HAMILTON CYCLES IN RANDOM GRAPHS AND DIRECTED GRAPHS

Colin Cooper
School of Mathematical Sciences,
University of North London,
London N7 8DB, U.K.

Alan Frieze*
Department of Mathematical Sciences,
Carnegie Mellon University,
Pittsburgh PA15213, U.S.A.

April 26, 2000

Abstract

We prove that almost every digraph $D_{2-in,2-out}$ is Hamiltonian. As a corollary we obtain also that almost every graph G_{4-out} is Hamiltonian.

1 Introduction

The random digraph $D_{k-in,\ell-out}$ is defined as follows: It has vertex set $V = [n]$ where $[n] = \{1, 2, \ldots, n\}$ and each $v \in [n]$ chooses a set $in(v)$ of k random edges directed into v and a set $out(v)$ of ℓ random edges directed out of v. We call such a digraph a k-in, ℓ-out digraph. For our purposes it is not important if v chooses edges with or without replacement and we shall assume that they are chosen without replacement. Thus $D_{k-in,\ell-out}$ has $(k+\ell)n$ edges. The probability space for $D_{k-in,\ell-out}$ will be denoted by $\mathcal{D}_{k-in,\ell-out}$. This model was

*Supported by an NSF grant

introduced by Fenner and Frieze [7] who discussed the strong connectivity of $D_{k-in,k-out}$ for $k \geq 2$. The remaining case, where $k = 1$ was discussed by Cooper and Frieze [3], and by McDiarmid and Reed [17]. In an earlier paper Cooper and Frieze [4] proved that $D_{3-in,3-out}$ is Hamiltonian **whp**, *with high probability* i.e., probability $1 - o(1)$ as $n \to \infty$.

The main result of this paper is an improvement of this to

Theorem 1

$$D_{2-in,2-out} \textit{ is Hamiltonian } \textbf{whp}.$$

This result is best possible since

$$\mathbf{Pr}(D_{1-in,2-out} \text{ is Hamiltonian}) = o(1).$$

This follows from the fact that **whp**$D_{1-in,2-out}$ contains 2 vertices of indegree 1 sharing a common in-neighbour.

If $k = 0$ then we write $D_{\ell-out}$. If we drop the orientation in D_{k-out} then we obtain the underlying *undirected* graph G_{k-out}. This has been the object of considerable study, and the main outstanding question, is how large should k be for G_{k-out} to have a Hamilton cycle **whp**.

It was previously known that $k \geq 5$ is sufficient, (Frieze and Łuczak [8]), and it is conjectured that the correct lower bound for k is 3. By ignoring orientation in Theorem 1 we obtain an improvement on [8]:

Corollary 2

$$G_{4-out} \textit{ is Hamiltonian } \textbf{whp}.$$

From now on we use D_n to denote $D_{2-in,2-out}$.

A *permutation digraph* is a set of vertex disjoint directed cycles that cover all n vertices. Its *size* is the number of cycles.

To prove the theorem, we will use a *three phase* method as outlined below, where we prove that each phase succeeds **whp**.

Phase 1. We show that D_n contains a permutation digraph Π_1 of size at most $2 \log n$.

Phase 2. We increase the minimum cycle length in the permutation digraph to at least $n_0 = \left\lceil \frac{1000n}{\log n} \right\rceil$.

Phase 3. We convert the **Phase 2** permutation digraph to a Hamilton cycle.

1.1 Chernoff bounds and some notation

Let $B(n,p)$ denote the Binomial random variable with parameters n,p. We use the following well known inequalities for the tails of the binomial distribution:

$$\mathbf{Pr}(|B(n,p)-np| \geq \epsilon np) \leq 2e^{-\epsilon^2 np/3}, \qquad 0 \leq \epsilon \leq 1, \tag{1}$$
$$\mathbf{Pr}(B(n,p) \geq anp) \leq (e/a)^{anp}. \tag{2}$$

Throughout the paper inequalities are only claimed to hold for n sufficiently large.

In addition to the notation **whp** referring to a sequence of events $\mathcal{E}_n$ we will use the following:

- **wlp**(x) or *with logarithmic probability* stands for $\mathbf{Pr}(\bar{\mathcal{E}}_n) = O((\log n)^{-x})$.
- **wpp**(x) or *with polynomial probability* stands for $\mathbf{Pr}(\bar{\mathcal{E}}_n) = O(n^{-x})$.
- **qs** or *quite surely* stands for $\mathbf{Pr}(\bar{\mathcal{E}}_n) = O(n^{-x})$ for any fixed x.

2 Phase 1. Making a permutation digraph with at most $2\log n$ cycles

With any digraph D on n vertices there is an associated bipartite graph $BIP(D)$ with $n+n$ vertices, which contains an edge (u,v) iff D contains the directed edge (u,v). It is well known that perfect matchings in $BIP(D)$ are in 1-1 correspondence with permutation digraphs of D. Let $BIP(D_n)$ be denoted by BIP.

Lemma 3 **Whp** *D_n contains a permutation digraph Π_1 with at most $2\log n$ cycles.*

Proof Walkup [18] has shown that **whp** BIP contains a perfect matching $\{(i,\phi(i)), i = 1,2,\ldots,n\}$. We can argue by symmetry (as in [8]) that we can take ϕ to be a random permutation. It is well known (e.g. Feller [6]), that **whp** a random permutation contains at most $2\log n$ cycles, and thus the permutation digraph has size at most $2\log n$. □

As we use some edges of D_n in Phases 2 and 3, we will need to understand the distribution of the edges not contained in Π_1. To do this we will consider a constructive version of Walkup's result. Karp, Rinnooy-Kan and Vohra [13] have described an algorithm for finding a perfect matching in BIP **whp**. The next sub-section describes the algorithm and subsequent sub-sections help us to understand the conditioning problems.

2.1 The matching algorithm

Let us write $A = \{a_1, a_2, \ldots, a_n\}$ and $B = \{b_1, b_2, \ldots, b_n\}$ for the two parts of the vertex partition of BIP. The edge set of BIP is $OUT \cup IN$ where $OUT = \{out(a_i),\ i = 1, 2, \ldots, n\}$. Here if $out(i) = \{k, \ell\}$ then $out(a_i) = \{b_k, b_\ell\}$. $IN = \bigcup_{i=1}^n in(b_i)$ is defined similarly.

We next consider two edge labeled multigraphs $IG = (A, \{e_b : b \in B\})$ and $OG = (B, \{f_a : a \in A\})$ where $e_b = (b, in(b))$ and $f_a = (a, out(a))$. Hence for example e_b is a pair $(b, \{x, y\})$ where $b \in B$ is the edge label and $x, y \in A$ are the endpoints of the edge.

We note that IG and OG are very close in distribution to $G_{n,n}$.

Informal description of Algorithm PAIR:

Let $H_1 = IG$ and $H_2 = (B, \emptyset)$. Consider an isolated tree T of H_1. Choose any vertex x as root. Orient the edges of T towards x. Let the directed edges of T now be $e_i = (b_i, \{a_i, a_i'\})$ for $1 \leq i \leq t$. If we match (a_i, b_i) together for $1 \leq i \leq t$ then only the root x is not matched. To deal with this we consider matching x with one of b, b' where $f_x = \{b, b'\}$. But if we match x with b then, to avoid later conflicts we should delete e_b from H_1. We thus go to and fro between H_1 and H_2, rooting trees in H_1, adding edges to H_2 and deleting edges in H_1. If successful, the algorithm transforms H_1 and H_2 into graphs in which every component is either a tree or unicyclic. A perfect matching M can be constructed from the rooted trees (and unicyclic components) as described previously. The roots of trees in H_1 will be *marked* and the *non*-roots of trees in H_2 will be *checked*.

The following conditions will be observed by the algorithm.

- In H_1, marked vertices appear in isolated trees only. An isolated tree contains at most one marked vertex.
$$A^\star = \{\text{marked vertices}\}.$$
- H_2 contains no complex components. (A *complex* component is one with more edges than vertices.) The algorithm fails if one is created. An isolated tree contains a unique unchecked vertex. All vertices of unicyclic components are checked.
$$B^\star = \{\text{checked vertices}\}.$$

Whp IG consists of a unique giant (size $\Omega(n)$) connected component **GIANT** plus no other complex components. We only consider running **PAIR** if a **GIANT** exists. In fact Karp, Rinnooy-Kan and Vohra described their algorithm with **GIANT** replaced by the **2-CORE**. However, their calculations were made in terms of **GIANT**.

It will be necessary for us to carry out **PAIR** in rounds, something not considered in [13].

Algorithm PAIR

The algorithm returns a perfect matching M of the input $BIP(D_n)$ **whp**.

$H_1 := IG$ and $H_2 := (B, \emptyset)$.
$A^\star := B^\star := \emptyset$;
If IG does not have a unique component **GIANT** of size $\geq n/1000$ then **FAIL**.
$E_g :=$ {edges of **GIANT** in H_1};
$B_g := \{b : e_b \in E_g\}$;
For $t = 1, 2, \ldots$ **do** [Round t]
Begin
If every isolated tree of H_1 contains a marked vertex, go to SUCCESS; otherwise
$A_t^\star := B_t^\star := \emptyset$;
For every isolated tree T of H_1 in which all vertices are unmarked **do**
Begin
Randomly choose vertex $a \in T$;
$A_t^\star := A_t^\star \cup \{a\}$; i.e. mark a. MARK STEP
End
$A^\star := A^\star \cup A_t^\star$;
Arbitrarily order $A_t^\star = \{f_1, f_2, \ldots\}$;
For $j = 1, 2, \ldots$ **do**
Begin
Add the edge f_j to H_2. Let K be the component of H_2 containing f_j.
If K is complex then **FAIL**.
Else
If K is unicyclic, choose the one remaining unchecked vertex of K.
X **If** K is a tree, there are two unchecked vertices. Choose one of them. Give preference to vertices in B_g, choosing randomly if this does not yield a unique choice.
Let b be the vertex chosen at this step.
$B_t^\star := B_t^\star \cup \{b\}$; i.e. check b CHECK STEP
End
$B^\star := B^\star \cup B_t^\star$;
Delete the edges e_b, $b \in B_t^\star$ from H_1; DELETION STEP
End of round t.;
Next t;
SUCCESS Get the output matching M using the following rules;
Isolated trees: Orient the edges of the isolated trees of H_1, H_2 so that all directed paths lead to the unique marked vertex (H_1) or to the unique unchecked vertex (H_2).
If $a \in T, a \notin A^\star$ for some tree T of H_1 then add (a, b) to M where e_b is the unique edge of T which has been oriented out of a. Call this an IN-edge.
If $b \in T, b \in B^\star$ for some tree T of H_2 then add (a, b) to M where f_a is the unique edge of T which has been oriented out of b. Call this an OUT-edge.

Unicyclic components: Deal with similarly, i.e. randomly orient their cycles and then orient the edges towards the cycles.

End of Algorithm **PAIR**.

Let the final value of H_1 (resp. H_2) be denoted by H_{IN} (resp. H_{OUT}) and the value of H_1 (resp. H_2) at the end of round t by $H_1(t)$ (resp. $H_2(t)$). If we want to keep the orientations given at the end of **PAIR** then we refer to them as $\vec{H}_{IN}, \vec{H}_{OUT}$ respectively. Note that it is legitimate to say:

$$\text{The matching } M \text{ produced by } \textbf{PAIR} \text{ is } \textit{defined by } \vec{H}_{IN} \text{ and } \vec{H}_{OUT}. \tag{3}$$

For input I to **PAIR** we will let $\rho = \rho(I)$ denote the number of rounds executed. Furthermore, in order to be specific, we will make all random choices in the Algorithm **PAIR** according to the following rule.

Rule (R1). Selection rule for random choices

We choose two random permutations $\alpha \in S_A$ and $\beta \in S_B$. (Here the set of permutations of a finite set X is denoted by S_X.) Assuming A is ordered $a_1 < a_2 < \cdots < a_n$ we find that α induces an ordering $<_\alpha$ on A where $a <_\alpha a'$ iff $\alpha(a) < \alpha(a')$. Similarly for B and β. Choices are made according to the orderings $<_\alpha, <_\beta$. Thus in a MARK STEP we mark the first vertex under ordering $<_\alpha$ which is in tree T. In a CHECK STEP, if we have to choose between two vertices, we take the first one under the ordering $<_\beta$. Finally, to orient a cycle after a SUCCESSFUL termination, we (i) choose the first vertex x of the cycle (under $<_\alpha$ or $<_\beta$) and (ii) orient the edges so that the edge out of x points to its lowest neighbour (under $<_\alpha$ or $<_\beta$).

We refer the reader to [13] for an elegant analysis of Algorithm **PAIR** and the proof of the following theorem. Let

$$B^\star = \bigcup_{t=1}^{\rho} B_t^*.$$

Theorem 4 *[13]* **PAIR whp** *ends successfully with* $.1n \leq |A^\star| = |B^\star| \leq .49n$.

2.2 An equivalence relation on the input

We regard $\mathcal{D}_{2-in,2-out}$ as $\mathcal{D}_{2-in} \times \mathcal{D}_{2-out}$. The probability space of inputs to **PAIR** is thus $\mathcal{D}_{2-in} \times \mathcal{D}_{2-out} \times S_A \times S_B$ with uniform measure.

