Combinatorics, Probability and Computing (2020), **29**, pp. 900–942
doi:10.1017/S0963548320000140

CAMBRIDGE
UNIVERSITY PRESS

ARTICLE

Hamiltonicity in random directed graphs is born resilient

Richard Montgomery

School of Mathematics, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK
Email: r.h.montgomery@bham.ac.uk

(Received 28 January 2019; revised 17 February 2020; accepted 13th February 2020; first published online 24 June 2020)

Abstract

Let $\{D_M\}_{M\geqslant 0}$ be the n-vertex random directed graph process, where D_0 is the empty directed graph on n vertices, and subsequent directed graphs in the sequence are obtained by the addition of a new directed edge uniformly at random. For each $\varepsilon > 0$, we show that, almost surely, any directed graph D_M with minimum in- and out-degree at least 1 is not only Hamiltonian (as shown by Frieze), but remains Hamiltonian when edges are removed, as long as at most $1/2 - \varepsilon$ of both the in- and out-edges incident to each vertex are removed. We say such a directed graph is $(1/2 - \varepsilon)$-*resiliently Hamiltonian*. Furthermore, for each $\varepsilon > 0$, we show that, almost surely, each directed graph D_M in the sequence is not $(1/2 + \varepsilon)$-resiliently Hamiltonian.

This improves a result of Ferber, Nenadov, Noever, Peter and Škorić, who showed, for each $\varepsilon > 0$, that the binomial random directed graph $D(n, p)$ is almost surely $(1/2 - \varepsilon)$-resiliently Hamiltonian if $p = \omega(\log^8 n/n)$.

2020 MSC Codes: 05C20, 05C80, 05C38, 05C45

1. Introduction

One of the most studied properties of graphs is that of *Hamiltonicity*, the property that a graph contains a cycle through every vertex, known as a *Hamilton cycle*. The natural extremal function for Hamiltonicity was studied by Dirac [8], whose celebrated theorem demonstrates that any graph with $n \geqslant 3$ vertices and minimum degree at least $n/2$ is Hamiltonian. An early question of Erdős and Rényi [9] in the study of the binomial random graph $G(n, p)$, where edges among n vertices are chosen independently at random with probability p, asked when such a graph is likely to be Hamiltonian. After work by Pósa [27] and by Korshunov [20], this was determined independently by Komlós and Szemerédi [19] and Bollobás [6], who proved that if $p = (\log n + \log\log n + \omega(1))/n$, then $G(n, p)$ is Hamiltonian with probability $1 - o(1)$. We say here that $G(n, p)$ is *almost surely* Hamiltonian. This is best possible, for if $p = (\log n + \log\log n - \omega(1))/n$, then $G(n, p)$ almost surely has vertices of degree 0 or 1, and as such is clearly not Hamiltonian.

In fact, in $G(n, p)$, Hamiltonicity is almost surely concurrent with the property that every vertex has at least two neighbours. This is most precisely shown by the following beautiful result, proved independently by Bollobás [7] and Ajtai, Komlós and Szemerédi [1]. Consider the n-vertex random graph process $G_0, \ldots, G_{\binom{n}{2}}$, where G_0 is an empty graph on n vertices and each subsequent graph in the sequence is formed by the addition of a non-edge uniformly at random. Almost surely, the first graph in the sequence with minimum degree at least 2 is Hamiltonian [1, 7]. Furthermore, we can strengthen this by showing that, in almost every random graph process,

© The Author(s), 2020. Published by Cambridge University Press

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

every graph with minimum degree at least 2 is not only Hamiltonian, but remains so despite the removal of any set of edges, subject only to a simple condition on the edges removed. That is, it is *resiliently* Hamiltonian.

The general study of resilience in random graphs, initiated by Sudakov and Vu [31] in 2008, has developed into an active area of research (see *e.g.* [4, 5, 13, 21, 22, 31] and the survey [30]). We study the resilience of a graph G with respect to some property $\mathcal{P}$ using the following definition.

Definition 1.1. A graph G is *α-resilient with respect to the property $\mathcal{P}$* if, for any $H \subset G$ with $d_H(v) \leqslant \alpha d_G(v)$ for each $v \in V(G)$, $G - H$ has property $\mathcal{P}$.

Note that Dirac's theorem is exactly that the complete graph on $n \geqslant 3$ vertices is $(1/2)$-resiliently Hamiltonian. A natural generalization to random graphs is to ask how resiliently Hamiltonian a typical random graph is. This was the subject of series of results (see [5, 13, 31]), before, in a key breakthrough, Lee and Sudakov [22] showed that, if $p = \omega(\log n/n)$, then $G(n, p)$ is almost surely $(1/2 - o(1))$-resiliently Hamiltonian. Here the constant $1/2$ is best possible as such a random graph can typically be disconnected while removing only $1/2 + o(1)$ of the edges around any one vertex. However, the bound on p can be improved slightly, and the result made best possible by considering the resilience of Hamiltonicity in the random graph process. Indeed, independently, the author [24] and Nenadov, Steger and Trujić [25], showed that in almost every n-vertex random graph process, each Hamiltonian graph is $(1/2 - o(1))$-resiliently Hamiltonian. In this paper we prove the corresponding result for the random *directed* graph process.

A Hamilton cycle in a directed graph (digraph) is a cycle through every vertex whose edges are oriented in the same direction around the cycle. The corresponding result to Dirac's theorem was shown by Ghouila-Houri [14], who proved that every digraph on $n \geqslant 3$ edges with minimum in- and out-degree at least $n/2$ contains a Hamilton cycle. The binomial random digraph $D(n, p)$ has n vertices and each possible edge chosen independently at random with probability p. For each vertex pair u, v, $\overrightarrow{uv}$ and $\overrightarrow{vu}$ may appear in $D(n, p)$. The techniques for studying Hamiltonicity in $G(n, p)$ do not immediately translate to the directed case, but an elegant general coupling argument of McDiarmid [23] shows that, if $p = (\log n + \log\log n + \omega(1))/n$, then $D(n, p)$ is almost surely Hamiltonian. However, the natural local impediment to Hamiltonicity in $D(n, p)$ is that every vertex must have in- and out-degree at least 1. This almost surely holds if $p = (\log n + \omega(1))/n$, and almost surely does not if $p = (\log n - \omega(1))/n$.

Frieze [12] showed that, if $p = (\log n + \omega(1))/n$, then $D(n, p)$ is almost surely Hamiltonian, and gave a corresponding result for the random digraph process. In the n-vertex random digraph process $D_0, D_1, \ldots, D_{n(n-1)}$, D_0 is the empty digraph on n vertices, and each subsequent digraph in the sequence is obtained by the addition of a new directed edge uniformly at random. Frieze [12] showed that, in almost every n-vertex random digraph process, every digraph with minimum in- and out-degree at least 1 is Hamiltonian.

To study resilience in directed graphs, we use the corresponding definition to resilience in graphs, as follows.

Definition 1.2. A directed graph D is *α-resilient with respect to the property $\mathcal{P}$* if, for any $H \subset D$ with $d^j_H(v) \leqslant \alpha d^j_D(v)$ for each $v \in V(D)$ and $j \in \{+, -\}$, $D - H$ has property $\mathcal{P}$.

Hefetz, Steger and Sudakov [17] showed that, if $p \gg \log n/\sqrt{n}$, then $D(n, p)$ is almost surely $(1/2 - o(1))$-resiliently Hamiltonian. As with the undirected case, the constant $1/2$ here is tight, but the bound on p is rather loose. Ferber, Nenadov, Noever, Peter and Škorić [11] showed that, if $p = \omega(\log^8 n/n)$, then $D(n, p)$ is almost surely $(1/2 - o(1))$-resiliently Hamiltonian. Here we will make a best possible improvement to the bound on p, and bring the known resilience of Hamiltonicity in random digraphs into line with that known for random graphs, as follows.

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

Theorem 1.3. *Let $\varepsilon > 0$. In almost every n-vertex random directed graph process $D_0, D_1, \ldots, D_{n(n-1)}$, the following is true for each $0 \leqslant M \leqslant n(n-1)$. If $\delta^{\pm}(D_M) \geqslant 1$, then D_M is $(1/2 - \varepsilon)$-resiliently Hamiltonian, but not $(1/2 + \varepsilon)$-resiliently Hamiltonian.*

Standard techniques easily infer from Theorem 1.3 that, if $p = (\log n + \omega(1))/n$, then $D(n, p)$ is almost surely $(1/2 - o(1))$-resiliently Hamiltonian (see *e.g.* Section 7).

The constant $1/2$ in Theorem 1.3 arises from the following. Almost surely, if $D(n, p)$ has minimum in- and out-degree at least 1, then it can be disconnected into two roughly equal halves by deleting only a little over half of the in- and out-edges at each vertex. This is easy to show when $p = \omega(\log n/n)$, and, with a little care, it is possible to show in almost every random digraph process for each digraph D_M with $\delta^{\pm}(D_M) \geqslant 1$ (see Section 7.6). Thus it is relatively straightforward to demonstrate the limits of the resilience of Hamiltonicity required for Theorem 1.3.

On the other hand, if we remove only at most a $(1/2 - \varepsilon)$-proportion of the in- and out-edges around each vertex, then we cannot disconnect the digraph D. In fact, typically we must retain two key properties. Firstly, if two large equal-sized vertex sets are chosen disjointly at random, then there is likely to be a matching directed from the first into the second. Secondly, given a small collection of pairs of vertices disjoint from a random small vertex subset, we can use the vertex subset to connect the pairs into a directed cycle. The first property allows us, by taking a sequence of random sets, to cover most of a typical random digraph by relatively few directed paths. The second property then allows us to join these paths together into a directed cycle, using a reserved random small set of vertices. This may, of course, not cover all the vertices, and hence we use the *absorbing method*. This is described in detail in Section 2, but, in short, we note that the key behind our progress compared to Ferber *et al.* [11], who used the same broad outline, is in our construction of the *reservoir*. In particular, each vertex in the reservoir is created by contracting a short directed path to create a new vertex. A Hamilton cycle in this altered digraph is found, before the contractions are undone to create a Hamilton cycle in the original digraph. This allows the use of a larger reservoir, and in combination with an adaptation of path connection methods by Glebov, Krivelevich and Johannsen [15] to the directed graph setting (see Section 4), and the careful division of vertex sets using the local lemma (see Section 3), makes the improvements required to show Theorem 1.3.

In the rest of this section we detail our notation. In Section 2 we give a sketch of our proof followed by an outline of the rest of the paper.

Notation

A digraph D has vertex set $V(D)$ and edge set $E(D)$, and we set $|D| = |V(D)|$ and $e(D) = |E(D)|$. For any set $A \subset V(D)$, we set

$$N_D^+(A) = \{v \in V(D) \setminus A \colon \exists u \in A \text{ s.t. } \overrightarrow{uv} \in E(D)\},$$
$$N_D^-(A) = \{v \in V(D) \setminus A \colon \exists u \in A \text{ s.t. } \overrightarrow{vu} \in E(D)\}.$$

We say $N_D^+(A)$ is the out-neighbourhood of A and $N_D^-(A)$ is the in-neighbourhood of A. Where A is a single vertex v, we let $d_D^j(v) = |N_D^j(v)|$ for each $j \in \{+, -\}$. Given a set of edges E, we let $V(E)$ be the set of vertices contained in these edges. For any disjoint vertex sets A and B in a digraph D, and each $j \in \{+, -\}$, $e_D^+(A, B)$ is the number of edges directed from A to B in D, and $e_D^-(A, B) = e_D^+(B, A)$. Where it is clear from context, we often drop the digraph D from the subscript. For a digraph D, $\Delta^-(D)$, $\Delta^+(D)$, $\delta^-(D)$ and $\delta^+(D)$ are the maximum in- and out-degree and the minimum in- and out-degree of D respectively. For any vertex set A in a digraph D, the digraph $D[A]$ has vertex set A and edges exactly those in D contained within A.

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

For convenience, we consider paths to have an inherent order, and thus treat them as an ordered sequence of vertices. In this sequence, we allow vertices to repeat consecutively without consequence. For example, if a path P has start vertex u and end vertex v, then we consider uPv to be the same path as P. Given a path P, $\overleftarrow{P}$ is the path on the same vertices as P but with the vertex order reversed. An *alternating path* is one whose vertices all have either in-degree 0 or out-degree 0 within the path. In a digraph D, for any disjoint vertex sets A and B, a *matching from A into B* is a set of $|A|$ independent edges oriented from A into B.

Where we use $\pm$ in an expression, we mean that this holds with $\pm$ replaced by both $+$ and $-$. We use log for the natural logarithm and, for each $k \geqslant 2$, we use $\log^{[k]} n$ to refer to the kth iterated logarithm of n, so that, for example, $\log^{[3]} n = \log\log\log n$. For each integer k, we let $[k] = \{1, \ldots, k\}$.

We use common asymptotic notation to relate functions of n, as follows. If $f = O(g)$ or $g = \Omega(f)$, then there exists a constant C such that $f(n) \leqslant Cg(n)$ for every $n \in \mathbb{N}$. When the implicit constant C depends on ε, we will denote this in the subscript, using, for example $f = O_\varepsilon(g)$. If $f = \omega(g)$ or $g = o(f)$, for a non-zero function g, then $f(n)/g(n) \to \infty$ as $n \to \infty$. When, for example, $\Omega(f)$ is used in expressions, we mean that this can be replaced by some function $g = \Omega(f)$ so that the expression holds. Many of our lemmas hold for $n \geqslant n_0(\varepsilon)$, for some function n_0 depending on ε. In the proofs we do not mention this explicitly, but only note that we take 'n sufficiently large' where our argument requires n to be large. Similarly, when $f = \omega(g)$ or $f = o(g)$ we mean that this is true for each fixed ε. If $f = O(g)$ and $g = O(f)$, then we say that $f = \Theta(g)$.

For clarity of presentation we do not include floor and ceiling symbols where they are not crucial.

2. Outline and proof of a key lemma

2.1 Proof sketch and outline

Pseudorandom digraphs. We will build a Hamilton cycle in any sufficiently large digraph which satisfies certain pseudorandom properties, before showing that random digraphs resiliently contain such a digraph. These properties are defined precisely in Definition 2.1, but, roughly speaking, they say the following.

- The minimum and maximum in- and out-degrees are bounded (see **A1**).
- Small sets with many incident edges expand well (see **A2** and **A3**).
- Medium-sized sets expand to more than one-half of the vertex set (see **A4**).

In our more informal discussion, we say that a set expands if its in- and out-neighbourhood is comfortably larger than the set itself. The exact parameters of the expansion we use are found in Definition 2.1.

The first two properties listed above are naturally resilient (if the minimum degree bounds are reduced by an appropriate factor). The third condition is naturally almost surely $(1/2 - o(1))$-resilient in $D(n, p)$ if $p = \omega(1/n)$. Typically, here, medium-sized sets will expand to almost all of the vertex set. Then removing at most $1/2 - \varepsilon$ of the in- and out-degrees around any vertex will only reduce the size of the in- and out-neighbourhood by at most a factor of $1/2 - \varepsilon/2$, so the third condition holds resiliently.

Boosting the minimum degrees. As we consider every Hamiltonian digraph in the random digraph process, we work with digraphs with very low minimum in- or out-degree. However, we use a natural modification to increase the minimum degree when there are a small number of vertices with low in- or out-degree. After the removal of edges, we take each low-degree vertex and assign it both an in- and out-neighbour, before deleting the low-degree vertex and merging its

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

assigned in-neighbour into its assigned out-neighbour (see Definition 2.5). This creates the pseudorandom digraph in which we find a Hamilton cycle. Taking this cycle, undoing the merging, and putting the low-degree vertices between their assigned neighbours, creates a Hamilton cycle in the original digraph.

Hamilton cycles in pseudorandom digraphs. We create a Hamilton cycle in a pseudorandom digraph D using the same broad outline as Ferber *et al.* [11]. We use the *absorbing method*, first given as a general method by Rödl, Ruciński and Szemerédi [28]. We find a directed path P in D in combination with a reservoir R in $V(D) \setminus V(P)$, so that, given any subset of vertices $R' \subset R$, we can find a directed path with vertex set $V(P) \cup R'$ and the same start and end vertices as P. Dividing the vertices $V(D) \setminus (V(P) \cup R)$ in the digraph into equal-sized sets at random, we find matchings between them to create a small number of directed paths which cover the remaining vertices. Using vertices in R, we then join these paths into a directed cycle with P – say the cycle is Q. This gives a cycle covering all the vertices except for $R \setminus V(Q)$. Using the absorbing property we then find a path with vertex set $V(P) \cup (R \setminus V(Q))$ and the same end vertices as P, and replace P with this path in Q to get a Hamilton cycle.

The improvements we make from the methods of Ferber *et al.* [11] come from three areas, as follows.

- We use a more efficient absorbing structure so that the reservoir may have size $\Omega(n \log^{[2]} n / \log n)$. Our reservoir in fact consists of disjoint directed paths, not vertices. We contract these paths into vertices in the obvious manner, and use these vertices as the reservoir. We then find a Hamilton cycle in the modified digraph, before replacing each contracted vertex by its original path to get a Hamilton cycle in the original digraph.
- To construct absorbers and join paths into a cycle we develop and use a directed graph version of some path connection techniques by Glebov, Krivelevich and Johannsen [15].
- We use the local lemma to randomly partition the vertex set into subsets and find matchings between them.

The first area represents the major innovation of this paper, while the subsequent two areas require quite a few technicalities. Due to this, we structure the paper so that the most important part of the argument appears first, in the rest of this section.

Outline. In the rest of this section we define our notion of pseudorandomness precisely, before defining a *good partition*. We then prove a key lemma, Lemma 2.7, that says any digraph with a good partition is Hamiltonian. This allows us to give the most important part of our argument, before embarking on the more technical aspects. Finally, we cover some useful results from the literature.

In Section 3 we give our use of the local lemma to find useful vertex partitions. In Section 4 we give a digraph version of techniques by Glebov, Krivelevich and Johannsen [15] for finding connecting paths. In Section 5 we divide the vertex set into subsets and find matchings between them in order to cover most of the digraph with a small number of directed paths. In Section 6 we combine this all to show that any sufficiently large pseudorandom digraph has a good partition. Finally, in Section 7, we find the pseudorandom properties in a random digraph needed to prove Theorem 1.3, and also show the promised limits of resilience.

2.2 Pseudorandom digraphs and good partitions

We will begin by defining a (d, ε)-*pseudorandom* digraph and a *good partition* of a digraph. A good partition is defined essentially as one with the properties needed to carry through our construction of a directed Hamilton cycle. On the other hand, the properties of a pseudorandom digraph more

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

naturally reflect those of a typical random digraph. For example, for each $\varepsilon > 0$, if $p = \omega(\log n/n)$, then $D(n, p)$ is typically (d, ε)-pseudorandom with $d = pn/2 \log n$.

Definition 2.1. An n-vertex digraph D is (d, ε)-*pseudorandom* if the following properties hold with $m = n \log^{[3]} n/d \log n$.

- **A1** $\delta^{\pm}(D) \geqslant d \log n$ and $\Delta^{\pm}(D) \leqslant 10^6 d \log n$.
- **A2** For each $j \in \{+, -\}$ and any disjoint sets $A, B \subset V(D)$, with $|A| \leqslant 2m$, and, for each $v \in A$, $d^j(v, B) \geqslant d \log^{[2]} n/\log^{[4]} n$, we have $|B| \geqslant 10|A|$.
- **A3** For each $j \in \{+, -\}$ and any disjoint sets $A, B \subset V(D)$, with $|A| \leqslant 2m$, and, for each $v \in A$, $d^j(v, B) \geqslant d(\log n)^{2/3}$, we have $|B| \geqslant (\log n)^{1/3}|A|$.
- **A4** Every set $A \subset V(D)$ with $|A| = m$ satisfies $|N^{\pm}(A)| \geqslant (1/2 + \varepsilon)n$.

We will show that every sufficiently large pseudorandom digraph is Hamiltonian, as follows.

Theorem 2.2. *For each $\varepsilon > 0$, there exists some $n_0 = n_0(\varepsilon)$ such that, for every $d \geqslant 10^{-5}$, every (d, ε)-pseudorandom digraph with at least n_0 vertices is Hamiltonian.*

To show this, we will show that any sufficiently large pseudorandom digraph has a *good partition* (see Lemma 6.2). This definition requires that directed cycles are found through particular edges. For this, we define the following weak and strong connection properties, where the key difference is that the latter property allows a cycle to be found through particular edges *in a given order*.

Definition 2.3. A vertex set U in a digraph D is *weakly connected* if, for any independent set E of directed edges in the complete digraph with vertex set U, there is a directed cycle in $D + E$ which contains every edge in E.

Definition 2.4. A vertex set U in a digraph D is *strongly connected* if, for any ℓ and any independent set $E = \{e_1, \ldots, e_\ell\}$ of directed edges in the complete digraph with vertex set U, there is a directed cycle in $D + E$ which contains the edges $e_1, \ldots, e_\ell$ in that order.

We also merge vertices using the following definition.

Definition 2.5. In a digraph D, we merge a vertex x into a vertex y by deleting x and y and creating a new vertex z with in-neighbourhood $N_D^-(x) \setminus \{y\}$ and out-neighbourhood $N_D^+(y) \setminus \{x\}$.

Using these definitions, we define a good partition as follows.

Definition 2.6. In a digraph D, a vertex partition $V(D) = A \cup B_1 \cup B_2 \cup R_1 \cup R_2 \cup R_3 \cup R_4$ is an (ℓ, r)-*good partition* if the following hold.

- **B1** Any set $U \subset V(D) \setminus A$ with $|U| \leqslant 4r$ is strongly connected in $D[A \cup U]$.
- **B2** If $B' \subset V(D)$ contains $B_1 \cup B_2$, and $u, v \in B' \setminus (B_1 \cup B_2)$ with $u \neq v$, then there is a collection of at most ℓ disjoint directed paths with length at least 1 in $D + \overrightarrow{uv}$ which cover B' exactly, each start and end in B_2, and one of which contains the edge $\overrightarrow{uv}$.
- **B3** There are matchings M_1, M_2 and M_3 from R_2 into R_1, R_2 into R_3 and R_4 into R_3 in D, respectively, and $|R_i| = r$ for each $i \in [4]$, so that the following holds.
- **B4** Let $f\colon R_1 \to R_4$ be such that, for each $v \in R_1$, $f(v)$ and v are the end vertices of an alternating path in $M_1 \cup M_2 \cup M_3$. Merge each vertex $v \in R_1$ into $f(v)$ in D to get the digraph D'. Let R

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

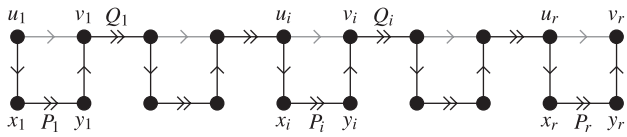

Figure 1. The directed path in (2.1), with the additional edges $\overrightarrow{u_iv_i}$, $i \in [r]$, in grey.

