Combinatorics, Probability and Computing (2026), **35**, pp. 486–523
doi:10.1017/S0963548326100418

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Trees and treelike structures in dense digraphs

Richard Mycroft[1] and Tássio Naia[2]

[1]University of Birmingham, UK and [2]Centre de Recerca Matemàtica, Barcelona, Spain
Corresponding author: Richard Mycroft; Email: r.mycroft@bham.ac.uk

(Received 17 December 2020; revised 23 December 2025; accepted 28 January 2026; first published online 4 June 2026)

Abstract

We prove that every oriented tree on n vertices with bounded maximum degree appears as a spanning subdigraph of every directed graph on n vertices with minimum semidegree at least $n/2 + o(n)$. This can be seen as a directed graph analogue of a well-known theorem of Komlós, Sárközy, and Szemerédi. Our result for trees follows from a more general result, allowing the embedding of arbitrary orientations of a much wider class of spanning 'tree-like' structures, such as collections of at most $O(n^{0.99})$ pairwise vertex-disjoint cycles and subdivisions of graphs H with $|H| < \exp\left(\sqrt{O(\log n)}\right)$ in which each edge is subdivided at least once.

Keywords: Trees; directed graphs

2020 MSC Codes: Primary: 05C20

1. Introduction

A celebrated theorem of Komlós, Sárközy, and Szemerédi [13] states that if G is a graph of order n with $\delta(G) \geq n/2 + o(n)$, then G contains every tree of order n with bounded maximum degree.

Theorem 1. [13] *For all $\Delta \in \mathbb{N}$ and $\alpha > 0$ there exists n_0 such that every graph G of order $n \geq n_0$ with $\delta(G) \geq \left(\frac{1}{2} + \alpha\right)n$ contains every tree T of order n with $\Delta(T) \leq \Delta$.*

Komlós, Sárközy, and Szemerédi later strengthened Theorem 1, replacing the constant bound Δ by $cn/\log n$, where c is some constant depending on α [14]. Many variations and extensions of Theorem 1 have been investigated, e.g., [2, 4–6, 15]. We prove the following directed graph (digraph) analogue of Theorem 1, where minimum degree is replaced by minimum semidegree $\delta^0(\cdot)$ (the minimum of in- and outdegrees over all vertices) and the maximum degree is replaced by the maximum total degree $\Delta(\cdot)$ (maximum degree in the underlying tree).

Theorem 2. *For all $\Delta \in \mathbb{N}$ and $\alpha > 0$ there exists n_0 such that every digraph G of order $n \geq n_0$ with $\delta^0(G) \geq \left(\frac{1}{2} + \alpha\right)n$ contains every oriented tree T of order n with $\Delta(T) \leq \Delta$.*

In fact we prove a stronger result (Theorem 4), which allows the embedding of a large class of treelike graphs, and which implies both Theorem 2 and the more general embedding result for trees below. For this we define a bare path $P = p_1p_2 \cdots p_n$ in a (di)graph G to be a path whose internal vertices $p_2, \ldots, p_{n-1}$ each have degree 2 in (the underlying graph of) G.

*Richard Mycroft acknowledges support from EPSRC (Standard Grant EP/R034389/1).
**Tássio Naia acknowledges support from CNPq (201114/2014-3) and FAPESP (2019/04375-5).

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

Theorem 3. *Suppose* $\frac{1}{n} \ll \lambda \ll \alpha$. *If* G *is a digraph of order* n *with* $\delta^0(G) \geq (1/2+\alpha)n$, *then* G *contains every oriented tree* T *of order* n *with* $\Delta(T) \leq \exp(\sqrt{\log n})$ *such that* T *contains either*

(i) at least λn *pairwise vertex-disjoint bare paths of order* 7, *or*

(ii) at least λn *pairwise disjoint edges incident to leaves.*

Note that for all positive ε we have $\mathrm{polylog}(n) \asymp \exp\big(\Theta(\log\log n)\big) \lesssim \exp\big(\sqrt{O(\log n)}\big) \lesssim n^{\varepsilon}$. Our main result allows embedding families of sparse graphs which arise from an arbitrary small graph by numerous applications of the following operations:

(A) append a leaf (i.e., add a new vertex connected to the graph by a single edge);

(B) subdivide an edge (i.e., replace some edge uv by a path uxv, where x is a new vertex).

Throughout the text, vertices of degree one are called *leaves* (even in graphs other than trees), and $|G|$ denotes the order of the graph G.

Theorem 4. *Suppose* $\frac{1}{n} \ll \lambda \ll \alpha$. *Fix a graph* Q_0 *and let* Q *be a graph of order* n *obtained from* Q_0 *by a sequence of operations (A) and (B) in which each edge of* Q_0 *is subdivided at least once. Suppose additionally that* $|Q_0| \leq n^{0.99}$ *and* $\Delta(Q) \leq \exp(\sqrt{\log n})$, *and let* G *be a digraph with* $\delta^0(G) \geq (1/2+\alpha)|G|$.

(1) If $|G| \geq (1+\alpha)n$, *then* G *contains every orientation of* Q.

(2) If $|G| = n$ *and* Q *contains either* λn *pairwise vertex-disjoint bare paths of order* 7 *or* λn *pairwise disjoint edges incident to leaves, then* G *contains every orientation of* Q.

Theorem 4 can be used to embed a wide range of spanning treelike subdigraphs in a digraph of high minimum semidegree. For example, it implies that every digraph of order n with minimum semidegree at least $n/2 + o(n)$ contains every orientation of a Hamilton cycle. This gives an asymptotic version of recent results by DeBiasio and Molla [8] and by DeBiasio, Kühn, Molla, Osthus, and Taylor [7], which can be stated jointly as the following theorem (the statement for directed cycles had previously been obtained by Ghouila-Houri [10]).

Theorem 5. [7, 8] *There exists* $n_0 \in \mathbb{N}$ *such that the following holds for every digraph* G *of order* $n \geq n_0$.