Let $\mathcal{S}$ denote the subset of the input space for which Algorithm **PAIR** terminates successfully.

We now prove a lemma showing that the matching M produced by **PAIR** is random, and that the set $A^\star$ is independent of M. The result is intuitively obvious, but important enough to demand a detailed proof.

Lemma 5 *(a) Let M_1, M_2 be perfect matchings of A with B. Then*

$$\mathbf{Pr}(M = M_1) = \mathbf{Pr}(M = M_2).$$

(b) Let $S, S' \subseteq A$ be of the same cardinality. Then

$$\mathbf{Pr}(A^\star = S \mid M) = \mathbf{Pr}(A^\star = S' \mid M).$$

Proof (a) Let $M_1 = \{(a_i, b_i) : i \in [n]\}$ and $M_2 = \{(a_i, \pi(b_i)) : i \in [n]\}$ where π is a permutation of B. It is enough to consider this case. Consider an input $\mathcal{I}_1 = (D^1_{\text{in}}, D^1_{\text{out}}, \alpha_1, \beta_1)$ to **PAIR** that yields $M = M_1$.

To obtain D^2, the instance of D_n in the input $\mathcal{I}_2 = (D^2_{\text{in}}, D^2_{\text{out}}, \alpha_2 = \alpha_1, \beta_2 = \beta_1\pi^{-1})$ which leads to $M = M_2$ we make the following substitutions in D^1.

A1 If $out(a) = \{b_1, b_2\}$ in $\mathcal{I}_1$ then let $out(a) = \{\pi(b_1), \pi(b_2)\}$ in $\mathcal{I}_2$.

A2 If $in(b) = \{a_1, a_2\}$ in $\mathcal{I}_1$ then let $in(\pi(b)) = \{a_1, a_2\}$ in $\mathcal{I}_2$.

With these substitutions IG is unchanged except for a permutation of edge labels and OG is unchanged except for a permutation of vertex labels. In $\mathcal{I}_2$, $\pi(b)$ takes on the role of b in $\mathcal{I}_1$. Conditions **A1** and **A2** enforce this in terms of edge and vertex labels and $\beta_2 = \beta_1\pi^{-1}$ ensures that the position of $\pi(b)$ under ordering $<_{\beta_2}$ is that of b in $<_{\beta_1}$. The set B_g of $\mathcal{I}_1$ is replaced by $\pi(B_g)$ in $\mathcal{I}_2$; so if a unique choice to check $b \in B_1$ exists for $\mathcal{I}_1$ at Step X then $\mathcal{I}_2$ checks $\pi(b) \in \pi(B_g)$. The mapping from inputs $\mathcal{I}_1$ to inputs $\mathcal{I}_2$ is measure preserving, and so (a) follows.

(b) Let M_1, M_2 be as in (a). We first argue that

$$\mathbf{Pr}(A^\star = S \mid M = M_1) = \mathbf{Pr}(A^\star = S \mid M = M_2). \tag{4}$$

Let $\mathcal{I}_1$ denote an input which produces $A^\star = S$ and $M = M_1$. If $\mathcal{I}_2$ is constructed from $\mathcal{I}_1$ as in (a) then **PAIR** will yield $A^\star = S$ and $M = M_2$. This proves (4). Thus

$$\mathbf{Pr}(A^\star = S \mid M) = \mathbf{Pr}(A^\star = S).$$

We just have to show that

$$\mathbf{Pr}(A^\star = S) = \mathbf{Pr}(A^\star = S').$$

Now let π be any permutation of A for which $\pi(S) = S'$. Let $\mathcal{I}_S = (D^S_{\text{in}}, D^S_{\text{out}}, \alpha, \beta)$ be an input which produces $A^\star = S$. Let $\mathcal{I}_{S'} = (D^{S'}_{\text{in}}, D^{S'}_{\text{out}}, \alpha\pi^{-1}, \beta)$ where instead of **a1** and **a2** we make the substitutions given below to transform the input graph.

A3 If $out(a) = \{b_1, b_2\}$ in $\mathcal{I}_S$ then let $out(\pi(a)) = \{b_1, b_2\}$ in $\mathcal{I}_{S'}$.

A4 If $in(b) = \{a_1, a_2\}$ in $\mathcal{I}_S$ then let $in(b)) = \{\pi(a_1), \pi(a_2)\}$ in $\mathcal{I}_{S'}$.

The substitutions ensure that $\pi(a)$ takes on the role of a at each step of the algorithm. Thus $A^\star = S'$, and we have defined a measure preserving map from inputs for which $A^\star = S$ to those for which $A^\star = S'$. □

We need the following notation for $I \in \mathcal{S}$: Here each piece of notation implicitly depends on the input e.g. $DEL = DEL(I)$.

- DEL denotes the set of edges of IG that are deleted by **PAIR** on instance I.
- $e \in DEL$ consists of a pair $(\Xi(e), endpts(e))$ where $\Xi : DEL \to B^\star$ is a bijection.
- $DEL_{t,1} = \{e \in DEL : \ \Xi(e) \in B_g \cap B_t^\star\}$ for $t = 1, 2, \dots, \rho$.
- $DEL_{t,2} = \{e \in DEL : \ \Xi(e) \in B_t^\star \setminus B_g\}$ for $t = 1, 2, \dots, \rho$.
- $ENDPTS_{t,i} = \{endpts(e) : \ e \in DEL_{t,i}\}$ as multi-sets for $i = 1, 2, t = 1, 2, \dots \rho$.
- $B_{t,1}^\star = B_t^\star \cap B_g$ and $B_{t,2}^\star = B_t^\star \setminus B_g$ for $t = 1, 2, \dots \rho$.
- $B_{0,1}^\star = B_g \setminus B^\star$ and $B_{0,2}^\star = B \setminus (B_g \cup B^\star)$.

Given a perfect matching M, the *matching function* $\mu : A \longrightarrow B$ is defined by $M = \{(a, \mu(a)) : \ a \in A\}$. Note that

$$\mu(A^\star) = B^\star. \tag{5}$$

For if $a \in A^\star$ then f_a is an edge of H_{OUT} and a will be matched to a checked vertex as described at the final SUCCESS phase of Algorithm **PAIR**.

We now define an equivalence relation $\sim$ on the set $\mathcal{S}$ of inputs I leading to a successful completion of the Algorithm **PAIR**. We define it by starting with a fixed representative $\tilde{I}$ and then describing all inputs which are related to it.

Our idea is to condition the input to be a random member of a fixed equivalence class. Within an equivalence class M will be invariant. The structure of the class will be sufficiently simple that we can analyse the distribution of edges not contained in M.

We will use ˜ to denote quantities associated with $\tilde{I}$. We start with the digraphs $\tilde{D}_1 = \vec{H}_{IN}(\tilde{I}), \tilde{D}_2 = \vec{H}_{OUT}(\tilde{I})$.

$\tilde{D}_1, \tilde{D}_2$ together contain n edges $\tilde{E} = \tilde{E}_1 \cup \tilde{E}_2$. Let π be an arbitrary bijection between $\tilde{E}$ and M. We use π to relabel the edges of $\tilde{D}_1, \tilde{D}_2$ as follows:

Labels of edges and non-roots of trees: If $e = \underset{\tilde{b}}{(\tilde{a}_1, \tilde{a}_2)} \in \tilde{E}_1$ and $\pi(e) = (a, b)$ then vertex label $\tilde{a}_1$ of e is replaced by a and the edge label $\tilde{b}$ of e is replaced by b. If

$e = (\tilde{b}_1, \tilde{b}_2) \in \tilde{E}_2$ and $\pi(e) = (a, b)$ then vertex label $\tilde{b}_1$ of e is replaced by b and the edge label $\tilde{a}$ of e is replaced by a. We see immediately from this that two labelled digraphs $D_{1,\pi}, D_{2,\pi}$ produced by this process will also define the matching M, in the sense of (3).

The first part of our procedure for randomly sampling from $\mathcal{E}$ is then:

G$_1$: Construct a random bijection π from $\tilde{E}$ to M.

So far we have shown how π produces new random labels for all edges of $\vec{H}_{IN}, \vec{H}_{OUT}$ and for some of the vertices. We need to describe how to put in new labels for the roots of trees. We will need to further fill in the missing parts of instance I and then show that the digraphs $\vec{H}_{IN}(I), \vec{H}_{OUT}(I)$ produced by **PAIR** are the same as the ones produced by our relabelling.

Labels of roots of trees If $\tilde{x}$ is the label of the root of a tree in $\tilde{D}_1$ (resp. $\tilde{D}_2$) then $\tilde{D}_2$ (resp. $\tilde{D}_1$) contains an edge with label $\tilde{x}$. If π changes this to $a \in A$ (resp. $b \in B$) then the vertex label $\tilde{x}$ of the root will be replaced by a (resp. b).

Construction of I: We have now partially specified a new instance I.

1. If $D_{1,\pi}$ contains an edge (a_1, a_2) with edge label $b = \mu(a_1)$ then $in(b) = \{a_1, a_2\}$.
2. Similarly, if $D_{2,\pi}$ contains an edge (b_1, b_2) with edge label $a = \mu^{-1}(b_1)$ then $out(a) = \{b_1, b_2\}$.

We let $A^\star$ be those elements of A which have *out* specified in 2. and let $B^\star$ be those elements of B which have *not* had *in* specified in 1. (We confirm the consistency of this notation later). Note that

$$|A^\star| = |B^\star| = |A^\star(\tilde{I})| \text{ is independent of } \pi.$$

(a): $out(a), a \notin A^\star$: Since these sets are not exposed by **PAIR** they do not affect the construction of $\vec{H}_{IN}, \vec{H}_{OUT}$ and are unconditioned by **PAIR**. So in our equivalence class we leave them unconditioned as well.

G$_2$: For $a \notin A^\star$ choose $out(a) = \{b_1, b_2\}$ randomly and independently.

π induces a permutation $\pi_A \in S_A$ as follows: If $e = (\tilde{a}, \tilde{a}') \in \tilde{E}_1$ and $\pi(e) = (a, b)$ then $\pi_A(\tilde{a}) = a$. If $e \in \tilde{E}_2$ and e has edge label $\tilde{a}$ and $\pi(e) = (a, b)$ then $\pi_A(\tilde{a}) = a$.

(b): $in(b), b \in B^\star$: These edges do not produce edges of $\vec{H}_{IN}$ but nevertheless they do have an effect on its construction. If $b \in B^\star$ then $(\mu^{-1}(b), b)$ is the image under π of some edge $e = (\tilde{b}, \tilde{b}')$ of $\tilde{D}_2$. If $\tilde{b} \in \tilde{B}^\star_{t,i}$ then we place b into $B^\star_{t,i}$. In this way we define $B^\star_{t,i}$ for $t = 1, 2, \ldots, \tilde{\rho}$, $i = 1, 2$.

G$_3$: For $t = 1, 2, \ldots, \tilde{\rho}$, $i = 1, 2$ let $\phi_{t,i}$ be a random bijection between $B^\star_{t,i}$ and $\{\{\pi_A(a_1), \pi_A(a_2)\} : \{a_1, a_2\} \in ENDPTS_{t,i}(\tilde{I})\}$. Then $in(b) = \phi_{t,i}(b)$ for $b \in B^\star_{t,i}$.

We let $\pi_B = \mu\pi_A\mu^{-1}$, $B_g = \pi_B(\tilde{B}_g)$ and $\alpha = \tilde{\alpha}\pi_A^{-1}$ and $\beta = \tilde{\beta}\pi_B^{-1}$. If $b \in B_{t,i}^\star$ then there is an edge $e = (\tilde{b}, \tilde{b}') \in \tilde{E}_2$ where $\tilde{b} \in \tilde{B}_{t,i}^\star$ and $\pi_B(\tilde{b}) = b$ showing that

$$\pi_B(\tilde{B}_{t,i}^\star) = B_{t,i}^\star \text{ for } t = 1, 2, \ldots, \tilde{\rho},\ i = 1, 2,$$

and so

$$\pi_A(\tilde{A}^\star) = A^\star.$$

This completes our description of an equivalence class. We must show that the definition is consistent. To go from $\tilde{I}$ to I we have made the following replacements:

$$\tilde{b} \notin \tilde{B}^\star,\ in(\tilde{b}) = \{\tilde{a}_1, \tilde{a}_2\} \Longrightarrow in(\pi_B(\tilde{b})) = \{\pi_A(\tilde{a}_1), \pi_A(\tilde{a}_2)\}.$$

The edge $(\tilde{a}_1, \tilde{a}_2)$ with label $\tilde{b}$ is replaced by the edge $(\pi_A(\tilde{a}_1), \pi_A(\tilde{a}_2))$ with label $\pi_B(\tilde{b})$.

$$\tilde{a} \in \tilde{A}^\star,\ in_A(\tilde{a}) = \{\tilde{b}_1, \tilde{b}_2\} \Longrightarrow in_A(\pi_A(\tilde{a})) = \{\pi_B(\tilde{b}_1), \pi_B(\tilde{b}_2)\}.$$

The edge $(\tilde{b}_1, \tilde{b}_2)$ with label $\tilde{a}$ is replaced by the edge $(\pi_B(\tilde{b}_1), \pi_B(\tilde{b}_2))$ with label $\pi_A(\tilde{a})$.