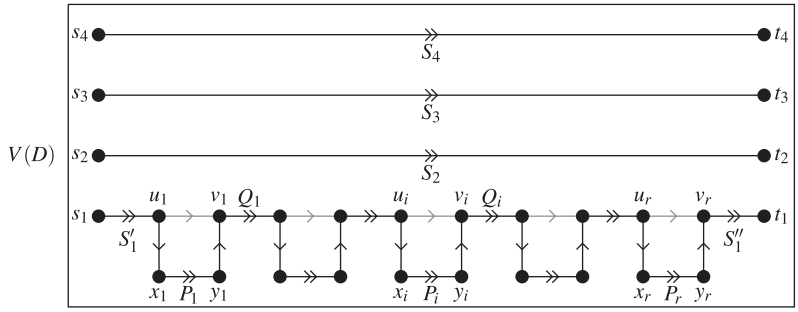

Figure 2. The partition of $V(D)$ given by the paths in (2.3). It remains to connect the s_i, t_i-paths in some order using some of the paths $x_jP_jy_j, j \in [r]$, to find a directed Hamilton cycle.

be the set of merged vertices in D'. Then any set $U \subset B_2$ with $|U| \leqslant 2\ell$ is weakly connected in $D'[R \cup U]$.

If a digraph has an (ℓ, r)-good partition for some $\ell, r > 0$, then we say D *has a good partition*.

We now give the main part of our argument, showing that any digraph with a good partition is Hamiltonian.

Lemma 2.7. *Any digraph with a good partition is Hamiltonian.*

Proof. Let D be a digraph and let $V(D) = A \cup B_1 \cup B_2 \cup R_1 \cup R_2 \cup R_3 \cup R_4$ be an (ℓ, r)-good partition of D, for some integers $\ell, r > 0$. Using **B3**, find a matching from R_2 into R_1, R_2 into R_3 and R_4 into R_3, so that **B4** holds. Use these matchings to label vertices so that $R_1 = \{x_1, \dots, x_r\}$, $R_2 = \{u_1, \dots, u_r\}$, $R_3 = \{v_1, \dots, v_r\}$ and $R_4 = \{y_1, \dots, y_r\}$, and, for each $i \in [r]$, $\overrightarrow{u_ix_i}, \overrightarrow{u_iv_i}, \overrightarrow{y_iv_i} \in E(D)$.

By **B1** applied with $E = \{\overrightarrow{u_ix_i}, \overrightarrow{y_iv_i} \colon i \in [r]\}$, we can find disjoint directed paths P_i and Q_i, $i \in [r-1]$, and P_r, in $D[A \cup V(E)]$ so that

$$u_1x_1P_1y_1v_1Q_1u_2x_2P_2y_2v_2Q_2 \cdots u_rx_rP_ry_rv_r \tag{2.1}$$

is a directed u_1, v_r-path in D (see Figure 1). Note that, for each $i \in [r]$, if $x_iP_iy_i$ is removed from the path in (2.1), then, as $\overrightarrow{u_iv_i} \in E(D)$, this is still a directed u_1, v_r-path in D.

Let $R = R_1 \cup R_2 \cup R_3 \cup R_4$ and note that every vertex in R appears in (2.1). Let A' be the set of vertices in A not appearing in (2.1). By **B2**, we can find some $m \in [\ell]$ and vertices and directed paths $s_iS_it_i$, $i \in [m]$, so that $\{s_iS_it_i \colon i \in [m]\}$ is a set of disjoint directed paths in $D + \overrightarrow{u_1v_r}$ with length at least 1 which exactly covers $A' \cup B_1 \cup B_2 \cup \{u_1, v_r\}$ and for which $\{s_i, t_i \colon i \in [m]\} \subset B_2$, and so that S_1 contains the edge $\overrightarrow{u_1v_r}$. Say that $S_1 = S_1'u_1v_rS_1''$. Thus the following set of paths forms a partition of $A' \cup B_1 \cup B_2 \cup \{u_1, v_r\}$:

$$\{s_1S_1'u_1\} \cup \{v_rS_1''t_1\} \cup \{s_iS_it_i \colon 2 \leqslant i \leqslant m\}. \tag{2.2}$$

Furthermore, then, the following paths form a partition of $V(D)$ (as depicted in Figure 2):

$$\{s_1S_1'u_1\} \cup \{v_rS_1''t_1\} \cup \{s_iS_it_i \colon 2 \leqslant i \leqslant m\} \cup \{x_iP_iy_i \colon i \in [r]\} \cup \{v_iQ_iu_{i+1} \colon 1 \leqslant i < r\}. \tag{2.3}$$

Let D' be the digraph formed from D by, for each $i \in [r]$, merging x_i into y_i to get the vertex z_i. Let $R' = \{z_i \colon i \in [r]\}$. Let $E' = \{\overrightarrow{s_it_i} \colon i \in [m]\}$. By **B4**, there are disjoint directed paths $T_1, \dots, T_m$ in

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

$D'[R' \cup V(E')]$ and a bijection $g\colon [m] \to [m]$ with $g(1) = 1$ so that

$$t_1 T_1 s_{g(2)} t_{g(2)} T_2 s_{g(3)} t_{g(3)} T_3 \cdots s_{g(m)} t_{g(m)} T_m s_1$$

is a directed path in $D'[R' \cup V(E')] + E'$.

For each $i \in [m]$, replace each vertex z_j, $j \in [r]$, of T_i with the corresponding directed path $x_j P_j y_j$ and call the resulting path T'_i. Note that, from the definition of D', T'_i is a directed path in D. Thus the path

$$t_1 T'_1 s_{g(2)} t_{g(2)} T'_2 s_{g(3)} t_{g(3)} T'_3 \cdots s_{g(m)} t_{g(m)} T'_m s_1$$

is a directed path in $D + E'$. Replacing each edge $s_i t_i$, $2 \leqslant i \leqslant m$ with the directed path $s_i S_i t_i$, and adding the paths $v_r S''_1 t_1$ and $s_1 S'_1 u_1$, we get that

$$C_1 \coloneqq v_r S''_1 t_1 T'_1 s_{g(2)} S_{g(2)} t_{g(2)} T'_2 s_{g(3)} t_{g(3)} T'_3 \cdots s_{g(m)} S_{g(m)} t_{g(m)} T'_m s_1 S'_1 u_1$$

is a directed path in D. As the paths in (2.2) form a partition of $A' \cup B_1 \cup B_2 \cup \{u_1, v_r\}$, C_1 has vertex set $A' \cup B_1 \cup B_2 \cup \{u_1, v_r\}$ with some sets $V(x_j P_j y_j)$ added (those appearing in some path T'_i).

For each $j \in [r]$, if $x_j P_j y_j$ is contained in C_1, then let Z_j be the empty path on no vertices, and otherwise let Z_j be $x_j P_j y_j$. Thus the path

$$C_2 \coloneqq u_1 Z_1 v_1 Q_1 u_2 Z_2 v_2 Q_2 \cdots u_r Z_r v_r$$

is the path in (2.1) with some paths $x_i P_i y_i$, $i \in [r]$, removed, which, by construction, is a directed path. The path C_2 contains exactly the vertices in the path in (2.1) except for those appearing in C_1, as well as u_1 and v_r. Thus, as the path C_1 contains all the vertices not in the path (2.1), the two paths C_1 and C_2 form a cycle with vertex set $V(D)$, as required. □

2.3 Preliminaries

We will use the following well-known version of Chernoff's lemma (see *e.g.* [18, Corollary 2.3]).

Lemma 2.8. *If X is a binomial variable with standard parameters n and p, denoted $X = \mathrm{Bin}(n, p)$, and ε satisfies $0 < \varepsilon \leqslant 3/2$, then*

$$\mathbb{P}(|X - \mathbb{E}X| \geqslant \varepsilon \mathbb{E}X) \leqslant 2 \exp(-\varepsilon^2 \mathbb{E}X/3). \qquad \square$$

To find a matching from one set into another, we will use the following simple proposition (for undirected graphs).

Proposition 2.9. *Let G be a bipartite graph with vertex classes A and B with size n each, such that, for each $U \subset A$ or $U \subset B$ with $|U| \leqslant \lceil n/2 \rceil$, $|N(U)| \geqslant |U|$. Then there is a matching from A to B in G.*

Proof. Let $U \subset A$ with $\lceil n/2 \rceil < |U| \leqslant n$. By considering a subset $U' \subset U$ with size $\lceil n/2 \rceil$, we have that $|N(U)| \geqslant |N(U')| \geqslant \lceil n/2 \rceil$. Thus $|B \setminus N(U)| \leqslant n - \lceil n/2 \rceil = \lfloor n/2 \rfloor$, so that, by the property in the lemma, $|N(B \setminus N(U))| \geqslant |B \setminus N(U)|$. Thus, as there are no edges between U and $B \setminus N(U)$, we have

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

$$|U| \leqslant n - |N(B \setminus N(U))| \leqslant n - |B \setminus N(U)| = n - (n - |N(U)|),$$

and therefore $|N(U)| \geqslant |U|$. By the property in the lemma, this is also true for all $U \subset A$ with $|U| \leqslant \lceil n/2 \rceil$. Thus Hall's matching condition is satisfied, and there is a matching from A into B. □

3. Set division using the local lemma

We will take vertex partitions using the following version of the Erdős–Lovász local lemma, due to Lovász [29, Theorem 1.1].

Theorem 3.1. *Let $A_1, \ldots, A_n$ be events in a probability space Ω with dependence graph G. Suppose there exist $0 < x_1, \ldots, x_n < 1$ such that, for each $i \in [n]$,*

$$\mathbb{P}(A_i) \leqslant x_i \prod_{j: ij \in E(G)} (1 - x_j).$$

Then no such event A_i occurs with strictly positive probability.

Throughout the following lemma we use Theorem 3.1 as follows. Given a vertex set A in a digraph, where every vertex has plenty of in- and out-neighbours in A, we partition A so that every vertex has at least some in- and out-neighbours in each subset in the partition. We use this in a similar manner to Hefetz, Krivelevich and Szabó [16] in their work on the sharp threshold of certain spanning trees in $G(n, p)$.

Lemma 3.2. *Let $\ell, m, n, \delta, \Delta \in \mathbb{N}$ with $2 \leqslant \ell \leqslant \log n$, and let $\varepsilon \in (0, 1)$. Let D be a digraph with n vertices and $A \subset V(D)$ so that the following hold.*

C1 *For each $v \in V(D)$, $d^{\pm}(v, A) \geqslant \delta$ and $d^{\pm}(v) \leqslant \Delta$.*
C2 *For each $U \subset V(D)$ with $|U| = m$, $|N^{\pm}(U, A)| \geqslant (1/2 + \varepsilon)|A|$.*

Let $a = |A|$, and suppose that a_i, $i \in [\ell]$, are integers with $\sum_{i \in [\ell]} a_i \leqslant a$ such that the following hold for each $i \in [\ell]$.

D1 $\dfrac{\varepsilon^2 a_i^2}{a} \geqslant 10^3 \ell^3.$
D2 $\dfrac{\varepsilon^2 a_i^2}{a} \cdot \exp\left(\dfrac{a_i \delta}{24a}\right) \geqslant 10^5 \ell^3 n.$
D3 $\exp\left(\dfrac{a_i \delta}{24a}\right) \geqslant 320 \ell \Delta^2.$
D4 $\dfrac{\varepsilon^2 a_i}{10^3} \geqslant m \log\left(\dfrac{en}{m}\right).$

Then there are disjoint sets $A_1, \ldots, A_\ell \subset A$ such that $|A_i| = a_i$ for each $i \in [\ell]$ and the following hold.

E1 *For each $v \in V(D)$ and $i \in [\ell]$, $a_i \delta / 4a \leqslant d^{\pm}(v, A) \leqslant 4 a_i \Delta / a$.*
E2 *For each $U \subset V(D)$, with $|U| = m$, and $i \in [\ell]$, $|N^{\pm}(U, A_i)| \geqslant (1/2 + \varepsilon/2) a_i$.*

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

Proof. Without loss of generality, assume that $a_1 \geqslant a_2 \geqslant \cdots \geqslant a_\ell$. Let $p = \varepsilon a_\ell / 10\ell a$. For each $i \in [\ell]$, let

$$p_i = (a_i/a) - p \geqslant (1 - \varepsilon/10)a_i/a. \tag{3.1}$$

Let

$$p_0 = \min\left\{1 - \sum_{i\in[\ell]} p_i, (\ell+1)p\right\} \leqslant (\ell+1)p \leqslant \frac{\varepsilon a_\ell}{5a} \overset{(3.1)}{\leqslant} p_\ell. \tag{3.2}$$

Noting that $\sum_{i=0}^{\ell} p_i \leqslant 1$, pick random disjoint sets

$$B_0, B_1, B_2, \ldots, B_\ell \subset A$$

so that, for each vertex $v \in V(D)$, $\mathbb{P}(v \in B_i) = p_i$ for each $0 \leqslant i \leqslant \ell$, and whether v appears in one of the sets, and which set it appears in, is independent of the location of all the other vertices in D.

We will show, using Theorem 3.1, that with positive probability the partition satisfies the following properties.

F1 For each $i \in [\ell]$, $|B_i| \leqslant a_i$, and $|\cup_{i=0}^{\ell} B_i| \geqslant \sum_{i\in[\ell]} a_i$.
F2 Every subset $U \subset V(D)$ with $|U| = m$ satisfies $|N^{\pm}(U, B_i)| \geqslant (1/2 + \varepsilon/2)a_i$ for each $i \in [\ell]$.
F3 For each $v \in V(D)$ and $i \in [\ell]$, $a_i\delta/4a \leqslant d^{\pm}(v, B_i) \leqslant 2a_i\Delta/a$.
F4 For each $v \in V(D)$, $d^{\pm}(v, B_0) \leqslant 2a_\ell\Delta/a$.

This will be sufficient to prove the lemma. Indeed, there will thus exist some partition in which **F1**–**F4** hold. Then let $A_1, \ldots, A_\ell \subset A$ be disjoint subsets such that $B_i \subset A_i \subset B_i \cup B_0$ and $|A_i| = a_i$ hold for each $i \in [\ell]$; this is possible by **F1**. As **E1** follows from **F3** and **F4**, and **E2** follows from **F2**, we have the required partition.

Then let B be the event that **F1** or **F2** does not hold. For each $v \in V(D)$, let $B(v)$ be the event that **F3** or **F4** does not hold for v. Let

$$q_B = 1/2 \quad \text{and} \quad q = 40\ell \exp(-a_\ell\delta/24a). \tag{3.3}$$

Note that each event $B(v)$ has some dependence on B and at most $4\Delta^2$ other events $B(v')$.

We will show the following two claims.

Claim 1. *For each $v \in V(D)$, $\mathbb{P}(B(v)) \leqslant q(1 - q_B)(1 - q)^{4\Delta^2}$.*

Claim 2. $\mathbb{P}(B) \leqslant q_B(1 - q)^n$.

Thus, by Theorem 3.1 and Claims 1 and 2, with positive probability some partition exists for which no event B or $B(v)$, $v \in V(D)$, holds, and thus for which **F1**–**F4** hold, as required. It remains then to prove the two claims.

Proof of Claim 1. Let $v \in V(D)$, $j \in \{+, -\}$ and $d = d^j(v, A)$, so that $\delta \leqslant d \leqslant \Delta$, by **C1**. For each $i \in [\ell]$, noting that $a_i/2a \leqslant p_i \leqslant a_i/a$, by (3.1), we have, using Lemma 2.8,

$$\begin{aligned}\mathbb{P}(d^j(v, B_i) \notin (a_i\delta/4a, 2a_i\Delta/a)) &\leqslant \mathbb{P}(|d^j(v, B_i) - p_i d| > p_i d/2)\\ &\leqslant 2\exp(-p_i d/12)\\ &\leqslant 2\exp(-a_i\delta/24a)\\ &\leqslant 2\exp(-a_\ell\delta/24a).\end{aligned} \tag{3.4}$$

As $p_0 \leqslant p_\ell$, by (3.2), we also have

$$\mathbb{P}(d^j(v, B_0) > 2a_\ell\Delta/a) \leqslant \mathbb{P}(d^j(v, B_\ell) > 2a_\ell\Delta/a) \overset{(3.4)}{\leqslant} 2\exp(-a_\ell\delta/24a).$$

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

Thus, for each $v \in V(D)$, we have

$$\mathbb{P}(B(v)) \leqslant 4(\ell+1)\exp(-a_\ell\delta/24a) \leqslant q/5.$$

By **D3**, we have that $4q\Delta^2 \leqslant 1/2$. Thus, as $q_B = 1/2$,

$$\mathbb{P}(B(v)) \leqslant q\cdot(1-q_B)\cdot(1-4q\Delta^2) \leqslant q\cdot(1-q_B)\cdot(1-q)^{4\Delta^2}. \qquad \square$$

We will prove Claim 2 using two further claims.

Claim 3. $\mathbb{P}($***F1*** *holds*$) \leqslant 4\ell\exp(-\varepsilon^2 a_\ell^2/300\ell^2 a)$.

Proof of Claim 3. By (3.1) and Lemma 2.8, for each $i \in [\ell]$,

$$\begin{aligned}\mathbb{P}(|B_i| > a_i) &= \mathbb{P}(|B_i| - p_i a > pa)\\ &= \mathbb{P}(|B_i| - p_i a > (p/p_i)\cdot p_i a)\\ &\leqslant 2\exp(-(p/p_i)^2\cdot p_i a/3)\\ &= 2\exp(-p^2 a/3p_i)\\ &\leqslant 2\exp(-p^2 a/3)\\ &= 2\exp(-\varepsilon^2 a_\ell^2/300\ell^2 a). \end{aligned} \tag{3.5}$$

Note that, for each $v \in A$, by (3.1) and (3.2),

$$\mathbb{P}(v \in \cup_{i=0}^{\ell} B_i) = \sum_{i=0}^{\ell} p_i = \min\left\{1, \left(\sum_{i\in[\ell]} \frac{a_i}{a}\right) + p\right\} =: \bar{p}.$$

If $\bar{p} = 1$, then, with probability 1,

$$|\cup_{i=0}^{\ell} B_i| = a \geqslant \sum_{i\in[\ell]} a_i,$$

as required, so assume that $\bar{p} = (\sum_{i\in[\ell]} a_i/a) + p$. Then, by Lemma 2.8,

$$\begin{aligned}\mathbb{P}\left(|\cup_{i=0}^{\ell} B_i| < \sum_{i\in[\ell]} a_i\right) &= \mathbb{P}(|\cup_{i=0}^{\ell} B_i| - \bar{p}a < -pa)\\ &= \mathbb{P}(|\cup_{i=0}^{\ell} B_i| - \bar{p}a < -p(\bar{p}a)/\bar{p})\\ &\leqslant 2\exp(-(p/\bar{p})^2\cdot(\bar{p}a)/3)\\ &\leqslant 2\exp(-p^2 a/3)\\ &= 2\exp(-\varepsilon^2 a_\ell^2/300\ell^2 a). \end{aligned} \tag{3.6}$$

Thus the claim follows from (3.6) and, for each $i \in [\ell]$, (3.5). $\square$

Claim 4. *For each $U \subset V(D)$ with $|U| = m$, $i \in [\ell]$, and $j \in \{+,-\}$,*

$$\mathbb{P}(|N^j(U, B_i)| < (1/2+\varepsilon/2)a_i) \leqslant 2\exp(-\varepsilon^2 a_i/400).$$

Proof of Claim 4. Let $U \subset V(D)$ with $|U| = m$, $i \in [\ell]$, and $j \in \{+,-\}$, and take $U' = N^j(U, A)$. By **C2**, $|U'| \geqslant (1/2+\varepsilon)a$, so that

$$\mathbb{E}|U' \cap B_i| \overset{(3.1)}{\geqslant} (1/2+\varepsilon)a\cdot(1-\varepsilon/10)a_i/a \geqslant (1/2+3\varepsilon/4)a_i \geqslant \frac{1/2+\varepsilon/2}{1-\varepsilon/8}a_i.$$

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

Thus we have both $\mathbb{E}|U' \cap B_i| \geqslant a_i/2$ and $(1/2+\varepsilon/2)a_i \leqslant (1-\varepsilon/8)\mathbb{E}|U' \cap B_i|$. Therefore, by Lemma 2.8,

$$\begin{aligned}\mathbb{P}(|U' \cap B_i| < (1/2+\varepsilon/2)a_i) &\leqslant 2\exp\left(-(\varepsilon/8)^2 \cdot \mathbb{E}|U' \cap B_i|/3\right)\\ &\leqslant 2\exp\left(-(\varepsilon/8)^2 \cdot a_i/6\right)\\ &\leqslant 2\exp\left(-\varepsilon^2 a_i/400\right).\end{aligned}$$

□

With these two claims, we can now prove Claim 2, as follows.

Proof of Claim 2. By Claim 3 and Claim 4, and taking into account that there are at most $\binom{n}{m} \leqslant (en/m)^m$ subsets with size m of $V(D)$, we have

$$\begin{aligned}\mathbb{P}(B) &\leqslant 4\ell \exp\left(-\frac{\varepsilon^2 a_\ell^2}{300\ell^2 a}\right) + 4\ell \cdot \left(\frac{en}{m}\right)^m \cdot \exp\left(-\frac{\varepsilon^2 a_\ell}{400}\right)\\ &\overset{\mathbf{D4}}{\leqslant} 4\ell \exp\left(-\frac{\varepsilon^2 a_\ell^2}{300\ell^2 a}\right) + 4\ell \exp\left(-\frac{\varepsilon^2 a_\ell}{10^3}\right)\\ &\leqslant 8\ell \exp\left(-\frac{\varepsilon^2 a_\ell^2}{300\ell^2 a}\right).\end{aligned}$$

Note that, by (3.3) and **D3**, $q \leqslant 1/2$. Therefore, as $q_B = 1/2$, we have $q_B(1-q)^n \geqslant e^{-2qn}/2$. Thus, for Claim 2, it is sufficient to show that

$$8\ell \exp\left(-\frac{\varepsilon^2 a_\ell^2}{300\ell^2 a}\right) \leqslant e^{-2qn}/2,$$

or, equivalently,

$$\frac{\varepsilon^2 a_\ell^2}{300\ell^2 a} - \log(16\ell) \geqslant 2qn.$$

By **D1**, then, it is sufficient to show that

$$\frac{\varepsilon^2 a_\ell^2}{300\ell^2 a} \geqslant 4qn.$$

However, as $q = 40\ell \exp(-a_\ell \delta/24a)$, this holds directly from **D2**. □

4. Path connection in pseudorandom digraphs

In order to connect edges into a cycle, we develop a directed version of techniques of Glebov, Krivelevich and Johannsen [15]. In [15], a concept of (d, m)-extendability is defined and used to find trees in any graph with certain expansion properties. In short, when a tree S is (d, m)-extendable in a graph G and $v \in V(S)$, then, subject to certain simple conditions, we can add a leaf to v in S so that the subsequent tree remains (d, m)-extendable. As shown in [15], this gives a flexible framework for embedding trees, but also allows paths to be found between vertex sets in (d, m)-extendable graphs. This latter property is what we want, except we will adapt this to work with directed graphs.