(i) If $\delta^0(G) \geq n/2+1$, *then* G *contains every orientation of a Hamilton cycle.*

(ii) If $\delta^0(G) \geq n/2$, *then* G *contains every orientation of a Hamilton cycle, except perhaps for the anti-directed orientation in which each vertex has either no inneighbours or no outneighbours.*

In the same way we can embed every orientation of a disjoint union of $O(n^{0.99})$ cycles, a result which may be of independent interest.

Corollary 6. *For all* $\alpha > 0$ *there exists* n_0 *such that the following holds for every digraph* G *of order* $n \geq n_0$ *with* $\delta^0(G) \geq (1/2+\alpha)n$. *If* H *is a graph of order at most* n *consisting of at most* $\frac{1}{4}n^{0.99}$ *pairwise vertex-disjoint cycles, then* G *contains every orientation of* H.

Proof. Cycles of length at most 5 cover at most $5n^{0.99} < \alpha n/2$ vertices of H, so we may embed all such cycles greedily, whereupon the subdigraph G' induced by the n' uncovered vertices satisfies $\delta^0(G) \geq (1/2+\alpha/2)n'$. The remaining cycles each have length at least 6 and so are subdivisions of triangles where each edge is subdivided at least once. We may therefore apply Theorem 4(2) with $\alpha/2$ and n' in place of α and n respectively to embed these cycles in G', completing the embedding of H in G. □

We also consider embeddings of random trees. Moon [18] showed that a uniformly-random labelled n-vertex tree T has sub-polylogarithmic maximum degree with high probability. It is not

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

difficult to check that with high probability T also satisfies condition (ii) of Theorem 3 (see, e.g., [19]). Together these observations imply the following corollary, for which we denote by $\mathcal{T}_n$ the set of oriented trees with vertex set $[n]$.

Corollary 7. *Fix $\alpha > 0$. If T is chosen uniformly at random from $\mathcal{T}_n$, then with high probability we have $T \subseteq G$ for every digraph G of order n with $\delta^0(G) \geq (1/2 + \alpha)n$.*

While this manuscript was under review, Kathapurkar and Montgomery [12] announced a stronger version of Theorem 2, in which the constant bound on the maximum degree $\Delta(T)$ is replaced by a best-possible bound of $\Delta(T) \leq cn/\log n$. This impressive breakthrough uses very different methods to those used in this paper, which cannot hope to succeed for trees with maximum degree even close to that size. However, the methods Kathapurkar and Montgomery used appear quite specific to trees, whereas our approach allows a much wider class of 'treelike graphs' to be handled similarly, as in Theorem 4 and its applications. For this reason we believe the methods and results of this paper should still be of widespread use and interest even in the light of this new advance.

In the next section of this paper we outline the key ideas used to prove our main results. Section 3 then introduces the main definitions and results that we combine to give the full proofs in Section 4.

2. Key proof ideas

Very broadly speaking, we use the following approach to prove Theorem 4, where our aim is to embed a treelike oriented graph Q into a directed graph G with high minimum semidegree. First, we *allocate* each vertex of Q to a cluster of a reduced graph R of G obtained by an application of the Szemerédi regularity lemma for digraphs; this allocation should respect the directions of edges in G and Q by having the property that if x is an outneighbour of y in Q, then the cluster V_x to which x is allocated should be an outneighbour in R of the cluster V_y to which y is allocated (in other words, the allocation of vertices to clusters should be a homomorphism from Q to R). Having done this, we then *embed* each vertex of Q within the cluster to which it was allocated so as to form a copy of Q in G. This two-step process of first allocating vertices to clusters of the reduced graph, then embedding within the clusters, was previously used by Kühn, Mycroft, and Osthus [16] and subsequently further developed by Mycroft and Naia [19], in both cases to embed trees within tournaments. For our present application a more significant development of this approach is required, with many new ideas, to reflect the setting of a digraph of high minimum semidegree and the fact that the graph to be embedded may no longer be a tree (in particular, the previous works relied heavily on the fact that each cluster within a tournament induced a subtournament, giving far greater freedom to embed vertices within a cluster; this is no longer possible in our more general setting as each cluster could be an independent set).

2.1. Sketch proof for non-spanning trees

Consider the case of Theorem 4(1) in which Q is a tree on n vertices and G is a directed graph on $(1+\alpha)n$ vertices with $\delta^0(G) \geq (1/2+\alpha)|G|$. Let T be an orientation of Q; our goal is then to embed T within G. In this case we may proceed as follows.

(i) Apply the regularity lemma to partition $V(G)$ into clusters $V_1, \dots, V_k$ of equal size, as well as a small set of exceptional vertices V_0 which can safely be ignored, so that almost all pairs of clusters form regular pairs in each direction. In particular each cluster has size $|V_i| \geq (1+\alpha/2)n/k$. Form a reduced graph $R^\star$ with vertex set $[k]$ in which $i \to j$ is an edge if and only if the graph $G[V_i \to V_j]$ of edges directed from V_i to V_j is regular and dense. The reduced graph $R^\star$ then inherits an analogous minimum semidegree condition from G, namely that $\delta^0(R^\star) \geq (1/2+\alpha/2)k$. The full definitions and details for this step are presented in a more general form in Section 3.3.

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

(ii) The minimum semidegree of $R^\star$ is enough to ensure that $R^\star$ contains a spanning subgraph R which is a d-regular expander, meaning that every proper nonempty subset $S \subseteq V(R)$ has $|N^+(S)|$, $|N^-(S)| > |S|$ and every vertex $i \in V(R)$ has both indegree and outdegree of precisely d. A more general version of this statement, along with a key mixing property of expander digraphs, is presented in Section 3.7.