Furthermore,

$$\tilde{a} <_{\tilde{\alpha}} \tilde{a}_2 \leftrightarrow \pi_A(\tilde{a}) <_\alpha \pi_A(\tilde{a}_2) \text{ and } \tilde{b} <_{\tilde{\beta}} \tilde{b}_2 \leftrightarrow \pi_B(\tilde{b}) <_\beta \pi_B(\tilde{b}_2).$$

Suppose that in addition we add

$$\tilde{b} \in \tilde{B}^\star,\ in(\tilde{b}) = \{\tilde{a}_1, \tilde{a}_2\} \Longrightarrow in(\pi_B(\tilde{b})) = \{\pi_A(\tilde{a}_1), \pi_A(\tilde{a}_2)\}. \tag{6}$$

Then as far as **PAIR** is concerned $\pi_A(\tilde{a}), \pi_B(\tilde{b})$ are just "new names" for $\tilde{a}, \tilde{b}$ and so **PAIR** will produce the labelled digraph given in the description of the equivalence class.

If (6) does not hold then we are just changing the bijections $\phi_{t,i}$. But changing such a bijection does not change the final digraphs $\vec{H}_{IN}, \vec{H}_{OUT}$. This is because in the DELETION STEP of round t we delete the edges $e_b, b \in B_t^\star$ and changing $\phi_{t,i}$ changes the edge labels around but does not change the endpoints of the edges that are deleted.

In our analysis of Phase 2 we will condition on our input being chosen randomly from some fixed equivalence class $\mathcal{E}$. To facilitate working in this model we assume that we have chosen a fixed representative $\tilde{I} \in \mathcal{E}$. We will continue to use a *tilde* to refer to quantities associated with $\tilde{I}$.

Let $\beta_{t,i}^\star = |\tilde{B}_{t,i}^\star|$ for $i = 1, 2$, $t = 0, 1, \ldots, \rho$ and let $\beta_t^\star = \beta_{t,1}^\star + \beta_{t,2}^\star$ for $t = 0.1, \ldots, \rho$.

2.3 Sizes

In this section we prove some facts about the sizes of various objects in H_{IN}, H_{OUT}.

Lemma 6 *There is an absolute constant* $0 < \gamma < 1/2$ *such that* **whp**

$$\gamma \leq \frac{\beta^\star_{t,i}}{\beta^\star_t} \leq 1-\gamma$$

simultaneously for all t, i *such that* $\beta^\star_t \geq n^{9/10}$.

Proof Suppose that $H_2(t)$ has $\alpha_t n/2$ edges. Here $\alpha_t \leq .98$ **whp**, see Theorem 4. Let an isolated tree of $H_2(t)$ be a g-tree if all of its vertices are in B_g. Note that as observed in [13], at the end of a round, the number of g-trees is precisely the number of edges left in **GIANT**. Next let

$$\psi(x,y) = \sum_{k=1}^{\infty} \frac{x^k k^{k-2}}{k!} y^{k-1} e^{-yk}.$$

The number of isolated trees contained in a fixed set of size xn in $G_{n,yn/2}$ is **qs** equal to $\psi(x,y)+o(n^{3/4})$. A calculation involving known results on the size of the giant component yields that there is a constant $.75 \leq \xi_g \leq .8$. such that **qs** $|B_g|/n = \xi_g + o(1)$.

Now we have

$$\beta^\star_t = (\alpha_t - \alpha_{t-1})n \tag{7}$$

and by the above **qs**

$$\begin{aligned} \beta^\star_{t,1} &= (\psi(\xi_g,\alpha_{t-1}) - \psi(\xi_g,\alpha_t))n + o(n^{3/4}) \\ &= (\alpha_{t-1}-\alpha_t)f'(\eta)n + o(n^{3/4}) \end{aligned} \tag{8}$$

where $f(\alpha) = \psi(\xi_g,\alpha)$ and $\alpha_{t-1} \leq \eta \leq \alpha_t$.

But

$$\begin{aligned} f'(\eta) &= \frac{1}{\eta}\sum_{k=1}^{\infty} \frac{\xi_g^k k^{k-2}(k-1)}{k!}\eta^{k-1}e^{-\eta k} - \sum_{k=1}^{\infty}\frac{\xi_g^k k^{k-1}}{k!}\eta^{k-1}e^{-\eta k} \\ &= \eta^{-1}\zeta_1 - \zeta_2 \end{aligned}$$

where

$$n\zeta_1 + o(n) = \text{the expected number of edges of isolated trees of } G_{n,\eta/n} \text{ in } [\xi_g n].$$

$$n\zeta_2 + o(n) = \text{the expected number of vertices of isolated trees of } G_{n,\eta/n} \text{ in } [\xi_g n].$$

Assuming $\eta < 1$ we have

$$\zeta_1 = \eta\xi_g^2/2 \text{ and } \zeta_2 = \xi_g. \tag{9}$$

The lemma follows from (7)–(9) with $\gamma = (\xi_g - \frac{1}{2}\xi_g^2)/2$. □

Lemma 7 *There exists an absolute constant* $\kappa > 0$ *such that* **whp**, *on termination of* **PAIR**, *the graph* H_{OUT}:

1. Contains no complex components.

2. Has at most $\kappa \log n$ vertices on unicyclic components.

3. Has maximum component size at most $\kappa \log n$.

4. Has at most $n/(\log n)^{20}$ vertices in components of size $\geq \kappa \log\log n$.

Proof We know from Theorem 4 that **whp** H_{OUT} will be composed of αn, $\alpha \leq .49$, random edges. Erdős and Rényi [5] show that **whp** such a graph satisfies 1,2,3 above. We have to account for repeated edges since [5] deals with $G_{n,m}$. But **whp** there will be $O(\log n)$ such edges and they will not upset the desired conclusion.

A simple calculation verifies 4. Indeed, if $\alpha \leq .49$ and Z is the number of vertices on isolated trees of size $k \in [k_0 = \kappa \log\log n, k_1 = \kappa \log n]$ and

$$N = \binom{n}{2}$$

then

$$\begin{aligned} \mathbf{E}(Z) &\leq \sum_{k=k_0}^{k_1} \binom{n}{k} k^{k-2} \frac{(\alpha n)_{k-1}}{N^{k-1}} \left(1 - \frac{k(n-k)}{\binom{n}{2}}\right)^{\alpha n} \\ &\leq \alpha^{-1} \sum_{k=k_0}^{k_1} n(2\alpha e^{1-2\alpha})^k \\ &\leq n/(\log n)^{40} \end{aligned}$$

for large enough κ. So Property 4 follows from the Markov inequality. □

We now do a similar analysis of H_{IN}. Observe first that after $t < \rho$ rounds of **PAIR** we find that $H_1(t) = H_1^g(t) \cup H_1^{\neg g}(t)$ where the superscript g will denote the subgraph induced by **GIANT** and the superscript $\neg g$ will denote the subgraph induced by the remaining vertices. More precisely

$H_1^g(t)$ is IG^g after the deletion of $\beta_{1,1}^\star + \cdots + \beta_{t,1}^\star$ random edges. (Note that there is some implied conditioning viz. that after the deletion of this number of random edges there is still at least one complex component.)

$H_1^{\neg g}(t)$ is $IG^{\neg g}$ after the deletion of $\beta_{1,2}^\star + \cdots + \beta_{t,2}^\star$ random edges.

To better understand H_{IN}^g we imagine a graph process $G_0, G_1, \ldots, G_m = ([n], E_m), \ldots$ where E_{i+1} is obtained from E_i by adding a random edge $\mathbf{e}_{i+1}$. (Note that $\mathbf{e}_{i+1} \in E_i$ is allowed here.) Thus IG can be identified with G_n. Now define

$$m_t = \max\{m : G_m^g \text{ has exactly } \beta_{1,1}^\star + \cdots + \beta_{t,1}^\star \text{ fewer edges than } G_n^g\}$$

and

$$\epsilon_t = \frac{2m_t}{n} - 1.$$

Also let

$$m^\star = \max\{m : G^g_m \text{ has no complex components}\}.$$

Then $m_1 > m_2 > \cdots > m_\rho = m^\star$ and $H^g_{IN} = G^g_{m^\star}$. We cannot claim that $H_{IN} = G_{m^\star}$ since the former will tend to have more edges outside of **GIANT**. In general

$$H_1(t) = G^g_{m_t} \cup G^{\neg g}_{\hat{m}_t} \tag{10}$$

where $G^{\neg g}_{\hat{m}_t}$ has the same number of edges as $H_1^{\neg g}(t)$.

We note next that Łuczak, Pittel and Wierman [15] have shown that **whp**

$$\tfrac{1}{2}n + n^{2/3}/\log n < m^\star < \tfrac{1}{2}n + n^{2/3}\log n. \tag{11}$$

In the following lemma we will define $\epsilon = \epsilon(m)$ by

$$m = \frac{1+\epsilon}{2}n$$

where $n^{-1/3}/\log n \leq \epsilon \leq 1$. We let $\epsilon^\star = \epsilon(m^\star)$.

The function $\eta = \eta(\epsilon)$ is defined by $0 < \eta < 1$ and

$$(1-\eta)e^{-(1-\eta)} = (1+\epsilon)e^{-(1+\epsilon)}.$$

A simple calculation yields

$$\eta = \epsilon - \frac{2}{3}\epsilon^2 + O(\epsilon^3).$$

We also need the following notation: $T_{k,m}$ denotes the number of edges of G_m which lie in isolated trees containing at least k vertices. U_m is the number of edges lying in unicyclic components, C_m is the number of edges which lie in complex components. The edges of a complex component are divided into *mantle* edges and *2-core* edges. The former are distinguished by their deletion producing an isolated tree. $\hat{C}_m$ is the number of edges which are mantle edges of a complex component of G_m.

The lemma is not meant to be best possible, but merely sufficient to our purposes. Property (c) is particularly "crude". It is possible that what is needed could be gleaned from papers dealing with the fine detail of the growth of the giant component e.g. Bollobás [2], Łuczak, Pittel and Wierman [15], Janson, Knuth, Łuczak and Pittel [11]. The reader who is happy with this statement is encouraged to skip the proof.

Lemma 8 **Whp** *the following conditions hold for* G_m *in the specified ranges: Let* $k_0 = \lceil 1000\epsilon^{-2}\log n\rceil$.

(a) :	$T_{k_0,m} = 0,$	$n^{-1/3}/\log n \le \epsilon \le 1.$
(b) :	$T_{k,m} \le \frac{n}{\sqrt{k}},$	$n^{-1/3}/\log n \le \epsilon \le 1,$ $1 \le k \le n^{1/4}.$
(c) :	$U_m \le n^{3/4},$	$n^{-1/10} \le \epsilon \le 1.$
(d) :	$C_m = \left(\frac{1+\epsilon}{2} - \frac{(1-\eta)^2}{2(1+\epsilon)}\right) n + O(n^{.7+o(1)}) \approx (2\epsilon - \frac{8}{3}\epsilon^2 + O(\epsilon^3))n$	$n^{-1/10} \le \epsilon \le 1.$
(e) :	$\hat{C}_m = (1-\eta)\left(1 - \frac{1-\eta}{1+\epsilon}\right) + O(n^{.7+o(1)}) \approx (2\epsilon - \frac{14}{3}\epsilon^2 + O(\epsilon^3))n$	$n^{-1/10} \le \epsilon \le 1.$

Proof When ϵ is constant we can find these results in [5]. We have to account for repeated edges since [5] deals with $G_{n,m}$. But **whp** there will be $O(\log n)$ such edges and they will not upset the desired conclusion.

So for the rest of the proof assume that $\epsilon \le \epsilon_0$ where ϵ_0 is a sufficiently small absolute constant.

Let $X_{k,m}$ denote the number of edges of G_m contained in trees with k vertices. Then

$$\mathbf{E}(X_{k,m}) = \binom{n}{k}(k-1)k^{k-2}\frac{(m)_{k-1}}{N^{k-1}}\left(1 - \frac{a_k}{N}\right)^{m-k+1} \tag{12}$$

where $a_k = k(n-k) + \binom{k}{2}$ and $N = \binom{n}{2}$.