To do this, we work with two bipartite graphs, H_1 and H_2 say, with the same vertex classes, A_1 and A_2 say. When we apply the results to a digraph D, H_1 will typically be the edges in D directed from A_1 into A_2 (with the directions removed) while H_2 will be the edges in D directed from A_2 into A_1. If the edges of a path alternate between H_1 and H_2 then this will be a directed path in the digraph.

To define our version of extendability we need the following definition.

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

Definition 4.1. Given a forest S, an edge $e \in E(S)$ and a vertex set $X \subset V(S)$ with exactly one vertex in each tree in S, let $d(e, X)$ be the distance of the shortest path from any vertex in e to any vertex in X.

Our extendable subgraph S will be a forest lying in the union of the two bipartite graphs H_1 and H_2 mentioned above. We use the set X, with exactly one vertex per component tree of S, to record which edges of S are in H_1 and which are in H_2. The edges will alternate between H_1 and H_2 working out from the vertex in X in each component of S. The choice of H_1 and H_2 we later use, arising from a digraph D as mentioned above, will imply that S with the edges of each component tree directed out from the vertex in X will lie in D.

We define extendability in a pair of bipartite graphs as follows.

Definition 4.2. Suppose that H_1 and H_2 are two bipartite graphs with the same vertex classes A_1 and A_2, and that S is a forest containing $X \subset V(S) \cap A_1$ in which exactly one vertex in each tree in S is in X. For each $i \in [2]$, let S_i be the subgraph of S with edge set $\{e \in E(S) \colon d(e, X) \equiv i + 1 \mod 2\}$.

Then we say (S, X) is *(d, m)-extendable in* (H_1, H_2) if the following hold.

G1 $\Delta(S) \leqslant d$.
G2 For each $i \in [2]$, $S_i \subset H_i$.
G3 For each $i \in [2]$ and $U \subset A_i$ with $0 < |U| \leqslant 2m$,

$$|N_{H_i}(U) \setminus V(S)| \geqslant d|U| - e_{S_i}(U, A_{3-i}).$$

G4 For each $i \in [2]$ and $U \subset A_i$ with $|U| \geqslant m$, $|N_{H_i}(U)| \geqslant |A_{3-i}|/2$.

Given a (d, m)-extendable forest that is not too large, we can add an edge to any vertex with degree less than d in the forest while remaining (d, m)-extendable, as follows.

Lemma 4.3. *Let $d \geqslant 2$ and $m \geqslant 1$. Suppose that H_1 and H_2 are two bipartite graphs with the same vertex classes A_1 and A_2, and that S is a forest containing $X \subset V(S) \cap A_1$ in which exactly one vertex in each tree in S is in X. Suppose that (S, X) is (d, m)-extendable in (H_1, H_2), and*

$$|S| \leqslant \min\{|A_1|, |A_2|\}/2 - 2dm - 2. \tag{4.1}$$

Then, for each $i \in [2]$ and $x \in V(S) \cap A_i$ with $d_S(x) < d$, there exists some $y \in N_{H_i}(x) \setminus V(S)$ so that $(S + xy, X)$ is (d, m)-extendable in (H_1, H_2).

Proof. Suppose, to the contrary, that there is some $i \in [2]$ and $x \in V(S) \cap A_i$ with $d_S(x) < d$ for which no such y exists. For each $y \in N_{H_i}(x) \setminus V(S)$, **G1**, **G2** and **G4** for $(S + xy, X)$ to be (d, m)-extendable in (H_1, H_2) hold directly from the same statements for the extendability of (S, X) and as $y \in N_{H_i}(x) \setminus V(S)$. Furthermore, as $V(S + xy) \cap A_i = V(S) \cap A_i$, **G3** for the index $3 - i$ holds for $(S + xy, X)$.

Therefore, for each $y \in N_{H_i}(x) \setminus V(S)$, there is some set $U_y \subset A_i$ such that $|U_y| \leqslant 2m$ and

$$|N_{H_i}(U_y) \setminus V(S + xy)| < d|U_y| - e_{S_i+xy}(U_y, A_{3-i}). \tag{4.2}$$

From simple set relations and the extendability of (S, X) in (H_1, H_2), we have, for each $y \in N_{H_i}(x) \setminus V(S)$, that

$$\begin{aligned} |N_{H_i}(U_y) \setminus V(S + xy)| + \mathbf{1}_{\{y \in N_{H_i}(U_y)\}} &= |N_{H_i}(U_y) \setminus V(S)| \\ &\overset{\textbf{G3}}{\geqslant} d|U_y| - e_{S_i}(U_y, A_{3-i}) \\ &= d|U_y| - e_{S_i+xy}(U_y, A_{3-i}) + \mathbf{1}_{\{x \in U_y\}}. \end{aligned} \tag{4.3}$$

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

Thus, as (4.2) holds, we must have that $\mathbf{1}_{\{y\in N_{H_i}(U_y)\}}=1$, $\mathbf{1}_{\{x\in U_y\}}=0$, and that equality holds throughout (4.3). That is, we have the following.

G5 For each $y\in N_{H_i}(x)\setminus V(S)$, we have $y\in N_{H_i}(U_y)$, $|N_{H_i}(U_y)\setminus V(S)|=d|U_y|-e_{S_i}(U_y,A_{3-i})$ and $x\notin U_y$.

We will now show that, in fact, each such U_y must have size at most m, using the following claim.

Claim 5. *If $U\subset A_i$ and $m\leqslant |U|\leqslant 2m$, then $|N_{H_i}(U)\setminus V(S+xy)|>d|U|$.*

Proof of Claim 5. Let $U\subset A_i$ with $m\leqslant|U|\leqslant 2m$. Then, from **G4** and (4.1), we have

$$|N_{H_i}(U)\setminus V(S+xy)|\geqslant |A_{3-i}|/2-|S|-1\geqslant 2dm+1>d|U|. \qquad \square$$

Thus, by (4.2) and Claim 5, for each $y\in N_{H_i}(x)\setminus V(S)$, we have $|U_y|<m$.

Claim 6. *For each $Y\subset N_{H_i}(x)\setminus V(S)$, $|\cup_{y\in Y}U_y|<m$ and*

$$|N_{H_i}(\cup_{y\in Y}U_y)\setminus V(S)|=d|\cup_{y\in Y}U_y|-e_{S_i}(\cup_{y\in Y}U_y,A_{3-i}).$$

Proof of Claim 6. We prove this by induction on $|Y|$. We know this to be true if $|Y|=1$ by **G5**, so assume that $|Y|>1$, and, picking $y_0\in Y$, that the claim is true for $Y'\coloneqq Y\setminus\{y_0\}$ and $\{y_0\}$. Note first that $|\cup_{y\in Y}U_y|\leqslant|\cup_{y\in Y'}U_y|+|U_{y_0}|\leqslant 2m$. Then, by the induction hypothesis and simple set relations (in particular, that for all sets A,B in a graph G, $|N(A\cup B)|+|N(A\cap B)|\leqslant|N(A)|+|N(B)|$), we have

$$\begin{aligned}
&|N_{H_i}(\cup_{y\in Y}U_y)\setminus V(S)|+|N_{H_i}((\cup_{y\in Y'}U_y)\cap U_{y_0})\setminus V(S)|\\
&\quad\leqslant|N_{H_i}(\cup_{y\in Y'}U_y)\setminus V(S)|+|N_{H_i}(U_{y_0})\setminus V(S)|\\
&\quad=d|\cup_{y\in Y'}U_y|-e_{S_i}(\cup_{y\in Y'}U_y,A_{3-i})+d|U_{y_0}|-e_{S_i}(U_{y_0},A_{3-i})\\
&\quad=d|\cup_{y\in Y}U_y|+d|(\cup_{y\in Y'}U_y)\cap U_{y_0}|-e_{S_i}(\cup_{y\in Y}U_y,A_{3-i})-e_{S_i}((\cup_{y\in Y'}U_y)\cap U_{y_0},A_{3-i}).
\end{aligned}\tag{4.4}$$

Now, by **G3** applied to $(\cup_{y\in Y'}U_y)\cap U_{y_0}$ (noting that this also holds if the set is empty), we have that

$$|N_{H_i}((\cup_{y\in Y'}U_y)\cap U_{y_0})\setminus V(S)|\geqslant d|(\cup_{y\in Y'}U_y)\cap U_{y_0}|-e_{S_i}((\cup_{y\in Y'}U_y)\cap U_{y_0},A_{3-i}).$$

Therefore, in combination with (4.4), we have

$$|N_{H_i}(\cup_{y\in Y}U_y)\setminus V(S)|\leqslant d|\cup_{y\in Y}U_y|-e_{S_i}(\cup_{y\in Y}U_y,A_{3-i}). \tag{4.5}$$

Thus, from **G3** applied to $\cup_{y\in Y}U_y$, we have that equality holds in (4.5), as required. By Claim 5, we also then have that $|\cup_{y\in Y}U_y|<m$. $\square$

From Claim 6 with $Y=N_{H_i}(x)\setminus V(S)$, we have $|\cup_{y\in Y}U_y|<m$ and

$$|N_{H_i}(\cup_{y\in Y}U_y)\setminus V(S)|=d|\cup_{y\in Y}U_y|-e_{S_i}(\cup_{y\in Y}U_y,A_{3-i}). \tag{4.6}$$

By **G5**, we have $y\in N_{H_i}(U_y)$ for each $y\in Y=N_{H_i}(x)\setminus V(S)$, and thus

$$|N_{H_i}((\cup_{y\in Y}U_y)\cup\{x\})\setminus V(S)|=|N_{H_i}(\cup_{y\in Y}U_y)\setminus V(S)|. \tag{4.7}$$

As, by **G5**, $x\notin U_y$ for each $y\in Y$, and $d_S(x)<d$, we have

$$d|(\cup_{y\in Y}U_y)\cup\{x\}|-e_{S_i}((\cup_{y\in Y}U_y)\cup\{x\},A_{3-i})>d|\cup_{y\in Y}U_y|-e_{S_i}(\cup_{y\in Y}U_y,A_{3-i}). \tag{4.8}$$

Combining (4.6), (4.7) and (4.8), we have

$$d|(\cup_{y\in Y}U_y)\cup\{x\}|-e_{S_i}((\cup_{y\in Y}U_y)\cup\{x\},A_{3-i})>|N_{H_i}((\cup_{y\in Y}U_y)\cup\{x\})\setminus V(S)|.$$

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

As $|(\cup_{y\in Y} U_y)\cup\{x\}| < m+1 \leqslant 2m$, this contradicts **G3** in definition of the (d,m)-extendability of (S,X) in (H_1,H_2). Thus some such y as required by the lemma must exist. □

Applying Lemma 4.3 repeatedly, we can build an extendable copy of a tree, as follows.

Lemma 4.4. *Let $d \geqslant 2$ and $m \geqslant 1$. Let A_1 and A_2 be disjoint sets and let $X \subset A_1$ and $X' \subset X$. Let T be a forest containing X, in which each component has exactly one vertex in X and each vertex in X' has degree 0. Let T_x, $x \in X'$, be vertex-disjoint trees with maximum degree at most d such that T_x contains x. Let $T' = \cup_{x\in X'} T_x$. Let H_1 and H_2 be bipartite graphs with vertex classes A_1 and A_2. Suppose (T,X) is (d,m)-extendable in (H_1,H_2) and that*

$$|T| + |T'| \leqslant \min\{|A_1|,|A_2|\}/2 - 2dm - 1. \tag{4.9}$$

Then there is a copy S of T' so that each vertex $x \in X'$ is copied to itself, $V(T)\cap V(S) = X'$, and $(T+S,X)$ is (d,m)-extendable in (H_1,H_2).

Proof. We will prove this by induction on $|E(T')|$. If $|E(T')| = 0$, then let S be the forest with vertex set X' and no edges, noting this satisfies the lemma as $T+S=T$. Suppose then that $|E(T')| \geqslant 1$. Pick $x_0 \in X'$ with $|E(T_{x_0})| \geqslant 1$, let y_0 be a leaf of T_{x_0} which is not equal to x_0 and let z_0 be the neighbour of y_0 in T_{x_0}. By the induction hypothesis, there is some copy $S' \subset H_1 \cup H_2$ of $T' - y_0$ such that x is copied to x, for each $x \in X'$, $V(T)\cap V(S') = X'$, and $(T+S',X)$ is (d,m)-extendable in (H_1,H_2).

Let z_0' be the copy of z_0 in S'. Suppose $z_0' \in A_1$, where the case where $z_0' \in A_2$ follows similarly. As T_{x_0} has maximum degree at most d, the degree of z_0 in S' is at most $d-1$. By (4.9),

$$|T+S'| \leqslant |T| + |T'| - 1 \leqslant \min\{|A_1|,|A_2|\}/2 - 2dm - 2.$$

Thus, by Lemma 4.3, there is some vertex y_0' in $A_2 \setminus V(T+S')$ so that $z_0'y_0' \in E(H_1)$ and, if $S := S' + z_0'y_0'$, $(T+S,X)$ is (d,m)-extendable in (H_1,H_2). Noting that $V(T)\cap V(S) = V(T)\cap V(S') = X'$, and S is a copy of T' in which each vertex in X' is copied to itself, completes the proof. □

A critical part of the method used by Glebov, Krivelevich and Johannsen [15] is that not only can we add a leaf and remain (d,m)-extendable, but we can also remove a leaf and remain (d,m)-extendable, as follows.

Lemma 4.5. *Suppose that H_1 and H_2 are two bipartite graphs with the same vertex classes A_1 and A_2, and that S is a forest containing $X \subset V(S)\cap A_1$ in which exactly one vertex in each tree in S is in X. Suppose $j \in [2]$, $x \in V(S)\cap A_i$ and $y \in A_{3-j}\setminus V(S)$ so that $(S+xy,X)$ is (d,m)-extendable in (H_1,H_2).*

Then (S,X) is (d,m)-extendable in (H_1,H_2).

Proof. That **G1**, **G2** and **G4** hold for (S,X) to be (d,m)-extendable in (H_1,H_2) follows directly from the same conditions for the (d,m)-extendability of $(S+xy,X)$, so we need only check **G3**.

For each $i \in [2]$, let S_j be the subgraph of S with edge set $\{e \in E(S)\colon d(e,X) \equiv i+1 \mod 2\}$ and let S_i' be the subgraph of $S+xy$ with edge set $\{e \in E(S+xy)\colon d(e,X) \equiv i+1 \mod 2\}$. Note that $S_j' = S_j + xy$ and $S_{3-j}' = S_{3-j}$.

Let $U \subset A_i$ with $0 < |U| \leqslant 2m$. If $x \notin U$, then, as $(S+xy,X)$ is (d,m)-extendable in (H_1,H_2),

$$\begin{aligned}|N_{H_j}(U)\setminus V(S)| &\geqslant |N_{H_j}(U)\setminus V(S+xy)|\\ &\geqslant d|U| - e_{S_j+xy}(U,A_{3-j})\\ &= d|U| - e_{S_j}(U,A_{3-j}).\end{aligned}$$

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

On the other hand, if $x \in U$, then $y \in N_{H_j}(U)$, so that

$$\begin{aligned}|N_{H_j}(U) \setminus V(S)| &= |N_{H_j}(U) \setminus V(S+xy)| + 1\\ &\geqslant d|U| - e_{S_j+xy}(U, A_{3-j}) + 1\\ &= d|U| - e_{S_j}(U, A_{3-j}).\end{aligned}$$

Finally, let $U \subset A_{3-j}$ with $0 < |U| \leqslant 2m$, then, as $y \in A_{3-j}$, and $S'_{3-j} = S_{3-j}$,

$$\begin{aligned}|N_{H_{3-j}}(U) \setminus V(S)| &= |N_{H_{3-j}}(U) \setminus V(S+xy)|\\ &\geqslant d|U| - e_{S'_{3-j}}(U, A_{3-j})\\ &= d|U| - e_{S_{3-j}}(U, A_{3-j}).\end{aligned}$$

Thus **G3** holds so that (S, X) is (d, m)-extendable in (H_1, H_2). □

4.1 Path connections

We will now build paths between vertex sets, using the work in this section so far. The next lemma is the key lemma we use to show weak and strong connectivity in pseudorandom digraphs. We work with two extendable forests S and T in two vertex-disjoint pairs of bipartite graphs (G_1, G_2) and (H_1, H_2). Given vertex sets $X' \subset V(S)$ and $Y' \subset V(T)$, we add a $(d-1)$-ary tree disjointly to each vertex in X' and Y' while retaining the respective extendability properties. The resulting trees attached to X' and Y' are, together, very large, and this will allow us to connect two trees from X' and Y' respectively using another set Z. This allows us to find a path from a vertex in X' to a vertex in Y' which initially alternates between edges in G_1 and G_2, then passes through Z, before alternating between edges in H_1 and H_2. The edges in these graphs will be chosen so that this will correspond to a directed path in our initial digraph. Crucially, we can then use Lemma 4.5 to remove the vertices we added but did not use in this connection, while remaining extendable. This allows us to efficiently repeat the argument to find more connections.

We emphasize again that this is a fairly direct adaptation of the work of Glebov, Krivelevich and Johannsen [15], only in a form applicable to digraphs.

Lemma 4.6. *Let $m \geqslant 1$, $d \geqslant 3$ and k satisfy $k = \lceil \log m / \log(d-1) \rceil$, and let $0 \leqslant j \leqslant k$. Let G be a graph containing disjoint vertex sets A_1, A_2, B_1, B_2 and Z. Let $X \subset A_1$ and $Y \subset B_1$. Let S and T be forests so that there is exactly one vertex of X and Y in each component of S and T respectively, and so that*

$$|S|, |T| \leqslant \frac{1}{2}\min\{|A_1|, |A_2|, |B_1|, |B_2|\} - 10dm - 1. \tag{4.10}$$

Let G_1 and G_2 be bipartite subgraphs of G with vertex classes A_1 and A_2 so that (S, X) is (d, m)-extendable in (G_1, G_2). Let H_1 and H_2 be bipartite subgraphs of G with vertex classes B_1 and B_2 so that (T, Y) is (d, m)-extendable in (H_1, H_2).

For any $U \subset A_1 \cup A_2$ or $U \subset B_1 \cup B_2$ with $|U| = m$, suppose that $|N_G(U, Z)| > |Z|/2$. Suppose that $X' \subset X$ and $Y' \subset Y$ with $|X'|, |Y'| \geqslant m/(d-1)^j$ are sets of vertices with degree 0 in S and T respectively. Then there are some $x \in X'$, $y \in Y'$, $x_0 \in A_1 \cup A_2$ and $y_0 \in B_1 \cup B_2$ and paths $P \subset G_1 \cup G_2$ and $Q \subset H_1 \cup H_2$ such that the following hold.

- *P is an x, x_0-path with length at most j and no vertices in $V(S) \setminus \{x\}$, and $(S + P, X)$ is (d, m)-extendable in (G_1, G_2).*
- *Q is a y, y_0-path with length at most j and no vertices in $V(T) \setminus \{y\}$, and $(T + Q, Y)$ is (d, m)-extendable in (H_1, H_2).*
- *$N_G(x_0) \cap N_G(y_0) \cap Z \neq \emptyset$.*

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

Proof. By removing vertices from X' and Y' if necessary, assume that $|X'| = |Y'| = \lceil m/(d-1)^j \rceil$. Let S_x, $x \in X'$, be a collection of disjoint $(d-1)$-ary trees with depth j so that S_x has root x for each $x \in X'$, and let $S' = \cup_{x \in X'} S_x$. Let T_y, $y \in Y'$, be a collection of disjoint $(d-1)$-ary trees with depth j so that T_y has root y for each $y \in Y'$, and let $T' = \cup_{y \in Y'} T_y$.

Note that, by (4.10),

$$|S'| \leqslant 4md \leqslant \min\{|A_1|, |A_2|\}/2 - 2dm - 1 - |S|.$$

Therefore, by Lemma 4.4, there is a copy S'' of S' such that x is copied to x for each $x \in X'$ and $(S + S'', X)$ is (d, m)-extendable in (G_1, G_2). Similarly, by Lemma 4.4, there is a copy T'' of T' such that y is copied to y for each $y \in Y'$ and $(T + T'', Y)$ is (d, m)-extendable in (H_1, H_2).

Noting that $|S''|, |T''| \geqslant m$, we can find some $x_0 \in V(S'')$ and $y_0 \in V(T'')$ such that $N_G(x_0) \cap N_G(y_0) \cap Z \neq \emptyset$. Let x be the vertex in X' for which there is a path, P say, in S'' with length at most j from x to x_0. Note that, by iteratively removing leaves not in X, $S + S''$ can be turned into $S + P$. Thus, by repeated application of Lemma 4.5, we have that $(S + P, X)$ is (d, m)-extendable in (G_1, G_2).

Similarly, there is a y, y_0-path, Q say, in T'' with length at most j, so that $(T + Q, Y)$ is (d, m)-extendable in (H_1, H_2). Thus the vertices x, y, x_0, y_0 and paths P and Q satisfy the requirements in the lemma. □

4.2 Strong connection in pseudorandom digraphs

We now use Lemma 4.6 to show a strong connection property in pseudorandom digraphs. We first use Lemma 3.2 to find the sets to which we can then apply Lemma 4.6. In our application, we connect pairs of vertices with paths of length $O(\log n/\log^{[2]} n)$.

In the proof of the following theorem, and several times later in the paper, we make various calculations, often at length, to ensure that Lemma 3.2 can be applied. For the reader who wishes to take these calculations as read, each such calculation begins with 'We will now check the conditions', and we proceed after the calculation with a new paragraph continuing 'Having checked the appropriate conditions'.

Theorem 4.7. *For each $\varepsilon > 0$, there exists $n_0 = n_0(\varepsilon)$ such that the following holds for every $d \geqslant 10^{-5}$ and $n \geqslant n_0$. Let D be an n-vertex (d, ε)-pseudorandom digraph containing a set A with $|A| \geqslant n/\log^{[3]} n$ and in which the following hold with $m = n \log^{[3]} n/d \log n$.*

(1) *For each $v \in V(D)$, $d^{\pm}(v, A) \geqslant d(\log n)^{3/4}$.*
(2) *For each $U \subset V(D)$ with $|U| = m$, $|N^{\pm}(U, A)| \geqslant (1/2 + \varepsilon/2)|A|$.*

Then any set $V \subset V(D) \setminus A$ with

$$|V| \leqslant \frac{|A| \log^{[2]} n}{\log n \cdot \log^{[7]} n}$$

is strongly connected in $D[A \cup V]$.

Proof. Let $\alpha = \lfloor |A|/5 \rfloor$, $\delta = d(\log n)^{3/4}$ and $\Delta = 10^6 d \log n$ so that, for each $v \in V(D)$, $d^{\pm}(v, A) \geqslant \delta$, and, by **A1** in the definition of (d, ε)-pseudorandomness, we have $\Delta^{\pm}(D) \leqslant \Delta$.