(iii) *Allocation:* We now allocate the vertices of T to clusters using a randomised allocation algorithm, which builds a homomorphism $\varphi : T \to R$ one vertex at a time. First, choose a root r of T and define $\varphi(r) \in [k]$ arbitrarily. Next iterate the following step: choose a vertex $x \in V(T)$ for which $\varphi(x)$ has been defined, say $\varphi(x) = u_x$, but for which $\varphi(y)$ has not been defined for any child y of x. Choose vertices $u_x^+ \in N_R^+(u_x)$ and $u_x^- \in N_R^-(u_x)$ uniformly at random, and set $\varphi(y) = u_x^+$ for every child y of x which is an outneighbour of x, and $\varphi(y) = u_x^-$ for every child y of x which is an inneighbour of x. Proceed in this manner until $\varphi(x)$ is defined for every vertex of T, and observe that the resulting map φ is then a homomorphism $\varphi{:}T \to R$. Moreover, and crucially, so long as the maximum degree $\Delta(T)$ of T is not too large, the mixing property of expander digraphs ensures that with high probability φ allocates the vertices of T approximately uniformly to the clusters of R, with, say, for each $i \in [k]$ at most $(1 + \alpha/3)n/k$ vertices $x \in V(T)$ having $\varphi(x) = i$. The proof of this statement, for a more general version of this algorithm, is presented in Section 3.8.

(iv) *Embedding:* Finally we form a copy of T in G by greedily embedding each vertex x of T within the cluster $V_{\varphi(x)}$ of G to which it was allocated, starting with the root r. Each time we embed a vertex x in a cluster V_{u_x} we reserve sets of size $O(\sqrt{n})$ in the clusters $V_{u_x^-}$ and $V_{u_x^+}$ in which the children of x which are respectively inneighbours and outneighbours of x will be embedded; no other vertices may be embedded within these sets until all children of x have been embedded. In this way we avoid treading on our toes' by occupying all of the inneighbours or outneighbours of x whilst they are still required for children of x; moreover the fact that edges of R correspond to dense regular pairs in G is sufficient for us to successfully choose appropriate embeddings within the specified sets. For this process to succeed we need the order in which we proceed through the vertices of T to have the property that at any time there are not too many vertices which have been embedded but have a child vertex yet to be embedded; the notion of a *tidy ancestral order* presented in Section 3.1 captures what we need for this, whilst the way in which we choose the reserved sets is explained in Section 3.4.

A bound on $\Delta(T)$ is crucial for this argument to succeed, since we need the randomised allocation algorithm to distribute the vertices of T approximately uniformly among the k clusters (by contrast, the vertices of a star would be allocated in a highly-unbalanced way). To achieve this, we specifically need the distance between almost all pairs of vertices of T to be significantly larger than k, and indeed the bound on $\Delta(T)$ given in Theorem 4 is chosen to achieve precisely this property (see Sections 3.7 and 3.8 for more details).

2.2. Sketch proof for spanning trees

Let us now consider the case of Theorem 4(2) in which Q is a tree on n vertices and G is a directed graph on n vertices with $\delta^0(G) \geq (1/2 + \alpha)n$. Again, let T be an orientation of Q. The difference with the previous case is that our aim is now a spanning embedding of T in G; there is no 'room to spare'. However, Theorem 4(2) does provide a linear number of either bare paths on seven vertices or edges incident to leaves; for the sake of this discussion we assume the former (the arguments for the latter are quite similar and somewhat simpler).

We begin by splitting T into two not-too-small subtrees T_1 and T_2 with one vertex v in common, which we take as the root of both trees. Without loss of generality T_1 contains a set $\mathcal{P}$ of

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

bare paths on seven vertices such that $\mathcal{P}$ has small linear size; and moreover we may insist that all paths in $\mathcal{P}$ have the same 'pattern' (the sequence of directions of edges along the path). As before we apply the regularity lemma to obtain clusters $V_1, \dots, V_k$ of equal size and a small set V_0 of bad vertices, and similarly as there we may define a reduced graph $R^\star$ with vertex set $[k]$ whose vertices correspond to clusters, and also obtain a subgraph R of $R^\star$ on vertex set $[k]$ which is a d-regular expander. However, we now additionally insist that R contains a Hamilton cycle H (this is possible by the minimum degree condition that R inherits from G); without loss of generality we assume that H has edges $1 \to 2 \to 3 \to \cdots \to 1$. We also insist that for each edge $i \to j$ of H the pair $G[i \to j]$ is dense and superregular; this can be achieved by deleting a small number of vertices from each cluster and adding them to V_0.

Next we apply a modified version of the allocation algorithm presented above to T_1: the change is that if x is a non-initial vertex of a bare path in $\mathcal{P}$, and y is the parent of x, then we instead set $\varphi(x)$ to be $\varphi(y)+1$ if $y \in N^+(x)$ and $\varphi(y)-1$ if $y \in N^-(x)$. The effect is that internal edges of bare paths in $\mathcal{P}$ are allocated deterministically within the cycle H rather than randomly within the regular expander R. Despite this change, the allocation algorithm still allocates the vertices of T_1 approximately uniformly across $[k]$ with high probability; moreover, the initial vertices of bare paths in $\mathcal{P}$ are also allocated approximately uniformly across $[k]$ with high probability.

We now choose pairwise disjoint sets $\mathcal{P}^H, \mathcal{P}^0, \mathcal{P}^\diamond \subseteq \mathcal{P}$ such that $|\mathcal{P}^0| = |V^0|$, whilst $\mathcal{P}^H$ and $\mathcal{P}^\diamond$ each have small linear size. Each of these collections of bare paths will play a distinct role in completing the embedding of T in G, as follows.

(i) We reallocate the internal vertices of paths $P \in \mathcal{P}^0$ in such a way that the middle vertex of each path P can be embedded to a corresponding vertex of V_0. Note that, since in the allocation step we are forming a homomorphism $\varphi: V(T) \to R^\star$, for notational convenience we actually also include each bad vertex in V_0 as a vertex of $R^\star$, so $V(R^\star) = [k] \cup V_0$, and for each vertex $v \in V_0$ our allocation will have $\varphi(x) = v$ for precisely one vertex x, which will then be embedded to v at the embedding step (x is the middle vertex of a path $P \in \mathcal{P}^0$). In this way we ensure that all vertices of V_0 are covered by our embedding of T.