We use the following estimates:

$$\binom{n}{k} = \frac{n^k}{k!}\exp\left\{-\sum_{t=1}^{\infty}\frac{k^{t+1}}{t(t+1)n^t} + O\left(\frac{k}{n} + \frac{1}{n-k}\right)\right\} \tag{13}$$

$$\frac{(m)_{k-1}}{N^{k-1}} = \left(\frac{1+\epsilon}{n}\right)^{k-1}\exp\left\{-\sum_{t=1}^{\infty}\frac{k^{t+1}}{t(t+1)m^t} + O\left(\frac{k}{n} + \frac{1}{n-k}\right)\right\} \tag{14}$$

$$\left(1 - \frac{a_k}{N}\right)^{m-k+1} = \exp\left\{-(m-k+1)\sum_{t=1}^{\infty}\left(\frac{2k}{n} - \frac{k^2}{n^2} + O\left(\frac{k}{n^2}\right)\right)^t t^{-1}\right\} \tag{15}$$

It follows from (12)–(15) that

$$\mathbf{E}(X_{k,m}) = n\frac{(k-1)k^{k-2}}{k!}(1+\epsilon)^{k-1}e^{\phi(k,n,\epsilon)}\left(1 + O\left(\frac{k}{n} + \frac{1}{n-k}\right)\right) \tag{16}$$

where

$$\phi(k,n,\epsilon) = -k - \epsilon k + (\epsilon - 2\epsilon^2 + O(\epsilon^3))\frac{k^2}{2n} - \frac{k^3}{6n^2} + O\left(\frac{k^4}{n^3} + \epsilon\frac{k^3}{n^2}\right).$$

For constant ϵ we know from [5] that **qs** G_m has a giant component of size

$$\left(1 - \frac{1-\eta}{1+\epsilon}\right)n + O(n^{.5+o(1)}) = (2\epsilon - \tfrac{8}{3}\epsilon^2 + O(\epsilon^3) + O(n^{-.5+o(1)}))n. \tag{17}$$

The remaining parts of the lemma also follow from [5] when ϵ is constant. So for ϵ_0 sufficiently small we can restrict our attention to $k \leq \epsilon_0 n$ and then we can assume that the terms $-\frac{k^3}{6n^2} + O\left(\frac{k^4}{n^3} + \epsilon\frac{k^3}{n^2}\right)$ in the definition of ϕ are bounded above by $-\frac{k^3}{7n^2}$ and that $\epsilon - 2\epsilon^2 + O(\epsilon^3) \leq \epsilon$. So we write

$$\phi(k,n,\epsilon) \leq -k - \epsilon k + \epsilon\frac{k^2}{2n} - \frac{k^3}{7n^2}.$$

Going back to (16) and using Stirling's approximation we have

$$\begin{aligned}
\mathbf{E}(X_{k,m}) &\leq \frac{n}{\sqrt{2\pi}k^{3/2}}\exp\left\{-\left(\frac{\epsilon^2}{2}-\frac{\epsilon^3}{3}\right)k + \epsilon\frac{k^2}{2n} - \frac{k^3}{7n^2}\right\} \\
&= \frac{n}{\sqrt{2\pi}k^{3/2}}\exp\left\{\frac{\epsilon^3}{3}k - \frac{k^3}{56n^2} - \frac{k}{2}\left(\epsilon - \frac{k}{2n}\right)^2\right\} \\
&\leq \frac{n}{\sqrt{2\pi}k^{3/2}} \times \begin{cases} \exp\left\{\frac{1}{3}\epsilon^3 k - \frac{1}{8}\epsilon^2 k\right\} & k \leq \epsilon n \\ \exp\left\{\frac{1}{3}\epsilon^3 k - \frac{1}{56}\epsilon^2 k\right\} & k \geq \epsilon n \end{cases} \\
&= o(n^{-2}),
\end{aligned}$$

for $k \geq k_0$. Property (a) follows immediately.

We see next that since $T_{k,m} = X_{k,m} + X_{k+1,m} + \cdots$,

$$\mathbf{E}(T_{k,m}) \leq \sum_{\ell=k}^{\infty} \frac{n}{\sqrt{2\pi}\ell^{3/2}} \leq (1+o(1))\sqrt{\frac{2}{\pi k}}n.$$

Now changing one edge of G_m can only change $T_{k,m}$ by at most $2k$ and so applying the Azuma-Hoeffding martingale tail inequality we obtain that for any $u > 0$

$$\mathbf{Pr}(|T_{k,m} - \mathbf{E}(T_{k,m})| \geq u) \leq 2\exp\left\{-\frac{2u^2}{4mk^2}\right\} = o(n^{-2})$$

if $u = \frac{n}{10\sqrt{k}}$ and (b) follows.

To prove (c) we use Rényi's formula

$$R_k \approx \sqrt{\pi/8}k^{k-1/2}$$

for the number R_k of unicyclic connected graphs with vertex set $[k]$.

If $Y_{k,m}$ denotes the number of edges in unicyclic components of G_m (including trees with one edge doubled) then arguing as for (16) we obtain

$$\mathbf{E}(Y_{k,m}) = O\left(\frac{k^{k+1/2}}{k!}(1+\epsilon)^k e^{\phi(k,n,\epsilon)}\right)$$

$$= O\left(\exp\left\{-\left(\frac{\epsilon^2}{2}-\frac{\epsilon^3}{3}\right)k+\epsilon\frac{k^2}{2n}-\frac{k^3}{7n^2}\right\}\right).$$
$$= \begin{cases} O\left(\exp\left\{\frac{1}{3}\epsilon^3 k - \frac{1}{8}\epsilon^2 k\right\}\right) & k \leq \epsilon n \\ O\left(\exp\left\{\frac{1}{3}\epsilon^3 k - \frac{1}{56}\epsilon^2 k\right\}\right) & k \geq \epsilon n \end{cases}$$

We deduce that

$$\mathbf{E}(U_m) = O(\epsilon^{-2}) \tag{18}$$

and that

$$\mathbf{Pr}(\exists k \geq k_0 :\ Y_{k,m} \neq 0) = O(n^{-10}). \tag{19}$$

Let

$$\hat{U}_m = Y_{k,m} + \cdots + Y_{k_0,m}.$$

Then **wpp**(10) we have

$$U_m = \hat{U}_m \tag{20}$$

We see from (18) and (19) that

$$\mathbf{E}(\hat{U}_m) = O(\epsilon^{-2}). \tag{21}$$

Applying the Azuma-Hoeffding martingale tail inequality we see that, since changing one edge changes $\hat{U}_m$ by at most k_0,

$$\mathbf{Pr}(|\hat{U}_m - \mathbf{E}(\hat{U}_m)| \geq u) \leq 2\exp\left\{-\frac{2u^2}{mk_0^2}\right\} \tag{22}$$

for any $u > 0$. Putting $u = n^{1/2}k_0 \log n$ we see that with probability **wpp**(10) $\hat{U}_m$ will not deviate from its mean by more than $n^{1/2}k_0(\log n)^2 = o(n^{3/4})$ which together with (21) implies (c).

To prove (d) we go back to (16) and write

$$\begin{aligned} \mathbf{E}(T_{1,m}) &= (1+o(n^{-3/4}))\sum_{k=1}^{k_0}\frac{n}{1+\epsilon}\frac{(k-1)k^{k-2}}{k!}((1+\epsilon)e^{-(1+\epsilon)})^k \\ &= (1+o(n^{-3/4}))\sum_{k=1}^{\infty}\frac{n}{1+\epsilon}\frac{(k-1)k^{k-2}}{k!}((1-\eta)e^{-(1-\eta)})^k \\ &= (1+o(n^{-3/4}))\frac{1-\eta}{2(1+\epsilon)}n. \end{aligned}$$

(The final equation follows indirectly from the fact that almost all edges of $G_{n,(1-\eta)/n}$ are in trees.)

Putting $\hat{T}_{1,m} = X_{k,m} + \cdots + X_{k_0,m}$ we see that **wpp**(10) we have

$$\hat{T}_{1,m} = T_{1,m}.$$

Applying the Azuma-Hoeffding martingale tail inequality we see that, since changing one edge changes $\hat{T}_{1,m}$ by at most k_0,

$$\mathbf{Pr}(|\hat{T}_{1,m} - \mathbf{E}(\hat{T}_{1,m})| \geq u) \leq 2\exp\left\{-\frac{2u^2}{mk_0^2}\right\} \tag{23}$$

for any $u > 0$. Putting $u = n^{1/2}k_0 \log n$ we see that with probability $1 - O(n^{-10})$ $T_{1,m}$ will not deviate from its mean by more than $n^{1/2}k_0(\log n)^2 = o(\epsilon^2 n)$. Property (d) now follows from this and (c).

To prove (e) observe that $\mathbf{e}_m$ is a mantle edge iff one of its endpoints is a complex component of G_{m-1} and its other endpoint is in an isolated tree of G_{m-1}. By a similar argument as to that given for (d) we can prove that the LHS expression (17) for the number of vertices in the giant component of G_{m-1} remains **whp** $\left(1 - \frac{1-\eta}{1+\epsilon}\right) n + O(n^{.5+o(1)})$. This gives the correct estimate for the expectation of $\hat{C}_m$. Concentration follows from the fact that changing one edge of G_m can only change $\hat{C}_m$ by $2k_0$, assuming there are no isolated or mantle trees of size greater than k_0. This is true **whp** and we can now use a martingale tail inequality as in (c), (d). □

Let $LC_1(I, m)$ denote the subset of A appearing in G_m, (i) as vertex labels in tree components of size greater than $(\log n)^{1/2}$ or (ii) in unicyclic components. Let $LC_1(I) = LC_1(I, m^\star)$. The above lemma and (11) imply that for $\frac{1}{2}n + n^{2/3}/\log n \leq m \leq n$, **whp**

$$|LC_1(I, m)| \leq 2n/(\log n)^{1/4}. \tag{24}$$

Note that the sizes of $LC_1(I, m)$ are invariants of the equivalence class $\mathcal{E}$.

Next let $LB^\star_{t,i} = \{b \in B^\star_{t,i} : \ in(b) \cap LC_1 \neq \emptyset\}$.

Lemma 9 *Let γ be as in Lemma 6. Conditional on (24),* **qs** *for all t, i such that $\beta^\star_{t,i} \geq n^{9/10}$ we have*

$$|LB^\star_{t,i}| \leq 10\gamma\beta^\star_{t,i}/(\log n)^{1/4}.$$

Proof Let $E_{m',s} = \{\mathbf{e}_{m'-s+1} \dots, \mathbf{e}_{m'}\}$ and for $m < m' - s$ let $E_{m,s,m'} = \{e \in E_{m',s} : e \cap LC_1(m) \neq \emptyset\}$. E_m and $E_{m',s}$ are independently chosen when $m < m' - s$. Also, assuming the condition in (24) holds we see that $|E_{m,s,m'}|$ is stochastically dominated by the binomial $B(s, 5/(\log n)^{1/4})$ and so if $\mathcal{B}_{m,s,m'}$ denotes the event $\{|E_{m,s,m'}| \geq 10s/(\log n)^{1/4}\}$ we see that

$$\mathbf{Pr}(\mathcal{B}_{m,s,m'}) \leq e^{-s/(\log n)^{1/4}}.$$

To complete the proof note that the event of the lemma is contained in $\bigcup_{m',s,m} \mathcal{B}_{m,s,m'}$, $s \geq n^{9/10}$. □

Now let

$$B^\star(s) = \bigcup_{t,i:\ \beta^\star_{t,i}<s} B^\star_{t,i}$$

and let $A^\star(s) = \mu^{-1}(B^\star(s))$.

Lemma 10 *Suppose* $0 \le \theta = O(\log\log\log n)$. *Then conditional on* $|A^\star| \le .49n$, **wpp***(10)*

$$|B^\star(n/(\log n)^{3\theta+4})| \le \frac{n}{(\log n)^{\theta+1}},$$

Proof Observe first that $\beta^\star_t$ is monotone decreasing with t. This is because the deletion of an edge creates at most one tree with an unmarked vertex. On the other hand if $\epsilon_t \ge n^{-1/10}$ and $\beta^\star_t \ge n^{9/10}$ then **qs**

$$\beta^\star_{t+1} \le (1 - 2\gamma\epsilon_t^2 + O(\epsilon_t^3))\beta^\star_t. \tag{25}$$

where γ is as in Lemma 6. To see this consider the deletion of $e_b, b \in B^\star_t$ during algorithm **PAIR**. $\beta^\star_t - \beta^\star_{t+1}$ is precisely the number of $b \in B^\star_{t,1}$ such that $in(b)$ lies in the 2-core of a current complex component. Thus **qs**

$$\beta^\star_t - \beta^\star_{t+1} = \frac{|C_{m_t}| - |\hat{C}_{m_t}|}{m_{t,g}}\beta^\star_{t,1} + O(n^{.5+o(1)})$$

where $m_{t,g}$ is the number of edges of G_{m_t} which are contained in **GIANT**. A calculation based on known results about the size of the giant component shows that **qs** $m_{0,g} \ge .95n$ and so $m_{t,g} \ge .46n$ for $1 \le t \le \rho$. (25) now follows from Lemma 8 and Lemma 6.

Let $s = n/(\log n)^\theta$ and $s' = n/(\log n)^{\theta+4}$. Let $t_0 = \max\{t : \beta^\star_t \ge s\}$, $t_1 = \max\{t : \beta^\star_t \ge s'/\gamma\}$ and $t_2 = \max\{t : \epsilon_t \ge s/(n\log n)\}$. Note that Lemma 6 implies that **qs**

$$|B^\star(s')| \le \sum_{t=t_1+1}^{\rho} \beta^\star_t.$$

Furthermore **qs**

$$\sum_{t=t_2+1}^{\rho} \beta^\star_t \le m_{t_2} - \tfrac{1}{2}n \le \frac{s}{2\log n}.$$

So we are done if $t_2 \le t_1$. Assume therefore that $t_2 > t_1$. Then applying (25)

$$\begin{aligned}
\sum_{t=t_1+1}^{t_2} \beta^\star_t &\le \beta^\star_{t_1+1}\sum_{t=0}^{\infty}\left(1 - \gamma\left(\frac{s}{n\log n}\right)^2\right)^t \\
&\le \frac{s'}{\gamma^2}\left(\frac{n\log n}{s}\right)^2 \\
&\le \frac{s}{3\log n}.
\end{aligned}$$

$\square$

Finally, for this section we have the following simple lemma.

