(1).$$

The Markov inequality implies that w.h.p. such components span at most $\log n = o(n)$ vertices.

2 Proof of Theorem 1.1

For $\vec{T} \in \vec{\mathcal{T}}$, let $\vec{X}_T$ be the set obtained by contracting each path $\vec{P}$ of $\vec{\mathcal{Q}}(\vec{T})$ to a vertex $v_{\vec{P}}$ with blue in-neighbors in V_1 equal to the blue in-neighbors in V_1 of the start vertex of $\vec{P}$ and red out-neighbors in V_1 equal to the red out-neighbors in V_1 of the end vertex of $\vec{P}$. Note that the colors of the internal edges of a path $\vec{P}$ do not play a role here. Let $\vec{X}^* = \bigcup_{\vec{T}\in\vec{\mathcal{T}}}\vec{X}_T$. By construction, the digraph induced by V_1 contains a copy of $D_{5-in,5-out}$ with $N = |V_1|$ vertices. Indeed, the blue edges contributing the 5-in edges and the red edges contributing the 5-out edges. For each $v \in V_1$, the blue in-neighbors form a random set of size at least five, independent of the other vertices in V_1. Similarly for the red out-neighbors.

We let D^* be the digraph with vertex set $V_1^* = V_1 \cup \vec{X}^*$ and a copy of $D_{5-in,5-out}$ on V_1 and for each $x \in \vec{X}^*$ five red edges joining x to V_1 and five blue edges from V_1 to x.

Our next task is to prove that the random digraph D^* defined in the previous section contains a Hamilton cycle. Let H denote such a cycle through V_1^*. We obtain a Hamilton cycle of V^* (defined following Theorem 1.1) by uncontracting each path $\vec{P}$ of $\vec{\mathcal{Q}}(\vec{T})$. This will complete the proof of Theorem 1.1. Our proof of the existence of H will be very similar to the proof in Cooper and Frieze [7]. It doesn't really offer any new technical insights and so we have placed the proof into an appendix.

3 Proof of Theorem 1.3

For $\vec{T} \in \vec{\mathcal{T}}$ we let $v_0(\vec{T})$ denote the set of vertices in $\vec{T}$ that do not have neighbors outside $\vec{T}$. For $v \in K_1$ we let $\phi(v) = \phi(\vec{T})/|v_0(\vec{T})|$ if $v \in v_0(T)$ for some $\vec{T} \in \vec{\mathcal{T}}$ and $\phi(v) = 0$ otherwise.

Thus

$$\sum_{T\in\vec{\mathcal{T}}}\phi(\vec{T})=\sum_{v\in K_1}\phi(v).$$

Hence (1) can be rewritten as,

$$\vec{L}_{c,n}\approx|K_1|-\sum_{v\in K_1}\phi(v). \tag{12}$$

Let $k_1=k_1(\epsilon,c)$ be the smallest positive integer such that

$$\sum_{k=k_1-1}^{\infty}(e^9 2^{11}ce^{-c/5})^k<\frac{\epsilon}{3}.$$

Note that for $\varepsilon\leq 1/2$ and and $c\geq 200$, we have

$$k_1\leq\frac{30}{c}\log\frac{1}{\varepsilon}. \tag{13}$$

as

$$\sum_{k=k_1-1}^{\infty}(e^9 2^{11}ce^{-c/5})^k\leq 2((e^9 2^{11}c)^{5/c}e^{-1})^{-6\log\varepsilon}\leq 2((e^9 2^{11}200)^{5/200}e^{-1})^{-6\log\varepsilon}<\frac{\epsilon}{3}.$$

To begin let $\vec{K}_{5,5}$ denote the complete bipartite digraph with ten vertices, five in each part of the partition. The arcs inside $\vec{K}_{5,5}$ are consider to have both colors, red and blue. For $v\in K_1$ let D_v be the digraph consisting of the vertices of $D=D_{n,p}=D_{blue}\cup D_{red}$ that are within distance k_1 from v, where for every vertex u in the k_1 neighborhood of v we introduce a new copy of $\vec{K}_{5,5}$ and join u to each vertex of the same one part of the bipartition of its $\vec{K}_{5,5}$ by a blue in-arc and a red out-arc from u. Distance here is graph distance in the undirected graph underlying D. We consider the algorithm for the construction of Γ_L on G_v, the graph underlying D_v. Let $K_{1,v},\Gamma_{L,v},V_{1,v},S_{L,v},\upsilon_{0,v}(\vec{T})$ be the corresponding sets/quantities.

For a tree $\vec{T}\in S_{L,v}$ let $\vec{f}(\vec{T})$ be equal to $|\vec{T}|$ minus the maximum number of vertices that can be covered by a set of vertex disjoint paths with endpoints in $V_{2,v}$ (we allow paths of length 0). For $v\in K_1$, if v belongs to some tree $\vec{T}\in S_{L,v}$ set $\vec{f}(v)=\vec{f}(\vec{T})/\upsilon_{0,v}(\vec{T})$, otherwise set $\vec{f}(v)=0$.

For $v\in K_1$ let $t(v)=1$ if $v\in V_1$ or if $v\in S_L$ and in Γ_L, v lies in a component with at most k_1-2 vertices in Γ_L. Set $t(v)=0$ otherwise. Observe that if $t(v)=1$ then $\phi(v)=\vec{f}(v)$. Otherwise $|\phi(v)-\vec{f}(v)|\leq 1$.

By repeating the arguments used to prove (11) and (10) it follows that if $t(v)=0$ then v lies on a subgraph spanned by some set of vertices K of size at most $\log n$. In addition at least

$(|K|-1)/9$ vertices in $K\setminus\{v\}$ either do not have blue in-neighbors or red out-neighbors outside K. Thus the expected number of vertices v satisfying $t(v)=0$ is bounded by

$$\begin{aligned}
&\sum_{k=k_1-1}^{\log n}\sum_{j=k}^{9k}\binom{n}{j}\binom{j}{k}j^{j-2}(2p)^{j-1}\times\left(2\left(1-\frac{p}{2}\right)^{(n-j)}\right)^k\\
&\leq 2n\sum_{k=k_1-1}^{\log^2 n}9k\left(\frac{e}{9k}\right)^{9k}2^{9k}(9k)^{9k-2}(2c)^{k-1}2^ke^{-ck/5}\\
&\leq 2n\sum_{k=k_1-1}^{\infty}(e^9 2^{11}ce^{-c/5})^k<\frac{\epsilon n}{3}.
\end{aligned}$$

A vertex $v\in[n]$ is *good* if the ith level of its Breadth First Search (BFS) neighborhood has size at most $3(2c)^i k_1/\epsilon$ for every $i\leq k_1$ and it is *bad* otherwise. Here the BFS is done on the graph underlying D. Because the expected size of the i^{th} neighborhood is $\approx(2c)^i$ we have by the Markov inequality that v is bad with probability at most $(1+o(1))\varepsilon/3\leq\varepsilon/2$ and so the expected number of bad vertices is bounded by $\varepsilon n/2$. Thus

$$\begin{aligned}
\mathbb{E}\left(\left|\sum_{v\in V}\phi(v)-\sum_{v\text{ is good}}\vec{f}(v)\right|\right)&\leq\mathbb{E}\left(\left|\sum_{v\in V}\phi(v)-\sum_{v\in V}\vec{f}(v)\right|\right)+\mathbb{E}\left(\left|\sum_{v\text{ is bad}}\vec{f}(v)\right|\right)\\
&\leq\mathbb{E}\left(\left|\sum_{v:t(v)=0}|\phi(v)-\vec{f}(v)\right|\right)+\mathbb{E}\left(\sum_{v\text{ is bad}}1\right)\\
&\leq\mathbb{E}\left(\sum_{v:t(v)=0}1\right)+\frac{\epsilon n}{2}\\
&\leq\frac{\epsilon n}{3}+\frac{\epsilon n}{2}<\epsilon n.
\end{aligned}$$