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

We will now check the conditions **D1**–**D4** to apply Lemma 3.2 with δ, Δ, m, n unchanged, $\ell = 5$, $a_1 = \cdots = a_5 = \alpha$ and $\varepsilon/2$ in place of ε. For **D1**–**D4** in turn, we have that

$$\frac{\varepsilon^2\alpha^2}{4|A|} = \Omega_\varepsilon(\alpha) = \omega(1),$$

$$\frac{\varepsilon^2\alpha^2}{4|A|} \cdot \exp\left(\frac{\alpha\delta}{24|A|}\right) \geqslant \frac{\varepsilon^2\alpha^2}{4|A|} \cdot \exp\left(\frac{\delta}{200}\right) = \Omega_\varepsilon\left(\frac{n}{\log^{[3]} n} \cdot \exp\left(\frac{d(\log n)^{3/4}}{200}\right)\right) = \omega(n),$$

$$\exp\left(\frac{\alpha\delta}{24|A|}\right) \geqslant \exp\left(\frac{\delta}{200}\right) = \exp\left(\Omega(d(\log n)^{3/4})\right) = \omega(d^2 \log^2 n) = \omega(\Delta^2),$$

and, as $\log(en/m) = O(\log(d \log n))$,

$$\begin{aligned}\frac{\varepsilon^2\alpha}{4 \cdot 10^3} = \Omega_\varepsilon(\alpha) &= \Omega_\varepsilon\left(\frac{n}{\log^{[3]} n}\right)\\ &= \Omega_\varepsilon\left(m \cdot \frac{d \log n}{(\log^{[3]} n)^2}\right)\\ &= \omega(m \cdot \log(d \log n))\\ &= \omega\left(m \cdot \log\left(\frac{en}{m}\right)\right).\end{aligned}$$

Having checked the appropriate conditions, by Lemma 3.2, for sufficiently large n, there are disjoint vertex sets A_1', A_2, B_1', B_2, Z in A, each with size α, such that

H1 For each $v \in V(D)$ and $B \in \{A_1', A_2, B_1', B_2, Z\}$,

$$d^{\pm}(v, B) \geqslant \alpha\delta/4|A| \geqslant d(\log n)^{3/4}/40.$$

H2 For each $U \subset V(D)$ with $|U| = m$ and $B \in \{A_1', A_2, B_1', B_2, Z\}$,

$$|N^{\pm}(U, B)| \geqslant (1/2 + \varepsilon/4)|B|.$$

Now let

$$\ell_0 := \frac{|A| \log^{[2]} n}{\log n \cdot \log^{[7]} n}, \tag{4.11}$$

and take an arbitrary set $E = \{\overrightarrow{y_i x_i} : i \in [\ell]\}$ of $\ell \leqslant \ell_0$ vertex-disjoint edges in the complete digraph with vertex set $V(D) \setminus A$. To show the lemma, it is sufficient to find a directed cycle in $D[A \cup V(E)] + E$ which contains the edges $\overrightarrow{y_1 x_1}, \ldots, \overrightarrow{y_\ell x_\ell}$ in order.

Let $X = \{x_1, \ldots, x_\ell\}$ and $Y = \{y_1, \ldots, y_\ell\}$. Let G_1 and G_2 be the bipartite (undirected) graphs with vertex set $A_1 := A_1' \cup X$ and A_2 and the edges in D from A_1 to A_2 and A_2 to A_1 respectively, but without their directions. Let H_1 and H_2 be the bipartite (undirected) graphs with vertex set $B_1 := B_1' \cup Y$ and B_2 and the edges in D from B_2 to B_1 and B_1 to B_2 respectively, again without their directions. Let G have vertex set $A \cup X \cup Y$ and consist of the edges in G_1, G_2, H_1, H_2 as well as the edges from $A_1 \cup A_2$ to Z and Z to $B_1 \cup B_2$. Let $d_0 = (\log n)^{1/3}$ and note that $d_0 m = o(a)$. Let I_X and I_Y respectively be the graphs with vertex set X and Y which each have no edges.

Claim 7. *For sufficiently large n, (I_X, X) is (d_0, m)-extendable in (G_1, G_2) and (I_Y, Y) is (d_0, m)-extendable in (H_1, H_2).*

Proof of Claim 7. We will show that (I_X, X) is (d_0, m)-extendable in (G_1, G_2). That (I_Y, Y) is (d_0, m)-extendable in (H_1, H_2) follows similarly. Note that **G1** and **G2** in Definition 4.2 are immediate as I_X and I_Y have no edges. Furthermore, from **H2**, for each $U \subset A_2$ with $|U| \geqslant m$, we have

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

$$\begin{aligned}|N_{G_2}(U)\setminus X| &= |N_D^+(U, A_1\setminus X)|\\ &= |N_D^+(U, A_1')|\\ &\geqslant (1/2+\varepsilon/4)|A_1'|\\ &= (1/2+\varepsilon/4)(|A_1|-|X|)\\ &> |A_1|/2,\end{aligned}$$

where we have used that $|X| = \ell = o(a)$. Thus, as for each $U \subset A_1$ with $|U| \geqslant m$ we have, from **H2**, that $|N_{G_1}(U)| = |N_D^+(U, A_2)| > |A_2|/2$, we have that **G4** holds.

Let $U \subset A_1$ with $0 < |U| \leqslant 2m$. Each vertex in U has at least $d(\log n)^{3/4}/40 \geqslant d(\log n)^{2/3}$ out-neighbours in A_2 in D by **H1**, for sufficiently large n. Thus, as $A_2 \cap X = \emptyset$, by **A3** in the definition of (d, ε)-pseudorandomness, we have

$$|N_{G_1}(U, A_2\setminus X)| = |N_D^+(U, A_2)| \geqslant d_0|U|. \tag{4.12}$$

Furthermore, for each $U \subset A_2$ with $0 < |U| \leqslant 2m$, each vertex in U has at least $d(\log n)^{3/4}/40 \geqslant d(\log n)^{2/3}$ out-neighbours in A_1' in D by **H1**, for sufficiently large n. Thus, by **A3** in the definition of (d, ε)-pseudorandomness, we have

$$|N_{G_2}(U, A_1\setminus X)| = |N_D^+(U, A_1')| \geqslant d_0|U|. \tag{4.13}$$

In combination, (4.12) and (4.13) show that **G3** holds to complete the proof of the (d_0, m)-extendability of (I_X, X) in (G_1, G_2). □

Let

$$k = \left\lceil \frac{\log m}{\log(d_0-1)} \right\rceil = O\left(\frac{\log n}{\log^{[2]} n}\right) \overset{(4.11)}{=} o\left(\frac{a}{\ell_0}\right). \tag{4.14}$$

Take a maximal set $I \subset [\ell]$ for which there is a vertex-disjoint set of paths $P_i \subset G_1 \cup G_2$ and $Q_i \subset H_1 \cup H_2$, $i \in I$, with length at most k each and distinct vertices z_i, $i \in I$, in Z such that the following hold (with addition modulo ℓ in the indices).

I1 For each $i \in I$, P_i is a path with start vertex x_i and no vertices in $X \setminus \{x_i\}$, and $(I_X + \sum_{i\in I} P_i, X)$ is (d_0, m)-extendable in (G_1, G_2).
I2 For each $i \in I$, Q_i is a path with start vertex y_{i+1} and no vertices in $Y \setminus \{y_{i+1}\}$, and $(I_Y + \sum_{i\in I} Q_i, Y)$ is (d_0, m)-extendable in (H_1, H_2).
I3 For each $i \in I$, $P_i z_i \overleftarrow{Q_i}$ is a directed x_i, y_{i+1}-path in D.

Note that, by Claim 7, $I = \emptyset$ satisfies **I1**–**I3**, so that such a set I exists.

If $I = [\ell]$, then, by **I3**, $\sum_{i\in I} (y_i x_i P_i z_i \overleftarrow{Q}_i)$ is a directed cycle in $E + D[A \cup V(E)]$ containing the edges $y_i x_i$, $i \in I$, in order, as required. Thus assume for contradiction that there is some $j \in [\ell] \setminus I$.

Let the paths P_i and Q_i, and disjoint vertices z_i, $i \in I$, satisfy the conditions above. Let $Z' = Z \setminus (\cup_{i\in I} z_i)$, and note that $|Z'| \geqslant |Z| - \ell$. Thus, by **H2**, the definition of G, and as $\ell \leqslant \ell_0 = o(a)$, we have the following.

I1 For any $U \subset A_1 \cup A_2$ or $U \subset B_1 \cup B_2$ with $|U| = m$, we have

$$|N_G(U, Z')| \geqslant (1/2+\varepsilon/4)|Z| - \ell > |Z|/2 \geqslant |Z'|/2.$$

Let $S = I_X + \sum_{i\in I} P_i$ and $T = I_Y + \sum_{i\in I} Q_i$, so that, by **I1** and **I2**, (S, X) and (T, Y) are (d_0, m)-extendable in (G_1, G_2) and (H_1, H_2) respectively. Note that, by (4.14), as $d_0 m = o(a)$, we have

$$|S|, |T| \leqslant \ell(k+1) = o(a) = o\left(\frac{1}{2}\min\{|A_1|, |A_2|, |B_1|, |B_2|\} - 10d_0 m - 1\right). \tag{4.15}$$

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

Thus, by (4.14), (4.15) and **I1**, we can apply Lemma 4.6 with $X' = \{x_j\}$ and $Y' = \{y_{j+1}\}$ to get vertices $x' \in A_1 \cup A_2$, $y' \in B_1 \cup B_2$ and $z_j \in Z'$ and paths $P_j \subset G_1 \cup G_2$ and $Q_j \subset H_1 \cup H_2$ such that the following hold.

J1 P_j is an x_j, x'-path with length at most k and no vertices in $V(S) \setminus \{x_j\}$, and $(S + P_j, X)$ is (d_0, m)-extendable in (G_1, G_2).
J2 Q_j is a y_{j+1}, y'-path with length at most k and no vertices in $V(T) \setminus \{y_j\}$, and $(T + Q_j, Y)$ is (d_0, m)-extendable in (H_1, H_2).
J3 $x'z_j, z_jy' \in E(G)$.

Note that, by the definition of (d_0, m)-extendability, for each i, the ith edge of P_j, counting from x_j, is in $G_{2-(i \bmod 2)}$. Thus, by the choice of the graphs G_1 and G_2, P_j is a directed path in D. Similarly, Q_j is, when reversed, a directed path in D. By **J3**, and the choice of G, then $P_jz_j\overleftarrow{Q_j}$ is a directed x_jy_{j+1}-path in D.

Then P_i, Q_i and z_i, $i \in I \cup \{j\}$, satisfy **I1**–**I3**, contradicting the maximality of I. □

4.3 Weak connection in pseudorandom digraphs

We now use Lemma 4.6 to connect edges into a cycle using directed paths with, on average, length $O(1)$ (see Theorem 4.8), rather than the (potential) length $\Omega_\varepsilon(\log n/\log^{[2]} n)$ used in Theorem 4.7. For this to be feasible we remove the restriction that the original edges appear in the cycle in a specified order; this is the difference between weak and strong connection (see Definitions 2.3 and 2.4).

Similarly to the proof of Theorem 4.7, we prove Theorem 4.8 by first applying Lemma 3.2 to find sets before, for any appropriate set of edges E, selecting a maximal set of paths satisfying some conditions. Lemma 4.6 will then show that we have as many paths as possible, and thus have the cycle we need. However, for discussion it is more convenient to think about finding paths one by one, each time connecting two of the edges in E and moving closer to the cycle. At the start, there are many potential pairs of edges we could connect together, allowing us to apply Lemma 4.6 with initially large sets X' and Y'. When we have connected most of the edges into a cycle we will need to use longer paths, but we will still have used, on average, short paths.

Theorem 4.8. *For each $\varepsilon > 0$, there exists $n_0 = n_0(\varepsilon)$ such that the following holds for every $d \geqslant 10^{-5}$ and $n \geqslant n_0$. Let D be an n-vertex (d, ε)-pseudorandom digraph containing disjoint vertex sets A and V so that*

$$\frac{n \log^{[2]} n}{\log n \cdot \log^{[4]} n} \leqslant |A| \leqslant \frac{n(\log^{[2]} n)^3}{\log n} \tag{4.16}$$

and $|V| \leqslant |A|/\log^{[6]} n$, and the following hold with $m = n \log^{[3]} n/d \log n$.

(1) *For each $v \in A \cup V$, $d^{\pm}(v, A) \geqslant 40d \log^{[2]} n/\log^{[4]} n$ and $d^{\pm}(v, A \cup V) \leqslant d(\log^{[2]} n)^3$.*
(2) *For each $U \subset A \cup V$ with $|U| = m$, $|N^{\pm}(U, A)| \geqslant (1/2 + \varepsilon/8)|A|$.*

Then V is weakly connected in $D[A \cup V]$.

Proof. Let $\alpha = \lfloor |A|/5 \rfloor$, $\delta = 40d \log^{[2]} n/\log^{[4]} n$, $\Delta = d(\log^{[2]} n)^3$ and $n' = |A \cup V| \leqslant 6a$.

We will now check the conditions **D1**–**D4** to apply Lemma 3.2 to $D[A \cup V]$ with n' in place of n, δ, Δ, m unchanged, $\ell = 5$, $a_1 = \cdots = a_5 = \alpha$ and $\varepsilon/8$ in place of ε. For sufficiently large n, we have for **D1**–**D3** in turn that

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

$$\left(\frac{\varepsilon}{8}\right)^2 \cdot \frac{\alpha^2}{|A|} = \Omega_\varepsilon(\alpha) = \omega(1),$$

$$\left(\frac{\varepsilon}{8}\right)^2 \cdot \frac{\alpha^2}{|A|} \cdot \exp\left(\frac{\delta\alpha}{24|A|}\right) = \Omega_\varepsilon\left(\alpha \cdot \exp\left(\frac{d \log^{[2]} n}{5 \log^{[4]} n}\right)\right) = \omega(\alpha) = \omega(n'),$$

and

$$\exp\left(\frac{\delta\alpha}{24|A|}\right) = \exp\left(\Omega\left(\frac{d \log^{[2]} n}{\log^{[4]} n}\right)\right) = \omega(d^2(\log^{[2]} n)^6) = \omega(\Delta^2).$$

Using (4.16), for **D4**, note that, as $n' \leqslant 6\alpha \leqslant 6n(\log^{[2]} n)^3/\log n$, $\log(en'/m) = O(\log(d \log^{[2]} n))$, and hence

$$\begin{aligned}\left(\frac{\varepsilon}{8}\right)^2 \cdot \frac{\alpha}{10^3} &= \Omega_\varepsilon(\alpha) \\ &\overset{(4.16)}{=} \Omega_\varepsilon\left(\frac{n \log^{[2]} n}{\log n \cdot \log^{[4]} n}\right) \\ &= \Omega_\varepsilon\left(m \cdot \frac{d \log^{[2]} n}{\log^{[3]} n \cdot \log^{[4]} n}\right) \\ &= \omega\left(m \cdot \log\left(\frac{en'}{m}\right)\right).\end{aligned}$$

Having checked the appropriate conditions, by Lemma 3.2, for sufficiently large n, there are disjoint vertex sets A_1', A_2, B_1', B_2, Z in A, each with size α, such that

K1 For each $v \in V(D)$ and $B \in \{A_1', A_2, B_1', B_2, Z\}$,

$$d^{\pm}(v, B) \geqslant d \log^{[2]} n/\log^{[4]} n.$$

K2 For each $U \subset V(D)$ with $|U| = m$ and $B \in \{A_1', A_2, B_1', B_2, Z\}$,

$$|N^{\pm}(U, B)| \geqslant (1/2 + \varepsilon/16)|B|.$$

Now take an arbitrary set $E = \{\overrightarrow{y_i x_i} \colon i \in [\ell]\}$ of $\ell \leqslant |V|/2 = o(a)$ vertex-disjoint edges in the complete digraph with vertex set V. To show the lemma, it is sufficient to find a directed cycle in $D[A \cup V(E)] + E$ which contains the edges $\overrightarrow{y_1 x_1}, \ldots, \overrightarrow{y_\ell x_\ell}$ in any order.

Let $X = \{x_1, \ldots, x_\ell\}$ and $Y = \{y_1, \ldots, y_\ell\}$. Let G_1 and G_2 be the bipartite (undirected) graphs with vertex set $A_1 := A_1' \cup X$ and A_2 and the edges in D from A_1 to A_2 and A_2 to A_1 respectively, but without their directions. Let H_1 and H_2 be the bipartite (undirected) graphs with vertex set $B_1 := B_1' \cup Y$ and B_2 and the edges in D from B_2 to B_1 and B_1 to B_2 respectively. Let G have vertex set $A \cup X \cup Y$ and consist of the edges in G_1, G_2, H_1, H_2 as well as the edges from $A_1 \cup A_2$ to Z and Z to $B_1 \cup B_2$, again without their directions. Let I_X and I_Y respectively be the graphs with vertex set X and Y which each have no edges. Let $d_0 = 10$.

Claim 8. *For sufficiently large n, (I_X, X) is (d_0, m)-extendable in (G_1, G_2) and (I_Y, Y) is (d_0, m)-extendable in (H_1, H_2).*

Proof of Claim 8. We will show that (I_X, X) is (d_0, m)-extendable in (G_1, G_2). That (I_Y, Y) is (d_0, m)-extendable in (H_1, H_2) follows similarly. Note that **G1** and **G2** in Definition 4.2 hold as I_X and I_Y have no edges. Furthermore, from **K2**, we have, for each $U \subset A_2$ with $|U| \geqslant m$, that

$$|N_{G_2}(U, A_1) \setminus X| = |N_D^+(U, A_1')| \geqslant (1/2 + \varepsilon/16)|A_1'| > |A_1|/2,$$

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

where we have used that $|X| = \ell = o(a)$. Thus, as, for each $U \subset A_1$ with $|U| \geqslant m$, we have, from **K2**, that $|N_{G_1}(U, A_2)| = |N_D^+(U, A_2)| > |A_2|/2$, we have that **G4** holds.

Let $U \subset A_1$ with $0 < |U| \leqslant 2m$. By **K1**, for sufficiently large n, each vertex in U has at least $d \log^{[2]} n / \log^{[4]} n$ out-neighbours in A_2 in D. Thus, as $A_2 \cap X = \emptyset$, by **A2** in the definition of (d, ε)-pseudorandomness, we have

$$|N_{G_1}(U, A_2 \setminus X)| = |N_D^+(U, A_2)| \geqslant d_0|U|. \tag{4.17}$$

Furthermore, for sufficiently large n, for each $U \subset A_2$ with $0 < |U| \leqslant 2m$, each vertex in U has at least $d \log^{[2]} n / \log^{[4]} n$ out-neighbours in A_1' in D by **K1**. Thus, by **A2** in the definition of (d, ε)-pseudorandomness, we have

$$|N_{G_2}(U, A_1 \setminus X)| = |N_D^+(U, A_1')| \geqslant d_0|U|. \tag{4.18}$$

In combination, (4.17) and (4.18) show that **G3** holds to complete the proof of the (d_0, m)-extendability of (I_X, X) in (G_1, G_2). □

We will now cover the edges $\overrightarrow{y_i x_i}$, $i \in [\ell]$, using as few directed paths as possible, subject to some conditions (**L1**–**L4** below). We will then use Lemma 4.6 to show that in fact we have one directed path. Applying Lemma 4.6 again will then allow us to complete this path into a cycle.

To govern the length of the covering paths, we use a function g defined as follows. For each $r \in [\ell]$, let

$$g(r) = \sum_{i=1}^{r} \left\lceil \log\left(\frac{4m}{\ell + 1 - i}\right) \Big/ \log(d_0 - 1) \right\rceil. \tag{4.19}$$

As before, let I_X and I_Y be the graphs with no edges and vertex sets X and Y respectively.

Now, for the smallest possible $r \in [\ell]$, find vertex-disjoint directed paths R_i, $i \in [r]$, in $D[A \cup V(E)] + E$ satisfying the following properties.

- **L1** Each edge $\overrightarrow{y_i x_i}$ appears in some path R_j, $j \in [r]$.
- **L2** In total, the paths R_i, $i \in [r]$, have length at most $\ell + 4 \cdot g(\ell - r)$ and contain at most $\ell - r$ vertices in Z.
- **L3** Each path R_i, $i \in [r]$, starts with some vertex y_j and ends with some vertex $x_{j'}$.
- **L4** Letting P and Q be the graphs of the edges in the paths R_i which appear (without their directions) in $G_1 \cup G_2$ and $H_1 \cup H_2$ respectively, $(I_X + P, X)$ is (d_0, m)-extendable in (G_1, G_2) and $(I_Y + Q, Y)$ is (d_0, m)-extendable in (H_1, H_2).

Note that the ℓ paths consisting of just the edges $\overrightarrow{y_i x_i}$, $i \in [\ell]$, satisfy these properties, so such an r and such paths R_i, $i \in [r]$, exist.

We will show, by contradiction, that $r = 1$. Let us assume then that $r \geqslant 2$. Let $r' = \lfloor r/2 \rfloor \geqslant 1$. Let X' be a set of r' end vertices of some of the paths R_i, $i \in [r]$, and let Y' be a set of r' start vertices of some of the paths R_i, $i \in [r]$, so that no path R_i, $i \in [r]$, has a vertex in both X' and Y'. This is possible as $2r' \leqslant r$. Note that, by **L3**, $X' \subset X$ and $Y' \subset Y$. Furthermore, each vertex in X' appears only in some edge in E in the paths R_i, $i \in [r]$, and therefore has degree 0 in $I_X + P$. Similarly, each vertex in Y' has degree 0 in $I_Y + Q$.

We will apply Lemma 4.6 to $I_X + P$ and $I_Y + Q$, so we need a bound on their size, which we get from the following claim.

Claim 9. *We have*

$$g(\ell) = \sum_{i=1}^{\ell} \left\lceil \log\left(\frac{4m}{\ell + 1 - i}\right) \Big/ \log(d_0 - 1) \right\rceil = o(a).$$

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

Proof of Claim 9. For any positive integer $s \geqslant 2$ and each $i \in [\ell]$, if

$$s = \lceil \log(4m/(\ell+1-i))/\log(d_0-1)\rceil,$$

then

$$(d_0-1)^{s-1} < \frac{4m}{\ell+1-i} \leqslant (d_0-1)^s,$$

and thus

$$\ell+1-\frac{4m}{(d_0-1)^{s-1}} < i \leqslant \ell+1-\frac{4m}{(d_0-1)^s}.$$

Certainly, there are at most $1+4m/(d_0-1)^{s-1}$ integers i which satisfy this.