Lemma 11 **Qs** *IG does not contain a set $S \subseteq A, |S| \geq n_0 = n(\log n)^{-100 \log\log n}$ such that S is incident with $\geq |S|(\log\log n)^2$ edges.*

Proof The expected number of sets violating the condition is at most

$$\sum_{s=n_0}^{n} \binom{n}{s} \binom{n}{s(\log\log n)^2} \left(\frac{2s}{n}\right)^{s(\log\log n)^2} \leq \sum_{s=n_0}^{n} \left(\frac{ne}{s}\left(\frac{2e}{(\log\log n)^2}\right)^{(\log\log n)^2}\right)^s = O(n^{-K})$$

for any constant $K > 0$. $\square$

3 Phase 2. Removing small cycles

We condition on $I \in \mathcal{E}$. We assume that we have a permutation digraph Π_1 containing at most $2\log n$ cycles and that the likely events of Lemmas 6, 7, 8 9, 10 and 11 hold. We note that the discussion of the previous sections is in terms of A and B, but the vertex set of D_n is V. Elements $a, a_i, a' \in V$ etc. will simultaneously refer to elements $a, a_i, a' \in A$ as well. Similarly for elements of B. We hope that this does not lead to confusion.

We say that a cycle C of Π_1 is *small* if $|C| < n_0 = \left\lceil \frac{1000n}{\log n} \right\rceil$ and large otherwise. The set of vertices on small cycles is denoted by SMALL and the remaining vertices are placed in LARGE. It is easy to see that $\mathbf{E}(|SMALL|) = n_0 - 1$. We will therefore assume from now on that $\mathcal{E}$ is such that

$$|SMALL| \leq n_0 \log\log n.$$

This is true **whp**.

At the start of Phase 2 we choose a set $X = \{x_1, x_2, \ldots, x_{2r}\}$ where (i) each $e_i = (x_{2i-1}, x_{2i})$ is an edge of Π_1, (ii) each small cycle C contains one such edge ($|C| \leq \gamma_1$) or $\gamma_1/2$ such edges ($|C| > \gamma_1$), (iii) the x_i are distinct except for those on cycles of length one. Here γ_1 is an even positive integer defined in Section 3.2.1.

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 $[n] \setminus V(P(\Gamma))$. In the associated bipartite graph BIP it cooresponds to a matching of size $n-1$.

We describe a process which removes a small cycle C from a permutation digraph Π. We start by choosing an edge $(a^{(0)}, b^{(0)})$ of C and deleting it to obtain an NPD $\Gamma^{(0)}$ with $P_0 = P(\Gamma^{(0)}) \in \mathcal{P}(b^{(0)}, a^{(0)})$, where $\mathcal{P}(x, y)$ denotes the set of paths from x to y in D_n. Here

$(a^{(0)}, b^{(0)}) = e_i$ for some $1 \leq i \leq r$. The aim of the process is to produce a large set S of NPD's such that for each $\Gamma \in S$, (i) $P(\Gamma)$ has a least n_0 edges and (ii) the small cycles of $NPD(\Gamma)$ are a subset of the small cycles of Π. We will then 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. The process consists of a single Out-Phase followed by a set of In-Phases.

If this process succeeds then we remove another small cycle, if necessary. Otherwise, if C is not too small we try again with a different starting edge of X for $(a^{(0)}, b^{(0)})$. Indeed, if $|C| \geq \gamma_1$ then we will try a number of times with a different edge before giving up.

3.1 Out-Phase

The *basic step* in an *Out-Phase* of this process is to take an NPD Γ with $P(\Gamma) \in \mathcal{P}(b^{(0)}, a)$ and to examine the edges $out(a)$ of D_n leaving (i.e. edges going *out* from) the end of the path. Let b be the terminal vertex of such an edge and assume that Γ contains an edge (a', b). Then $\Gamma' = \Gamma \cup \{(a, b)\} \setminus \{(a', b)\}$ is also an NPD. We will find use for the notation $\Gamma' = NPD(\Gamma; a, b, a')$ and describe this basic step as $bs(\Gamma; a, b, a')$. Note that in BIP it also means adding one edge (a, b) and deleting edge (a', b) and so represents the use of an alternating path of length 2. We use a sequence of such paths to build up longer alternating paths.

Γ' is *acceptable* if:

(C1) $P(\Gamma')$ contains least n_0 edges.

(C2) Any new cycle created (i.e. in Γ' and not Γ) also has at least n_0 edges.

If Γ contains no edge (a', b) then $b = b^{(0)}$ and we could close the cycle *provided* the cycle produced has at least n_0 edges. We will not accept the edge in the interest of making a simple uniform definition of acceptance. It is anyway, an unlikely event.

As mentioned previously, we create our first NPD by deleting an edge $(a^{(0)}, b^{(0)})$ of a small cycle C. We create a collection of NPD's by repeatedly making basic steps. This leads naturally to a tree $\mathcal{T}$ of NPD's where the children of a node are those NPD's obtainable by making an acceptable basic step. We let

$$\Gamma^{(0)} \text{ denote the root of } \mathcal{T}$$

and

$$a(\Gamma) \text{ denote the endpoint of } P(\Gamma) \text{ other than } b^{(0)}.$$

Ignoring acceptability, a node Γ of $\mathcal{T}$ has one or two descendants. Γ has one descendant whenever $a(\Gamma) \in A^\star$. (Now and again an NPD Γ for which $a(\Gamma) \notin A^\star$ could have only one descendant due to one basic step being acceptable and the other not. To simplify matters we make the other step unacceptable so that Γ is a leaf of $\mathcal{T}$.

Let Z be the root of $\mathcal{T}$ plus the set of nodes with two descendants plus the leaves. Now contract any path in $\mathcal{T}$ joining two nodes of Z to get another tree $\hat{\mathcal{T}}$ where every internal vertex has exactly two descendants. If Γ is the parent of Γ' in $\hat{\mathcal{T}}$ then we say that Γ' is obtained from Γ by a *composite* step.

We now examine a composite step $\Gamma \to \Gamma'$ in detail. Suppose first that $\Gamma \neq \Gamma^{(0)}$. Then $a = a(\Gamma) \notin A^\star$, (as will be evident from the following construction). Let $out(a) = \{b, b'\}$. We will use the notation

$$out(\Gamma) = \{b, b'\}.$$

Suppose b lies in a tree T of the graph $\vec{D}_2$. Let the path from b to the root of T be $b = b_1, b_2, \ldots, b_{k+1}$ where $b_1, b_2, \ldots, b_k \in B^\star$ and $b_{k+1} \notin B^\star$. Let the edge label of (b_i, b_{i+1}) be a_i for $i = 1, 2, \ldots, k$. Thus $b_i = \mu(a_i)$ and $out(a_i) = \{b_i, b_{i+1}\}$ for $1 \leq i \leq k$. If we restrict ourselves to adding and removing OUT-edges then we remove the edges (a_i, b_i) from M and replace them by (a_i, b_{i+1}) for $1 \leq i \leq k$. Let $a_{k+1} = \mu^{-1}(b_{k+1}) \notin A^\star$. We remove the edge (a_{k+1}, b_{k+1}) from M as well. Thus we make the sequence of basic steps

$$bs(\Gamma; a, b, a_1), bs(\Gamma_i; a_i, b_{i+1}, a_{i+1}) \text{ for } 1 \leq i \leq k$$

where

$$\Gamma_i = NPD(\Gamma_{i-1}; a_i, b_{i+1}, a_{i+1}) \text{ for } 1 \leq i \leq k \text{ and } \Gamma_0 = \Gamma.$$

Finally let

$$out(a_{k+1}) = \{\beta, \beta'\}.$$

If the composite step is successful then $\Gamma' = \Gamma_k$ will be a descendant of Γ in $\hat{T}$ and we will have $out(\Gamma') = \{\beta, \beta'\}$. At this point we see that $a(\Gamma') = a_{k+1} \notin A^\star$ justifying our previous assumption $a(\Gamma) \notin A^\star$ (modulo dealing with the descendants of $\Gamma^{(0)}$). The other descendant Γ'' will be obtained from b'.

If $\Gamma = \Gamma^{(0)}$ and $a^{(0)} \notin A^\star$ then we carry on as above with $out(\Gamma^{(0)}) = \{b, b'\} = out(a^{(0)})$. If $\Gamma = \Gamma^{(0)}$ and $a^{(0)} \in A^\star$ then $out(a^{(0)}) = \{b^{(0)}, b\}$ and we continue as above, working through the tree T containing vertex b. In this case we let $out(\Gamma^{(0)}) = \{b\}$.

We build $\hat{\mathcal{T}}$ in a breadth-first fashion and each non-leaf vertex Γ (other than possibly the root $\Gamma^{(0)}$) gives rise to *two* NPD children Γ', Γ'' by composite steps. The set of nodes at depth t is denoted by L_t. The construction of $\hat{\mathcal{T}}$ ends when we first have $\nu = \left\lceil \sqrt{n \log n} \right\rceil$ leaves. We show subsequently that we achieve this goal **whp**. The construction of $\hat{\mathcal{T}}$ constitutes an Out-Phase of our procedure to eliminate small cycles. Having constructed $\hat{\mathcal{T}}$ we need to do a further *In-Phase*, which is similar to *a set of* Out-Phases.

3.1.1 Probabilistic analysis of a composite step in an Out-Phase

The algorithm requires us to expose some of the random choices in G_1, G_2, G_3 which define our instance. These exposures will fix parts of π and the mappings $\phi_{t,i}$ etc.. We follow the *method of deferred decisions* [14], and generate the necessary parts of them as we go.

The history $\mathcal{H}$ of Phase 2 consists of sets (i) $M_{\mathcal{H}} = \{e \in M$ such that $\pi^{-1}(e)$ has been specified$\}$, (ii) $B^{\star}_{\mathcal{H}} = \{b \in B^{\star}_{t,i}$ for some t, i such that $\phi_{t,i}(b)$ has been specified$\}$,and (iii) $A_{\mathcal{H}} = \{a \notin A^{\star}$ such that $out(a)$ has been specified$\}$. We will **whp** manage to keep $|\mathcal{H}| = |M_{\mathcal{H}}| + |B_{\mathcal{H}}| + |A_{\mathcal{H}}| \leq n^{.5+o(1)}$ throughout Phase 2.

In what follows our probabilities are all *implicitly* computed conditional on the current value of $\mathcal{H}$.

We start Phase 2 with M and the digraphs $\tilde{D}_1, \tilde{D}_2$. As we learn more about the instance $I \in \mathcal{E}$, we will change some of the labels on $\tilde{D}_1, \tilde{D}_2$ and let us write $\tilde{D}^{\mathcal{H}}_1, \tilde{D}^{\mathcal{H}}_2$ to indicate that changes that have been made.

A component of $\tilde{D}^{\mathcal{H}}_1$ or $\tilde{D}^{\mathcal{H}}_2$ is *clean* if none of its labels have been changed by Phase 2 and *dirty* otherwise.

Let $\Gamma \in \hat{\mathcal{T}}$. We estimate the probability that Γ has 2 children. First consider the case $\Gamma \neq \Gamma^{(0)}$. Let $a = a(\Gamma)$ and $out(\Gamma) = \{b, b'\}$. From the description of a composite step we see that b, b' were chosen randomly through G_2.

Let $a_1 = \mu^{-1}(b)$ and $a'_1 = \mu^{-1}(b')$, $e = (a_1, b)$ and $e' = (a'_1, b')$. Focusing on e we see that

$$\mathbf{Pr}(e \in M_{\mathcal{H}}) = \frac{|M_{\mathcal{H}}|}{n} = O(n^{-.5+o(1)}). \tag{26}$$

Assume $e \notin M_{\mathcal{H}}$. Then we choose $\pi^{-1}(e)$ randomly from $\pi^{-1}(M \setminus M_{\mathcal{H}})$ and relabel $\tilde{D}^{\mathcal{H}}_2$ accordingly. Let K be the component of $\tilde{D}^{\mathcal{H}}_2$ containing $\pi^{-1}(e)$. Then by Lemma 7

$$\mathbf{Pr}(|K| \geq \kappa \log\log n \text{ or } K \text{ is unicyclic}) = O((\log n)^{-20}). \tag{27}$$

There are at most $O(|M_{\mathcal{H}}| \log\log n + n(\log n)^{-20})$ vertices on dirty components and so

$$\mathbf{Pr}(K \text{ is dirty}) = O((\log n)^{-20}). \tag{28}$$

We can assume therefore that (i) $e \notin M_{\mathcal{H}}$, (ii) K is a tree, (iii) $|K| \leq \log\log n$, and (iv) K is clean. Also

$$\mathbf{Pr}(bs(\Gamma; a, b, a_1) \text{ is not successful} \mid (i) - -(iv)) = O(1/\log n).$$

This is mainly the probability that adding the edge (a, b) creates a short path or cycle.