(ii) We reallocate the internal vertices of paths $P \in \mathcal{P}^\diamond$ to structures in R called 'diamonds' the purpose of which is that for each such path P we have two choices for the allocation of the middle vertex v_4^P of P, where each choice is consistent with the allocation of the neighbours v_3^P and v_5^P of v_4^P in P. This gives us the flexibility to reallocate the middle vertex v_4^P of each path $P \in \mathcal{P}^\diamond$ after all other vertices have been allocated, so as to slightly adjust the number of vertices allocated to each cluster. (Details of diamonds and this reallocation process can be found in Section 3.6.)

(iii) Finally, we leave the allocation of paths in $\mathcal{P}^H$ unchanged; the internal vertices of these paths will be the vertices we embed at the very end to 'finish off' our spanning embedding of T in G.

Unfortunately, it is unavoidable that the reallocation of vertices in paths in $\mathcal{P}^0$ and $\mathcal{P}^\diamond$ may create greater imbalances in the allocation of vertices of T_1 to clusters than can be corrected by reallocating the middle vertices of paths in $\mathcal{P}^\diamond$. This is the purpose of the tree T_2: after we have applied the randomised allocation algorithm to T_1, and subsequently reallocated vertices of paths in $\mathcal{P}^0$ and $\mathcal{P}^\diamond$, we now apply a biased version of the allocation algorithm to T_2, which with high probability allocates slightly more vertices to the clusters which were underoccupied by T_1, and slightly fewer vertices to the clusters which were overoccupied by T_1. The end result is that the overall allocation of vertices of T to clusters is very close to uniform; specifically it is close enough that it can be made uniform by reallocating the middle vertices of paths in $\mathcal{P}^\diamond$. That is, after this reallocation exactly the same number of vertices are allocated to each cluster, meaning that the allocation 'fits' perfectly.

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

Finally, we embed the vertices of T in G by the same embedding approach as described previously, with two modifications. Firstly, we take special care with the embeddings of paths $P \in \mathcal{P}^0$, whose middle vertices v_4^P must be embedded to the vertex $v \in V_0$ to which they were allocated; to do this we embed the remaining vertices of P so that the neighbours v_3^P and v_5^P of v_4^P are embedded to inneighbours/outneighbours (as appropriate) of the vertices in V_0. Secondly, we do not embed the internal vertices of paths $P \in \mathcal{P}^H$ at this stage, but leave these until the very end. This ensures that there is a little 'room to spare' when embedding all vertices of T other than those in paths in $\mathcal{P}^H$, so the previous approach of reserving sets for the children of each vertex as it is embedded is still valid. Finally, to complete the embedding we need to embed the internal vertices of the paths in $\mathcal{P}^H$ to the vertices still unoccupied within each cluster. Since each path in $\mathcal{P}^H$ is allocated along a sequence of superregular pairs (corresponding to edges of H), this can be done by applying the four-layer theorem of Komlós, Sárközy, and Szemerédi (see Section 3.3).

2.3. Sketch of the full proof

For the full version of the lemma, the graph Q that we seek to embed may not be a tree. However the conditions on Q do ensure that we can find a small 'ground set' $V_{\text{ground}} \subseteq V(Q)$ so that V_{ground} contains all vertices of Q_0 as well as all short paths between vertices of Q_0. In particular, $Q \setminus V_{\text{ground}}$ is a forest F. We embed the vertices of V_{ground} greedily in G, and then proceed similarly as in the previous case, dividing our forest F into two subforests F_1 and F_2 and allocating and embedding these similarly as we described for T_1 and T_2. Some additional difficulties arise here in doing this: for example, components of F may contain two vertices with neighbours in V_{ground}, and whilst one can serve as the root of the component, the other (which we call a secondary attachment) must be handled carefully to ensure that it is allocated and embedded appropriately given the embedding of its neighbour in V_{ground}. However, these difficulties can be overcome through careful modification to the allocation and embedding procedures (in particular, this is the purpose of the set Z of bad vertices in the general allocation algorithm in Section 3.8).

3. Auxiliary concepts and results

The following concepts and results play an important role in our proofs. We follow standard graph-theoretical notation (see, e.g., [9]). For clarity, we define some of our notation (mostly related to digraphs) below. More specific terms are defined in later sections.

A *directed graph* G, or *digraph* for short, is a pair $\big(V(G), E(G)\big)$ of sets: a vertex set $V(G)$ and an edge set $E(G)$, where each edge $e \in E(G)$ is an ordered pair of distinct vertices. The *order* of G, denoted $v(G)$ or $|G|$, is defined to be $|V(G)|$ and the *size* of G is $e(G) = |E(G)|$. We think of the edge (u, v) as being directed from u to v, and write $x \to y$ or $y \leftarrow x$ to denote the edge (x, y); if the orientation of the edge does not matter, we write $\{u, v\}$ (or $\{v, u\}$) instead. In either case, u and v are said to be the *endvertices* of $\{u, v\}$, and we also call u (respectively v) a *neighbour* of v (respectively u).

In a digraph G, the *outneighbourhood* $N_G^+(x)$ of a vertex x is the set $\{\, y : x \to y \in E(G) \,\}$; the *inneighbourhood* $N_G^-(x)$ of x is $\{\, y : x \leftarrow y \in E(G) \,\}$. The *outdegree* and *indegree* of x in G are respectively $\deg_G^+(x) := \big|N_G^+(x)\big|$ and $\deg_G^-(x) := \big|N_G^-(x)\big|$, and the *semidegree* $\deg_G^0(x)$ of x is the minimum of the outdegree and indegree of x. We say that G is *r-regular* if for all $x \in V(G)$ we have $\deg^-(x) = \deg^+(x) = r$. The *minimum semidegree* $\delta^0(G)$ of G is the minimum of $\deg_G^0(x)$ over all $x \in V(G)$. For any subset $Y \subseteq V(G)$, we write $\deg_G^-(x, Y)$ for $|N_G^-(x) \cap Y|$, the *indegree of x in Y*; the *outdegree of x in Y*, denoted by $\deg_G^+(x, Y)$, is defined similarly. The *semidegree of x in Y*, denoted by $\deg_G^0(x, Y)$, is the minimum of those two values. We drop the subscript when there is no danger of confusion, writing $N^-(x)$, $\deg^0(x)$, and so forth. We sometimes symbols such as $\bullet$ or

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

$\diamond$ as placeholder for either $-$ or $+$, as in 'if $x \in N^{\bullet}(y)$ and $y \in N^{\bullet}(x)$, then x and y form a cycle of length 2', in which both occurrences of $\bullet$ are meant to be either $-$ or $+$.