Let $\mathcal{H}_\varepsilon$ be the set of BFS neighborhoods that are good i.e. whose ith levels are of size at most $3(2c)^i k_1/\epsilon$ for every $i\leq k_1$. Every element of $\mathcal{H}_\varepsilon$ corresponds to a pair (H,o_H) where H is a digraph and o is a distinguished vertex of H, that is considered to be the root. Also for $v\in K_1$ let $D(N_{k_1}(v))$ be the subdigraph induced by the ${k_1}^{th}$ neighborhood of v. For $(H,o_H)\in\mathcal{H}_\varepsilon$ let $int(H)$ be the set of vertices incident to the first k_1-1 neighborhoods of o_H and let $Aut(H,o_H)$ be the number of automorphisms of H that fix o_H. Note that each good vertex v is associated with a pair $(H,o_H)\in\mathcal{H}_\varepsilon$ from which we can compute $\vec{f}(v)$, since $\vec{f}(v)=\vec{f}(o_H)$. Thus, if now

$$M=|E(K_1)|, N=|K_1|,$$

$$\mathbb{E}\left(\sum_{v\text{ is good}}\vec{f}(v)\middle|M,N\right)=\sum_v\sum_{k\geq 1}\sum_{\substack{(H,o_H)\in\mathcal{H}_\varepsilon\\(D(N_{k_1}(v)),v)=(H,o_H)\\|V(H)|=k}}\rho_{H,o_H}\vec{f}(o_H)$$

$$= o(n) + \sum_v \sum_{k\geq 1} \sum_{\substack{(H,o_H)\in\mathcal{H}_\varepsilon \\ H \text{ is a tree} \\ (D(N_{k_1}(v)),v)=(H,o_H) \\ |V(H)|=k}} \rho_{H,o_H}\vec{f}(o_H), \tag{14}$$

where ρ_{H,o_H} is the probability $(D(N_{k_1}(v)), v) = (H, o_H)$ in K_1. We show in Section 3.1 that

$$\rho_{H,o_H} \approx \frac{1}{Aut(H, o_H)} \left(\frac{N}{M}\right)^{k-1} \lambda^{2k-2} \frac{e^{2k\lambda}}{f_1(\lambda)^{2k}}, \tag{15}$$

where f_1 is defined in (18) below and λ satisfies (19) below.

Finally observe that with the exception of the $o(n)$ term, all the terms in (14) are independent of n. We let

$$\vec{f}_\varepsilon(c) = \sum_{k\geq 1} \sum_{\substack{(H,o_H)\in\mathcal{H}_\varepsilon \\ H \text{ is a tree} \\ |V(H)|=k}} \frac{\vec{f}(o_H)}{Aut(H, o_H)} \left(\frac{N}{M}\right)^{k-1} \lambda^{2k-2} \frac{e^{2k\lambda}}{f_1(\lambda)^{2k}}. \tag{16}$$

Then for a fixed c, we see that $\vec{f}_\varepsilon(c)$ is monotone increasing as $\varepsilon \to 0$. This is simply because $\mathcal{H}_\varepsilon$ grows. Furthermore, $\vec{f}_\varepsilon(c) \leq 1$ and so the limit $\vec{f}(c) = \lim_{\varepsilon\to 0} f_\varepsilon(c)$ exists. Let $S_{\varepsilon,n}$ be the number of vertices in $D_{n,p}$ (i) whose first k_1 neighborhoods are good and so total at most $4(2c)^{k_1}k_1/\varepsilon - 1$ vertices, and (ii) span a cycle in the underlying graph. The $o(n)$ term in (14) is bounded by $S_{\varepsilon,n}$. Hence, with $s = 4(2c)^{k_1}k_1/\varepsilon$, the $o(n)$ term is bounded by

$$\sum_{i=1}^{s} i\binom{n}{i} i^{i-2}\binom{i}{2}(2p)^i \leq \sum_{i=1}^{s} i\left(\frac{en}{i}\right)^i i^i(2p)^i \leq 2s(2ec)^s \leq \log\frac{1}{\varepsilon} \times e^{\log\frac{1}{\varepsilon}} \leq \frac{1}{\varepsilon^2},$$

which depends only on ε.

This verifies part (a) of Theorem 1.3. For part (b), we prove, (see (30)),

Lemma 3.1.

$$\mathbb{P}(|\vec{L}_{c,n} - \mathbb{E}(\vec{L}_{c,n})| \geq \varepsilon n + n^{3/4}) = O(n^{-2}).$$

Proof. To prove this we show that if $\nu(H)$ is the number of copies of H in K_1 then $H \in \mathcal{H}_\varepsilon$ implies that

$$\mathbb{P}(|\nu(H) - \mathbb{E}(\nu(H))| \geq n^{3/5}) = O(n^{-3}). \tag{17}$$

The inequality follows from a version of Azuma's inequality (see (30)), and the lemma follows from taking a union bound over

$$\exp\left\{O\left(\frac{c^{k_1(\epsilon,c)}k_1(\epsilon,c)}{\epsilon}\right)\right\} = \exp\left\{O\left(\frac{c^{\frac{30}{c}\log\frac{1}{\varepsilon}}\frac{30}{c}\log\frac{1}{\varepsilon}}{\varepsilon}\right)\right\}$$
$$= \exp\left\{O\left(\frac{(1/\varepsilon)^{\frac{30}{c}\log c}\log\frac{1}{\varepsilon}}{\varepsilon}\right)\right\} = \exp\left\{O((1/\varepsilon)^3)\right\}$$

graphs H. Note also that the $o(n)$ term in (14) is bounded by $S_{\varepsilon,n}$ and the probability that this exceeds $n^{1/2}$ is certainly at most the RHS of (17). We will give details of our use of the Azuma inequality in Section 3.1. □

Part (b) of Theorem 1.3 follows by letting $\varepsilon \to 0$ and from the Borel-Cantelli lemma.

3.1 A Model of K_1

K_1 induces a random digraph with minimum in-degree and out-degree at least one. K_1 is distributed as a random strongly connected digraph with N vertices and M edges. This follows from the fact that each such digraph has the same number of extensions to a digraph with n vertices and m edges where K_1 is the unique giant strongly connected component. Most vertices of K_1 will have in-degree and out-degree close to c, since c is large. It follows from Theorem 3 of Cooper and Frieze [7] that a random digraph with this degree sequence has a probability of being strongly connected that is asymptotic to $e^{-\beta}$ where $\beta = \beta(c) \to 0$ as $c \to \infty$. It follows from this that we can model the digraph induced by K_1 as a random digraph with N vertices and M edges. The probability of any event will be inflated by at most $(1 + o(1))e^{\beta}$ by conditioning on strong connectvity. We denote this model by $D^{\pm 1}_{N,M}$.

3.1.1 Random Sequence Model

This is essentially a repeat of Section 3.1.1 of [1]. The differences are minor, but we feel we need to include the argument. We must now take some time to explain the model we use for $D^{\pm 1}_{N,M}$. We use a variation on the pseudo-graph model of Bollobás and Frieze [3] and Chvátal [4]. Given a sequence $\mathbf{x} = (x_1, x_2, \ldots, x_{2M}) \in [n]^{2M}$ of $2M$ integers between 1 and N we can define a (multi)-digraph $D_{\mathbf{x}} = D_{\mathbf{x}}(N, M)$ with vertex set $[N]$ and edge set $\{(x_{2i-1}, x_{2i}) : 1 \leq i \leq M\}$. The in-degree $d_{\mathbf{x},-}(v)$ of $v \in [N]$ and the out-degree $d_{\mathbf{x},+}(v)$ of $v \in [N]$ are given by

$$d_{\mathbf{x},-}(v) = |\,\{j \in [M] : x_{2j} = v\}\,| \text{ and } d_{\mathbf{x},+}(v) = |\,\{j \in [M] : x_{2j-1} = v\}\,|.$$

If $\mathbf{x}$ is chosen randomly from $[N]^{2M}$ then $D_{\mathbf{x}}$ is close in distribution to $D_{N,M}$. Indeed, conditional on being simple, $D_{\mathbf{x}}$ is distributed as $D_{N,M}$. To see this, note that if $D_{\mathbf{x}}$ is simple then it has vertex set $[N]$ and M edges. Also, there are $M!$ distinct equally likely values of $\mathbf{x}$ which yield the same digraph.