We now have to expose the path $b = b_1, b_2, \ldots, b_{k+1}$ from b to the root of K along with the matching edges $(a_i, b_i), i = 1, 2, \ldots, k$. Assume that we have chosen $(a_j, b_j), j = 1, 2, \ldots, i$. We have already dealt with the case $i = 1$. We now choose (a_{i+1}, b_{i+1}) randomly from $M \setminus M_{\mathcal{H}}$. Then for the basic step $bs(\Gamma; a_i, b_{i+1}, a_{i+1})$ we have

$$\mathbf{Pr}(\neg\mathbf{C1} \text{ or } \neg\mathbf{C2}) = O(n_0/n) \tag{29}$$

and so

$$\mathbf{Pr}(\exists i : \ bs(\Gamma; a_i, b_{i+1}, a_{i+1}) \text{ is not acceptable}) = O(\log\log n/\log n).$$

Also, $out(\Gamma' = \Gamma_k)$ will be a random pair since $a_{k+1} \notin A^\star$.

In summary, if $\Gamma \neq \Gamma^{(0)}$, then under the assumption that $out(\Gamma)$ is a random pair,

$$\mathbf{Pr}(\Gamma \text{ does not have two children } \Gamma', \Gamma'') = O(\log\log n/\log n). \tag{30}$$

Furthermore, $out(\Gamma), out(\Gamma'')$ will both be random pairs.

Now consider the case $\Gamma = \Gamma^{(0)}$. Assume that $(a^{(0)}, b^{(0)}) \notin M_{\mathcal{H}}$. We discuss the probability of this later. We need to know whether or not $a^{(0)} \in A^\star$. Up to this point we have decided on $\nu \leq n^{.5+o(1)}$ members of A as to whether or not they are in $A^\star$. To decide if $a^{(0)} \in A^\star$ we need to compute $\pi^{-1}(a^{(0)}, b^{(0)})$. We don't necessarily need this much information. So we flip a biassed coin and with probability $p = \frac{|A^\star| - \nu}{n - |M_{\mathcal{H}}|}$ we decide $a^{(0)} \in A^\star$ and with probability $1 - p$ we decide $a^{(0)} \notin A^\star$.

If $a^{(0)} \notin A^\star$ then $out(\Gamma^{(0)}) = \{b, b'\}$ is a random pair and we can proceed as above to show that $\Gamma^{(0)}$ has two descendants with probability bounded below as in (30). In this calculation we take account of the following: The first basic steps of the Out-phase will be unacceptable if b or b' lie on a small cycle. But

$$\mathbf{Pr}(\{b, b'\} \cap SMALL \neq \emptyset) = O\left(\frac{\log\log n}{\log n}\right)$$

and only the first basic steps from $\{b, b'\}$ have to take this into account.

If $a^{(0)} \in A^\star$ then we first choose $\pi^{-1}(a^{(0)}, b^{(0)})$ randomly from $\tilde{E}_2 \setminus \pi^{-1}(M_{\mathcal{H}})$. (We remind the reader that $\tilde{E}_2$ is the edge set of $\tilde{D}_2$). Taking $a_1 = a^{(0)}, b_1 = b^{(0)}$ we proceed as in the paragraph just before equation (29). Thus we see that in this case

$$\Gamma^{(0)} \text{ has one descendant with probability } 1 - O(\log\log n/\log n). \tag{31}$$

We can now discuss the expected growth rate of $\hat{\mathcal{T}}$.

Lemma 12 *Let C be a small cycle and let $\nu = \left\lceil\sqrt{n\log n}\right\rceil$. Assume, $|\mathcal{H}| \leq n^{.5+o(1)}$ and condition on its value at the start of the construction of $\mathcal{T}$. Then with conditional probability* $1 - O(1/(\log n)^{1-o(1)})$

(a) *There exists t such that $\nu \leq |L_t| \leq 2\nu$.*

(b) *In the process $\mathcal{H}$ grows by at most $n^{.5+o(1)}$.*

Proof (a) We consider the following three events: Here $\epsilon_0 = 1/\log_2\log_2 n$,

$$\mathcal{E}_0 = \{L_1 = \emptyset\}.$$

$$\mathcal{E}_1 = \{\exists 2 \leq t \leq t_0 = \lceil 4\log_2\log_2\log_2 n\rceil : |L_t| < 2|L_{t-1}|\}.$$

$$\mathcal{E}_2 = \{\exists t_0 \leq t \leq t_1 = \lceil \log_{1.999}\nu\rceil : |L_t| \leq (2-\epsilon_0)|L_{t-1}|\}.$$

We show that

$$\mathbf{Pr}(\mathcal{E}_0 \cup \mathcal{E}_1) = O\left(\frac{(\log\log n)^5}{\log n}\right) \tag{32}$$

$$\mathbf{Pr}(\mathcal{E}_2) = O((\log n)^{-3}). \tag{33}$$

Part (a) of the lemma follows from (32) and (33), for if neither $\mathcal{E}_1$ nor $\mathcal{E}_2$ occur then $|L_{t_0}| \geq 2^{t_0-1} \geq (\log_2\log_2 n)^4/2$ and then $|L_t|$ grows at a rate between $2-\epsilon_0$ and 2 until it reaches the target size.

Equation (32) follows from (31) and (30) since the latter implies

$$\mathbf{Pr}(\mathcal{E}_1) = O\left(2^{t_0}\frac{\log\log n}{\log n}\right).$$

Equation (33) also follows from (30) and Chernoff bounds for tails of the binomial. This is because (30) implies that given $|L_t|$, $|L_{t+1}|$ stochastically dominates $2B(|L_t|, 1-o(\epsilon_0))$.

(b) At the time when L_t reaches its target size in $[\nu, 2\nu]$, $\hat{\mathcal{T}}$ will have $O(\nu\log n)$ nodes. Each attempt at growing children increases $\mathcal{H}$ by at most $O(\log n)$ and clearly $O(\nu(\log n)^2) = n^{.5+o(1)}$. □

In summary, as long as $|\mathcal{H}| \leq n^{.5+o(1)}$, if C is a small cycle and the Out-Phase starts with $X \cap M_{\mathcal{H}} = \emptyset$ then with probability $1 - O(1/(\log n)^{1-o(1)})$ we succeed in producing $\nu_1 \in [\nu, 2\nu]$ NPD's $\Gamma^{(1)}, \Gamma^{(2)}, \ldots, \Gamma^{(\nu_1)}$ in which each path ends at a distinct vertex $a \notin A^\star$ and $out(a)$ is unexposed.

The total contribution to $\mathcal{H}$ from Out-Phases will be $n^{.5+o(1)}\log n = n^{.5+o(1)}$. (We try this process once for $C \in SMALL, |C| \leq \gamma_1$ and up to $\gamma_1/2$ times for $C \in SMALL, |C| \geq \gamma_1$.)

3.2 In-Phase

After an Out-Phase we execute an In-Phase. This involves the construction of trees $\hat{\mathcal{T}}_i, i = 1, 2, \ldots \nu_1$. Assume that $P(\Gamma^{(i)}) \in \mathcal{P}(b^{(0)}, a^{(i)})$. We start with $\Gamma^{(i)}$ and build $\hat{\mathcal{T}}_i$ in a similar way to $\hat{\mathcal{T}}$ except that here all paths generated end with $a^{(i)}$. This is done as follows: If a current NPD Γ has $P(\Gamma) \in \mathcal{P}(b, a^{(i)})$ then we consider adding an edge (a, b), $a \in in(b)$ and deleting the edge $(a, b') \in \Gamma$. 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.

We now examine a composite step $\Gamma \to \Gamma'$ in detail. Let

$$b(\Gamma) \text{ denote the endpoint of } P(\Gamma) \text{ other than } a^{(i)}.$$

Suppose first that $\Gamma \neq \Gamma^{(i)}$. Then $b = b(\Gamma) \in B^\star$, as will be evident from the following construction. Let $in(b) = \{a, a'\}$. We will use the notation $in(\Gamma) = \{a, a'\}$. Suppose a lies in a tree T of the graph $\vec{D}_1$. Let the path from a to the root of T be $a = a_1, a_2, \dots, a_{k+1}$ where $a_1, a_2, \dots, a_k \notin A^\star$ and $a_{k+1} \in A^\star$. Let the edge label of (a_i, a_{i+1}) be b_i for $i = 1, 2, \dots, k$. Thus $b_i = \mu(a_i)$ and $in(b_i) = \{a_i, a_{i+1}\}$ for $1 \leq i \leq k$. If we restrict ourselves to adding and removing IN-edges then we remove the edges (a_i, b_i) from M and replace them by (a_{i+1}, b_i) for $1 \leq i \leq k$. Let $b_{k+1} = \mu(a_{k+1}) \in B^\star$. We have to remove the edge (a_{k+1}, b_{k+1}) from M as well. Thus we make the sequence of basic steps

$$bs(\Gamma; b, a, b_1), bs(\Gamma_i; b_i, a_{i+1}, b_{i+1}) \text{ for } 1 \leq i \leq k$$

where

$$\Gamma_i = NPD(\Gamma_{i-1}; b_i, a_{i+1}, b_{i+1}) \text{ for } 1 \leq i \leq k \text{ and } \Gamma_0 = \Gamma.$$

Finally let

$$in(b_{k+1}) = \{\alpha, \alpha'\}. \tag{34}$$

If the composite step is successful then $\Gamma' = \Gamma_k$ will be a descendant of Γ in $\hat{\mathcal{T}}_i$ and we will have $in(\Gamma') = \{\alpha, \alpha'\}$. At this point we see that $b(\Gamma') = b_{k+1} \in B^\star$ justifying our previous assumption $b(\Gamma) \in B^\star$ (modulo dealing with the descendants of $\Gamma^{(i)}$). The other descendant Γ'' will be obtained from b'. We use the notation

$$a_1 = a_1(\Gamma'). \tag{35}$$

If $\Gamma = \Gamma^{(i)}$ and $b^{(i)} \in B^\star$ then we carry on as above with $in(\Gamma^{(i)}) = \{a, a'\} = in(b^{(i)})$. If $\Gamma = \Gamma^{(i)}$ and $b^{(i)} \notin B^\star$ then $in(b^{(i)}) = \{a^{(i)}, a\}$ and we continue as above, working through the tree T containing vertex a. In this case we let $in(\Gamma^{(i)}) = \{a\}$.

3.2.1 Probabilistic analysis of a composite step in an In-Phase

We focus first on a fixed $\hat{\mathcal{T}}_i$ e.g. $\hat{\mathcal{T}}_1$. Let $\Gamma \in \hat{\mathcal{T}}_1$. First consider the case $\Gamma \neq \Gamma^{(1)}$. We estimate the probability that Γ has 2 children. Let $b \in B^\star$ be the endpoint of $P(\Gamma)$ other than $a^{(1)}$. We define $b(\Gamma) = b$ and let $in(\Gamma) = in(b) = \{a, a'\}$. From the description of a composite step we see that a, a' were chosen randomly through G_3. Thus a, a' are not completely random but we will show that we can assume that they are always both chosen randomly from a set of of $n^{1-o(1)}$ pairs.

Let $b_1 = \mu(a)$ and $b_1' = \mu(a')$, $f = (a, b_1)$ and $f' = (a_1', b')$. Focusing on f we see that

$$\mathbf{Pr}(f \in M_{\mathcal{H}}) = \frac{|M_{\mathcal{H}}|}{n^{1-o(1)}} = O(n^{-.5+o(1)}). \tag{36}$$

Assume $f \notin M_{\mathcal{H}}$. Then we choose $\pi^{-1}(f)$ randomly from $\pi^{-1}(M \setminus M_{\mathcal{H}})$ and relabel $\tilde{D}_1^{\mathcal{H}}$ accordingly. Let K be the component of $\tilde{D}_1^{\mathcal{H}}$ containing $\pi^{-1}(f)$. Then by (24) we see that

$$\mathbf{Pr}(|K| \geq \kappa(\log n)^{1/2} \text{ or } K \text{ is unicyclic}) = O((\log n)^{-1/4}). \tag{37}$$

Let DC denote the set of vertices lying in dirty components of $\tilde{D}_1^{\mathcal{H}}$. Then using Lemma 8 we can assume

$$|DC| \leq 3n^{.9} + n^{.75} + n^{.875} + n^{.5+o(1)}n^{.25} \leq 4n^{.9}. \tag{38}$$

We can do our calculations for G_m where $\epsilon \approx n^{-.1}$ i.e. on a *supergraph* of $\tilde{H}_I$. The first term on the RHS of (38) bounds the number of vertices in complex components, the second bounds the number of vertices in unicyclic components, the third bounds the number of components in dirty trees of size at least $n^{.25}$ and the last bounds the number in trees of size at most $n^{.25}$.

So

$$\mathbf{Pr}(K \text{ is dirty}) = O(n^{-.1}). \tag{39}$$

We can assume therefore that (i) $e \notin M_{\mathcal{H}}$, (ii) K is a tree, (iii) $|K| \leq (\log n)^{1/2}$, and (iv) K is clean. Also

$$\mathbf{Pr}(bs(\Gamma; b, a, b_1) \text{ is not successful} \mid (i) - -(iv)) = O(1/\log n).$$

This is the probability of creating a short path or cycle.