Thus we have, as $\ell \leqslant |V|/2 \leqslant |A|/\log^{[6]} n = o(a)$, that

$$\begin{aligned} g(\ell) &\overset{(4.19)}{=} \sum_{i=1}^{\ell} \left\lceil \log\left(\frac{4m}{\ell+1-i}\right) \Big/ \log(d_0-1) \right\rceil \\ &\leqslant \ell + \sum_{s=2}^{\lceil \log(4m)/\log(d_0-1)\rceil} \left(1+\frac{4m}{(d_0-1)^{s-1}}\right)\cdot s \\ &\leqslant \ell + \left\lceil \frac{\log(4m)}{\log(d_0-1)} \right\rceil^2 + 4m\cdot \sum_{s=2}^{\infty} \frac{s}{(d_0-1)^{s-1}} \\ &\leqslant \ell + \log^2 n + 4m\cdot O(1) \\ &= o(a). \end{aligned}$$

□

As $\ell = o(a)$, by Claim 9 and **L2**, we have

$$|I_X+P|+|I_Y+Q| \leqslant 3\ell + \sum_{i=1}^{r} 4\left\lceil \log\left(\frac{4m}{\ell+1-i}\right) \Big/ \log(d_0-1) \right\rceil = o(a).$$

Let $Z' = Z \setminus (\cup_{i\in[r]} V(R_i))$, noting that, by **L2**, **K2** and the definition of G, and as $\ell = o(a)$, we have, for sufficiently large n, that

K2′ For each $U \subset A_1 \cup A_2$ and $U \subset B_1 \cup B_2$ with $|U| = m$,

$$|N_G(U, Z')| \geqslant (1/2+\varepsilon/16)|Z| - \ell > |Z'|/2.$$

Let $k' = \lceil \log(4m/r)/\log(d_0-1)\rceil$, so that

$$g(\ell-r+1) = g(\ell-r) + k', \tag{4.20}$$

and $|X'| = |Y'| = r' \geqslant r/2 \geqslant 2m/(d_0-1)^{k'}$. Thus, as $d_0 m = o(a)$ by (4.16), by Lemma 4.6, there are some $j, j' \in [\ell]$, vertices $x' \in A_1 \cup A_2$ and $y' \in B_1 \cup B_2$, and paths $P' \subset G_1 \cup G_2$ and $Q' \subset H_1 \cup H_2$ such that the following hold.

M1 $x_j \in X'$ and $y_{j'} \in Y'$.
M2 P' is an x_j, x'-path with length at most k', no vertices in $(X \cup V(P)) \setminus \{x_j\}$, and for which $(I_X + P + P', X)$ is (d_0, m)-extendable in (G_1, G_2).
M3 Q' is a $y_{j'}, y'$-path with length at most k' and no vertices in $(Y \cup V(Q)) \setminus \{y_{j'}\}$, and $(I_Y + Q + Q', Y)$ is (d_0, m)-extendable in (H_1, H_2).
M4 $N_G(x') \cap N_G(y') \cap Z' \neq \emptyset$.

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

Note that, by definition of the (d_0, m)-extendability, the ith edge of P', counting from x_j, is in $G_{2-(i \bmod 2)}$. Thus, by the choice of the graphs G_1 and G_2, P' is a directed x_j, x'-path in D. Similarly, Q' is, when reversed, a directed $y_{j'}, y'$-path in D. Using **M4**, and noting that, by the choice of G,

$$N_G(x') \cap N_G(y') \cap Z' = N_D^+(x') \cap N_D^-(y') \cap Z',$$

select a vertex $z \in Z'$ such that $P'z\overleftarrow{Q'}$ is a directed $x_j y_{j'}$-path in D.

Note that, by **M1** and the choice of X' and Y', x_j is the end vertex of a path different to the path of which $y_{j'}$ is the start vertex. Assume, then, by relabelling if necessary, that x_j is the end vertex of R_{r-1} and $y_{j'}$ is the start vertex of R_r. We will show that the $r-1$ paths $R'_i = R_i$, $i \in [r-2]$, and $R'_{r-1} = R_{r-1}P'z\overleftarrow{Q'}R_r$ satisfy **L1**–**L4**, contradicting the definition of r.

By **M2** and **M3**, the choice of Z', and as the paths R_i, $i \in [r]$, are vertex-disjoint, the paths R'_i, $i \in [r-1]$, are vertex-disjoint. By **L1** for the paths R_i, $i \in [r]$, and as the paths R'_i contain the paths R_i, **L1** holds for the paths R'_i, $i \in [r-1]$. Each path R'_i shares a start vertex with some path $R_{i'}$ and an end vertex with some (potentially different) path $R_{i''}$, and therefore, as **L3** holds for the paths R_i, $i \in [r]$, **L3** holds for the paths R'_i, $i \in [r-1]$.

Note that $P + P'$ and $Q + Q'$ are exactly the graphs of edges in the paths R'_i which appear (without direction) in $G_1 \cup G_2$ and $H_1 \cup H_2$ respectively. Thus **L4** holds for the paths R'_i, $i \in [r-1]$, by **M2** and **M3**.

The paths R'_i, $i \in [r-1]$, contain one additional vertex in Z compared to the paths R_i, $i \in [r]$, so that, in total, they have at most $\ell - (r-1)$ vertices in Z by **L2** for the paths R_i. As P' and Q' have length at most k', we have, by **L2** for the paths R_i again, that the paths R'_i have total length at most

$$\ell + 4g(\ell - r) + 2k' + 2 \leqslant \ell + 4g(\ell - r) + 4k' \overset{(4.20)}{\leqslant} \ell + 4g(\ell - r + 1).$$

Therefore **L2** holds for the paths R'_i, $i \in [r-1]$. This completes the proof that **L1**–**L4** hold for the paths R'_i, $i \in [r-1]$, contradicting the choice of r.

Therefore we have that $r = 1$. That is, with relabelling, there is a single path R in $D[A \cup V(E)] + E$ containing each edge in E, with start vertex y_1 and end vertex x_2, with length at most $o(a)$ (using Claim 9) and at most ℓ vertices in Z such that the following holds. If R' and R'' are the graphs of the edges in the path R which appear (without direction) in $G_1 \cup G_2$ and $H_1 \cup H_2$ respectively, then $(I_X + R', X)$ and $(I_Y + R'', Y)$ are (d_0, m)-extendable in (G_1, G_2) and (H_1, H_2) respectively.

Let $Z'' = Z \setminus V(R)$, and note that, as $|Z \cap V(R)| \leqslant \ell = o(a)$, for sufficiently large n, by **K2**, we have that

K2$''$ For each $U \subset A_1 \cup A_2$ and $U \subset B_1 \cup B_2$ with $|U| = m$,

$$|N_G(U, Z'')| \geqslant (1/2 + \varepsilon/16)|Z| - \ell > |Z''|/2.$$

Note that, R' is a collection of paths in $G_1 \cup G_2$, where each path contains exactly one vertex in X, which is furthermore one of its end-points. The other end-point of such a path has an in- or out-neighbour in Z'' in R. Similarly, R'' is a collection of paths in $H_1 \cup H_2$, where each path contains exactly one vertex in Y, which is one of its end-points. The other end-point of such a path has an in- or out-neighbour in Z'' in R.

Now R has the start vertex y_1 and contains the edge $\overrightarrow{y_1x_1}$, so y_1 is not in R''. Similarly, R has the end vertex x_2 and contains the edge $\overrightarrow{y_2x_2}$, so x_2 is not in R. Therefore x_2 and y_1 have degree 0 in $I_X + R'$ and $I_Y + R''$, respectively.

Let $k = \lceil \log(2m)/\log(d_0 - 1) \rceil$, so that $m/(d_0 - 1)^k \leqslant 1$. By Claim 9, **L2** and **L4**, and as $d_0 m = o(a)$, for sufficiently large n, by Lemma 4.6 applied with $X' = \{x_2\}$ and $Y' = \{y_1\}$, there are vertices $x'' \in A_1 \cup A_2$ and $y'' \in B_1 \cup B_2$ and paths $P'' \subset G_1 \cup G_2$ and $Q'' \subset H_1 \cup H_2$ such that the following hold.

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

N1 P'' is an x_2, x''-path with length at most k and no vertices in $V(R) \setminus \{x_2\}$, and for which $(I_X + R' + P'', X)$ is (d_0, m)-extendable in (G_1, G_2).
N2 Q'' is a y_1, y''-path with length at most k and no vertices in $V(R) \setminus \{y_1\}$, and $(I_Y + R'' + Q'', Y)$ is (d_0, m)-extendable in (H_1, H_2).
N3 $N_G(x'') \cap N_G(y'') \cap Z' \neq \emptyset$.

Note that, by the definition of (d_0, m)-extendability, the ith edge of P'', counting from x_2, is in $G_{2-(i \bmod 2)}$. Thus, by the choice of the graphs G_1 and G_2, P'' is a directed x_2, x''-path in D. Similarly, Q'' is, when reversed, a directed y'', y_1-path in D. Using **N3**, and noting that, by the choice of G,

$$N_G(x'') \cap N_G(y'') \cap Z'' = N_D^+(x'') \cap N_D^-(y'') \cap Z'',$$

select a vertex $z' \in Z''$ such that $P''z'\overleftarrow{Q''}$ is a directed x_2y_1-path in $D[A \cup V]$.

Therefore $RP''z'\overleftarrow{Q''}$ is a directed cycle in $D[A \cup V] + E$ containing each edge in E, as required. □

5. Covering vertices with directed paths

To cover most of the vertices with few directed paths in a pseudorandom digraph, we divide the vertex set into random sets using the local lemma and then find matchings between the sets. It would be nice to do this in one application of Lemma 3.2, but this is not possible. Essentially, this would attempt to track the in- and out-degrees of all vertices into each random set, which is too much for our methods. To see this, consider the following. We wish to cover at least $n/2$ vertices in a pseudorandom digraph D using paths of length $k = \Theta(\log n \cdot \log^{[5]} n / \log^{[2]} n)$, and therefore we take sets with size $\Theta(n/k)$ and find matchings between them. Looking at **D3**, to apply Lemma 3.2 to $V(D)$ in D to get a partition including a set of size $\Theta(n/k)$, we need to have $\exp(\Theta(\delta(D)/k)) \geqslant (\Delta(D))^2$. However, if D is (d, ε)-pseudorandom with $d = 1$ (as we must consider), so that $\delta(D) = \Theta(\log n)$ and $\Delta(D) = \Theta(\log n)$, this corresponds to $\exp(\Theta(\log^{[2]} n / \log^{[5]} n)) \geqslant \Theta(\log^2 n)$. Thus we cannot apply Lemma 3.2 in this manner. There is also a very similar issue that the number of sets in the partition will be too large for **D3** if we take k sets of such a size.

However, these calculations are not far from working. Considering still $d = 1$, the calculations above suggest we could use Lemma 3.2 to split the digraph into sets with size $\Theta(n \log^{[2]} n/k)$, as then $\exp(\Theta(\delta(D) \cdot \log^{[2]} n/k)) \geqslant (\Delta(D))^2$ can hold. By that lemma, this would give sets whose induced digraph in D has maximum degree at most $\Theta(\Delta(D) \cdot \log^{[2]} n/k) = O(\log^{[2]} n)$ and minimum degree at least $\Theta(\delta(D) \cdot \log^{[2]} n/k) = \Theta((\log^{[2]} n)^2 / \log^{[5]} n)$. If we then apply Lemma 3.2 to such an induced digraph, some D' say, to divide into $\log^{[2]} n$ different sets, **D3** requires that $\exp(\Theta(\delta(D')/\log^{[2]} n)) \geqslant (\Delta(D'))^2$, which now (rather easily) does hold. Doing this to the digraph induced on each set from the first application of Lemma 3.2 then gives k sets. Dividing into two stages also solves the similar issue with the number of sets in the partition being too large.

This double application lies behind our proof of Lemma 5.1, though the proof is (only slightly) more complicated, as for each vertex we need to track its in- and out-degree into the sets we want to match it with.

Lemma 5.1. *For each $\varepsilon > 0$, there exists $n_0 = n_0(\varepsilon)$ such that the following holds for every $d \geqslant 10^{-5}$ and $n \geqslant n_0$. Suppose an n-vertex (d, ε)-pseudorandom digraph D contains a set B with $|B| \geqslant n/2$ such that the following hold with $m = n \log^{[3]} n/d \log n$.*

O1 *For each $v \in V(D)$, $d^{\pm}(v, B) \geqslant d \log n/8$.*
O2 *For each $U \subset V(D)$ with $|U| = m$, $|N^{\pm}(U, B)| \geqslant (1/2 + \varepsilon/2)|B|$.*

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

Then there is a collection of at most $n\log^{[2]} n/(20\log n\cdot\log^{[5]} n)$ directed paths (with single vertices permitted) which partition B, so that any vertex $v\in V(D)$ has at most $d(\log^{[2]} n)^2$ in- or out-neighbours among their start and end vertices.

Proof. Let $\ell = n\log^{[2]} n/(50\log n\cdot\log^{[5]} n)$ and $k=\lfloor |B|/\ell\rfloor$. Noting that $k=\Theta(\log n\cdot\log^{[5]} n/\log^{[2]} n)$, take integers r and $k_1,\dots,k_r$ such that $\log^{[2]} n\leqslant k_i\leqslant 2\log^{[2]} n$ and $\sum_{i\in[r]} k_i = k$, and note that, for sufficiently large n, $r\leqslant \log n$.

We will now check that the conditions **D1**–**D4** hold for an application of Lemma 3.2 (using r in place of ℓ where it appears in that lemma, and $\varepsilon/2$ in place of ε) with $a_i = k_i\ell$ for each $2\leqslant i\leqslant r$ and $a_1 = |B| - a_2 - \cdots - a_r \geqslant k_1\ell$. Note that if these conditions hold for $k_1\ell$, then they also hold for $a_1\geqslant k_1\ell$.

Let $\Delta = 10^6 d\log n$, so that from the definition of a (d,ε)-pseudorandom digraph, we have $\Delta^{\pm}(D)\leqslant\Delta$. Let $\delta = d\log n/8$, so that, for each $v\in V(D)$, $d^{\pm}(v,B)\geqslant\delta$ by **O1**. For **D1**, note that, for each $i\in[r]$,

$$\left(\frac{\varepsilon}{2}\right)^2\cdot\frac{(k_i\ell)^2}{|B|} = \Omega_\varepsilon\left(\frac{n}{\log^2 n}\right) = \omega(r^3).$$

Now, for each $i\in[r]$, note that

$$\frac{\delta(k_i\ell)}{24|B|}\geqslant\frac{d\log n\cdot(k_i\ell)}{192n} = \Omega\left(\frac{dk_i\log^{[2]} n}{\log^{[5]} n}\right) = \omega\left(d\log^{[2]} n\right). \tag{5.1}$$

For **D2**, then, for each $i\in[r]$ and sufficiently large n, we have

$$\left(\frac{\varepsilon}{2}\right)^2\cdot\frac{(k_i\ell)^2}{|B|}\cdot\exp\left(\frac{\delta(k_i\ell)}{24|B|}\right)\overset{(5.1)}{\geqslant}\frac{\varepsilon^2\ell^2}{4n}\cdot\exp\left(\omega(d\log^{[2]} n)\right) = \Omega_\varepsilon\left(\frac{n}{\log^2 n}\cdot\log^6 n\right) = \omega(nr^3).$$

Furthermore, for **D3**, for each $i\in[r]$, as $r\leqslant\log n$ and $\Delta = 10^6 d\log n$, we have

$$\exp\left(\frac{\delta(k_i\ell)}{24|B|}\right)\overset{(5.1)}{=}\exp\left(\omega(d\log^{[2]} n)\right) = \omega(r\Delta^2).$$

Finally, as $\log(en/m) = O(\log(d\log n))$, we have, for each $i\in[r]$,

$$\left(\frac{\varepsilon}{2}\right)^2\cdot\frac{k_i\ell}{10^3} = \Omega_\varepsilon\left(\frac{n(\log^{[2]} n)^2}{\log n\cdot\log^{[5]} n}\right) = \Omega_\varepsilon\left(m\cdot\frac{d(\log^{[2]} n)^2}{\log^{[3]} n\cdot\log^{[5]} n}\right) = \omega\left(m\log\left(\frac{en}{m}\right)\right).$$

Having checked the appropriate conditions, for sufficiently large n, by Lemma 3.2, we can take disjoint sets $B_1,\dots,B_r$ in B so that

P1 For each $2\leqslant i\leqslant r$, $|B_i| = k_i\ell$, and $k_1\ell\leqslant|B_1| = |B| - a_2 - \cdots - a_r\leqslant(k_1+1)\ell$.

P2 For each $i\in[r]$ and $v\in V(D)$,

$$\frac{\delta k_i\ell}{4|B|}\leqslant d^{\pm}(v,B_i)\leqslant\frac{4\Delta(k_i+1)\ell}{|B|}\leqslant\frac{16\Delta k_i\ell}{n} = o(d(\log^{[2]} n)^2).$$

P3 For each $U\subset B$ with $|U| = m$, we have $|N^{\pm}(U,B_i)|\geqslant(1/2+\varepsilon/4)|B_i|$.

The directed paths we find will all start and end in $B_1\cup B_r$, so that, by **P2**, for sufficiently large n, every vertex in $V(D)$ has at most $d(\log^{[2]} n)^2$ in- and out-neighbours among these vertices.

Let $D_1 = D[B_1\cup B_2]$, $D_r = D[B_{r-1}\cup B_r]$, and, for each $2\leqslant i\leqslant r-1$, let $D_i = D[B_{i-1}\cup B_i\cup B_{i+1}]$. For each $i\in[r]$, let

$$n_i = |D_i|\leqslant 5k_i\ell+\ell\leqslant\frac{n(\log^{[2]} n)^2}{\log n}\leqslant m\cdot d(\log^{[2]} n)^2. \tag{5.2}$$

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

Let $\bar{\Delta} = d(\log^{[2]} n)^2/2$, so that, by **P2**, for sufficiently large n, $\Delta(D_i) \leqslant \bar{\Delta}$ for each $i \in [r]$.

Let $d_0 = d \log^{[2]} n / \log^{[4]} n$, so that $\exp(d_0) = \omega(k_i \bar{\Delta}^2)$ for each $i \in [r]$. Furthermore, for each $i \in [r]$, let $\delta_i = \delta k_i \ell / 4|B|$, so that, by **P2**, for each $v \in V(D_i)$,

$$d^{\pm}(v, B_i) \geqslant \frac{\delta k_i \ell}{4|B|} = \delta_i \geqslant \frac{(d \log n) \cdot k_i \cdot \ell}{32n} \geqslant \frac{d k_i \log^{[2]} n}{10^4 \log^{[5]} n} = \omega(k_i d_0). \tag{5.3}$$

We will now check the conditions **D1**–**D4** to apply Lemma 3.2 with $\varepsilon/4$ in place of ε to the set B_i in the digraph D_i, for each $i \in [r]$, to find, in B_i, k_i disjoint sets with size ℓ. Note that, for each $i \in [r]$, $|B_i| \leqslant k_i \ell + \ell$. First, for **D1**, we have, for each $i \in [r]$,

$$\left(\frac{\varepsilon}{4}\right)^2 \cdot \frac{\ell^2}{k_i \ell + \ell} = \Omega_\varepsilon\left(\frac{\ell}{k_i}\right) = \omega\left(\frac{n}{\log^2 n}\right) = \omega(k_i^3).$$

Secondly, for **D2**, we have, for each $i \in [r]$,

$$\begin{aligned}\left(\frac{\varepsilon}{4}\right)^2 \cdot \frac{\ell^2}{k_i \ell + \ell} \cdot \exp\left(\frac{\ell \delta_i}{24(k_i \ell + \ell)}\right) &= \Omega_\varepsilon\left(\frac{\ell}{k_i} \cdot \exp\left(\frac{\delta_i}{48 k_i}\right)\right) \\ &\overset{(5.3)}{=} \Omega_\varepsilon\left(\frac{n}{\log n \cdot \log^{[5]} n} \exp(d_0)\right) \\ &= \Omega_\varepsilon\left(\frac{n}{\log n} \cdot (\log^{[2]} n)^6\right) \\ &\overset{(5.2)}{=} \omega(n_i k_i^3).\end{aligned}$$

Furthermore, for **D3**, we have, for each $i \in [r]$,

$$\exp\left(\frac{\ell \delta_i}{24(k_i \ell + \ell)}\right) \overset{(5.3)}{=} \Omega(\exp(d_0)) = \omega(k_i \bar{\Delta}^2).$$

Finally, for **D4**, for each $i \in [r]$, by (5.2), we have $\log(e n_i/m) = O(\log(d \log^{[2]} n))$. Thus

$$\left(\frac{\varepsilon}{4}\right)^2 \cdot \frac{\ell}{10^3} = \Omega_\varepsilon\left(\frac{n \log^{[2]} n}{\log n \cdot \log^{[5]} n}\right) = \Omega_\varepsilon\left(m \cdot \frac{d \log^{[2]} n}{\log^{[5]} n \cdot \log^{[3]} n}\right) = \omega\left(m \log\left(\frac{e n_i}{m}\right)\right).$$

Having checked the appropriate conditions, for sufficiently large n, by Lemma 3.2 applied to the set B_i in the digraph D_i, for each $i \in [r]$, we can find in B_i disjoint sets $B_{i,1}, \ldots, B_{i,k_i}$ so that the following hold.

Q1 For each $i \in [r]$ and $j \in [k_i]$, $|B_{i,j}| = \ell$.
Q2 For each $i \in [r]$, $j \in [k_i]$ and $v \in V(D_i)$, $d^{\pm}(v, B_{i,j}) \geqslant \delta_i \ell / 4(k_i \ell + \ell) \geqslant d_0$ (using (5.3)).
Q3 For each $i \in [r]$, $j \in [k_i]$ and $U \subset V(D_i)$ with $|U| = m$, we have

$$|N^{\pm}(U, B_{i,j})| \geqslant (1/2 + \varepsilon/8)|B_{i,j}|.$$

Note that, for $2 \leqslant i \leqslant r$, the sets $B_{i,1}, \ldots, B_{i,k_i}$ partition B_i, and the sets $B_{1,1}, \ldots, B_{1,k_1}$ cover all but at most ℓ vertices in B_1.

Recall that $\sum_{i \in [r]} k_i = k$. Relabelling the sets $B_{1,1}, \ldots, B_{1,k_1}, B_{2,1}, \ldots, B_{2,k_2}, \ldots, B_{r,1}, \ldots, B_{r,k_r}$ as $C_1, \ldots, C_k$ respectively, from **Q1**–**Q3**, we have that $|C_i| = \ell$ for each $i \in [k]$ and the following hold.

R1 For each $1 \leqslant i \leqslant k-1$ and $v \in C_i$, $d^+(v, C_{i+1}) \geqslant d_0$.
R2 For each $2 \leqslant i \leqslant k$ and $v \in C_i$, $d^-(v, C_{i-1}) \geqslant d_0$.

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

R3 For each $1 \leqslant i \leqslant k-1$ and $U \subset C_i$ with $m \leqslant |U| \leqslant \lceil \ell/2 \rceil$, we have
$$|N^+(U, C_{i+1})| \geqslant (1/2 + \varepsilon/8)\ell \geqslant |U|.$$
R4 For each $2 \leqslant i \leqslant k$ and $U \subset C_i$ with $m \leqslant |U| \leqslant \lceil \ell/2 \rceil$, we have
$$|N^-(U, C_{i-1})| \geqslant (1/2 + \varepsilon/8)\ell \geqslant |U|.$$

By **R1**, **R2** and **A2** in the definition of (d, ε)-pseudorandomness, the following hold.