Our situation is complicated by there being a lower bound of one on the minimum in-degree and out-degree. So we let

$$[N]^{2M}_{\delta\pm\geq 1} = \{\mathbf{x} \in [N]^{2M} : d_{\mathbf{x},\pm}(j) \geq 1 \text{ for } j \in [N]\}.$$

Let $D_{\mathbf{x}}$ be the multi-graph $D_{\mathbf{x}}$ for $\mathbf{x}$ chosen uniformly from $[N]^{2M}_{\delta\pm\geq 1}$. It is clear then that conditional on being simple, $D_{\mathbf{x}}$ has the same distribution as $D^{\pm 1}_{N,M}$. It is important therefore to estimate the probability that this graph is simple. For this and other reasons, we need to have an understanding of the degree sequence $d_{\mathbf{x}}$ when $\mathbf{x}$ is drawn uniformly from $[N]^{2M}_{\delta\pm\geq 1}$. Let

$$f_1(\lambda) = e^{\lambda} - 1. \tag{18}$$

Lemma 3.2. *Let* $\mathbf{x}$ *be chosen randomly from* $[N]^{2M}_{\delta\pm\geq 1}$*. Let* $Y_j, Z_j, j = 1, 2, \ldots, N$ *be independent copies of a* truncated Poisson *random variable* $\mathcal{P}$*, where*

$$\mathbb{P}(\mathcal{P} = t) = \frac{\lambda^t}{t! f_1(\lambda)}, \qquad\qquad t \geq 1.$$

Here λ *satisfies*

$$\frac{\lambda e^\lambda}{f_1(\lambda)} = \frac{M}{N}. \tag{19}$$

Then $\{d_{\mathbf{x},-}(j)\}_{j\in[N]}$ *is distributed as* $\{Y_j\}_{j\in[N]}$ *conditional on* $Y = \sum_{j\in[n]} Y_j = M$ *and* $\{d_{\mathbf{x},+}(j)\}_{j\in[N]}$ *is distributed as* $\{Z_j\}_{j\in[N]}$ *conditional on* $Z = \sum_{j\in[n]} Z_j = M$*.*

Proof. This can be derived as in Lemma 4 of [2]. □

We note that w.h.p.

$$N \geq n(1 - 2e^{-c/2}) \text{ and } M \in (1 \pm \varepsilon_1)cN, \tag{20}$$

where $\varepsilon_1 = c^{-1/3}$. The bound on N follows from (2) and (7) and the bound on M follows from the fact that in $G_{n,p}$,

$$\mathbb{P}\left(\exists S : |S| = N, e(S) \notin (1 \pm \varepsilon_1)N(N-1)p\right) \leq 2\binom{n}{N} \exp\left\{-\frac{\varepsilon_1^2 N(N-1)p}{3}\right\} = o(1).$$

It follows from (19) and (20) and the fact that $e^\lambda / f_1(\lambda) \to 1$ as $c \to \infty$ that for large c,

$$\lambda = c\left(1 + O(e^{-c})\right). \tag{21}$$

We note that the variance σ^2 of $\mathcal{P}$ is given by

$$\sigma^2 = \frac{\lambda(\lambda+1)e^\lambda f_1(\lambda) - \lambda^2 e^{2\lambda}}{f_1^2(\lambda)}.$$

Furthermore,

$$\mathbb{P}\left(\sum_{j=1}^N Y_j = M\right) = \frac{1}{\sigma\sqrt{2\pi N}}(1 + O(N^{-1}\sigma^{-2})) \tag{22}$$

and

$$\mathbb{P}\left(\sum_{j=2}^N Y_j = M - d\right) = \frac{1}{\sigma\sqrt{2\pi N}}\left(1 + O((d^2+1)N^{-1}\sigma^{-2})\right). \tag{23}$$

This is an example of a local central limit theorem. See for example, (5) of [2]. It follows by repeated application of (22) and (23) that if $k = O(1)$ and $d_1^2 + \cdots + d_k^2 = o(N)$ then

$$\mathbb{P}\left(Y_i = d_i, i = 1, 2, \ldots, k \mid \sum_{j=1}^N Y_j = M\right) \approx \prod_{i=1}^k \frac{\lambda^{d_i}}{d_i! f_1(\lambda)}. \tag{24}$$

Let $\nu_{\mathbf{x},-}(s)$ denote the number of vertices of in-degree s in $D_\mathbf{x}$ and let $\nu_{\mathbf{x},+}(s)$ denote the number of vertices of out-degree s in $D_\mathbf{x}$.

Lemma 3.3. *Suppose that* $\log N = O((N\lambda)^{1/2})$. *Let* $\mathbf{x}$ *be chosen randomly from* $[N]^{2M}_{\delta\geq 2}$. *Then as in equation (7) of [2], we have that with probability* $1 - o(N^{-10})$,

$$\left|\nu_{\mathbf{x},\pm}(j) - \frac{N\lambda^j}{j!f_1(\lambda)}\right| \leq \left(1 + \left(\frac{N\lambda^j}{j!f_1(\lambda)}\right)^{1/2}\right)\log^2 N,\ \ 1 \leq j \leq \log N. \tag{25}$$

$$\nu_{\mathbf{x}}(j) = 0, \quad j \geq \log N. \tag{26}$$

We can now show that $D_{\mathbf{x}}$, $\mathbf{x} \in [N]^{2M}_{\delta\pm\geq 1}$ is a good model for $D^{\pm 1}_{N,M}$. For this we only need to show now that

$$\mathbb{P}(D_{\mathbf{x}} \text{ is simple}) = \Omega(1). \tag{27}$$

Again, this follows as in [2].

Given a tree H with k vertices of in-degrees $y_1, y_2, ..., y_k$ and out-degrees $z_1, z_2, ..., z_k$ and a fixed vertex v we see that if ρ_H is the probability that $D(N_{k_1}(v)) = H$ in $D_{\mathbf{x}}$ then we have

$$\rho_H \approx \binom{N}{k-1}\frac{(k-1)!}{Aut(H,o_H)}\sum_{D^-,D^+=k-1}^{\infty}$$

$$\sum_{\substack{d_1^-\geq y_1,\dots,d_k^-\geq y_k\\ d_1^-+\cdots+d_k^-=D^-\\ d_1^+\geq z_1,\dots,d_k^+\geq z_k\\ d_1^++\cdots+d_k^+=D^+}} \prod_{i=1}^{k}\frac{\lambda^{d_i^-+d_i^+}}{d_i^-!d_i^+!f_1(\lambda)^2}\binom{M}{k-1}(k-1)!\prod_{i=1}^{k}\frac{d_i^-!d_i^+!}{(d_i^- - y_i)!(d_i^+ - z_i)!}\frac{1}{M^{2k-2}} \tag{28}$$

$$\approx \left(\frac{N}{M}\right)^{k-1}\frac{\lambda^{2k-2}}{Aut(H,o_H)f_1(\lambda)^{2k}}\sum_{\substack{d_1^-+\cdots+d_k^-=D^-\\ d_1^++\cdots+d_k^+=D^+}}\prod_{i=1}^{k}\frac{\lambda^{d_i^-+d_i^+-y_i-z_i}}{(d_i^- - y_i)!(d_1^+ - z_i)!}$$

$$= \left(\frac{N}{M}\right)^{k-1}\frac{\lambda^{2k-2}}{Aut(H,o_H)f_1(\lambda)^{2k}}\left(\sum_{D=k-1}^{\infty}\frac{(k\lambda)^{D-(k-1)}}{(D-(k-1))!}\right)^2 \tag{29}$$

$$\approx \frac{1}{Aut(H,o_H)}\left(\frac{N}{M}\right)^{k-1}\lambda^{2k-2}\frac{e^{2k\lambda}}{f_1(\lambda)^{2k}}.$$

Explanation for (28)**:** We use (24) to obtain the probability that the in-degrees and out-degrees of $[k]$ are $d_1^-, d_1^+, \dots, d_k^-, d_k^+$. This accounts for the term $\prod_{i=1}^{k}\frac{\lambda^{d_i^-+d_i^+}}{d_i^-!d_i^+!f_1(\lambda)^2}$. Implicit here is that $d_i^-, d_i^+ = O(\log n)$, from (26). The contributions to the sum of $D^-, D^+ \geq k\log n$ can therefore be shown to be negligible. We use the fact that k is small to argue that w.h.p. H is induced. We choose the vertices, other than v in $\binom{N}{k-1}$ ways and then $\frac{(k-1)!}{Aut(H,o_H)}$ counts the number of copies of H in K_k. We then choose the place in the sequence to put these edges in $\binom{M}{k-1}(k-1)!$ ways. Finally note that the probability the y_i occurrences of the ith vertex are as claimed is asymptotically equal to $\frac{d_i^-(d_i^- -1)\cdots(d_i^- -y_i+1)}{M^{z_i}}$ and this explains the factor $\prod_{i=1}^{k}\frac{d_i^-!d_i^+!}{(d_i^- -y_i)!(d_i^+ -z_i)!}\frac{1}{M^{2k-2}}$.

Explanation for (29)**:** We use the identity

$$\sum_{\substack{d_1,\ldots,d_k\\ d_1+\cdots+d_k=D}} \frac{D!}{d_1!\cdots d_k!} = k^D.$$

It only remains to verify (17). It follows from the above that $\mathbb{E}(\nu(H) \mid M, N) = \Omega(N)$. We first condition on a degree sequence $\mathbf{x}$ satisfying (25). Then we condition on no element $\log n$ times or more in $\mathbf{x}$. The latter occurs with probability

$$O\left(n^{1/2}e^{-\lambda}\frac{\lambda^{\log n}}{\log n!}\right) = O\left(n^{1/2}e^{-\lambda}\left(\frac{e\lambda}{\log n}\right)^{\log n}\right) = O(n^{-3}).$$

Interchanging two elements in a permutation can only change $\nu(H)$ by $(\log n)^{k_1} = n^{o(1)}$. We can therefore apply Azuma's inequality to show that

$$\mathbb{P}(|\nu(H) - \mathbb{E}(\nu(H))| \geq n^{3/5}) = O(e^{-\Omega(n^{1/5-o(1)})}) + O(n^{-3}) = O(n^{-3}). \tag{30}$$

(Specifically we can use Lemma 11 of Frieze and Pittel [11] or Section 3.2 of McDiarmid [15].) This verifies (17).