We now have to expose the path $a = a_1, a_2, \ldots, a_{k+1}$ from a to the root of K along with the matching edges $(a_i, b_i), i = 1, 2, \ldots, k$. Assume that we have chosen $(a_j, b_j), j = 1, 2, \ldots, i$. We have already dealt with the case $i = 1$. We now choose (a_{i+1}, b_{i+1}) randomly from $M \setminus M_{\mathcal{H}}$. Then for the basic step $bs(\Gamma; a_i, b_{i+1}, a_{i+1})$ we have

$$\mathbf{Pr}(\neg\mathbf{C1} \text{ or } \neg\mathbf{C2}) = O(n_0/n) \tag{40}$$

and so

$$\mathbf{Pr}(\exists i: \ bs(\Gamma; a_i, b_{i+1}, a_{i+1}) \text{ is not acceptable}) = O(1/(\log n)^{1/2}).$$

Also, $in(\Gamma' = \Gamma_k)$ will be a random pair chosen randomly from a set of $n^{1-o(1)}$ pairs as will discussed shortly.

In summary, if $\Gamma \neq \Gamma^{(1)}$, then under the assumption that $in(\Gamma)$ is chosen randomly from a set of $n^{1-o(1)}$ pairs,

$$\mathbf{Pr}(\Gamma \text{ does not have two children } \Gamma', \Gamma'') = O(1/(\log n)^{1/4}). \tag{41}$$

Furthermore, $in(\Gamma'), in(\Gamma'')$ will both be chosen randomly from a set of $n^{1-o(1)}$ pairs.

Now consider the case $\Gamma = \Gamma^{(1)}$. Assume once again that $(a^{(0)}, b^{(0)}) \notin M_{\mathcal{H}}$. If $b^{(0)} \in B^\star$ then $in(\Gamma^{(1)}) = \{a, a'\}$ is chosen randomly from a set of $n^{1-o(1)}$ pairs and we can proceed as above to show that $\Gamma^{(0)}$ has two descendants with probability bounded below as in (41).

If $b^{(0)} \notin B^\star$ then we first choose $\pi^{-1}(a^{(0)}, b^{(0)})$ randomly from $\tilde{E}_1 \setminus \pi^{-1}(M_{\mathcal{H}})$. (This was not done in the Out-phase!) Taking $a_1 = a^{(0)}, b_1 = b^{(0)}$ we proceed as in the paragraph just before equation (40). Thus we see that in this case

$$\Gamma^{(0)} \text{ has one descendant with probability } 1 - O(1/(\log n)^{1/4}). \tag{42}$$

For $S \subseteq A$ let comp(S) denote the set of vertices of $\tilde{D}_1$ which lie in components which contain a member of S or a member of $\mathcal{H}$. It follows from calculations similar to that for (38) (once again deducing the result from $\epsilon \approx n^{-.1}$) that provided $|S| \geq n^{9/10}$,

$$\begin{aligned} |\text{comp}(S)| &\leq 3n^{.9} + n^{.75} + \frac{n}{\omega^{1/2}} + |S|\omega \\ &= O(n^{2/3}|S|^{1/3}) \end{aligned} \tag{43}$$

on taking $\omega = (n/|S|)^{2/3} \leq n^{1/15}$. Next let $\bar{a}$ be the root of the first component K examined by the first composite step of the construction of $\hat{\mathcal{T}}_1$. We deduce from (43) and our random choice of K through the random choice of $\pi^{-1}(a^{(0)}, b^{(0)})$ that since $|A^\star(n/(\log n)^4))| \leq n/\log n$ – Lemma 10 –

$$\mathbf{Pr}(\bar{a} \in A^\star(n/(\log n)^4)) \leq 1/(\log n)^{1/3}. \tag{44}$$

Now consider a general composite step. Suppose the root $\bar{a}$ of K does not lie in $A^\star(n/(\log n)^\kappa)$ for some $\kappa > 0$. We argue next that with α', α'' as in (34) we have

$$\mathbf{Pr}(\alpha' \in A^\star(n/(\log n)^{9\kappa+10})) = O((\log\log n)^2/\log n). \tag{45}$$

First of all Lemma 10 with $\theta = 3\kappa + 2$ implies that

$$|A^\star(n/(\log n)^{9\kappa+10})| \leq \frac{n}{(\log n)^{3(\kappa+1)}}.$$

Then, (43) implies that

$$|\text{comp}(A^\star(n/(\log n)^{9\kappa+10}))| = O(n/(\log n)^{\kappa+1}).$$

Thus Lemmas 10 and 11 imply that there are at most $O(n(\log\log n)^2/(\log n)^{\kappa+1})$ edges incident with $\text{comp}(A^\star(n/(\log n)^{9\kappa+10}))$. As $\bar{a} \notin A^\star(n/(\log n)^\kappa))$ step G_3 in the generation of instance I chooses the endpoints $in(\mu(\bar{a}))$ randomly from a set of size $\Omega(n/(\log n)^\kappa)$. This implies (45).

It follows from Lemma 9 that

$$\mathbf{Pr}(in(\mu(\bar{a})) \cap LC_1(I) \neq \emptyset) = O((\log n)^{-1/4}),$$

where $LC_1(I)$ is defined just before equation (24). This (partially) confirms the proposition that we can assume $in(\Gamma)$ is always chosen randomly from a set of of $n^{1-o(1)}$ pairs. The problem at the moment is that the $o(1)$ term grows with depth. We deal with this subsequently. Call this the "$o(1)$ problem".

Let $\ell_0 = \lceil \log\log\log n \rceil$. We consider the probability that an In-Phase succeeds in producing a complete ℓ_0-level binary tree below a given node Γ, given that $b(\Gamma) \notin LC_1$. By the above analysis, this is

$$1 - O(2^{\ell_0}/(\log n)^{1/4}) = 1 - O(\log\log n/(\log n)^{1/4}).$$

Note that in this calculation we can assume by (45) that no root of any tree examined in this process lies in $A^\star(n/(\log n)^{\kappa_0})$ where $\kappa_0 = 10^{\ell_0}$. The important thing here is that $n/(\log n)^{\kappa_0} = n^{1-o(1)} \geq n^{9/10}$ and we can use previous lemmas.

We now deal with the $o(1)$ problem. Let us call the (attempted) construction of this ℓ_0-level tree a *superstep*. Thus starting with $b(\Gamma) \notin LC_1$ we succeed in a superstep with probability $1 - O(\log\log n/(\log n)^{1/4})$.

This is a rather pessimistic view of the process and it will suffice for ℓ_0 iterations. We also need the following:

Lemma 13 *There exists a constant $\theta > 0$ such that* **qs** *for each t, i with $\beta^\star_{t,i} \geq n^{9/10}$, at least $\theta\beta^\star_{t,i}$ of the sets $in(b), b \in B^\star_{t,i}$ contain a vertex which is not in* $\text{comp}(A^\star(n/\log n))$.

Proof Consider first the case when $i = 2$. Initially, **qs** IG contains $(1 - o(1))e^{-4}n$ isolated edges Z. In the first round of **PAIR** each such edge will have one endpoint marked and then the corresponding edge will be added to H_2. **Qs** at least an e^{-4} proportion of these edges will stay isolated throughout the execution of **PAIR**. Let Y denote these edges of H_2. **Qs** at least a proportion $1/100$ of Y will join two vertices not in B_g. Call this set Y_1. Let $Z_1 \subseteq Z$ be the set of $\geq (1 - o(1))e^{-8}n/100$ edges which correspond in this way to Y_1. Algorithm **PAIR** deletes a random subset of them. By Theorem 4 each such edge has a greater than $1/2$ (conditional) probability of not being deleted. Thus **qs**, when the edges corresponding to $B^\star_{t,2}$ are chosen for deletion, there will be at least $(1-o(1))e^{-8}n/200$ edges of Z_1 still to choose from. Then **qs** a proportion $\geq e^{-8}/201$ of members of Z_1 will be chosen for deletion. Such an edge has the form (a_1, a_2) where $a_2 \in A^\star_{1,2}$.

The case $i = 1$ is similar except that one has to start with the second round and argue that the deletion of edges $e_b, b \in B^\star_{1,1}$ produces $\theta' n$ isolated edges which once belonged to **GIANT**. This is not difficult to prove: **GIANT** contains $\theta'' n$ paths (x, y, z) where z has degree 1 and y has degree 2. The deletion of (x, y) produces an isolated edge (y, z). The rest of the argument is similar to the case $i = 2$. $\square$

To use Lemma 13 we consider the 2^{ℓ_0} nodes at the bottom level of a superstep. Each such node Γ has a $\geq \theta$ chance of having $a_1(\Gamma)$ (see (35) for definition) lying in a tree with root not in $A^\star(n/\log n)$. Call such a node *helpful*. Applying (1) we see that the probability that level ℓ_0 contains fewer than $\theta 2^{\ell_0 - 1}$ helpful nodes is at most $e^{-\theta 2^{\ell_0 - 4}} \leq (\log n)^{-\gamma_0}$ for some absolute constant $\gamma_0 > 0$.

Call a successful superstep *entirely* successful if it contains at least $\theta 2^{\ell_0 - 1}$ helpful nodes. Thus a superstep is entirely successful with probability at least $1 - O((\log n)^{-\gamma_0})$.

In growing our trees we only consider the (entirely successful) supersteps growing from the helpful nodes of a previous superstep. We let h_t denote the number of helpful nodes at superstep level t (which corresponds to ordinary level $\ell_0 t$) in the constructed tree. We know from the above discussion that

$$h_1 \geq \theta 2^{\ell_0 - 1} \qquad\qquad \mathbf{wlp}(\gamma_0).$$

Now let $\mathcal{E}_3(t) = \{h_{t+1} \leq \theta 2^{\ell_0 - 1} h_t\}$ so that, by (1) and the above,

$$\mathbf{Pr}(\mathcal{E}_3(t) \mid h_t) \leq e^{-\theta 2^{\ell_0 - 4} h_t} \leq (\log n)^{-\gamma_0 h_t},$$

provided $|\mathcal{H}| \leq n^{.5+o(1)}$. It follows that if

$$t_2 = \left\lceil \frac{\frac{1}{2}\log n + \log\log n}{\log\theta + (\ell_0 - 4)\log 2} \right\rceil \text{ and } \mathcal{E}_3 = \bigcup_{t=1}^{t_2} \mathcal{E}(t)$$

then $\bar{\mathcal{E}}_3$ occurs $\mathbf{wlp}(\gamma_0)$. But if $\mathcal{E}_3$ does not occur, then the process can be stopped after t_2 steps with between $n^{1/2}\log n$ and a maximum $\nu_2 = n^{.5+o(1)}$ leaves. In which case the size of $\mathcal{H}$ will grow by $O(n^{.5+o(1)}\sqrt{\log n}) = n^{.5+o(1)}$ during the construction, as required.

Some further technical changes are necessary. Before considering them let us examine the probability that $(a^{(0)}, b^{(0)}) \in M_{\mathcal{H}}$ at the start of an In-phase or Out-Phase. This can only occur if at some previous step we choose $\pi(e) \in X$ for some edge e of $\tilde{D}_1 \cup \tilde{D}_2$, other than when we specifically compute $\pi^{-1}(a^{(0)}, b^{(0)})$ at the beginning of an In-phase or Out-phase. Call the former event $\mathcal{B}$. Then

$$\mathbf{Pr}(\mathcal{B}) = O(|X| n^{.5+o(1)}/n) = o(1)$$

as required.

We consider the construction of our ν_1 trees in two stages. First of all we grow the trees without enforcing acceptability and thus allow the formation of small cycles and paths. We enforce conditions involving cleanliness and not choosing vertices of LC_1 and we grow in supersteps. The growth of the ν_1 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, $i = 1, 2 \ldots, \nu_1, \ell = 0, 1, \ldots, t_2$. Thus $L_{i,0} = \{b^{(0)}\}$ 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 used edges have the same set of initial vertices and since all of the deleted edges are Π_1-edges we have $L_{i,\ell+1} = L_{1,\ell+1}$. This explains why we temporarily drop acceptability. Acceptability of an edge varies with $\hat{\mathcal{T}}_j$. We continue this growing process until $|L_{1,\ell}| = \nu_2$. This provides room for the pruning process described next.

We now consider the fact that in some of the trees some of the leaves may have been constructed in violation of acceptability. We imagine that we prune the trees $\hat{\mathcal{T}}_1, \hat{\mathcal{T}}_2, \ldots \hat{\mathcal{T}}_{\nu_1}$ by disallowing any supernode that was constructed in violation of acceptability. Let a tree be BAD if after pruning it has less than ν leaves and GOOD otherwise. Now consider the pruning of $\hat{\mathcal{T}}_1$. From the analysis of an In-Phase we see that

$$\mathbf{Pr}(\hat{\mathcal{T}}_1 \text{ is BAD}) = O\left(\frac{1}{(\log n)^{\gamma_0}}\right).$$

Therefore

$$\mathbf{E}(\text{number of BAD trees}) = O\left(\frac{\nu}{(\log n)^{\gamma_0}}\right)$$

and

$$\mathbf{Pr}(\exists \geq \nu_1/2 \text{ BAD trees}) = O\left(\frac{1}{(\log n)^{\gamma_0}}\right).$$

Thus

$$\begin{aligned}
&\mathbf{Pr}(\text{there are less than } \nu_1/2 \text{ GOOD trees after pruning}) \\
&\quad \leq \mathbf{Pr}(\text{failure to construct } \hat{\mathcal{T}}_1, \hat{\mathcal{T}}_2, \ldots \hat{\mathcal{T}}_{\nu_1}) + \mathbf{Pr}(\exists \geq \nu_1/2 \text{ BAD trees}) \\
&\quad = O\left(\tfrac{1}{(\log n)^{\gamma_0}}\right).
\end{aligned}$$

Thus with probability 1-$O(1/(\log n)^{\gamma_0})$ we end up with $\nu_1/2$ sets of ν paths, each of length at least $1000n/\log n$ where the ith set of paths all terminate in $a^{(i)} \notin A^\star$. The sets $out(a^{(i)})$ are still unconditioned and hence

$$\begin{aligned}
\mathbf{Pr}(\text{no } out(a^{(i)}) \text{ edge closes one of these paths}) &\leq \left(1 - \frac{2\nu}{n}\right)^{\nu_1/2} \\
&= O(n^{-1}).
\end{aligned}$$

Consequently the probability that we fail to eliminate a particular small cycle C after breaking an edge is $O(1/(\log n)^{\gamma_0})$. If $|C| \geq \gamma_1 = 4\lceil \gamma_0^{-1} \rceil$ then we try $2\lceil \gamma_0^{-1} \rceil$ times using independent edges of C and so the probability we fail to eliminate a given small cycle

C is certainly $O(1/(\log n)^2)$ for $|C| \geq \gamma_1$ (remember that we calculated all probabilities conditional on previous outcomes and assuming $|\mathcal{H}| \leq n^{.5+o(1)}$.)