R1′ For each $1 \leqslant i \leqslant r-1$ and $U \subset C_i$ with $|U| \leqslant m$, $|N^+(U, C_{i+1})| \geqslant |U|$.
R2′ For each $2 \leqslant i \leqslant r$ and $U \subset C_i$ with $|U| \leqslant m$, $|N^-(U, C_{i-1})| \geqslant |U|$.

Thus, for each $i \in [r-1]$, by **R1′**, **R2′**, **R3**, **R4** and Proposition 2.9 applied to the bipartite graph between C_i and C_{i+1} with (undirected) edges those directed from C_i to C_{i+1} in D, there is a matching from C_i into C_{i+1} in D. Combining such matchings gives ℓ vertex-disjoint paths covering $C_1 \cup \cdots \cup C_k$. These paths start in B_1 and end in B_r, and cover all the vertices in B except for $|B_1| - k_1\ell \leqslant \ell$ vertices in B_1. Taking these paths with the uncovered vertices in B_1, to get at most 2ℓ paths, thus gives the required partition of B. □

We wish to cover most of our digraph with few paths, all of which end in a certain subset B_2, in order to have **B2** in the definition of a good partition. To do this, we use Lemma 5.1 to cover the vertices outside of B_2 with few paths, and then use Lemma 3.2 in the same way as before to find directed paths covering B_2. By matching the end vertices of the paths outside B_2 into some of the start vertices of the paths in B_2, and similarly attaching the start vertices of the paths outside B_2 into some of the end vertices of the paths in B_2, we will get a set of paths which start and end in B_2.

Lemma 5.2. *For each $\varepsilon > 0$, there exists $n_0 = n_0(\varepsilon)$ such that the following holds for every $d \geqslant 10^{-5}$ and $n \geqslant n_0$. Suppose an n-vertex (d, ε)-pseudorandom digraph D contains disjoint sets B_1 and B_2 such that the following hold with*
$$d_0 = \frac{d \log^{[2]} n}{\log^{[4]} n}, \quad \ell = \frac{n \log^{[2]} n}{\log n \cdot \log^{[5]} n} \quad \text{and} \quad k = \log^{[2]} n.$$

P1 *$|B_1| \geqslant n/2$ and $|B_2| = k\ell$.*
P2 *For each $v \in V(D)$, $d^{\pm}(v, B_1) \geqslant d \log n/8$ and, for each $U \subset V(D)$ with $|U| = m$,*
$$|N^{\pm}(U, B_1)| \geqslant (1/2 + \varepsilon/2)|B_1|.$$
P3 *For each $v \in V(D)$, $4d_0k \leqslant d^{\pm}(v, B_2) \leqslant d(\log^{[2]} n)^2$ and, for each $U \subset V(D)$ with $|U| = m$,*
$$|N^{\pm}(U, B_2)| \geqslant (1/2 + \varepsilon/4)|B_2|.$$

Then, for any set $V \subset V(D) \setminus (B_1 \cup B_2)$ with $|V| \leqslant \varepsilon|B_1|/20 - 2$ and $u, v \in V(D) \setminus (V \cup B_1 \cup B_2)$, there is a set of at most ℓ directed paths with length at least 1 which partition $V \cup B_1 \cup B_2 \cup \{u, v\}$, each start and end in B_2, and one of which contains $\overrightarrow{uv}$.

Proof. Let $B = V \cup B_1 \cup \{u, v\}$ and note that, by **P2**, as $|V \cup \{u, v\}| \leqslant \varepsilon|B|/20$, we have the following.

S1 For each $U \subset V(D)$ with $|U| = m$, we have
$$|N^{\pm}(U, B)| \geqslant (1/2 + \varepsilon/2)|B_1| \geqslant (1/2 + \varepsilon/4)|B|.$$

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

S2 For each $v \in V(D)$, we have $d^{\pm}(v, B) \geqslant d \log n/8$.

Thus, by **S1**, **S2**, **P1** and Lemma 5.1, there is a collection of directed paths $P_1, \ldots, P_r$, for some $r \leqslant \ell/20$, in D which partition B (allowing single vertices as paths) and so that every vertex in D has at most $d(\log^{[2]} n)^2$ in- and out-neighbours among their start and end vertices. Let P_{r+1} be the path with length 1 consisting of the edge $\overrightarrow{uv}$. For each $i \in [r+1]$, label vertices so that P_i is an x_i, y_i-path (noting that we may have $x_i = y_i$). Let $X = \{x_1, \ldots, x_{r+1}\}$ and $Y = \{y_1, \ldots, y_{r+1}\}$, and let D' be the digraph on the vertex set $B_2 \cup X \cup Y$ whose edges are the edges in $D[B_2 \cup X \cup Y]$ with at least one vertex in B_2.

Note that every vertex has at most $(\log^{[2]} n)^2 + 2$ in- or out-neighbours in $X \cup Y$. Combining this with **P3**, letting $\bar{\Delta} = d(\log^{[2]} n)^3$, we have, for sufficiently large n, that $\Delta^{\pm}(D') \leqslant \bar{\Delta}$. Let $k_1 = \log^{[2]} n$, so that, $\log^{[2]} n \leqslant k_1 \leqslant 2 \log^{[2]} n$, and, by definition of k, $k_1 = k$. Note that, for each $v \in V(D')$, $d^{\pm}(v, B_2) \geqslant 4k_1 d_0$ by **P3**. Let $n_1 = |D'|$, so that $n_1 \leqslant 2k_1\ell$. The conditions **D1**–**D4** for an application of Lemma 3.2 to D' to partition B_2 into k sets with size ℓ hold very similarly to the same conditions in the second application of this lemma in the proof of Lemma 5.1 (from (5.2) onwards); all that differs is a factor of 50 in the value of ℓ.

Thus, for sufficiently large n, by Lemma 3.2 there is a partition $C_1 \cup \cdots \cup C_k$ of B_2 such that $|C_i| = \ell$ for each $i \in [\ell]$, and the following hold.

T1 For each $i \in [\ell]$ and $U \subset V(D')$ with $|U| = m$, we have $|N^{\pm}(U, C_i)| \geqslant (1/2 + \varepsilon/8)|C_i|$.
T2 For each $v \in V(D')$ and $i \in [\ell]$, we have $d^{\pm}(v, C_i) \geqslant 4kd_0 \cdot \ell/4k\ell = d_0$.

Similarly to the reasoning in the proof of Lemma 5.1, for each $i \in [k-1]$, there is a matching from C_i into C_{i+1}. Combine these matchings to get directed paths Q_i, $i \in [\ell]$, which cover B_2. Similarly to the reasoning in the proof of Lemma 5.1, by **T1** and **T2**, for each $U \subset X$ we have $|N^-(U, C_k)| \geqslant |U|$, and for each $U \subset Y$ we have $|N^+(U, C_1)| \geqslant |U|$. Note that here it is important that $|X| \leqslant |C_k|/2$ and $|Y| \leqslant |C_1|/2$. Thus, as Hall's matching condition is satisfied, we can find vertex-disjoint edges e_i, $i \in [r+1]$, directed from Y into C_1, and vertex-disjoint edges f_i, $i \in [r+1]$ directed from C_k into X. Renaming if necessary, assume that, for each $i \in [r+1]$, $x_i \in f_i$ and $y_i \in e_i$.

Note that combining the paths Q_i, $i \in [\ell]$, and paths $P_i + e_i + f_i$, $i \in [r+1]$, gives a collection of at most ℓ directed paths and cycles in D which cover $V \cup B_1 \cup B_2 \cup \{u, v\}$ and so that each path starts and ends in B_2, each cycle contains some path Q_i, and as the edge $\overrightarrow{uv}$ is contained in P_{r+1}, it is contained in one path or cycle. Breaking an edge in some Q_i in each cycle then gives the required set of paths. □

6. Finding an (ℓ, r)-good partition

To find a good partition of a pseudorandom digraph, we first apply Lemma 3.2 twice to find the necessary sets for the good partition and record the properties we get, as follows.

Lemma 6.1. *For each $\varepsilon > 0$, there exists $n_0 = n_0(\varepsilon)$ such that the following holds for every $d \geqslant 10^{-5}$ and $n \geqslant n_0$ with*

$$k = \log^{[2]} n, \quad \ell = \frac{n \log^{[2]} n}{\log n \cdot \log^{[5]} n}, \quad d_0 = \frac{d \log^{[2]} n}{\log^{[4]} n}, \quad m = \frac{n \log^{[3]} n}{d \log n}, \quad r = \frac{n \log^{[2]} n}{\log n \cdot \log^{[6]} n}.$$

Every n-vertex (d, ε)-pseudorandom digraph D has a partition $V(D) = A \cup B_1 \cup B_2 \cup R_1 \cup R_2 \cup R_3 \cup R_4$ such that the following hold.

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

U1 $|A| = \varepsilon n/40$, $|B_2| = k\ell$ *and* $|R_1| = |R_2| = |R_3| = |R_4| = r$.

U2 *For each* $v \in V(D)$, $d^{\pm}(v, A) \geqslant d(\log n)^{3/4}$ *and, for each* $U \subset V(D)$ *with* $|U| = m$,

$$|N^{\pm}(U, A)| \geqslant (1/2 + \varepsilon/2)|A|.$$

U3 *For each* $v \in V(D)$, $d^{\pm}(v, B_1) \geqslant d \log n/8$ *and, for each* $U \subset V(D)$ *with* $|U| = m$,

$$|N^{\pm}(U, B_1)| \geqslant (1/2 + \varepsilon/2)|B_1|.$$

U4 *For each* $v \in V(D)$, $4d_0 k \leqslant d^{\pm}(v, B_2) \leqslant d(\log^{[2]} n)^2$ *and, for each* $U \subset V(D)$ *with* $|U| = m$,

$$|N^{\pm}(U, B_2)| \geqslant (1/2 + \varepsilon/4)|B_2|.$$

U5 *For each* $v \in B_2 \cup R_1 \cup R_2 \cup R_3 \cup R_4$ *and* $i \in [4]$, $40d_0 \leqslant d^{\pm}(v, R_i) \leqslant d(\log^{[2]} n)^2$ *and, for each* $U \subset B_2 \cup R_1 \cup R_2 \cup R_3 \cup R_4$ *with* $|U| = m$,

$$|N^{\pm}(U, R_i)| \geqslant (1/2 + \varepsilon/4)|R_i|.$$

Proof. We will apply Lemma 3.2 twice, again so that the second application may be with a stronger maximum degree condition.

We will check the conditions **D1**–**D4** for an application of Lemma 3.2 to find a partition $V(D) = A \cup B_1 \cup B_2' \cup B_3'$. Let

$$a_1 = \frac{\varepsilon n}{40}, \quad a_3 = (1 - \varepsilon/5)k\ell, \quad a_4 = \frac{\varepsilon k\ell}{5} + 4r \quad \text{and} \quad a_2 = n - a_1 - a_3 - a_4 \geqslant \frac{n}{2}, \tag{6.1}$$

where the last inequality follows for sufficiently large n. Let $\delta = d \log n$ and $\Delta = 10^6 d \log n$. From the definition of (d, ε)-pseudorandomness, we have that $\delta^{\pm}(D) \geqslant \delta$ and $\Delta^{\pm}(D) \leqslant \Delta$.

Note that $a_1, a_2, a_3 \geqslant a_4$, so that we need only check **D1**–**D4** for a_4. First, for **D1**, we have

$$\frac{\varepsilon^2 a_4^2}{n} = \Omega_\varepsilon\left(\frac{n}{\log^2 n}\right) = \omega(1). \tag{6.2}$$

Note that $k\ell\delta/n = \omega(d \log^{[2]} n)$. Thus, for **D2**, we have

$$\frac{\varepsilon^2 a_4^2}{n} \cdot \exp\left(\frac{a_4\delta}{24n}\right) \overset{(6.2)}{=} \Omega_\varepsilon\left(\frac{n}{\log^2 n} \cdot \exp\left(\frac{\varepsilon k\ell\delta}{10^3 n}\right)\right) = \frac{n}{\log^2 n} \cdot \exp\left(\omega(d \log^{[2]} n)\right) = \omega(n).$$

Furthermore, for **D3**, we have

$$\exp\left(\frac{a_4\delta}{24n}\right) \geqslant \exp\left(\frac{\varepsilon k\ell\delta}{10^3 n}\right) = \exp\left(\omega(d \log^{[2]} n)\right) = \omega(d^2 \log^2 n) = \omega(\Delta^2).$$

Finally, we have that $\log(en/m) = O(\log(d \log n))$, so that, for **D4**,

$$\frac{\varepsilon^2 a_4}{10^3} = \Omega_\varepsilon(k\ell) = \Omega_\varepsilon\left(\frac{n(\log^{[2]} n)^2}{\log n \cdot \log^{[5]} n}\right) = \Omega_\varepsilon\left(m \cdot \frac{d(\log^{[2]} n)^2}{\log^{[3]} n \cdot \log^{[5]} n}\right) = \omega\left(m \cdot \log\left(\frac{en}{m}\right)\right).$$

As $a_1 = \varepsilon n/40$, $a_1\delta/4n = \varepsilon d \log n/160 = \omega(d(\log n)^{3/4})$. As $a_2 \geqslant n/2$, $a_2\delta/4n \geqslant d \log n/8$.

Having checked the appropriate conditions, for sufficiently large n, by Lemma 3.2, $V(D)$ has a partition $A \cup B_1 \cup B_2' \cup B_3'$ so that $|A| = a_1 = \varepsilon n/40$, $|B_1| = a_2$, $|B_2'| = a_3 = (1 - \varepsilon/5)k\ell$, $|B_3'| = a_4 = \varepsilon k\ell/5 + 4r$, and **U2** and **U3** hold along with the following.

V1 For each $v \in V(D)$, $a_3\delta/4n \leqslant d^{\pm}(v, B_2') \leqslant 4a_3\Delta/n$, and, for each $U \subset V(D)$ with $|U| = m$,

$$|N^{\pm}(U, B_2')| \geqslant (1/2 + \varepsilon/2)|B_2'|.$$

V2 For each $v \in V(D)$, $a_4\delta/4n \leqslant d^{\pm}(v, B_3') \leqslant 4a_4\Delta/n$, and, for each $U \subset V(D)$ with $|U| = m$,

$$|N^{\pm}(U, B_3')| \geqslant (1/2 + \varepsilon/2)|B_3'|.$$

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

Let $D' = D[B_2' \cup B_3']$. Let $n' = |D'| = a_3 + a_4 = k\ell + 4r \leqslant 2k\ell$, for sufficiently large n, and let $\bar{\Delta} = d(\log^{[2]} n)^2$. Note that

$$\frac{4(a_3+a_4)\Delta}{n} \leqslant \frac{8k\ell\Delta}{n} = O\left(\frac{d(\log^{[2]} n)^2}{\log^{[5]} n}\right) = o(\bar{\Delta}). \tag{6.3}$$

Thus, by **V1** and **V2**, for sufficiently large n we have $\Delta(D') \leqslant \bar{\Delta}$. Using (6.1), let

$$\delta' = \frac{a_4\delta}{4n} \geqslant \frac{\varepsilon k\ell\delta}{20n} = \Omega_\varepsilon\left(k\ell \cdot \frac{d\log^{[2]} n}{r\log^{[6]} n}\right) = \Omega_\varepsilon\left(\frac{k\ell}{r} \cdot \frac{d_0\log^{[4]} n}{\log^{[6]} n}\right) = \omega\left(\frac{k\ell}{r} \cdot d_0\right). \tag{6.4}$$

By **V2**, for each $v \in V(D')$, we have $d^{\pm}(v, B_3') \geqslant \delta'$.

We will now check the conditions **D1**–**D4** to apply Lemma 3.2 to D' to get 4 sets, R_1, R_2, R_3 and R_4 in B_3', each with size r. First, for **D1**, as $n' \leqslant 2k\ell$, we have

$$\left(\frac{\varepsilon}{2}\right)^2 \cdot \frac{r^2}{n'} = \Omega_\varepsilon\left(\frac{n}{\log n}\right) = \omega(1).$$

Next, for **D2**, we have, as $n' \leqslant 2k\ell$,

$$\begin{aligned}\left(\frac{\varepsilon}{2}\right)^2 \cdot \frac{r^2}{n'} \cdot \exp\left(\frac{r\delta'}{24n'}\right) &= \Omega_\varepsilon\left(n' \cdot \left(\frac{r}{k\ell}\right)^2 \cdot \exp\left(\frac{r\delta'}{48k\ell}\right)\right)\\ &\overset{(6.4)}{=} \Omega_\varepsilon\left(n' \cdot \left(\frac{1}{\log^{[2]} n}\right)^2 \cdot \exp\left(\omega(d_0)\right)\right)\\ &= \omega(n').\end{aligned}$$

Furthermore, for **D3**, we have

$$\exp\left(\frac{r\delta'}{24n'}\right) \geqslant \exp\left(\frac{r\delta'}{24k\ell}\right) \overset{(6.4)}{=} \exp\left(\omega(d_0)\right) = \omega(\bar{\Delta}^2).$$

Finally, we have that $\log(en'/m) \leqslant \log(2ek\ell/m) = O(\log(d\log^{[2]} n))$, so that

$$\left(\frac{\varepsilon}{2}\right)^2 \cdot \frac{r}{10^3} = \Omega_\varepsilon\left(\frac{n\log^{[2]} n}{\log n \cdot \log^{[6]} n}\right) = \Omega_\varepsilon\left(m \cdot \frac{d\log^{[2]} n}{\log^{[3]} n \cdot \log^{[6]} n}\right) = \omega\left(m \cdot \log\left(\frac{en'}{m}\right)\right).$$

Having checked the appropriate conditions, by Lemma 3.2 and (6.4), for sufficiently large n, there are disjoint sets R_1, R_2, R_3, R_4 in B_3' so that the following holds.

V3 For each $v \in B_2' \cup B_3'$ and $i \in [4]$, $d^{\pm}(v, R_i) \geqslant 40d_0$, and, for each $U \subset B_2' \cup B_3'$ with $|U| = m$,

$$|N^{\pm}(U, R_i)| \geqslant (1/2 + \varepsilon/4)|R_i|.$$

Let $B_2 = (B_2' \cup B_3') \setminus (R_1 \cup R_2 \cup R_3 \cup R_4)$, so that $|B_2| = a_3 + a_3 - 4r = k\ell$. Note that we have chosen our set sizes so that **U1** holds. Note further that $B_2' \subset B_2$, so that, by **V1**, for each $U \subset V(D)$ with $|U| = m$,

$$|N^{\pm}(U, B_2)| \geqslant (1/2 + \varepsilon/2)|B_2'| = (1/2 + \varepsilon/2)(1 - \varepsilon/5)k\ell \geqslant (1/2 + \varepsilon/4)|B_2|. \tag{6.5}$$

By **V1**, **V2**, (6.3) and (6.4), and for each $v \in V(D)$ we have, for sufficiently large n, $4d_0k \leqslant d^{\pm}(v, B_2) \leqslant \bar{\Delta}$. Therefore, in combination with (6.5), we have that **U4** holds.

Note that, for each $v \in B_2 \cup R_1 \cup R_2 \cup R_3 \cup R_4 = V(D')$ and $i \in [4]$, we have $d^{\pm}(v, R_i) \leqslant \bar{\Delta} = d(\log^{[2]} n)^2$. Therefore, with **V3**, we have that **U5** holds. This completes the proof that **U1**–**U5** hold, so we have found the partition as required. □

We now combine the work in the last few sections to find good partitions.

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

Lemma 6.2. *For each $\varepsilon > 0$, there exists $n_0 = n_0(\varepsilon)$ such that the following holds for every $d \geqslant 10^{-5}$. Every (d, ε)-pseudorandom digraph D with at least n_0 vertices has a good partition.*

Proof. Let

$$k = \log^{[2]} n, \quad \ell = \frac{n \log^{[2]} n}{\log n \cdot \log^{[5]} n}, \quad d_0 = \frac{d \log^{[2]} n}{\log^{[4]} n}, \quad m = \frac{n \log^{[3]} n}{d \log n}, \quad r = \frac{n \log^{[2]} n}{\log n \cdot \log^{[6]} n}.$$

Let D be an n-vertex (d, ε)-pseudorandom digraph. Letting n be sufficiently large, by Lemma 6.1 we can find a partition $V(D) = A \cup B_1 \cup B_2 \cup R_1 \cup R_2 \cup R_3 \cup R_4$ such that **U1**–**U5** hold. By Theorem 4.7, **U1** and **U2**, and observing that

$$r = O\left(\frac{|A| \cdot \log^{[2]} n}{\log n \cdot \log^{[6]} n}\right) = o\left(\frac{|A| \cdot \log^{[2]} n}{\log n \cdot \log^{[7]} n}\right),$$

we have, for sufficiently large n, that **B1** holds in Definition 2.6. By Lemma 5.2, **U3** and **U4**, and as $|A \cup R_1 \cup R_2 \cup R_3 \cup R_4| \leqslant \varepsilon n/40 + 4r$, for sufficiently large n we have that **B2** holds in Definition 2.6.

Using **U5** and **A2**, by the simple reasoning at the end of the proof of Lemma 5.1, the conditions in Proposition 2.9 hold for the edges directed from R_i into R_j for any $j \neq i$. Therefore we can find matchings M_1, M_2 and M_3 from R_2 into R_1, R_2 into R_3 and R_4 into R_3 in D, respectively.

Let $f\colon R_1 \to R_4$ come from the matchings M_1, M_2 and M_3, and suppose each vertex $v \in R_1$ is merged into $f(v)$ in D to get the digraph D' (as in **B4**). Let R be the set of merged vertices in D'. For each $v \in R$, let $v^- \in R_1$ and $v^+ \in R_4$ be such that v^- is merged into v^+ to create v. For each $U \subset V(D')$, let

$$U^- = (U \setminus R) \cup \{v^- \colon v \in U \cap R\} \quad \text{and} \quad U^+ = (U \setminus R) \cup \{v^+ \colon v \in U \cap R\}.$$

By **U4** and **U5**, we have that the following hold.

W1 For each $v \in R \cup B_2$, we have $d^+_{D'}(v, R) = d^+_D(v^+, R_1) \geqslant 40 d_0$ and $d^-_{D'}(v, R) = d^-_D(v^-, R_4) \geqslant 40 d_0$.

W2 For each $v \in R \cup B_2$, we have

$$d^+_{D'}(v, R \cup B_2) = d^+_D(v^+, R_1 \cup B_2) \leqslant 2d(\log^{[2]} n)^2,$$
$$d^-_{D'}(v, R \cup B_2) = d^-_D(v^-, R_4 \cup B_2) \leqslant 2d(\log^{[2]} n)^2.$$

Thus $\Delta^{\pm}(D'[R \cup B_2]) \leqslant 2d(\log^{[2]} n)^2$.