References

[1] M. Anastos and A.M. Frieze, A scaling limit for the length of the longest cycle in a sparse random graph, to appear in JCTB.

[2] J. Aronson, A.M. Frieze and B.G. Pittel, Maximum matchings in sparse random graphs: Karp-Sipser re-visited, *Random Structures and Algorithms* 12 (1998) 111-178.

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

[4] V. Chvátal, Almost all graphs with 1.44n edges are 3-colourable, *Random Structures and Algorithms* 2 (1991) 11-28.

[5] C. Cooper and A.M. Frieze, Hamilton cycles in a class of random directed graphs, *Journal of Combinatorial Theory* B 62 (1994) 151-163.

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

[7] C. Cooper and A.M. Frieze, The size of the largest strongly connected component of a random digraph with a given degree sequence, *Combinatorics, Probability and Computing* 13 (2004) 319-338.

[8] P.Erdős and A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* 5 (1960) 17-61.

[9] A.M. Frieze, On large matchings and cycles in sparse random graphs, *Discrete Mathematics* 59 (1986) 243-256.

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

[11] A.M. Frieze and B. Pittel, Perfect matchings in random graphs with prescribed minimal degree, *Trends in Mathematics*, Birkhauser Verlag, Basel (2004) 95-132.

[12] R. Karp, The transitive closure of a random digraph, *Random Structures and Algorithms* 1 (1990) 73-93.

[13] M. Krivelevich, E. Lubetzky and B. Sudakov, Longest cycles in sparse random digraphs, *Random Structures and Algorithms* 43 (2013) 1-15.

[14] C. McDiarmid, Clutter percolation and random graphs, *Mathematical Programming* 13 (1980) 17-25.

[15] C. McDiarmid, Concentration, *Probabilistic Methods for Algorithmic Discrete Mathematics* (M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, B. Reed eds.), Springer, Berlin (1998) 1–46.

A Proof that D^* is Hamiltonian w.h.p.

The proof can be broken into three parts: suppose that $|V_1^*| = N = N_1 + N_2$ where

$$N_1 = |V_1| \geq N(1 - e^{-c/2}).$$

(a) Find a collection Π_1 of $O(\log N)$ vertex disjoint directed cycles that cover V_1^*.

(b) Transform Π_1 into a collection Π_2 of vertex disjoint cycles such that each cycle is of length at least $N_0 = \left\lceil \frac{200N}{\log N} \right\rceil$.

(c) Break up Π_2 and re-assemble it as a Hamilton cycle.

A.1 Constructing Π_1

Each vertex of D^* is associated with five blue and five red edges. We randomly select three of each color and make them light and the rest heavy. We let D_3 be the digraph spanned by the light edges. We now consider the bipartite graph H with bipartition made up of two copies A, B of V_1^* and an edge $\{v, w\}$ iff (v, w) is a light edge. We show that w.h.p. H contains a perfect matching. In the context of D^* this gives us the collection of vertex disjoint directed cycles that cover V_1^*. We refer to this as a permutation digraph. We will argue that w.h.p. the number of cycles in the collection is $O(\log N)$. The probability that H has no perfect matching can be bounded by

$$2\sum_{k=4}^{N/2}\sum_{k_1=0}^{k}\sum_{k_2=0}^{k}\binom{N_1}{k_1}\binom{N_1}{k_2}\binom{N_2}{k-k_1}\binom{N_2}{k-k_2}\left(\frac{k_2}{N_1}\right)^{3k}\left(1-\frac{k_1}{N_1}\right)^{3(N-k)} \tag{31}$$
$$\leq 2\sum_{k=4}^{N/2e^2}\sum_{k_1=0}^{k}\sum_{k_2=0}^{k}\binom{N}{k}\binom{N}{k}\left(\frac{k}{N_1}\right)^{3k}+2\sum_{N/2e^2}^{N/2}\sum_{k_1=0}^{k}\sum_{k_2=0}^{k}\binom{N}{k}\binom{N}{k}\left(\frac{k}{N_1}\right)^{3k}e^{-1.5k\times k_1/k}$$
$$\leq 2\sum_{k=4}^{N/2e^2}k^2\left(\frac{eN}{k}\right)^{2k}\left(\frac{k}{N_1}\right)^{3k}+2\sum_{N/2e^2}^{N/2}k^2\left(\frac{eN}{k}\right)^{2k}\left(\frac{k}{N_1}\right)^{3k}e^{-1.5k\times 0.9}$$
$$\leq 2\sum_{k=4}^{N/2e^2}k^2\left(\frac{k}{(1-e^{-c/2})N}\right)^{k}+2\sum_{k=N/2e^2}^{N/2}k^2\left(\frac{e^{0.65}k}{(1-e^{-c/2})N}\right)^{k}=o(1).$$

Explanation for (31)**:** we employ Hall's theorem. We choose a set $S\subseteq A$ of size $k\leq N/2$ and a set $T\subseteq B$ also of size k. (No need to make $|T|=k-1$ here.) We let $k_1=|S\cap V_1|$ and $k_2=|T\cap V_1|$. The number of ways of choosing these sets is given by the product of binomial coefficients. We then estimate the probability that $T\supseteq N(S)$. Each vertex in $S\cap A$ has probability at most $\left(\frac{k_2}{N_1}\right)^3$ of choosing all of its neighbors in $V_1\cap T$, explaining the factor $\left(\frac{k_2}{N_1}\right)^{3k}$. Each vertex in $B\setminus T$ has probability $\left(1-\frac{k_1}{N_1}\right)^3$ of not choosing any neighbors in $V_1\cap S$, explaining the term $\left(1-\frac{k_1}{N_1}\right)^{3(N-k)}$. In the third line of the above calculations we used the fact that if $k\geq N/2e^2$ then $k_1\geq k-e^{-c/2}n\geq k-e^{-c/2}N/(1-e^{-c/2})\geq 0.9k$.

This deals with $k\leq N/2$ and if $k>N/2$ then $B\setminus T$ and $A\setminus S$ can take the place of S,T respectively..

We now consider the number of cycles in cycle cover induced by a matching in H. Suppose we write $M=\{(m(i),i):i\in B\}$ for some permutation m of A. Further let $A=A_1\cup A_X$ where $A_1=\{a_1,a_2,\ldots,a_{N_1}\}$ corresponds to V_1 and A_X corresponds to $\vec{X}^*$. We assume an analogous decomposition for B. Given a permutation m we let $B_X(m)=\{b\in B:m(b)\in A_X\}\subseteq B_1$. The set inclusion follows from the fact that vertices in A_X only have neighbors in B_1. Suppose now that we assume after re-labelling that that A,B are disjoint copies of $[N_1]$ and that $B_X(m),A_X$ are disjoint copies of $[N_2]$. Thus m induces a permutation of $[N_2]$ and a permutation of $[N_2+1,N]$. We claim that conditional on this that m induces uniform random permutations on these two sets. Suppose now that m_1,m_2 are two permutations that satisfy $m_i([N_2])=[N_2]$ for $i=1,2$. For a permutation π of A that satisfies $\pi([N_2]))=[N_2]$ and graph H we let $\pi(H)$ be obtained from H by replacing edge $\{i,j\}$ by $\{\pi(i),j\}$. We note that H and $\pi(H)$ have the same distribution. But then where $\pi(a)=m_2(m_1^{-1}(a))$ for $a\in A$ we have

$$\mathbb{P}(m(H)=m_1)=\mathbb{P}(m(\pi(H))=m_2)=\mathbb{P}(m(H)=m_2), \tag{32}$$

justifying our uniformity claim.

Now a uniform random permutation on a set of size M has $O(\log M)$ cycles w.h.p. It follows that w.h.p. the number of cycles induced by the matching constructed in H has $O(\log N)$ cycles as claimed previously.

A.2 Constructing Π_2

We now show how to boost the minimum cycle size to at least N_0. We partition the cycles of the permutation digraph Π_1 into sets SMALL and LARGE, containing cycles C of length $|C| < N_0$ and $|C| \geq N_0$ respectively. We define a Near Permutation Digraph (NPD) to be a digraph obtained from a permutation digraph by removing one edge. Thus an NPD Γ consists of a path $P(\Gamma)$ plus a permutation digraph $PD(\Gamma)$ which covers $[N] \setminus V(P(\Gamma))$.