For digraphs G and H, we call H a *subgraph* of G if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$; H is said to be *spanning* if $V(H) = V(G)$. For any set $X \subseteq V(G)$, we write $G[X]$ for the subgraph of G *induced* by X, which has vertex set X and whose edges are all edges of G with both endvertices in X. If H is a subgraph of G then we write $G - H$ for $G\big[V(G) \setminus V(H)\big]$. Likewise, for a vertex v or set of vertices S, we write $G - v$ or $G - S$ for $G\big[V(G) \setminus \{v\}\big]$ or $G\big[V(G) \setminus S\big]$ respectively. For disjoint subsets $X, Y \subseteq V(G)$, where G is a digraph, we denote by $G[X \to Y]$, or equivalently by $G[Y \leftarrow X]$, the subdigraph of G with vertex set $X \cup Y$ and edge set $E\big(G[X \to Y]\big) := \{ x \to y \in E(G) : x \in X, y \in Y \}$.

An *oriented graph* is a digraph in which there is at most one edge between each pair of vertices. Equivalently, an oriented graph G can be formed by assigning an orientation to each edge $\{u, v\}$ of some (undirected) graph H, i.e. by replacing each $\{u, v\} \in E(H)$ by one of the possible ordered pairs (u, v) or (v, u). In this case we refer to H as the *underlying graph* of G, and say that G is an *orientation* of H. We refer to the *maximum degree* of an oriented graph G, denoted $\Delta(G)$, to mean the maximum degree of the underlying graph H.

A *directed path* of length k is an oriented graph with vertices $v_0, \dots, v_k$ and edges $v_{i-1} \to v_i$ for each $1 \leq i \leq k$. Likewise, a *directed cycle* of length k is an oriented graph with vertices $v_1, \dots, v_k$ and edges $v_i \to v_{i+1}$ for each $1 \leq i \leq k$ with addition taken modulo k.

A *tree* is an acyclic connected graph, and an *oriented tree* is an orientation of a tree. Where it is clear from the context that a tree is oriented, we may refer to it simply as a tree. A *leaf* in a graph or oriented graph G is a vertex $v \in V(G)$ which is incident to precisely one edge; v is an *in-leaf* if $\deg_T^+(v) = 1$ and an *out-leaf* otherwise (note that these definitions are only standard when G is a tree, but we define leaves, in-leaves and out-leaves for general graphs and digraphs as they are helpful concepts for working with treelike structures also). A *star* is a tree in which at most one vertex (the *centre*) is not a leaf. A *subtree* T' of a tree T is a subgraph of T which is also a tree, and we define subtrees of oriented trees similarly. A *forest* is a graph in which each component is a tree, and a *oriented forest* is an orientation of a forest, that is, a digraph in which each component is an oriented tree.

Given an integer d, a graph G is said to be *d-degenerate* if $\delta(H) \leq d$ for every subgraph $H \subseteq G$. Note that a graph is 1-degenerate if and only if it is a forest.

If H is an induced subgraph of a graph G, then we call $v \in V(H)$ an *attachment* of H if v has a neighbour in $V(G) \setminus V(H)$, and call v a *pendant vertex* otherwise. Attachments and pendant vertices in digraphs are defined similarly, considering the underlying graph.

Let $A_1, A_2, \dots$ be a sequence of events. We say that A_n holds *with high probability* if $\mathbb{P}(A_n) \to 1$ as $n \to \infty$. Likewise, all occurrences of the standard asymptotic notation $\mathrm{o}(f)$ refer to sequences $f(n)$ with parameter n as $n \to \infty$ (i.e., $g = \mathrm{o}(f)$ if $g(n)/f(n) \to 0$ as $n \to \infty$). We will often have sets indexed by $\{1, 2, \dots, k\}$, such as $V_1, \dots, V_k$, and addition of indices will always be performed modulo k. Also, if $\varphi \colon A \to B$ is a function from A to B and $A' \subseteq A$, then we write $\varphi(A')$ for the image of A' under φ. We omit floors and ceilings whenever they do not affect the argument, write $a = b \pm c$ to indicate that $b - c \leq a \leq b + c$, and write abc/def for the fraction $(abc)/(def)$. We write $\mathbb{N} := \{1, 2, \dots\}$ for the set of natural numbers. For each $k \in \mathbb{N}$ we denote by $[k]$ the set $\{1, 2, \dots, k\}$, and write $\binom{S}{k}$ to denote the set of all k-element subsets of a set S. For any two disjoint sets A and B, we write $A \dot\cup B$ for their union. We use the notation $x \ll y$ to indicate that for every positive y there exists a positive number x_0 such that for every $0 < x < x_0$ the subsequent statements hold. Such statements with more variables are defined similarly. We always write $\log x$ to mean the natural logarithm of x.

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

3.1. Trees and forests

Let T be a tree or oriented tree. It is often helpful to nominate a vertex r of T as the *root* of T; to emphasise this fact we sometimes refer to T as a *rooted tree*. If so, then every vertex x other than r has a unique *parent*; this is defined to be the (sole) neighbour p of x in the unique path in T from x to r, and x is said to be a *child* of p. In the same way we call w a *descendant* of v, and v an *ancestor* of w, if the path from r to w contains v. We denote the set of children of v by $C(v)$, and similarly write $C^-(v)$ for $N^-(v) \cap C(v)$ and $C^+(v)$ for $N^+(v) \cap C(v)$.