W3 For each $U \subset R \cup B_2$, with $|U| = m$, we have

$$|N^+_{D'}(U, R)| \geqslant |N^+_D(U^+, R_1)| \geqslant (1/2 + \varepsilon/8)r,$$
$$|N^-_{D'}(U, R)| \geqslant |N^-_D(U^-, R_4)| \geqslant (1/2 + \varepsilon/8)r.$$

Note that

$$\ell = \frac{r \log^{[6]} n}{\log^{[5]} n} = o\left(\frac{r}{\log^{[6]} n}\right).$$

Thus, for sufficiently large n, for any $V \subset B_2$ with $|V| \leqslant 2\ell$, by Theorem 4.8, **W1**, **W2** and **W3**, V is weakly connected in $D'[R \cup V]$. Therefore **B3** and **B4** hold in Definition 2.6.

Therefore **B1**–**B4** hold for the partition $V(D) = A \cup B_1 \cup B_2 \cup R_1 \cup R_2 \cup R_3 \cup R_4$, and thus D has an (ℓ, r)-good partition. □

Theorem 2.2 now follows immediately from Lemmas 2.7 and 6.2.

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

7. Pseudorandomness of random digraphs

In this section we study the pseudorandom properties of digraphs in the random digraph process, allowing us then to apply Theorem 2.2 to prove Theorem 1.3. This section is organized as follows. First, in Section 7.1, we give some simple results on maximum and minimum in- and out-degree, to later show that **A1** resiliently holds in Definition 2.1. Next, in Section 7.2, with **A4** in mind, we give a simple result concerning the edges between sets. Then, in Section 7.3, we prove a result showing that expansion will follow from minimum degree conditions (Lemma 7.6), which will allow us to show that **A2** and **A3** hold. In Section 7.4 we then study the vertices with low in- and out-degree in the digraphs early in the random digraph process. After recording together all the properties we use, in Section 7.5 we then prove the resilience of Hamiltonicity in the random digraph process needed for Theorem 1.3. Finally, in Section 7.6, we study the limits of the resilience of Hamiltonicity to complete the proof of Theorem 1.3.

It will often be convenient to show that properties are likely in $D(n,p)$, before inferring that these properties are also likely in the random digraph process. Let $D_{n,M}$ be the random digraph with n vertices and M edges, chosen uniformly at random from all such digraphs. Note that, in the n-vertex random digraph process $D_0, D_1, \ldots, D_{n(n-1)}$, for each $0 \leqslant M \leqslant n(n-1)$, D_M is distributed as $D_{n,M}$. We use the following standard proposition to relate $D_{n,M}$ and $D(n,p)$ (see *e.g.* [3, 26]).

Proposition 7.1. *Let $\mathcal{P}$ be a digraph property and let $p = M/n(n-1)$. If $M = M(n) \to \infty$ is any function such that $M(1-p) \to \infty$, then, for sufficiently large n,*

$$\mathbb{P}(D_{n,M} \text{ satisfies } \mathcal{P}) \leqslant 5\sqrt{M} \cdot \mathbb{P}(D(n,p) \text{ satisfies } \mathcal{P}).$$

7.1 Maximum and minimum degree

We will use the following standard result, which implies that, almost surely, each digraph D_M in the n-vertex random digraph process with $M \leqslant n(\log n - \log^{[2]} n)$ is not Hamiltonian.

Lemma 7.2. see [10]. *If $M = n(\log n - \log^{[2]} n)$, then $D = D_{n,M}$ almost surely satisfies $\delta^+(D) = 0$ or $\delta^-(D) = 0$.*

When $M \geqslant 50n \log n$, each digraph D_M in the random digraph process will likely have well-bounded minimum and maximum degrees, as follows.

Lemma 7.3. *In almost every n-vertex random digraph process $D_0, D_1, \ldots, D_{n(n-1)}$, if $M \geqslant 50n \log n$, then $\delta^{\pm}(D_M) \geqslant M/2n$ and $\Delta^{\pm}(D_M) \leqslant 2M/n$.*

Proof. For each $M \geqslant 50n \log n$, let $p_M = M/n(n-1)$ and $\bar{D}_M = D(n, p_M)$. For each $v \in V(D)$ and $j \in \{+,-\}$,

$$\mathbb{E}(d^j_{\bar{D}_M}(v)) = (n-1)p_M = M/n \geqslant 50 \log n,$$

so that, by Lemma 2.8,

$$\mathbb{P}(d^j_{\bar{D}_M}(v) < M/2n \text{ or } d^j_{\bar{D}_M}(v) > 2M/n) \leqslant 2\exp(-50 \log n/12) = o(n^{-4}).$$

Therefore, by a union bound, with probability $1 - o(n^{-3})$, $\delta^{\pm}(\bar{D}_M) \geqslant M/2n$ and $\Delta^{\pm}(\bar{D}_M) \leqslant 2M/n$.

Now, for $50n \log n \leqslant M \leqslant n(n-1) - \log n$, by Proposition 7.1, with probability $1 - o(n^{-2})$, $\delta^{\pm}(D_M) \geqslant M/2n$ and $\Delta^{\pm}(D_M) \geqslant 2M/n$. Thus, by a union bound, this property almost surely holds for each $50n \log n \leqslant M \leqslant n(n-1) - \log n$ in the random digraph process. Finally, note

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

that, for each $M \geqslant n(n-1) - \log n$, $\delta^{\pm}(D_M) \geqslant n - 1 - \log n \geqslant M/2n$, for sufficiently large n, and $\Delta^{\pm}(D_M) \leqslant n - 1 \leqslant 2M/n$. □

We also need a maximum in- and out-degree condition earlier in the random digraph process, as follows.

Lemma 7.4. *For each $n \log n/2 \leqslant M \leqslant 50n \log n$, if $D = D_{n,M}$, then $\mathbb{P}(\Delta^{\pm}(D) \leqslant 100M/n) = 1 - o(n^{-2})$. Furthermore, then, in almost every n-vertex random digraph process $D_0, D_1, \ldots, D_{n(n-1)}$, every digraph D_M with $M \geqslant n \log n/2$ has $\Delta^{\pm}(D) \leqslant 100M/n$.*

Proof. The required bounds on the maximum in- and out-degree almost surely hold for each $M \geqslant 50n \log n$ by Lemma 7.3. For each $n \log n/2 \leqslant M \leqslant 50n \log n$, let $p_M = M/n(n-1)$ and $\bar{D}_M = D(n, p_M)$. For each $v \in V(D)$ and $j \in \{+, -\}$, we have

$$\mathbb{P}\left(d^j_{\bar{D}_M}(v) \geqslant \frac{100M}{n}\right) \leqslant \binom{n-1}{100M/n} p_M^{100M/n} \leqslant \left(\frac{enp_M}{100M/n}\right)^{100M/n} \leqslant \left(\frac{e}{50}\right)^{100M/n} = o(n^{-4}).$$

Therefore, with probability $1 - o(n^{-3})$, $\Delta^{\pm}(\bar{D}_M) \leqslant 100M/n$. Thus, by Proposition 7.1, for each $n \log n/2 \leqslant M \leqslant 50n \log n$, with probability $1 - o(n^{-2})$, $\Delta^{\pm}(D_{n,M}) \leqslant 100M/n$, as required. Finally, by a union bound, this property almost surely holds for each $n \log n/2 \leqslant M \leqslant 50n \log n$ in the random digraph process. □

7.2 Edges between sets

We will use the following simple proposition on the typical number of edges between sets in $D(n, p)$.

Proposition 7.5. *Let $\varepsilon > 0$, $p \geqslant 1/n$ and $m = \log^{[4]} n/p$. Then, with probability $1 - o(n^{-3})$ in $D(n, p)$, if sets $A, B \subset V(D_M)$ satisfy $|A| \geqslant m/2$ and $|B| \geqslant n/10$, then*

$$(1 - \varepsilon)p|A||B| \leqslant e^{\pm}(A, B) \leqslant (1 + \varepsilon)p|A||B|.$$

Proof. For each such A and B, and each $j \in \{0, 1\}$, $e^j(A, B)$ is a binomial random variable with expectation $p|A||B|$. (Note that, even when A and B are not disjoint, each edge is counted exactly once in $e^j(A, B)$.) Thus, by Lemma 2.8, we have

$$\begin{aligned}\mathbb{P}(|e^j(A, B) - p|A||B|| > \varepsilon p|A||B|) &\leqslant 2\exp(-\varepsilon^2 p|A||B|/3)\\ &\leqslant 2\exp(-\varepsilon^2 n \log^{[4]} n/60)\\ &= 2\exp(-\omega(n)).\end{aligned}$$

There are at most $2(2^n)^2$ choices for $j \in \{0, 1\}$ and such sets A and B. Thus, by a union bound, the property in the proposition holds with probability $1 - o(n^{-3})$. □

7.3 Expansion from minimum degree conditions

We now prove a lemma used to show both **A2** and **A3** in Definition 2.1. The proof follows a section of the proof by Alon, Krivelevich and Sudakov of Lemma 3.1 in [2].

Lemma 7.6. *Let*

$$p \geqslant \frac{\log n}{10n}, \quad d = \frac{p(n-1)}{2 \cdot 10^3 \log n} \geqslant 10^{-5}, \quad m = \frac{n \log^{[3]} n}{d \log n} \quad \text{and} \quad f(n) = o(\log^{[3]} n).$$

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

Then, with probability $1 - o(n^{-3})$, *in* $D = D(n,p)$, *for any two disjoint sets* $A, B \subset [n]$, *with* $|A| \leqslant 4m$, *and any integer* k *with* $1 \leqslant k \leqslant n/8m$, *and any* $j \in \{+,-\}$, *if* $e^j_D(A,B) \geqslant dk|A| \log^{[2]} n/f(n)$, *then* $|B| \geqslant k|A|$.

Proof. For each $k \in [n/8m]$, let $d_k = dk \log^{[2]} n/f(n)$. If D does not have the property in the lemma then there is some $k \leqslant n/8m$, $j \in \{+,-\}$ and two disjoint sets $A, B \subset V(D)$, where $|A| \leqslant 4m$, $|B| = k|A|$ and $e^j_D(A,B) \geqslant d_k|A|$ (adding vertices to B if necessary to get equality). For each $r \in [4m]$, let $p_{r,k}$ be the probability no two such sets occur with $|A| = r \leqslant 4m$ (noting that this does not depend on j). Then

$$\begin{aligned} p_{r,k} &\leqslant \binom{n}{r}\binom{n}{kr}\binom{kr^2}{d_k r} p^{d_k r} \\ &\leqslant \left(\frac{en}{r}\left(\frac{en}{kr}\right)^k \left(\frac{ekrp}{d_k}\right)^{d_k}\right)^r \\ &\leqslant \left(\left(\frac{en}{r}\right)^{2k}\left(\frac{ekrp}{d_k}\right)^{d_k}\right)^r \\ &= \left(\left(\frac{e^2knp}{d_k}\right)^{2k}\left(\frac{ekrp}{d_k}\right)^{d_k-2k}\right)^r. \end{aligned} \tag{7.1}$$

Now

$$\frac{e^2knp}{d_k} = \frac{e^2np \cdot f(n)}{d \log^{[2]} n} = O\left(\frac{\log n \cdot f(n)}{\log^{[2]} n}\right) = o(\log n). \tag{7.2}$$

Furthermore,

$$\frac{ekrp}{d_k} = \frac{erp \cdot f(n)}{d \log^{[2]} n} = O\left(\frac{r \log n \cdot f(n)}{n \log^{[2]} n}\right) = o\left(\frac{r \cdot \log^{[3]} n \cdot f(n)}{md \cdot \log^{[2]} n}\right) = o\left(\frac{r}{md \cdot (\log^{[2]} n)^{1/2}}\right). \tag{7.3}$$

For sufficiently large n, we have $d_k \geqslant 4k$. Therefore, by (7.1), (7.2) and (7.3), we have, for sufficiently large n,

$$p_{r,k} \leqslant \left(\log^{2k} n\left(\frac{r}{md \cdot (\log^{[2]} n)^{1/2}}\right)^{d_k/2}\right)^r = \left(\log^2 n\left(\frac{r}{md \cdot (\log^{[2]} n)^{1/2}}\right)^{d_k/2k}\right)^{kr}. \tag{7.4}$$

If $r < \sqrt{n}$, then $r/md = O(\log n/\sqrt{n})$, and hence, as $d_k = \omega(k)$ and $kr \geqslant 1$, $p_{r,k} = o(n^{-4})$.

If $r \geqslant \sqrt{n}$, then, as $r \leqslant 4m$, we have, for large n, by (7.4), that

$$p_{r,k} \leqslant \left(\log^2 n\left(\frac{4}{d \cdot (\log^{[2]} n)^{1/2}}\right)^{d \log^{[2]} n/2f(n)}\right)^{kr} \leqslant \left(\log^2 n \cdot \exp\left(-\frac{d \log^{[2]} n \cdot \log^{[3]} n}{8f(n)}\right)\right)^{kr}.$$

As $d = \Omega(1)$ and $f(n) = o(\log^{[3]} n)$, we have that, for sufficiently large n, $p_{r,k} \leqslant 2^{-kr} \leqslant 2^{-k\sqrt{n}} = o(n^{-4})$.

Therefore $2\sum_{r,k} p_{r,k} = o(m \cdot (n/m) \cdot n^{-4}) = o(n^{-3})$. Thus the probability for some j, r and k that such a pair A, B exists is $o(n^{-3})$. □

7.4 Low-degree vertices

We will treat vertices with low in-degree or low out-degree separately. We will use that, typically, no vertex will have both low in-degree and low out-degree in the digraphs we consider.

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

Proposition 7.7. *In almost every n-vertex random digraph process $D_0, D_1, \ldots, D_{n(n-1)}$, if $M \geqslant 9n \log n/10$, then, for all $v \in V(D_M)$, $d^+_{D_M}(v) + d^-_{D_M}(v) \geqslant 2M/10^3 n$.*

Proof. Note that in almost every random digraph process this property holds for $M \geqslant 50n \log n$ by Lemma 7.3, so we need only show that this almost surely holds for every $9n \log n/10 \leqslant M \leqslant 50n \log n$.

Let $p = 7 \log n/8n$, $D = D(n, p)$ and $d = \log n/10$. Note that $20 \leqslant e^3$. For each $v \in V(D)$, the probability that $d^+_D(v) + d^-_D(v) \leqslant d$ is at most

$$\begin{aligned}
\sum_{i=0}^{d} \binom{2n-2}{i} p^i (1-p)^{2n-2-i} &\leqslant \sum_{i=0}^{d} \left(\frac{2enp}{i}\right)^i e^{-p(2n-2-i)} \\
&\leqslant e^{-pn(2-o(1))} \cdot \sum_{i=0}^{d} \left(\frac{2e \log n}{i}\right)^i \\
&\leqslant (d+1) \cdot e^{-pn(2-o(1))} \cdot \left(\frac{2e \log n}{d}\right)^d \\
&\leqslant 2d \cdot e^{-(7/4-o(1)) \log n} \cdot (20e)^{\log n/10} \\
&\leqslant 2d \cdot e^{-(7/4-2/5-o(1)) \log n} \\
&= o(n^{-1}).
\end{aligned}$$

Thus, by a union bound, almost surely, for each $v \in V(D)$, $d^+_D(v) + d^-_D(v) \geqslant d$.

An easy application of Lemma 2.8 demonstrates that D almost surely has at most $9n \log n/10$ edges. Furthermore, the property – $\mathcal{P}$ say – that, for each $v \in V(D)$, $d^+_D(v) + d^-_D(v) \geqslant d$, is an increasing property. Thus we have, with $M_0 = 9n \log n/10$,

$$1 - o(1) = \mathbb{P}(D(n,p) \in \mathcal{P}) \leqslant \mathbb{P}(e(D(n,p)) > M_0) + \mathbb{P}(D_{n,M_0} \in \mathcal{P}) = o(1) + \mathbb{P}(D_{n,M_0} \in \mathcal{P}).$$

Hence we have $\mathbb{P}(D_{n,M_0} \in \mathcal{P}) = 1 - o(1)$. Therefore, almost surely, if $M_0 = 9n \log n/10 \leqslant M \leqslant 50n \log n$, then, for every $v \in V(D_M)$, $d^+_{D_M}(v) + d^-_{D_M}(v) \geqslant d = \log n/10 \geqslant 2M/10^3 n$. □

We will collect the vertices of low in- or out-degree in the random digraph into a set S. We use the following definition to record that there will typically be no vertices in S which are close together (in a graph-theoretic sense).

Definition 7.8. For a vertex set S in a digraph D, an *S-path* is a path with length at most 4 in D (with any orientation on the edges) starting and ending in S. An *S-cycle* is a cycle with length at most 4 in D (with any orientations on the edges, and a cycle with length 2 permitted if it consists of two distinct edges) which contains a vertex in S.

We wish to show that, in almost every n-vertex random digraph process, each digraph D_M with $9n \log n/10 \leqslant M \leqslant 50 \log n$ has no S-paths or S-cycles, when S is the set of vertices with low in- or out-degree, and S is a small set (see Lemma 7.11). To do this, we cannot show this is likely for each such digraph and then take a union bound, as the property is not sufficiently likely. Instead, we start by showing that this property is likely in a certain random digraph $D(n, p)$.

Lemma 7.9. *If $p = 7 \log n/8n$, then, almost surely, the following holds for $D = D(n, p)$ with*

$$S = \{v \in V(D) \colon d^+(v) < \log n/20 \text{ or } d^-(v) < \log n/20\}.$$

There are no S-paths or S-cycles and $|S| \leqslant n^{1/3}$.

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

Proof. Let $d = \log n/20$. Note that $20 \leqslant e^3$. We have

$$\begin{aligned}
\mathbb{E}|S| &\leqslant n \cdot \sum_{i=0}^{d} 2\binom{n-1}{i} p^i (1-p)^{n-1-i} \\
&\leqslant 2n(d+1) \cdot \left(\frac{enp}{d}\right)^d \cdot e^{-(1-o(1))np} \\
&\leqslant n \log n \cdot (20e)^d \cdot e^{-(7/8-o(1))\log n} \\
&\leqslant \log n \cdot e^{4d} \cdot e^{(1/8+o(1))\log n} \\
&= \log n \cdot e^{(1/5+1/8+o(1))\log n} \\
&= o(n^{1/3}).
\end{aligned}$$

Thus, by Markov's inequality, we almost surely have that $|S| \leqslant n^{1/3}$.

Let X be the number of S-paths in $D(n,p)$. Then

$$\begin{aligned}
\mathbb{E}X &\leqslant \binom{n}{2} \sum_{k=0}^{3} (2p)^{k+1} n^k \cdot \sum_{i=0}^{2d} 4\binom{2n}{i} p^i (1-p)^{2n-5-i} \\
&\leqslant n(2np)^4 \cdot (2d+1) \cdot 4\left(\frac{2enp}{2d}\right)^{2d} e^{-(2-o(1))np} \\
&= O(n \log^5 n \cdot (20e)^{2d} e^{-(2-o(1))np}) \\
&= O(n \log^5 n \cdot \exp(8d - (7/4 - o(1))\log n)) \\
&= O(n \log^5 n \cdot \exp(5 \log n/4)) \\
&= o(1).
\end{aligned}$$

Thus, almost surely, there are no S-paths in D.

Let Y be the number of S-cycles in $D(n,p)$. Then, similarly,

$$\begin{aligned}
\mathbb{E}Y &\leqslant n \cdot \sum_{k=1}^{3} (2p)^{k+1} n^k \cdot \sum_{i=0}^{d} 2\binom{n}{i} p^i (1-p)^{n-4-i} \\
&= O\left((np)^4 \cdot (d+1) \cdot \left(\frac{enp}{d}\right)^d \cdot e^{-(1-o(1))np}\right) \\
&= O(\log^5 n \cdot (20e)^d \cdot e^{-(7/8-o(1))\log n}) \\
&= O(\log^5 n \cdot \exp(4d - (7/8 - o(1))\log n)) \\
&= o(1).
\end{aligned}$$

Thus, almost surely, there are no S-cycles in D. □

Lemma 7.9 shows only that a property exists with probability $1 - o(1)$, so we cannot use Proposition 7.1 as we did before. However, we can easily show that the property holds for some useful digraph in the random digraph process, as follows.

Corollary 7.10. *There exists some $M_0 = M_0(n)$ with $n \log n/2 \leqslant M_0 \leqslant 9n \log n/10$ such that the following almost surely holds for $D = D_{n,M_0}$ with*

$$S = \{v \in V(D) \colon d^+(v) < \log n/20 \text{ or } d^-(v) > \log n/20\}.$$

There are no S-paths or S-cycles and $|S| \leqslant n^{1/3}$.

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

Proof. Let $p = 7\log n/8n$, $N_0 = n\log n/2$ and $N_1 = 9n\log n/10$. By a simple application of Lemma 2.8, we have that, almost surely, $N_0 \leqslant e(D(n,p)) \leqslant N_1$. Thus, if $\mathcal{P}$ is the property of digraphs satisfying the condition in the corollary, then

$$\begin{aligned}
\mathbb{P}(D(n,p) \in \mathcal{P}) &= \sum_{M=0}^{n(n-1)} \mathbb{P}(D_{n,M} \in \mathcal{P}) \cdot \mathbb{P}(e(D(n,p) = M) \\
&\leqslant \mathbb{P}(e(D(n,p)) \notin (N_0, \ldots, N_1)) + \sum_{M=N_0}^{N_1} \mathbb{P}(D_{n,M} \in \mathcal{P}) \cdot \mathbb{P}(e(D(n,p)) = M) \\
&\leqslant o(1) + \left(\sup_{N_0 \leqslant M \leqslant N_1} \mathbb{P}(D_{n,M} \in \mathcal{P}) \right) \cdot \sum_{M=N_0}^{N_1} \mathbb{P}(e(D(n,p)) = M) \\
&\leqslant o(1) + \sup_{N_0 \leqslant M \leqslant N_1} \mathbb{P}(D_{n,M} \in \mathcal{P}).
\end{aligned}$$

Thus, as $\mathbb{P}(D(n,p) \in \mathcal{P}) = 1 - o(1)$, we must have

$$\sup_{N_0 \leqslant M \leqslant N_1} \mathbb{P}(D_{n,M} \in \mathcal{P}) = 1 - o(1).$$

Choosing M_0 to maximize $\mathbb{P}(D_{n,M_0} \in \mathcal{P})$ subject to $N_0 \leqslant M_0 \leqslant N_1$ thus gives the result. □

Next, by starting with the digraph D_{n,M_0} from Corollary 7.10, we show that if $O(n\log n)$ random edges are added then it is likely that no short paths between the vertices with small in- or out-degree are created. This will give us the following lemma.

Lemma 7.11. *Almost surely, in the n-vertex random digraph process $D_0, D_1, \ldots, D_{n(n-1)}$, the following holds for each M with $9n\log n/10 \leqslant M \leqslant 50n\log n$. Letting*

$$S_M = \{v \in V(D) \colon d^+_{D_M}(v) < \log n/20 \text{ or } d^-_{D_M}(v) < \log n/20\},$$

there are no S_M-paths or S_M-cycles, and $|S_M| \leqslant n^{1/3}$.