Now the number of cycles of length at most γ_1 in Π_1 is asymptotically Poisson with bounded mean and so there are fewer than $\log\log n$ **whp**. Hence, the probability we fail to eliminate all small cycles is $o(1)$.

We have now shown that **whp** a 2-in,2-out digraph contains a permutation digraph Π_2 in which the minimum cycle length is at least $n_0 = \lceil 1000n/\log n \rceil$.

Lemma 14 *Phase 2 produces a permutation digraph* Π_2 *with minimal cycle length at least* n_0 **whp**.

4 Phase 3. Patching the Phase 2 permutation digraph to a Hamilton cycle

In this section we will no longer need to condition on $\mathcal{E}$. We simply condition on the output matching M and we then use Lemma 5(b) to discuss the distribution of unmarked vertices over Π_2.

Let $C_1, C_2, \ldots, C_k$ be the cycles of Π_2, and let $\tilde{V} = V \setminus (\mathcal{H} \cup A^\star)$. If $v \in \tilde{V}$ then $out(v)$ is still unconditioned, and is still a random pair. Let $\tilde{C}_i = C_i \cap \tilde{V}$, $c_i = |\tilde{C}_i|$, $c_1 \leq c_2 \leq \cdots \leq c_k$.

Lemma 15 Whp

$$c_i \geq (.5051)|C_i| - n^{9/10} \geq \frac{505n}{\log n}, \qquad 1 \leq i \leq k.$$

Proof It follows from Theorem 4 and Lemma 5(b) that **whp** every sub-path of Π_1 of length $\ell \geq \lambda = \lceil (\log n)^2 \rceil$ contains at least $(.509)\ell$ members of $V \setminus A^\star$.

Fix $1 \leq i \leq k$. Observe that $|\Pi_2 \setminus \Pi_1| = O((\log n)^2)$. Delete those edges of C_i which are in $\Pi_2 \setminus \Pi_1$ or on a cycle of length at most λ in Π_1. The number of edges deleted $d = O((\log n)^3)$ **whp**. Delete any of the d paths formed which are of length less than λ. This leaves $\rho = |C_i| - O((\log n)^5)$ remaining edges R_i on sub-paths of Π_1 of length at least λ. Thus **whp** at least $(.509)\rho \geq (.5051)|C_i|$ of the vertices in R_i are in not $A^\star$. □

We also ensure that $c_i \leq 0.505|C_i|$ by selecting a random $0.505|C_i|$ subset of $\tilde{C}_i$ if necessary. 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 $v \in \tilde{C}_i$, and deleting the edge (v, u) in Π_2. Let $m = \sum_{i=1}^k m_i$ and relabel (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 in D_n, 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 **whp**. 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 $\Pi_2 \cup D_n$, but also in the *auxiliary digraph* Λ, whose edges are $(i, \lambda(i))$.

Lemma 16 $(m-2)! \leq |R_\phi| \leq (m-1)!$

Proof We grow a path $1, \lambda(1), \lambda^2(1), \dots, \lambda^r(1) \dots$ 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_n 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 $\lambda(1), \lambda^2(1), \dots, \lambda^r(1), 1 \leq r \leq m-3$ 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 D by setting $\lambda(h) = w, \lambda(w) = 1$. Completing the second case to a cycle in Π_2 means that

$$\lambda = (1, \lambda(1), \dots, \lambda^{m-2}(1))(w) \tag{46}$$

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 (46) which is a contradiction.

Thus there is a (unique) completion of the path in Λ. $\square$

We finish our proof:

Lemma 17 $\mathbf{Pr}$*(* D_n *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_n. We will use the inequality

$$\mathbf{Pr}(X > 0) \geq \frac{\mathbf{E}(X)^2}{\mathbf{E}(X^2)}. \tag{47}$$

Now the definition of the m_i yields that

$$\frac{(1.01)n}{a} - k \leq m \leq \frac{(1.01)n}{a} + k$$

and so

$$(1.009)\log n \leq m \leq (1.011)\log n.$$

Also

$$k \leq \log n/1000,\ m_i \geq 1009 \text{ and } \frac{c_i}{m_i} \geq \frac{a}{2.001}, \qquad 1 \leq i \leq k.$$

Let Ω denote the set of possible cycle re-arrangements. $\omega \in \Omega$ is a *success* if D_n contains the edges needed for the associated Hamilton cycle. Thus,

$$\begin{aligned}\mathbf{E}(X) &= \sum_{\omega\in\Omega} \mathbf{Pr}(\omega \text{ is a success})\\ &= \sum_{\omega\in\Omega} \left(1 - \left(1 - \frac{1}{n}\right)^2\right)^m\\ &\geq (1 - o(1)) \left(\frac{2}{n}\right)^m (m-2)! \prod_{i=1}^{k} \binom{c_i}{m_i}.\end{aligned}$$

Now, $n! = (n/e)^n\sqrt{2\pi n}e^{f(n)}$ where $1/(12n+1) \leq f(n) \leq 1/12n$. An application of this leads to

$$\binom{c_i}{m_i} \geq \frac{1}{\sqrt{2\pi}} \left(\frac{c_i e}{m_i^{1+(1/2m_i)}}\right)^{m_i} \left(1 - \frac{m_i^2}{c_i}\right),$$

where $m_i^{1/2m_i} < 1.004$. Using a $\sqrt{2\pi}$ factor to "mop up" factors of size 1-o(1) we obtain,

$$\begin{aligned}\mathbf{E}(X) &\geq \frac{1}{(2\pi)^{k/2} m\sqrt{m}} \left(\frac{2m}{en}\right)^m \prod_{i=1}^{k} \left(\frac{c_i e}{(1.004)m_i}\right)^{m_i}\\ &\geq \frac{1}{(2\pi)^{k/2} m\sqrt{m}} \left(\frac{2m}{en}\right)^m \left(\frac{ea}{2.001 \times 1.004}\right)^m\\ &\geq n^{.0035}.\end{aligned} \tag{48}$$

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

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

$$\begin{aligned}
\mathbf{E}(X^2) &\leq \mathbf{E}(X) + (1 + o(1)) \sum_{M \in \Omega} \left(\frac{2}{n}\right)^m \sum_{\substack{\Omega \\ N' \cap N = \emptyset}} \left(\frac{2}{n}\right)^m \\
&\quad + (1 + o(1)) \sum_{M \in \Omega} \left(\frac{2}{n}\right)^m \sum_{s=2}^{m} \sum_{t=1}^{s-1} \sum_{\substack{\Omega \\ <s,t>}} \left(\frac{2}{n}\right)^{m-t} \\
&= \mathbf{E}(X) + E_1 + E_2 \text{ say.}
\end{aligned} \tag{49}$$

Clearly

$$E_1 \leq (1 + o(1)) \mathbf{E}(X)^2. \tag{50}$$

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

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

Now

$$\begin{aligned}
\frac{\binom{c_i - m_i}{m_i - \sigma_i}}{\binom{c_i}{m_i}} &\leq \frac{\binom{c_i}{m_i - \sigma_i}}{\binom{c_i}{m_i}} \\
&\leq (1 + o(1)) \left(\frac{m_i}{c_i}\right)^{\sigma_i} \exp\left\{-\frac{\sigma_i(\sigma_i - 1)}{2m_i}\right\} \\
&\leq (1 + o(1)) \left(\frac{2.001}{a}\right)^{\sigma_i} \exp\left\{-\frac{\sigma_i(\sigma_i - 1)}{2m_i}\right\}
\end{aligned}$$

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

$$\begin{aligned}
\frac{E_2}{\mathbf{E}(X)^2} &\leq (1+o(1))e^{k/2} \sum_{s=2}^{m} \sum_{t=1}^{s-1} \binom{s}{t} \exp\left\{-\frac{s^2}{2m}\right\} \left(\frac{2.001}{a}\right)^s \binom{m}{s} \frac{(m-t-1)!}{(m-2)!} \left(\frac{n}{2}\right)^t \\
&\leq (1+o(1))n^{.001} \sum_{s=2}^{m} \sum_{t=1}^{s-1} \binom{s}{t} \exp\left\{-\frac{s^2}{2m}\right\} \left(\frac{2.001}{a}\right)^s \frac{m^{s-(t-1)}}{(s-1)!} \left(\frac{n}{2}\right)^t \\
&= (1+o(1))n^{.001} \sum_{s=2}^{m} \left(\frac{2.001}{a}\right)^s \frac{m^s}{s!} \exp\left\{-\frac{s^2}{2m}\right\} m \sum_{t=1}^{s-1} \binom{s}{t} \left(\frac{n}{2m}\right)^t \\
&\leq (1+o(1)) \left(\frac{2m^3}{n^{.999}}\right) \sum_{s=2}^{m} \left(\frac{(2.001)n \exp\{-s/2m\}}{2a}\right)^s \frac{1}{s!} \\
&= o(1) \qquad (51)
\end{aligned}$$

To verify that the RHS of (51) is $o(1)$ we can split the summation into

$$S_1 = \sum_{s=2}^{\lfloor m/2 \rfloor} \left(\frac{(2.001)n \exp\{-s/2m\}}{2a}\right)^s \frac{1}{s!}$$

and

$$S_2 = \sum_{s=\lfloor m/2 \rfloor+1}^{m} \left(\frac{(2.001)n \exp\{-s/2m\}}{2a}\right)^s \frac{1}{s!}.$$

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

$$\begin{aligned}
S_1 &\leq \sum_{s=2}^{\lfloor (.5055) \log n \rfloor} \frac{((1.0005) \log n)^s}{s!} \\
&= o(n^{9/10})
\end{aligned}$$

since this latter sum is dominated by its last term.

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

$$S_2 \leq n^{(1.0005)e^{-1/4}} < n^{9/10}.$$

The result follows from (47) to (51). □

Acknowledgement It is a pleasure to acknowledge a careful reading of the paper by Bruce Reed. He found errors and shed light on the whole understanding of Phase 2.

References

[1] N.Alon, J.Spencer and P.Erdős, The probabilistic method, John Wiley and Sons, 1992.

[2] B.Bollobás, Random graphs, Academic Press, New York, 1985.

[3] C.Cooper and A.M.Frieze, *The limiting probability that* $\alpha - in, \beta - out$ *is strongly is connected*, Journal of Combinatorial Theory B 48 (1990) 117-134.

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

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

[6] W.Feller, *An introduction to probability theory and its applications. (Vol. 1)*, John Wiley and Sons, New York, 1966.

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

[8] A.M.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, 1990, 53-59.

[9] G.R.Grimmett and D.R.Stirzaker, *Probability and Random Processes*, Clarendon Press, Oxford, 1983.

[10] W.Hoeffding, *Probability inequalities for sums of bounded random variables.* Journal of the American Statistical Association 58 (1963), 13-30.

[11] S.Janson, D.E.Knuth, T.Łuczak and B.Pittel, *The birth of the giant component*, Random Structures and Algorithms 3 (1993) 233-358.

[12] V.F.Kolchin, *Random mappings*, Optimization Software Inc., New York,1986.

[13] R.M.Karp, A.H.G.Rinnooy Kan and R.V.Vohra, *Average case analysis of a heuristic for the assignment problem*, Mathematics of Operations Research 19, (1994) 513-522.

[14] D.E.Knuth, R.Motwani and B.Pittel, *Stable marriage*, Random Structures and Algorithms 1 (1990) 1-14.

[15] T.Łuczak, B.Pittel and J.Wierman, *The structure of a random graph near the point of the phase transition*, Transactions of the American Mathematical Society 341 (1994) 721-748.

[16] B.Pittel, *On tree census and the giant component in sparse random graphs*, Random Structures and Algorithms 1 (1990) 311-342.

[17] B.Reed and C.J.H.McDiarmid, *The strongly connected components of 1-in,1-out*, Combinatorics, Probability and Computing 1 (1992) 265-274.

[18] D.W.Walkup, *Matchings in random regular bipartite digraphs*, Discrete Mathematics 31 (1980) 59-64.

This research was sponsored in part by National Science Foundation (NSF) grant no. CCR-0122581.