Now let F be a forest or oriented forest, so each component of F is a tree or oriented tree. If we choose a root vertex for each component tree, then the definitions of parent, child, descendant, and ancestor extend naturally to forests: we say that u is a child (respectively parent, descendant or ancestor) of v in F if u and v are in the same component T of F and u is a child (respectively parent, descendant or ancestor) of v in T.

An *ancestral order* of the vertices of a rooted tree T is an order of $V(T)$ in which every non-root vertex appears later than its parent; observe that this implies that the roof of T appears first in the order, and that each vertex appears prior to each of its descendants and later than each of its ancestors. In exactly the same way, given a forest F and a root vertex for each component tree in F, an *ancestral order* of the vertices of F is an order of $V(F)$ in which every non-root vertex appears later than its parent. Note that an ancestral order of the vertices of a tree (or forest) specifies the root(s) of the tree (or forest), since the root of each component must appear prior to every other vertex of that component. Consequently, when considering trees or forests equipped with ancestral orders we do not separately specify the roots of the tree or tree components.

We say that an ancestral order of a forest F is *tidy* if for any initial segment $\mathcal{I}$ of the order, at most $\log_2 |V(F)|$ vertices in $\mathcal{I}$ have a child not in $\mathcal{I}$. Such orders were considered by Kühn, Mycroft, and Osthus [16] for the purpose of embedding trees in tournaments; in particular they proved the following lemma for trees (the statement for forests follows immediately).

Lemma 8. [16, Lemma 2.11] *If F is a forest in which each component tree is rooted, then F admits a tidy ancestral order.*

Our proof strategy at various stages requires us to divide a tree into subtrees, which we capture by the following notion.

Definition 9. *Let T be a tree or oriented tree. A tree-partition of T is a collection $\{T_1, \dots, T_s\}$ of pairwise edge-disjoint subtrees of T such that $\bigcup_{i\in[s]} V(T_i) = V(T)$ and $\bigcup_{i\in[s]} E(T_i) = E(T)$.*

Note that distinct trees in a tree-partition $\mathcal{P}$ share at most one vertex; moreover, if $\mathcal{P}$ contains at least 2 trees, then each tree has at least one vertex in common with some other tree in $\mathcal{P}$. The following lemma shows that any tree admits a tree-partition into two subtrees which splits a given set of vertices somewhat evenly.

Lemma 10. [19, Lemma 5.7] *If T is a (possibly oriented) tree and $L \subseteq V(T)$, then T admits a tree-partition $\{T_1, T_2\}$ such that T_1 and T_2 each contain at least $|L|/3$ vertices of L.*

We will use the following straightforward corollary to obtain a small set of vertices which splits a forest into not-too-large parts.

Corollary 11. *Every forest F on n vertices admits a set $X \subseteq V(F)$ with $|X| \le 3n^{1/3}$ for which every component of $F - X$ has fewer than $n^{2/3}$ vertices.*

Proof. Let $\mathcal{T}_0$ be the set of components of F, so each $T \in \mathcal{T}_0$ is a subtree of F, and let $s := |\mathcal{T}_0|$. Also let $X_0 = \emptyset$. For each $i \ge 0$ in turn do the following: if $\mathcal{T}_i$ contains a tree T with $|T - X_i| \ge n^{2/3}$, then apply Lemma 10 to obtain a tree-partition $\{T', T''\}$ of T for which T' and T'' each contain at least $|T - X_i|/3 \ge n^{2/3}/3$ vertices of $V(T)\setminus X_i$, and let v_i be the unique vertex in $V(T') \cap V(T'')$. Set $\mathcal{T}_{i+1} := (\mathcal{T}_i \setminus \{T\}) \cup \{T', T''\}$, set $X_{i+1} := X_i \cup \{v_i\}$, and proceed to the next i. If instead $\mathcal{T}_i$ does

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

not contain a tree T with $|T - X_i| \geq n^{2/3}$ then we terminate the process, writing t for this terminal value of i, and setting $X := X_t$. So $\mathcal{T}_t$ does not contain a tree T with $|T - X| \geq n^{2/3}$; since each component of $F - X$ is a subtree of some tree in $\mathcal{T}_t$ it follows that each component of $F - X$ has fewer than $n^{2/3}$ vertices. Observe that our definition of $\mathcal{T}_{i+1}$ and X_{i+1} ensures that $|\mathcal{T}_{i+1}| = |\mathcal{T}_i| + 1$ and $|X_{i+1}| = |X_i| + 1$ for each $i \in [t]$, so by induction we have $|\mathcal{T}_t| = t + s$ and $|X| = t$. Now consider a tree $T \in \mathcal{T}_t$. Either T was one of the s trees in $\mathcal{T}_0$ or $T \in \mathcal{T}_{i+1} \setminus \mathcal{T}_i$ for some i. In the latter case, the way we chose T' and T'' at each step implies that $|T - X_i| \geq n^{2/3}/3$. It follows that $|T - X| \geq n^{2/3}/3 - 1$, since all vertices of $V(T) \cap X_i$ other than v_i are in $V(T) \cap X$ (this is because trees in $\mathcal{T}_t$ only intersect at vertices in X, and T was not subsequently split since $T \in \mathcal{T}_t$). We conclude that all but at most s trees T in $\mathcal{T}_t$ have $|T - X| \geq n^{2/3}/3 - 1$, so $\mathcal{T}_t$ contains at least $t = |X|$ trees with this property. Since each vertex of $V(F) \setminus X$ is in exactly one tree in $\mathcal{T}_t$, it follows that $|X|(n^{2/3}/3 - 1) \leq |V(F) \setminus X| = n - |X|$, so $|X| \leq 3n^{1/3}$, as required. □

The following proposition will be used at the very start of the proof of Theorem 4 to obtain the set V_{ground}; the stated properties will simplify the allocation and embedding process for the tree components that remain after this set is deleted. Note that in the proof we use the folklore result that every tree T contains a vertex x such that each component of $T - x$ has at most $|T|/2$ vertices; this can be proved by orienting each edge of T in the direction that leads to the most vertices (choosing arbitrarily if tied) and taking x to be a sink of the resulting orientation (which must exist since T contains no cycle).