Proof. Let M_0 be from Corollary 7.10. Let $N_0 = 9n\log n/10$ and $N_1 = 50n\log n$, and note that $M_0 \leqslant N_0$. Reveal the edges of D_{M_0} and let $S = S_{M_0}$, so that, almost surely, $|S| \leqslant n^{1/3}$ and there are no S-paths or S-cycles. Note that $S_M \subset S$ for each $M_0 \leqslant M \leqslant N_1$. Thus, if we can show that, almost surely, D_{N_1} has no S-paths or S-cycles, then we are done.

For each M, $M_0 \leqslant M < N_1$, let e_M be the edge added to D_M to get D_{M+1}. Let S'_M be the set of vertices within a graph distance 2 of any $u \in S_M$ in the underlying undirected graph of D_M. Let E_M be the event that e_M is contained within S'_M. Note that if no such event E_M, $M_0 \leqslant M < N_1$, occurs, then D_{N_1} has no S-paths or S-cycles.

Note that, for each M, $M_0 \leqslant M < N_1$, $|S'_M| \leqslant 2(\Delta^+(D_M) + \Delta^-(D_M))^2|S|$, and thus

$$\begin{aligned}
\mathbb{P}(E_M | (\Delta^{\pm}(D_M) \leqslant 10^4 \log n) \wedge (|S| \leqslant n^{1/3})) &= O\left(\frac{\log^4 n \cdot (n^{1/3})^2}{n(n-1) - M} \right) \\
&= O\left(\frac{n^{2/3} \log^4 n}{n^2} \right) \\
&= o\left(\frac{1}{n \log n} \right).
\end{aligned}$$

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

By Lemma 7.4, for each $M_0 \leqslant M < N_1$, $\mathbb{P}(\Delta^{\pm}(D_M) \leqslant 10^4 \log n) = 1 - o(n^{-2})$. Thus the probability that no event E_M, $M_0 \leqslant M < N_1$, occurs is at most

$$\mathbb{P}(|S| > n^{1/3}) + \sum_{M=M_0}^{N_1-1} (o(n^{-2}) + \mathbb{P}(E_M | (\Delta^{\pm}(D_M) \leqslant 10^4 \log n) \wedge (|S| \leqslant n^{1/3})) = o(1). \qquad \square$$

7.5 Resilience in the random digraph process

For convenience, we collect together the properties of the random digraph process that we have shown into the following corollary.

Corollary 7.12. *Let $\varepsilon > 0$. In almost every n-vertex random digraph process $D_0, D_1, \ldots, D_{n(n-1)}$, each digraph D_M with $\delta^{\pm}(D_M) \geqslant 1$ satisfies $M \geqslant n(\log n - \log^{[2]} n)$ and the following with*

$$d_M = \frac{M}{2 \cdot 10^3 n \log n}, \quad m_M = \frac{n \log^{[3]} n}{d_M \log n} \quad \text{and}$$
$$S_M = \{v \colon d^+_{D_M}(v) < 2d_M \log n \text{ or } d^-_{D_M}(v) < 2d_M \log n\}.$$

X1 $\Delta^{\pm}(D_M) \leqslant 100M/n \leqslant 10^6 d_M \log n/2$.
X2 *For each $v \in V(D_M)$, $d^+_{D_M}(v) + d^-_{D_M}(v) \geqslant 2M/10^3 n = 4d_M \log n$.*
X3 *There are no S_M-paths or S_M-cycles in D_M and $|S_M| \leqslant \sqrt{n}$.*
X4 *If sets $A, B \subset V(D_M)$ satisfy $|A| \geqslant m_M/2$ and $|B| \geqslant n/4$, then, for each $j \in \{+, -\}$, letting $p_M = M/n(n-1)$,*

$$(1 - \varepsilon/100)p_M|A||B| \leqslant e^j_{D_M}(A, B) \leqslant (1 + \varepsilon/100)p_M|A||B|.$$

X5 *For any disjoint sets $A, B \subset V(D_M)$ and $j \in \{+, -\}$ with $|A| \leqslant 4m_M$ and, for each $v \in A$, $d^j_{D_M}(v, B) \geqslant d_M \log^{[2]} n/4 \log^{[4]} n$, we have $|B| \geqslant 10|A|$.*
X6 *For any disjoint sets $A, B \subset V(D_M)$ with $|A| \leqslant 4m_M$ and $j \in \{+, -\}$, for each $v \in A$, $d^j_{D_M}(v, B) \geqslant d_M(\log n)^{2/3}/4$, we have $|B| \geqslant |A|(\log n)^{1/3}$.*

Proof. By Lemma 7.4, we almost surely have that $\Delta^{\pm}(D_M) \leqslant 100M/n$, and thus **X1** holds, for each $M \geqslant n \log n/2$. By Lemma 7.7, we almost surely have that **X2** holds for all $M \geqslant 9n \log n/10$. Almost surely, by Lemma 7.3, for each $M \geqslant 50n \log n$, $\delta^{\pm}(D_M) \geqslant M/2n \geqslant 10^3 d_M \log n$, so that $S_M = \emptyset$. Note that if $M \leqslant 50n \log n$ then $2d_M \log n \leqslant \log n/20$, and thus, combining this with Lemma 7.11, we have that, almost surely, **X3** holds for each $M \geqslant 9n \log n/10$. By Propositions 7.5 and 7.1, we have that **X4** almost surely holds for each $9n \log n/10 \leqslant M \leqslant n(n-1) - \log n$. Note that, when $M \geqslant n(n-1) - \log n$, there are at most $\log n$ missing edges, so **X4** easily holds for sufficiently large n.

Almost surely, by Lemma 7.6 with $f(n) = 40 \log^{[4]} n$ and taking the resulting property with $k = 10$, and using Proposition 7.1, **X5** holds for each $9n \log n/10 \leqslant M \leqslant n(n-1) - \log n$. Almost surely, by Lemma 7.6 with $f(n) = 4 \log^{[2]} n/(\log n)^{1/3}$ and taking the resulting property with $k = (\log n)^{1/3} = o(n/m_M)$, and using Proposition 7.1, **X6** holds for each $9n \log n/10 \leqslant M \leqslant n(n-1) - \log n$. Note that, if $M \geqslant n(n-1) - \log n$, then $d_M \log^{[2]} n/4 \log^{[4]} n \geqslant 40m_M$ and $d_M(\log n)^{2/3}/4 \geqslant 4m_M(\log n)^{1/3}$, so **X5** and **X6** hold.

Therefore **X1**–**X6** almost surely hold for each $M \geqslant 9n \log n/10$. By Lemma 7.2, almost surely, if $M \leqslant n(\log n - \log^{[2]} n)$, then $\delta^+(D_M) = 0$ or $\delta^-(D_M) = 0$. Thus, almost surely, if $\delta^{\pm}(D_M) \geqslant 1$, then $M \geqslant n(\log n - \log^{[2]} n)$ and **X1**–**X6** hold. $\square$

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

We now have the tools we need to prove the resilience in Theorem 1.3.

Proof of the resilience in Theorem 1.3. Almost surely, by Corollary 7.12, every digraph D_M in the random digraph process with $\delta^{\pm}(D_M) \geqslant 1$ satisfies $M \geqslant n(\log n - \log^{[2]} n)$ and **X1**–**X6** with

$$d_M = \frac{M}{2 \cdot 10^3 n \log n}, \quad m_M = \frac{n \log^{[3]} n}{d_M \log n} \quad \text{and}$$
$$S_M = \{v\colon d^+_{D_M}(v) < 2d_M \log n \text{ or } d^-_{D_M}(v) < 2d_M \log n\}.$$

We will show, for sufficiently large n, that each such D_M is $(1/2 - \varepsilon)$-resiliently Hamiltonian.

Fix $M \geqslant n(\log n - \log^{[2]} n)$ such that $\delta^{\pm}(D_M) \geqslant 1$, and suppose that $H \subset D_M$ with $d^j_H(v) \leqslant (1/2 - \varepsilon) d^j_{D_M}(v)$ for each $v \in V(D_M)$ and $j \in \{+, -\}$. Pick for each $v \in S_M$ an in-neighbour x_v and an out-neighbour y_v in $D_M - H$, noting that this is possible as $d^j_{D_M - H}(v) > d^j_{D_M}(v)/2 > 0$ for each $j \in \{+, -\}$. Note that, as there are no S_M-paths or S_M-cycles in D_M by **X3**, the vertices v, x_v and y_v, $v \in S_M$, are distinct. Form D from $D_M - H$ by deleting the vertices in S_M and, for each $v \in S_M$, merging x_v into y_v to get the new vertex z_v.

Claim 10. *For sufficiently large n, D is an* $(\varepsilon/100, d_M)$*-pseudorandom digraph.*

Proof. We will check the conditions **A1**–**A4**. Let $\bar{n} = |D| = n - 2|S_M| = (1 - o(1))n$, where we have used **X3**. Let $m = \bar{n} \log^{[3]} \bar{n}/d_M \log \bar{n}$, so that, for sufficiently large n, $m_M \geqslant m \geqslant m_M/2$. Note that, as there are no S_M-paths or S_M-cycles in D_M by **X3**, every vertex in $V(D_M) \setminus S_M$ has at most 1 in- or out-neighbour in $S_M \cup \{x_v, y_v\colon v \in S_M\}$ in D_M. Thus every vertex in D has in- and out- degree at least $(1/2 + \varepsilon) \cdot 2d_M \log n - 1 \geqslant d_M \log n \geqslant d_M \log \bar{n}$ in D, for sufficiently large n. Furthermore, by **X1**, $\Delta^{\pm}(D) \leqslant 10^6 d_M \log \bar{n}$, and therefore **A1** holds.

Now suppose $A, B \subset V(D)$ are disjoint sets with $|A| \leqslant 2m \leqslant 4m_M$ and, for each $v \in A$, $d^+_D(v, B) \geqslant d_M \log^{[2]} \bar{n}/\log^{[4]} \bar{n}$. Let $A' \subset V(D_M)$ be formed from A by replacing any vertex z_v, $v \in S_M$, with the vertex y_v, and let $B' \subset V(D_M)$ be formed from B by replacing any vertex z_v, $v \in S_M$, with the vertex x_v. Then, for each $v \in A'$, we have $d^+_{D_M}(v, B') \geqslant d_M \log^{[2]} n/4 \log^{[4]} n$, for sufficiently large n. Thus, using **X5**, we have $|B| = |B'| \geqslant 10|A'| = 10|A|$. Similar reasoning for $d^-_{D_M}(v, B')$, again using **X5**, completes the proof to show that **A2** holds. Similarly, **A3** follows from **X6**.

Therefore it is left only to show that **A4** holds. Suppose, for contradiction, there is some $U \subset V(D)$, with $|U| = m$, for which, without loss of generality $|N^+_D(U)| \leqslant (1/2 + \varepsilon/100)\bar{n}$. Let U' be the set U with any vertex z_v, $v \in S_M$, replaced by y_v, so that $|U'| = m$ and, using **X3**,

$$|N^+_{D_M - H}(U')| \leqslant |N^+_D(U)| + 2|S_M| \leqslant (1/2 + \varepsilon/100)\bar{n} + 2\sqrt{n}.$$

Let $V = V(D_M) \setminus (U' \cup N^+_{D_M - H}(U'))$, so that

$$|V| \geqslant n - m - (1/2 + \varepsilon/100)n - 2\sqrt{n} \geqslant (1/2 - \varepsilon/50)n, \tag{7.5}$$

for sufficiently large n. By **X4**, letting $p_M = M/n(n-1)$, we have

$$e^+_{D_M}(U', V) \geqslant (1 - \varepsilon/100)p_M|U'||V| \overset{(7.5)}{\geqslant} (1 - \varepsilon/100) \cdot (1/2 - \varepsilon/50) \cdot p_M|U'|n. \tag{7.6}$$

On the other hand, by the choice of V, we have $e^+_{D_M - H}(U', V) = 0$, so that

$$\begin{aligned} e^+_{D_M}(U', V) &= e^+_H(U', V) \\ &= \sum_{u \in U'} d^+_H(u, V) \end{aligned}$$

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

$$\begin{aligned}
&\leqslant \sum_{u\in U'} (1/2-\varepsilon)d^+_{D_M}(u)\\
&= (1/2-\varepsilon)e^+_{D_M}(U', V(D))\\
&\leqslant (1/2-\varepsilon)\cdot(1+\varepsilon/100)\cdot p_M|U'|n, \qquad (7.7)
\end{aligned}$$

where the last inequality follows by **X4**.

Thus we have, by (7.6) and (7.7),

$$(1-\varepsilon/100)\cdot(1/2-\varepsilon/50) \leqslant (1/2-\varepsilon)\cdot(1+\varepsilon/100),$$

a contradiction. Therefore **A4** holds, completing the proof that D is $(\varepsilon/100, d)$-pseudo-random. □

Thus, by Theorem 2.2, for sufficiently large n, D contains a directed Hamilton cycle, C say. For each vertex z_v, $v \in S_M$, in C, replace z_v by $x_v v y_v$. This gives a directed Hamilton cycle in $D_M - H$. Thus $D_M - H$ is Hamiltonian, and therefore D_M is $(1/2-\varepsilon)$-resiliently Hamiltonian. □

7.6 Non-resilience in the random digraph process

To prove the non-resilience of our random digraph, we divide the vertices without low in- or out-degree into two sets independently at random with equal probability. By Lemma 3.2, there must be such a partition where each vertex has roughly an equal number of in- and out-neighbours in each set. Carefully dividing the low-degree vertices between these sets, we reach a bipartition where deleting the edges across this partition does not remove substantially more than one-half of the in- and out-neighbours around any one vertex. This gives us a limit for the resilience of Hamiltonicity in the digraph.

Proof of the non-resilience in Theorem 1.3. Almost surely, by Corollary 7.12, every digraph D_M in the random digraph process with $\delta^{\pm}(D_M) \geqslant 1$ satisfies $M \geqslant n(\log n - \log^{[2]} n)$ and **X1**–**X6** with

$$d_M = \frac{M}{2\cdot 10^3 n \log n}, \quad m_M = \frac{n \log^{[3]} n}{d_M \log n} \quad \text{and}$$
$$S_M = \{v\colon d^+_{D_M}(v) < 2d_M \log n \text{ or } d^-_{D_M}(v) < 2d_M \log n\}.$$

Take the vertices in $V(D_M) \setminus S_M$ and partition them as $A \cup B$ so that each vertex is placed into A or B independently at random with probability 1/2. We will show that the following claim holds for sufficiently large n.

Claim 11. *With positive probability, the following holds.*

Y *For each $v \in V(D_M)$ and $j \in \{+,-\}$, if $d^j_{D_M}(v) \geqslant 2d_M \log n$, then $d^j_{D_M}(v, A), d^j_{D_M}(v, B) \geqslant (1/2-\varepsilon)d^j_{D_M}(v)$.*

Proof of Claim 11. For each $v \in V(D_M)$ and $j \in \{+,-\}$, let $E_{v,j}$ be the event that, if $d^j_{D_M}(v) \geqslant 2d_M \log n$, then either $d^j_{D_M}(v, A) < (1/2-\varepsilon)d^j_{D_M}(v)$ or $d^j_{D_M}(v, B) < (1/2-\varepsilon)d^j_{D_M}(v)$. Let G be the dependence graph of these events and note that, by **X1**, $\Delta(G) \leqslant (\Delta^+(D_M) + \Delta^-(D_M))^2 \leqslant 10^5 M^2/n^2$.

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

Let $q = 4\exp(-\varepsilon^2 d_M \log n/100)$, so that $q \cdot \Delta(G) = o(1)$. By **X3**, $d^j_{D_M - S_M}(v) \geqslant d^j_{D_M}(v) - 1 \geqslant 2d_M \log n - 1 \geqslant d_M \log n$. By Lemma 2.8, we then have

$$\begin{aligned}\mathbb{P}(E_{v,j}) &\leqslant 2\exp\left(-\frac{\varepsilon^2}{12} \cdot \frac{d^j_{D_M - S_M}(v)}{2}\right)\\ &\leqslant 2\exp\left(-\frac{\varepsilon^2 d_M \log n}{100}\right)\\ &= q/2\\ &\leqslant q(1 - q \cdot \Delta(G))\\ &\leqslant q(1-q)^{\Delta(G)}.\end{aligned}$$

Thus, by Theorem 3.1, no event $E_{v,j}$ occurs with positive probability. □

Thus, by Claim 11, there is some partition $V(D_M) \setminus S_M = A \cup B$ such that **Y** holds. Note that $A, B \neq \emptyset$. For each $v \in S_M$, by **X2**, there is j_v such that $d^{j_v}_{D_M}(v) \geqslant 2d_M \log n$, and thus, by **Y**, $d^{j_v}_{D_M}(v, A), d^{j_v}_{D_M}(v, B) \geqslant (1/2 - \varepsilon) d^{j_v}_{D_M}(v)$. Let $i_v \in \{+, -\}$ such that $i_v \neq j_v$. By **X3**, as $v \in S_M$, there are no edges from v to S_M in either direction. Thus there is some $X_v \in \{A, B\}$ such that $d^{i_v}_{D_M}(v, X_v) \geqslant d^{i_v}_{D_M}(v)/2$.

Let $A' = A \cup \{v \in S_M : X_v = A\}$ and $B' = B \cup \{v \in S_M : X_v = B\}$, and note that, for each $X \in \{A', B'\}$, $v \in X$ and $j \in \{+, -\}$, $d^j_{D_M}(v, X) \geqslant (1/2 - \varepsilon) d^j_{D_M}(v)$. Let H be the bipartite digraph with vertex classes A' and B', with edges exactly those edges in D_M between A' and B' in either direction. Thus we have, for each $v \in V(D_M)$ and $j \in \{+, -\}$, $d^j_H(v) \leqslant (1/2 + \varepsilon) d^j_{D_M}(v)$.

As $A, B \neq \emptyset$, $D_M - H$ is disconnected and hence not Hamiltonian. Therefore D_M is not $(1/2 + \varepsilon)$-resiliently Hamiltonian. □

References

[1] Ajtai, M., Komlós, J. and Szemerédi, E. (1985) First occurrence of Hamilton cycles in random graphs. *North-Holland Math. Studies* **115** 173–178.

[2] Alon, N., Krivelevich, M. and Sudakov, B. (2007) Embedding nearly-spanning bounded degree trees. *Combinatorica* **27** 629–644.

[3] Bal, D., Bennett, P., Cooper, C., Frieze, A. and Prałat, P. (2016) Rainbow arborescence in random digraphs. *J. Graph Theory* **83** 251–265.

[4] Balogh, J., Csaba, B. and Samotij, W. (2011) Local resilience of almost spanning trees in random graphs. *Random Struct. Algorithms* **38** 121–139.

[5] Ben-Shimon, S., Krivelevich, M. and Sudakov, B. (2011) On the resilience of Hamiltonicity and optimal packing of Hamilton cycles in random graphs. *SIAM J. Discrete Math.* **25** 1176–1193.

[6] Bollobás, B. (1983) Almost all regular graphs are Hamiltonian. *European J. Combin.* **4** 97–106.

[7] Bollobás, B. (1984) The evolution of sparse graphs. In *Graph Theory and Combinatorics: Cambridge Combinatorial Conference in Honour of Paul Erdős*, pp. 335–357, Academic Press.

[8] Dirac, G. (1952) Some theorems on abstract graphs. *Proc. London Math. Soc.* **3** 69–81.

[9] Erdős, P. and Rényi, A. (1959) On random graphs I. *Publ. Math. Debrecen* **6** 290–297.

[10] Erdős, P. and Rényi, A. (1964) On random matrices. *Magyar Tud. Akad. Mat. Kutató Int. Közl* **8** 455–461.

[11] Ferber, A., Nenadov, R., Noever, A., Peter, U. and Škorić, N. (2017) Robust Hamiltonicity of random directed graphs. *J. Combin. Theory Ser. B* **126** 1–23.

[12] Frieze, A. (1988) An algorithm for finding Hamilton cycles in random directed graphs. *J. Algorithms* **9** 181–204.

[13] Frieze, A. and Krivelevich, M. (2008) On two Hamilton cycle problems in random graphs. *Israel J. Math.* **166** 221–234.

[14] Ghouila-Houri, A. (1960) Une condition suffisante d'existence d'un circuit Hamiltonien. *CR Acad. Sci. Paris* **251** 495–497.

[15] Glebov, R. (2013) On Hamilton cycles and other spanning structures. PhD thesis.

[16] Hefetz, D., Krivelevich, M. and Szabó, T. (2012) Sharp threshold for the appearance of certain spanning trees in random graphs. *Random Struct. Algorithms* **41** 391–412.

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press

[17] Hefetz, D., Steger, A. and Sudakov, B. (2016) Random directed graphs are robustly Hamiltonian. *Random Struct. Algorithms* **49** 345–362.
[18] Janson, S., Łuczak, T. and Ruciński, A. (2011) *Random Graphs*, Wiley.
[19] Komlós, J. and Szemerédi, E. (1983) Limit distribution for the existence of Hamiltonian cycles in a random graph. *Discrete Math.* **43** 55–63.
[20] Korshunov, A. (1976) Solution of a problem of Erdős and Rényi on Hamilton cycles in non-oriented graphs. *Soviet Math. Dokl.* **17** 760–764.
[21] Krivelevich, M., Lee, C. and Sudakov, B. (2010) Resilient pancyclicity of random and pseudorandom graphs. *SIAM J. Discrete Math.* **24** 1–16.
[22] Lee, C. and Sudakov, B. (2012) Dirac's theorem for random graphs. *Random Struct. Algorithms* **41** 293–305.
[23] McDiarmid, C. (1983) General first-passage percolation. *Adv. Appl. Probab.* **15** 149–161.
[24] Montgomery, R. (2019) Hamiltonicity in random graphs is born resilient. *J. Comb. Theory Ser. B.* **139** 316–341.
[25] Nenadov, R., Steger, A. and Trujić, M. Resilience of perfect matchings and Hamiltonicity in random graph processes. *Random Struct. Algorithms*, **54** 797–819.
[26] Pittel, B. (1982) On the probable behaviour of some algorithms for finding the stability number of a graph. *Math. Proc. Cambridge Philos. Soc.* **92** 511–526.
[27] Pósa, L. (1976) Hamiltonian circuits in random graphs. *Discrete Math.* **14** 359–364.
[28] Rödl, V., Szemerédi, E. and Ruciński, A. (2008) An approximate Dirac-type theorem for k-uniform hypergraphs. *Combinatorica* **28** 229–260.
[29] Spencer, J. (1977) Asymptotic lower bounds for Ramsey functions. *Discrete Math.* **20** 69–76.
[30] Sudakov, B. (2017) Robustness of graph properties. In *Surveys in Combinatorics 2017*, Vol. 440 of London Mathematical Society Lecture Note Series, pp. 372–408, Cambridge University Press.
[31] Sudakov, B. and Vu, V. (2008) Local resilience of graphs. *Random Struct. Algorithms* **33** 409–433.

Cite this article: Montgomery R (2020). Hamiltonicity in random directed graphs is born resilient. *Combinatorics, Probability and Computing* **29**, 900–942. https://doi.org/10.1017/S0963548320000140

https://doi.org/10.1017/S0963548320000140 Published online by Cambridge University Press