We now give an informal description of a process which removes a small cycle C from a *current* permutation digraph Π. We start by choosing an (arbitrary) edge (v_0, u_0) of C and delete it to obtain an NPD Γ_0 with $P_0 = P(\Gamma_0) \in \mathcal{P}(u_0, v_0)$, where $\mathcal{P}(x, y)$ denotes the set of paths from x to y in D. The aim of the process is to produce a *large* set $\mathcal{S}$ of NPD's such that for each $\Gamma \in \mathcal{S}$, (i) $P(\Gamma)$ has a least N_0 edges and (ii) the small cycles of $PD(\Gamma)$ are a subset of the small cycles of Π. We will show that **whp** the endpoints of one of the $P(\Gamma)$'s can be joined by an edge to create a permutation digraph with (at least) one less small cycle.

We have so far used six of the edges available at each vertex of D^*, namely those in D_3. We now let D_4 denote the 1-in, 1-out digraph associated with an unused fourth in- and out-edge associated with each vertex of D^*. Each vertex $v \in V^*$ will be associated with a random in-neighbor $in_4(v)$ and a random out-neighbor $out_4(v)$.

The basic step in an *Out-Phase* of this process is to take an NPD Γ with $P(\Gamma) \in \mathcal{P}(u_0, v)$ and to examine the edges of D_4 leaving v i.e. edges going *out* from the end of the path. Let w be the terminal vertex of such an edge and assume that Γ contains an edge (x, w). Then $\Gamma' = \Gamma \cup \{(v, w)\} \setminus \{(x, w)\}$ is also an NPD. Γ' is acceptable if (i) $P(\Gamma')$ contains at least N_0 edges and (ii) any new cycle created (i.e. in Γ' and not Γ) also has at least N_0 edges.

If Γ contains no edge (x, w) then $w = u_0$. We accept the edge if P has at least N_0 edges. This would (prematurely) end an iteration, by closing a cycle, although it is unlikely to occur.

We do not want to look at very many edges of D_4 in this construction and we build a tree T_0 of NPD's in a natural breadth-first fashion where each non-leaf vertex $\Gamma \in T_0$ gives rise to NPD children Γ' as described above. The construction of T_0 ends when we first have $\nu = \left\lceil \sqrt{N \log N} \right\rceil$ leaves. The construction of T_0 constitutes an Out-Phase of our procedure to eliminate small cycles. Having constructed T_0 we need to do a further *In-Phase*, which is similar to a set of Out-Phases.

Then w.h.p. we close at least one of the paths $P(\Gamma)$ to a cycle of length at least N_0. If $|C| \geq 4$ and this process fails then we try again with a different independent edge of C in place of (u_0, v_0).

We now increase the the formality of our description. We start Phase 2 with a permutation digraph Π_0 and a general iteration of Phase 2 starts with a permutation digraph Π whose small cycles are a subset of those in Π_0. Iterations continue until there are no more small cycles. At the start of an iteration we choose some small cycle C of Π. There then follows an Out-Phase in which we construct a tree $T_0 = T_0(\Pi, C)$ of NPD's as follows: the root of T_0 is Γ_0 which is obtained by deleting an edge (v_0, u_0) of C.

We grow T_0 to a depth at most $\lceil 1.5 \log n \rceil$. The set of nodes at depth t is denoted by $\mathcal{S}_t$. Let $\Gamma \in \mathcal{S}_t$ and $P = P(\Gamma) \in \mathcal{P}(u_0, v)$. A *potential* child Γ' of Γ, at depth $t + 1$ is defined as follows.

Let w be the terminal vertex of an edge directed from v in D_4.
Case 1. w is a vertex of a cycle $C' \in PD(\Gamma)$ with edge $(x, w) \in C'$. Let $\Gamma' = \Gamma \cup \{(v, w)\} \setminus \{(x, w)\}$.
Case 2. w is a vertex of $P(\Gamma)$. Either $w = u_0$, or (x, w) is an edge of P. In the former case $\Gamma \cup \{(v, w)\}$ is a permutation digraph Π' and in the latter case we let $\Gamma' = \Gamma \cup \{(v, w)\} \setminus \{(x, w)\}$.

In fact we only admit to $\mathcal{S}_{t+1}$ those Γ' which satisfy the following conditions. We define a set W of *used* vertices. Initially all vertices are *unused* i.e. $W = \emptyset$. Whenever we examine an edge (v, w), we add both v and w to W. So if $v \notin W$ then $out_4(v)$ is still unconditioned and $in_4(v)$ is a random member of a set $U \supseteq V^* \setminus W$. We do not allow $|W|$ to exceed $N^{3/4}$.

C(i) The new cycle formed (Case 2 only) must have at least N_0 vertices, and the path formed (both cases) must either be empty or have at least N_0 vertices. When the path formed is empty we close the iteration and if necessary start the next with Π'.

C(ii) $x, w \notin W$.

An edge (v, w) which satisfies the above conditions is described as *acceptable*.

We let S_t be the set of endpoints of paths in $\mathcal{S}_t$ that are not u_0. If some $NPD \in \mathcal{S}_t$ is the union of cycles then we are done with the given iteration. Thus we may assume otherwise and therefore $|\mathcal{S}_t| = |S_t|$.

We also let $S_t^1 = S_t \cap V_1$ and $S_2^t = S_t \setminus S_1^t$.

Lemma A.1. *Let $C \in SMALL$. Then, where $\nu = \lceil \sqrt{N \log N} \rceil$,*

$$\mathbb{P}(\exists t < \lceil \log_{1.9} \nu + 1000 \log \log N \rceil \text{ such that } |S_t| \in [\nu, 3\nu]) = 1 - O((\log \log N)^3 / \log N).$$

Proof. We assume we stop an iteration, in mid-phase if necessary, when $|S_t| \in [\nu, 3\nu]$. Let us consider a generic construction in the growth of T_0. Thus suppose we are extending from Γ and $P(\Gamma) \in \mathcal{P}(u_0, v)$.

We consider S_{t+1} to be constructed in the following manner: we first examine $out_4(v), v \in S_t$ in the order that these vertices were placed in S_t to see if they produce acceptable edges.

We then add in those vertices $x \notin W$ which arise from (x, w) with $v = in_4(w) \in S_t, w \notin W$, (to avoid conditioning problems).

Let $Z(v)$ be the indicator random variable for $(v, out_4(v))$ being unacceptable and let $Z_t = \sum_{v \in S_t} Z(v)$. If $Z(v) = 1$ then either (i) $out_4(v)$ lies on $P(\Gamma)$ and is too close to an endpoint; this has probability bounded above by $2N_0/|V_1| \leq 401/\log N$, or (ii) the corresponding vertex x is in W; this has probability bounded above by $N^{3/4}/|V_1| \leq 2N^{-1/4}$, or (iii) $out_4(v)$ lies on a small cycle. Now in a random permutation the expected number of vertices on cycles of length at most N_0 is precisely N_0 ([12]). Thus, by the Markov inequality, w.h.p. Γ_0 contains at most $N_1 \log\log N_1/(2\log N_1) + N_2 \log\log N_2/(2\log N_2)$ vertices on small cycles. Condition on this event. Then $\mathbb{P}(Z(v) = 1) \leq 2\log\log N/\log N$ regardless of the history of the process and so Z_t is stochastically dominated by $B(|S_t|, 2\log\log N/\log N)$.

Next let $X(v)$ denote the number of vertices w in $V^* \setminus W$ such that $in_4(w) = v$, $x \notin W$ where (v, w) is acceptable and $(x, w) \in \Gamma$ (if there is no such x then the iteration can end early.) Let $X_t = \sum_{v \in S_t} X(v)$. Now assuming $|W| \leq N^{3/4}$ we see that there are $N' = N_1 - O(N \log\log N/\log N)$ vertices w which would produce an acceptable edge provided $v = in_4(w) \in S_t^1$. For these vertices $in_4(w)$ is a random choice from a set which contains S_t^1 and so X_t stochastically dominates $B(N', |S_t^1|/N)$.

Summing $1 - Z(v) + X(v)$ over $v \in S_t$ might seem to overestimate $|S_{t+1}|$. In principle we should subtract off the number Y_t of vertices of S_{t+1} that are counted more than once in this sum. But these arise in two ways. First there are the pairs $v_1, v_2 \in S_t$ with $out_4(v_1) = out_4(v_2)$. Suppose we examine v_1 before v_2. Then when we examine v_2 we find that $out_4(v_2) \in W$ and so we do not get a contribution to S_{t+1}. Secondly there is the possibility of their being $v_1, v_2 \in S_t$ and w such that $w = out_4(v_1)$ and $v_2 = in_4(w)$. But in this case w will only be counted once as $w \in W$ when it is time for $in_4(w)$ to be examined. We can then write

$$|S_{t+1}| = |S_t| - Z_t + X_t.$$

Now let $t_0 = \lceil 1000 \log\log N \rceil$, $t_1 = 10t_0$, $t_2 = \lceil \log_{1.9} \nu + 1000 \log\log N \rceil$, $s_0 = \lceil 1000 \log\log N \rceil$ and $s_1 = \lceil 1000 \log N \rceil$.