Proposition 12. *If T is a tree or oriented tree, and $X \subseteq V(T)$, then there exists a set $Y \subseteq V(T)$ with the following properties.*

(i) $X \subseteq Y$,

(ii) $1 \leq |Y| \leq \max(6|X|, 1)$,

(iii) *each component of $T - Y$ contains at most $|T|/2$ vertices, and*

(iv) *each component T' of $T - Y$ has either one or two attachments (i.e. vertices with neighbours in Y), each of which has only one neighbour in Y. Moreover, if T' has two attachments then these are not adjacent.*

Proof. Choose a vertex $x_0 \in V(T)$ such that every component of $T - x_0$ has at most $|T|/2$ vertices. If $|X| = 0$ then $Y := \{x_0\}$ has the desired properties, so assume that $|X| \geq 1$. Write $X = \{x_1, \ldots, x_t\}$ and for each $i \in [t]$ let P_i be the unique path in T from x_0 to x_i. Then $P^* = \bigcup_{i \in [t]} P_i$ is a subtree of T. Let A be the set of vertices of P^* with degree at least three in P^*; since P^* has at most $t + 1$ leaves we have $|A| \leq t - 1$. Now let $\mathcal{P}$ be the set of all paths P in P^* for which P has length at most three and for which the vertices of P which are in the set $\{x_0\} \cup X \cup A$ are precisely the endvertices of P. So in particular $|\mathcal{P}| \leq \big|\{x_0\} \cup X \cup A\big|$. Let B be the set of all vertices in paths in $\mathcal{P}$ which are not in $\{x_0\} \cup X \cup A$, so $|B| \leq 2|\mathcal{P}| \leq 2\big|\{x_0\} \cup X \cup A\big| \leq 4t$. Finally, set $Y = \{x_0\} \cup X \cup A \cup B$, so $|Y| \leq 6t = 6|X|$ and $X \subseteq Y$; observe also that $Y \subseteq V(P^*)$.

Since $x_0 \in Y$ each component of $T - Y$ contains at most $|T|/2$ vertices. It therefore remains only to verify (iv), so let T' be a component of $T - Y$. If T' contains no vertices of $V(P^*)$, then there is a unique vertex of T' with a neighbour in Y, and that neighbour is also unique. So we may assume that T' contains a vertex of $V(P^*)$, and therefore that $V(T') \cap V(P^*)$ induces a path P in P^* with at least 3 vertices (since vertices in shorter paths were in B so cannot be in T'). Note that each vertex of P has degree two in P^* (since higher degree vertices were in A so cannot be in T'). It follows that the two endvertices of P – which are not adjacent – are the only vertices of T' with a neighbour in Y, and moreover that each of these endvertices has only one neighbour in Y, as required. □

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

Our final result on trees is a simple proposition which enables us to assume without loss of generality that the oriented graph Q that we wish to embed has at most one component which is a tree.

Proposition 13. *Fix $\Delta \geq 4$. Let F be a forest with maximum degree $\Delta(F) \leq \Delta$ which contains either at least p pairwise vertex-disjoint bare paths of order 7 or at least p pairwise disjoint edges incident to leaves. Then we may add edges to F so that the resulting graph is a tree T^* with the same properties.*

Proof. Let $\mathcal{T}$ be the set of components of F, so each $T \in \mathcal{T}$ is a tree. We form T^* by adding to F the edges of a path which meets each $T \in \mathcal{T}$ in a single vertex v_T; note that we then have $\deg_{T^*}(v_T) \leq \deg_F(v_T) + 2$ for each $T \in \mathcal{T}$ and $\deg_{T^*}(v) = \deg_F(v)$ for every vertex $v \notin \{v^T : T \in \mathcal{T}\}$. Suppose first that F contains a set $\mathcal{P}$ of at least p pairwise vertex-disjoint bare paths of order 7, and for each $T \in \mathcal{T}$ choose v_T to be a leaf of T (or the unique vertex of T if $|T| = 1$). So each v_T has $\deg_F(v_T) \leq 1$, so $\Delta(T^*) \leq \Delta$, and $\mathcal{P}$ is a set of bare paths of order 7 in T^*, as required. Now suppose instead that F contains a set $\mathcal{E}$ of at least p pairwise disjoint edges incident to leaves, and consider each $T \in \mathcal{T}$. If T contains a vertex which is adjacent to at least two leaves of T, then at least one of these leaves is not in an edge of $\mathcal{E}$, and we choose this leaf as v_T. On the other hand, if every vertex of T is adjacent to at most one leaf of T and $|T| \geq 2$, then let x_1 and x_2 be the first and second vertices respectively of a longest path P in T and set $v_T = x_2$. Since x_1 is the only neighbour of x_2 which is a leaf, the extremality of P implies that x_2 has no neighbours outside P. Finally, if $|T| = 1$ then let v_T be the unique vertex of T. In each case we have $\deg_F(v_T) \leq 2$ and so $\Delta(T^*) \leq \Delta$. Moreover, $\mathcal{E}$ is a set of pairwise disjoint edges incident to leaves in T^*, since no leaf in an edge in $\mathcal{E}$ was selected for v_T. □

3.2. Estimates and bounds

We write $\mathbb{E}(X)$ for the expectation of a random variable X, and write $\mathbb{P}(A)$ for the probability of an event A. We use the following well-known Chernoff-type bounds.

Theorem 14 [11, Corollary 2.3 and Theorem 2.10]. *If $0 < a < 3/2$ and X has binomial or hypergeometric distribution, then $\mathbb{P}\big(|X - \mathbb{E}(X)| \geq a\mathbb{E}(X)\big) \leq 2\exp(-a^2\mathbb{E}(X)/3)$.*

We also use a concentration result of McDiarmid [17], in a form stated by Sudakov and Vondrák [20].