(a) $\mathbb{P}(\exists t \leq t_0 : |S_t| \leq s_0 \text{ and } Z_t > 0) = O((\log\log N)^3/\log N)$

(b) $\mathbb{P}(|\cup_{t \leq t_0} S_t^1| < 0.99|\cup_{t \leq t_0} S_t| \mid |S_t| \leq s_0 \text{ for } t \leq t_0) = O((\log\log N)^3/\log N)$.

(c) $\mathbb{P}(\sum_{t=1}^{t_0} X_t \leq s_0 \mid S_t \neq \emptyset \text{ and } |S_t| \leq s_0 \text{ for } t \leq t_0) = O((\log\log N)^3/\log N)$.

(d) $\mathbb{P}(\exists t \leq t_1 : |S_{t+1}^1| < 0.99|S_{t+1}| \mid S_t \geq 500 \log\log n) = O(1/\log N)$.

(e) $\mathbb{P}(\exists t \leq t_1 : 500 \log\log N \leq |S_t| \leq s_1 \text{ and } Z_t > X_t/100) = O(1/\log N)$.

(f) $\mathbb{P}(\exists t \leq t_1 : X_t < |S_t|/2 \mid |S_t| \geq 500 \log\log N) = O(1/\log N)$.

(g) $\mathbb{P}(\exists t \leq t_1 : |S_t| \leq s_1 \text{ and } X_t \geq 2s_1) = O(N^{-2})$.

(h) $\mathbb{P}(\exists t_1 \leq t \leq t_2 : |S_{t+1}^1| < 0.99|S_{t+1}| \mid S_t \geq s_1) = O(N^{-2})$.

(i) $\mathbb{P}(\exists t \leq t_2 : |S_t| \geq s_1 \text{ and } |X_t - Z_t - |S_t|| \geq |S_t|/10) = O(N^{-2})$.

Explanations:- we use the following standard inequalities for the tails of the binomial distribution:

$$\mathbb{P}(|B(n,p) - np| \geq \epsilon np) \leq 2e^{-\epsilon^2 np/3}, \qquad 0 \leq \epsilon \leq 1, \tag{33}$$

$$\mathbb{P}(B(n,p) \geq anp) \leq (e/a)^{anp}. \tag{34}$$

We let $\mathcal{E}_x, x \in \{a, b, \ldots, i\}$ be the low probability events described in (a)-(i) above.

(a) $\mathbb{P}(Z_t > 0 \mid \ |S_t| \leq 500 \log\log N) = O((\log\log N)^2/\log N)$ by the Markov inequality.

(b) Conditioned on $\mathcal{E}_a$ we have that $|\cup_{t \leq t_0} S_t| \geq t_0$ and $Z_t = 0$ for $t \leq t_0$. Let $v_1, v_2, ...$ be the order in which the vertices in $\cup_{t \leq t_0} S_t$ are examined. At step i with $w = out_4(v_i)$ we updated $\Gamma' = \Gamma \cup \{(v_i, w)\} \setminus \{(x, w)\}$ and added x to $\cup_{t \leq t_0} S_t$. x belongs to V_1 with probability $(1 + o(1))|N_1|/N > 0.999$. The rest follows from (33).

(c) Conditioned on $\mathcal{E}_a \cap \mathcal{E}_b$ we have that $|\cup_{t \leq t_0} S_t^1| \geq 0.99 t_0$. Thus $\sum_{t=1}^{t_0} X_t$ dominates $B(0.99 t_0 N', 1/N)$.

(d) Similar to (b).

(e) Condition on $|S_t| = s \geq 500 \log\log N$ and $\mathcal{E}_d$. Then $Z_t > X_t/100$ implies either that (i) $X_t \leq s/10 \leq 0.99|S_t^1|/10$ or (ii) $Z_t > 10s$. Both of these events have probability $O(1/(\log N)^3)$.

(f) Immediate from (33).

(g) Immediate from (33) and (34).

(h) Similar to (b).

(i) Similar to (c).

Assume the occurrence of $\bigcap_x \bar{\mathcal{E}}_x$. Then $\bar{\mathcal{E}}_a \cap \bar{\mathcal{E}}_c$ implies that $|S_t|$ reaches size at least $500 \log\log N$ before t reaches $t_0 + 1$. Once this happens, $\bar{\mathcal{E}}_e \cap \bar{\mathcal{E}}_f$ implies that $|S_t|$ then grows geometrically with t up to time t_1 at a rate of at least 1.49. Together with $\bar{\mathcal{E}}_g$ this proves that at some stage between 1 and t_1, $|S_t|$ reaches a size in the range $[s_0, 3s_0]$. $\bar{\mathcal{E}}_f$ then implies that $|S_t|$ increases at a rate $\lambda \in [1.9, 2.1]$ from then on. The lemma follows. □

The total number of vertices added to W in this way throughout the whole of Phase 2 is $O(\nu|SMALL|) = o(N^{3/4})$. (As we see later, we try this process once for $C \in SMALL, |C| \leq 3$ and once or twice for $C \in SMALL, |C| \geq 4$.)

Let t^* denote the value of t when we stop the growth of T_0. At this stage we have leaves Γ_i, for $i = 1, \ldots, \nu$, each with a path of length at least N_0, (unless we have already successfully made a cycle). We now execute an In-Phase. This involves the construction of trees $T_i, i = 1, 2, \dots \nu$. Assume that $P(\Gamma_i) \in \mathcal{P}(u_0, v_i)$. We start with Γ_i and build T_i in a similar way to T_0 except that here all paths generated end with v_i. This is done as follows: if a current NPD Γ has $P(\Gamma) \in \mathcal{P}(u, v_i)$ then we consider adding an edge $(w, u) \in D_4$ and deleting an edge $(w, x) \in \Gamma$. Thus our trees are grown by considering edges directed into the start vertex of each $P(\Gamma)$ rather than directed out of the end vertex. Some technical changes are necessary however.

We consider the construction of our ν trees in two stages. First of all we grow the trees only enforcing condition C(ii) of success and thus allow the formation of small cycles and paths. We try to grow them to depth t_2. The growth of the ν trees can naturally be considered to occur simultaneously. Let $L_{i,\ell}$ denote the set of start vertices of the paths associated with the nodes at depth ℓ of the i'th tree, $i = 1, 2 \dots, \nu, \ell = 0, 1, \dots, t_2$. Thus $L_{i,0} = \{u_0\}$ for all i. We prove inductively that $L_{i,\ell} = L_{1,\ell}$ for all i, ℓ. In fact if $L_{i,\ell} = L_{1,\ell}$ then the acceptable D_4 edges have the same set of initial vertices and since all of the deleted edges are D_3-edges (enforced by C(ii)) we have $L_{i,\ell+1} = L_{1,\ell+1}$.

The probability that we succeed in constructing trees $T_1, T_2, \dots T_\nu$ is, by the analysis of Lemma 3, $1 - O((\log\log N)^3/\log N)$. Note that the number of nodes in each tree is $O(2.1^{t_2+1}) = O(N^{.74\dots})$.

We now consider the fact that in some of the trees some of the leaves may have been constructed in violation of C(i). We imagine that we prune the trees $T_1, T_2, \dots T_\nu$ by disallowing any node that was constructed in violation of C(i). Let a tree be BAD if after pruning it has less than ν leaves and GOOD otherwise. Now an individual pruned tree has been constructed in the same manner as the tree T_0 obtained in the Out-Phase. (We have chosen t_2 to obtain ν leaves even at the slowest growth rate of 1.9 per node.) Thus

$$\mathbb{P}(T_1 \text{ is BAD}) = O\left(\frac{(\log\log N)^3}{\log N}\right)$$

and

$$\mathbb{E}(\text{number of BAD trees}) = O\left(\frac{\nu(\log\log N)^3}{\log N}\right)$$

and

$$\mathbb{P}(\exists \geq \nu/2 \text{ BAD trees}) = O\left(\frac{(\log\log N)^3}{\log N}\right).$$

Thus

$$\begin{aligned}
&\mathbb{P}(\exists < \nu/2 \text{ GOOD trees after pruning}) \\
&\leq \mathbb{P}(\text{failure to construct } T_1, T_2, \dots T_\nu) + \mathbb{P}(\exists \geq \nu/2 \text{ BAD trees}) \\
&= O\left(\frac{(\log\log N)^3}{\log N}\right).
\end{aligned}$$

Thus with probability 1-$O((\log\log N)^3/\log N)$ we end up with $\nu/2$ sets of ν paths, each of length at least $100n/\log N$ where the i'th set of paths all terminate in v_i. From these paths keep only those whose other endpoint u lies in V_1. Then, similarly to the proof of property (h) in Lemma A.1, w.h.p. from each set we keep at least 0.99ν paths. The $in_4(v_i)$ are still unconditioned and hence

$$\mathbb{P}(\text{no } D_4 \text{ edge closes one of these paths}) \leq \left(1 - \frac{0.99\nu}{n}\right)^{\nu/2} = O(N^{-1/2}).$$

Consequently the probability that we fail to eliminate a particular small cycle C after breaking an edge is $O((\log\log N)^3/\log N)$. If $|C| \geq 4$ then we try once or twice using independent edges of C and so the probability we fail to eliminate a given small cycle C is certainly $O(((\log\log N)^3/\log N)^2)$ for $|C| \geq 4$ (remember that we calculated all probabilities conditional on previous outcomes and assuming $|W| \leq N^{3/4}$.)

Now the number of cycles of length 1,2 or 3 in D_3 is asymptotically Poisson with mean O(1) and so there are fewer than $\log\log N$ w.h.p. Hence, since **whp** $|C| = O(\log N)$,

Lemma A.2. *The probability that Phase 2 fails to produce a permutation digraph with minimal cycle length at least N_0 is $o(1)$.*

At this stage we have shown that D^* almost always contains a permutation digraph Π_2 in which the minimum cycle length is at least N_0. We shall refer to Π_2 as the *Phase 2* permutation digraph.

A.3 Re-assembly

Let D_5 be the 1-in,1-out digraph left unused by the construction in the previous two sections. We will use the edges of D_5 to break-up and re-assemble the cycles of Π_2 into a Hamilton cycle. Let $C_1, C_2, \dots, C_k$ be the cycles of Π_2, and let $c_i = |C_i \cap V_1|$, $c_1 \leq c_2 \leq \cdots \leq c_k$. Note that $\vec{X}^*$ is an independent set of D^* and so at least half the vertices of each C_i are in V_1. If $k = 1$ we can skip this phase, otherwise let $a = \frac{N}{\log N}$. For each C_i we consider selecting a set of $m_i = 2\lfloor \frac{c_i}{a} \rfloor + 1$ vertices $v \in C_i \cap V_1$, and deleting the edge (v, u) in Π^*. Let $m = \sum_{i=1}^k m_i$ and re-label (temporarily) the broken edges as $(v_i, u_i), i \in [m]$ as follows: in cycle C_i identify the lowest numbered vertex x_i which loses a cycle edge directed out of it. Put $v_1 = x_1$ and then go round C_1 defining $v_2, v_3, \dots v_{m_1}$ in order. Then let $v_{m_1+1} = x_2$ and so on. We thus have m path sections $P_j \in \mathcal{P}(u_{\phi(j)}, v_j)$ in Π_2 for some permutation ϕ. We see that ϕ is an even permutation as all the cycles of ϕ are of odd length.

It is our intention to rejoin these path sections of Π_2 to make a Hamilton cycle using D_b, if we can. Suppose we can. This defines a permutation ρ where $\rho(i) = j$ if P_i is joined to P_j by $(v_i, u_{\phi(j)})$, where $\rho \in H_m$ the set of cyclic permutations on $[m]$. We will use the second moment method to show that a suitable ρ exists w.h.p. A technical problem forces a restriction on our choices for ρ. This will produce a variance reduction in a second moment calculation.

Given ρ define $\lambda = \phi\rho$. In our analysis we will restrict our attention to $\rho \in R_\phi = \{\rho \in H_m : \phi\rho \in H_m\}$. If $\rho \in R_\phi$ then we have not only constructed a Hamilton cycle in $\Pi_2 \cup D_5$, but also in the *auxillary digraph* Λ, whose edges are $(i, \lambda(i))$.

Lemma A.3. $(m-2)! \leq |R_\phi| \leq (m-1)!$

Proof. We grow a path $1, \lambda(1), \lambda^2(1), \ldots, \lambda^r(1) \ldots$ in Λ, maintaining feasibility in the way we join the path sections of Π_2 at the same time.

We note that the edge $(i, \lambda(i))$ of Λ corresponds in D_5 to the edge $(v_i, u_{\phi\rho(i)})$. In choosing $\lambda(1)$ we must avoid not only 1 but also $\phi(1)$ since $\lambda(1) = 1$ implies $\rho(1) = 1$. Thus there are $m - 2$ choices for $\lambda(1)$ since $\phi(1) \neq 1$ from the definition of m_1.

In general, having chosen $\lambda(1), \lambda^2(1), \ldots, \lambda^r(1), 1 \leq r \leq m-3$ our choice for $\lambda^{r+1}(1)$ is restricted to be different from these choices and also 1 and ℓ where u_ℓ is the initial vertex of the path terminating at $v_{\lambda^r(1)}$ made by joining path sections of Π_2. Thus there are either $m - (r+1)$ or $m - (r+2)$ choices for $\lambda^{r+1}(1)$ depending on whether or not $\ell = 1$.

Hence, when $r = m - 3$, there *may* be only one choice for $\lambda^{m-2}(1)$, the vertex h say. After adding this edge, let the remaining isolated vertex of Λ be w. We now need to show that we can complete λ, ρ so that $\lambda, \rho \in H_m$.

Which vertices are missing edges in Λ at this stage? Vertices $1, w$ are missing in-edges, and h, w out-edges. Hence the path sections of Π_2 are joined so that either

$$u_1 \to v_h, \quad u_w \to v_w \quad \text{or} \quad u_1 \to v_w, \quad u_w \to v_h.$$

The first case can be (uniquely) feasibly completed in both Λ and Π_2 by setting $\lambda(h) = w, \lambda(w) = 1$. Completing the second case to a cycle in Π_2 means that

$$\lambda = (1, \lambda(1), \ldots, \lambda^{m-2}(1))(w) \tag{35}$$

and thus $\lambda \notin H_m$. We show this case cannot arise.

$\lambda = \phi\rho$ and ϕ is even implies that λ and ρ have the same parity. On the other hand $\rho \in H_m$ has a different parity to λ in (35) which is a contradiction.

Thus there is a (unique) completion of the path in Λ. □

Let H stand for the union of the permutation digraph Π_2 and D_5. We finish our proof by proving

Lemma A.4. $\mathbb{P}$(*H does not contain a Hamilton cycle*) $= o(1)$.

Proof. Let X be the number of Hamilton cycles in H obtainable by deleting edges as above, rearranging the path sections generated by ϕ according to those $\rho \in R_\phi$ and if possible reconnecting all the sections using edges of D_5. We will use the inequality

$$\mathbb{P}(X > 0) \geq \frac{\mathbb{E}(X)^2}{\mathbb{E}(X^2)}. \tag{36}$$

Probabilities in (36) are thus with respect to the space of D_5 choices.

Now the definition of the m_i yields that

$$\frac{2N}{a} - k \le m \le \frac{2N}{a} + k$$

and so

$$(1.99)\log N \le m \le (2.01)\log N.$$

Also

$$k \le \frac{\log N}{200},\ m_i \ge 199 \text{ and } \frac{c_i}{m_i} \ge \frac{a}{2.01}, \qquad 1 \le i \le k.$$

Let Ω denote the set of possible cycle re-arrangements. $\omega \in \Omega$ is a *success* if D_5 contains the edges needed for the associated Hamilton cycle. Let b_i be the number of deleted edges (v_i, u_i) with $u_i \notin V_1$ and $b = \sum_{i=1}^k b_i$. Observe that if $u_i \in V_1$ then $(v_i, u_i) \in E(D_5) \setminus E(D_4)$ with probability $1 - \left(1 - \frac{1}{N_1}\right)^2$ while if $u_j \notin V_1$ then $(v_i, u_j) \in E(D_5) \setminus E(D_4)$ with probability $\frac{1}{N_1}$.

For a fixed $\alpha > 0$ we have

$$ne^{-c/2} \ge N - N_1 \ge b \ge \sum_{j: b_j \ge \alpha|C_j|} b_j \ge \alpha \sum_{j: b_j \ge \alpha|C_j|} |C_j|.$$

Putting $\alpha = 10^{-3}$ we see that at most $1000ne^{-c/2} \le e^{-c/3}N$ vertices lie on a cycle C_i with more than $0.001|C_i|$ vertices that do not lie in V_1. Therefore b is stochasticly dominated by $(1 + o(1))(e^{-c/3}m + Bin((1 - e^{-c/3})m, 10^{-3})$. Hence $\mathbb{P}(b > 0.01m) = o(1)$. Thus,

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{\omega\in\Omega} \mathbb{P}(\omega \text{ is a success}) \\
&= \sum_{\omega\in\Omega} \left(1 - \left(1 - \frac{1}{N_1}\right)^2\right)^{m-b(\omega)} \left(\frac{1}{N_1}\right)^{b(\omega)} \\
&\ge (1 - o(1)) \left(\frac{2}{N_1}\right)^m 2^{-0.01m} \cdot \mathbb{P}(b \le 0.01m)(m-2)! \prod_{i=1}^k \binom{c_i}{m_i} \\
&\ge \frac{1 - o(1)}{m\sqrt{m}} \left(\frac{2m}{eN_1}\right)^m \prod_{i=1}^k \left(\left(\frac{c_i e^{1-1/12m_i}}{m_i^{1+(1/2m_i)}}\right)^{m_i} \left(\frac{1 - 2m_i^2/c_i}{\sqrt{2\pi}}\right)\right) 2^{-0.01m} \\
&\ge \frac{(1 - o(1))(2\pi)^{-m/398} e^{-k/12}}{m\sqrt{m}} \left(\frac{2m}{eN_1}\right)^m \prod_{i=1}^k \left(\frac{c_i e}{(1.02)m_i}\right)^{m_i} 2^{-0.01m} \\
&\ge \frac{(1 - o(1))(2\pi)^{-m/398}}{n^{1/1200} m\sqrt{m}} \left(\frac{2m}{eN_1}\right)^m \left(\frac{ea}{2.01 \times 1.02}\right)^m 2^{-0.01m} \\
&\ge \frac{(1 - o(1))(2\pi)^{-m/398}}{N_1^{1/1200} m\sqrt{m}} \left(\frac{3.98}{2.0502}\right)^m 2^{-0.01m}
\end{aligned}$$

$$\geq \quad N_1^{1.3}. \tag{37}$$

Let A, A' be two sets of selected edges which have been deleted in Π_2 and whose path sections have been rearranged into Hamilton cycles according to ρ, ρ' respectively. Let B, B' be the corresponding sets of edges which have been added to make the Hamilton cycles. What is the interaction between these two Hamilton cycles?

Let $s = |A \cap A'|$ and $t = |B \cap B'|$. Now $t \leq s$ since if $(v, u) \in B \cap B'$ then there must be a unique $(\tilde{v}, u) \in A \cap A'$ which is the unique Π_2-edge into u. We claim that $t = s$ implies $t = s = m$ and $(A, \rho) = (A', \rho')$. (This is why we have restricted our attention to $\rho \in R_\phi$.) Suppose then that $t = s$ and $(v_i, u_i) \in A \cap A'$. Now the edge $(v_i, u_{\lambda(i)}) \in B$ and since $t = s$ this edge must also be in B'. But this implies that $(v_{\lambda(i)}, u_{\lambda(i)}) \in A'$ and hence in $A \cap A'$. Repeating the argument we see that $(v_{\lambda^k(i)}, u_{\lambda^k(i)}) \in A \cap A'$ for all $k \geq 0$. But λ is cyclic and so our claim follows.

We adopt the following notation. Let $< s, t >$ denote $|A \cap A'| = s$ and $|B \cap B'| = t$. So

$$\begin{aligned}
\mathbb{E}(X^2) &\leq \mathbb{E}(X) + (1+o(1)) \sum_{A \in \Omega} \left(\frac{2}{N_1}\right)^m \sum_{\substack{A' \in \Omega \\ B' \cap B = \emptyset}} \left(\frac{2}{N_1}\right)^m \\
&\quad + (1+o(1)) \sum_{A \in \Omega} \left(\frac{2}{N_1}\right)^m \sum_{s=2}^{m} \sum_{t=1}^{s-1} \sum_{\substack{A' \in \Omega \\ <s,t>}} \left(\frac{2}{N_1}\right)^{m-t} \\
&= \mathbb{E}(X) + E_1 + E_2 \ \text{ say.}
\end{aligned} \tag{38}$$

Clearly

$$E_1 \leq (1+o(1))\mathbb{E}(X)^2. \tag{39}$$

For given ρ, how many ρ' satisfy the condition $< s, t >$? Previously $|R_\phi| \geq (m-2)!$ and now given $< s, t >$, $|R_\phi(s,t)| \leq (m-t-1)!$, (consider fixing t edges of Λ').
Thus

$$E_2 \leq \mathbb{E}(X)^2 \sum_{s=2}^{m} \sum_{t=1}^{s-1} \binom{s}{t} \left[\sum_{\sigma_1 + \cdots + \sigma_k = s} \prod_{i=1}^{k} \frac{\binom{m_i}{\sigma_i}\binom{c_i - m_i}{m_i - \sigma_i}}{\binom{c_i}{m_i}} \right] \frac{(m-t-1)!}{(m-2)!} \left(\frac{N_1}{2}\right)^t.$$

For the above expression observe that given $A \cap A'$ there are $\binom{s}{t}$ choices for $B \cap B'$. Thereafter given A and σ_i there are $\binom{m_i}{\sigma_i}$ ways to choose $A \cap A' \cap C_i$ and $\binom{c_i - m_i}{m_i - \sigma_i}$ ways to choose the rest of $B'_i \cap C_i$.

Now

$$\begin{aligned}
\frac{\binom{c_i - m_i}{m_i - \sigma_i}}{\binom{c_i}{m_i}} &\leq \frac{\binom{c_i}{m_i - \sigma_i}}{\binom{c_i}{m_i}} \\
&\leq (1+o(1)) \left(\frac{m_i}{c_i}\right)^{\sigma_i} \exp\left\{-\frac{\sigma_i(\sigma_i - 1)}{2m_i}\right\} \\
&\leq (1+o(1)) \left(\frac{2.01}{a}\right)^{\sigma_i} \exp\left\{-\frac{\sigma_i(\sigma_i - 1)}{2m_i}\right\}
\end{aligned}$$

where the $o(1)$ term is $O((\log N)^3/N)$. Also

$$\sum_{i=1}^{k} \frac{\sigma_i^2}{2m_i} \geq \frac{s^2}{2m} \qquad \text{for } \sigma_1 + \cdots \sigma_k = s,$$

$$\sum_{i=1}^{k} \frac{\sigma_i}{2m_i} \leq \frac{k}{2},$$

and

$$\sum_{\sigma_1+\cdots+\sigma_k=s} \prod_{i=1}^{k} \binom{m_i}{\sigma_i} = \binom{m}{s}.$$

Hence

$$\begin{aligned}
\frac{E_2}{\mathbb{E}(X)^2} &\leq (1+o(1))e^{k/2} \sum_{s=2}^{m} \sum_{t=1}^{s-1} \binom{s}{t} \exp\left\{-\frac{s^2}{2m}\right\} \left(\frac{2.01}{a}\right)^s \binom{m}{s} \frac{(m-t-1)!}{(m-2)!} \left(\frac{N_1}{2}\right)^t \\
&\leq (1+o(1))N^{.005} \sum_{s=2}^{m} \sum_{t=1}^{s-1} \binom{s}{t} \exp\left\{-\frac{s^2}{2m}\right\} \left(\frac{2.01}{a}\right)^s \frac{m^{s-(t-1)}}{(s-1)!} \left(\frac{N_1}{2}\right)^t \\
&= (1+o(1))N^{.005} \sum_{s=2}^{m} \left(\frac{2.01}{a}\right)^s \frac{m^s}{s!} \exp\left\{-\frac{s^2}{2m}\right\} m \sum_{t=1}^{s-1} \binom{s}{t} \left(\frac{N_1}{2m}\right)^t \\
&\leq (1+o(1)) \left(\frac{2m^3}{N^{.99}}\right) \sum_{s=2}^{m} \left(\frac{(2.01)N_1 \exp\{-s/2m\}}{2a}\right)^s \frac{1}{s!} \\
&= o(1) \qquad (40)
\end{aligned}$$

To verify that the RHS of (40) is $o(1)$ we can split the summation into

$$S_1 = \sum_{s=2}^{\lfloor m/4 \rfloor} \left(\frac{(2.01)N_1 \exp\{-s/2m\}}{2a}\right)^s \frac{1}{s!}$$

and

$$S_2 = \sum_{s=\lfloor m/4 \rfloor+1}^{m} \left(\frac{(2.01)N_1 \exp\{-s/2m\}}{2a}\right)^s \frac{1}{s!}.$$

Ignoring the term $\exp\{-s/2m\}$ we see that

$$\begin{aligned}
S_1 &\leq \sum_{s=2}^{\lfloor (.5025)\log N \rfloor} \frac{((1.005)\log N)^s}{s!} \\
&= o(N^{9/10})
\end{aligned}$$

since this latter sum is dominated by its last term.

Finally, using $\exp\{-s/2m\} < e^{-1/8}$ for $s > m/4$ we see that

$$S_2 \leq N^{(1+o(1))1.005)e^{-1/8}} < N^{9/10}.$$

The result follows from (36) to (40). □