Lemma 15. [17, 20] *Fix $n \in \mathbb{N}$ and let $X_1, \ldots, X_n$ be random variables taking values in $[0, 1]$ such that $\mathbb{E}(X_i \mid X_1, \ldots, X_{i-1}) \leq a_i$ for each $i \in [n]$. If $\mu \geq \sum_{i=1}^n a_i$, then for every δ with $0 < \delta < 1$ we have*

$$\mathbb{P}\left(\sum_{i=1}^n X_i > (1+\delta)\mu\right) \leq \mathrm{e}^{-\delta^2\mu/3}.$$

3.3. Regularity

As well as the strict notion of regularity – that G is r-regular if every vertex in G has degree precisely r – we will also work with an approximate notion of regularity for bipartite graphs, where a graph is 'regular' if its edges are 'random-like' in the sense that they are distributed roughly uniformly. More formally, let G be a bipartite graph with vertex classes A and B. For any sets $X \subseteq A$ and $Y \subseteq B$, we write $G[X, Y]$ for the bipartite subgraph of G with vertex classes X and Y whose edges are the edges of G with one endvertex in each of the sets X and Y, and define the *density* $d_G(X, Y)$ of edges between X and Y to be

$$d_G(X, Y) := \frac{e\big(G[X, Y]\big)}{|X||Y|}.$$

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

Let $d, \varepsilon > 0$. We say that G is (d, ε)-*regular* if for all $X \subseteq A$ and all $Y \subseteq B$ such that $|X| \geq \varepsilon|A|$ and $|Y| \geq \varepsilon|B|$ we have $d_G(X, Y) = d \pm \varepsilon$. (To avoid confusion, note that the notion of being (d, ε)-regular will always have two parameters, whereas the notion of an r-regular graph or digraph has only one parameter.) The following well-known result is immediate from this definition.

Lemma 16 (Slicing lemma). *Fix $\alpha, \varepsilon, d > 0$ and let G be a (d, ε)-regular bipartite graph with vertex classes A and B. If $A' \subseteq A$ and $B' \subseteq B$ have sizes $|A'| \geq \alpha|A|$ and $|B'| \geq \alpha|B|$, then $G[A', B']$ is $(d, \varepsilon/\alpha)$-regular.*

We say that a bipartite graph G with vertex classes A and B is $(d_{\geq}, \varepsilon)$-*regular* if G is (d', ε)-regular for some $d' \geq d$. For small ε, if G is (d, ε)-regular then almost all vertices of A have degree close to $d|B|$ in B and vice-versa. We say that G is 'superregular' if no vertex has degree much lower than this. More precisely, G is (d, ε)-*superregular* if G is $(d_{\geq}, \varepsilon)$-regular and also for every $a \in A$ and $b \in B$ we have $\deg(a, B) \geq (d - \varepsilon)|B|$ and $\deg(b, A) \geq (d - \varepsilon)|A|$. The analogous statement to Lemma 16 for superregular graphs does not hold in general, since a subset of one vertex class may entirely avoid the neighbourhood of a vertex in the other vertex class. However, we can give a similar statement for uniformly-random subsets of each vertex class.

Lemma 17. *Suppose that $1/n \ll \varepsilon \ll d, \beta$ and that a, b, x, y are integers with $\beta n < x \leq a \leq n$ and $\beta n < y \leq b \leq n$. Let G be a (d, ε)-superregular bipartite graph with vertex sets A and B of size a and b. If we choose $X \subseteq A$ with $|X| = x$ and $Y \subseteq B$ with $|Y| = y$ uniformly at random among all subsets of these sizes, then $G[X, Y]$ is $(d, \varepsilon/\beta)$-superregular with high probability.*

Proof. For each $a \in A$ the random variable $\deg(a, Y)$ has hypergeometric distribution with expectation $\deg(a)|Y|/|B| \geq (d - \varepsilon)|Y|$, so by Theorem 14 the probability that $\deg(a, Y) < (d - \varepsilon/\beta)|Y|$ declines exponentially with n. A similar argument shows that for each $b \in B$ the probability that $\deg(b, X) < (d - \varepsilon/\beta)|X|$ also declines exponentially with n. Taking a union bound over these events for each of the at most $2n$ vertices of G we find that, with high probability, none of these events hold. Since $G[X, Y]$ is $(d_{\geq}, \varepsilon/\beta)$-regular by Lemma 16, we then have that $G[X, Y]$ is $(d, \varepsilon/\beta)$-superregular. □

The final step in the proof of Theorem 4 is to embed the last few vertices of Q into the equally-few remaining unused vertices of the host graph G. In the case where Q has many pairwise disjoint edges incident to leaves, we will achieve this by using the following well-known corollary of Hall's marriage theorem, that every balanced superregular bipartite graph contains a perfect matching.

Lemma 18. *If $d \geq 2\varepsilon$ and G is a (d, ε)-superregular bipartite graph with equally many vertices in each vertex class, then G contains a perfect matching.*

In the case where Q has many pairwise vertex-disjoint edges incident to leaves, we instead use the following lemma of Komlós, Sárközy, and Szemerédi [13] (they actually stated the lemma only for $\ell = 4$, but the statement for larger ℓ follows immediately by combining this with Lemma 18). For this, say that a graph G is an ℓ-layer (d, ε)-*superregular* graph if $V(G) = \bigcup_{i \in [\ell]} V_i$, where $V_1, \ldots, V_\ell$ are pairwise disjoint sets of equal size ℓ and $G[V_i, V_{i+1}]$ is (d, ε)-superregular for each $i \in [\ell - 1]$.

Lemma 19 ([13], Theorem 2.1). *For every integer $\ell \geq 4$ and every $d > 0$ there exist ε and m_0 such that the following holds for all $m \geq m_0$. Let G be an ℓ-layer (d, ε)-superregular graph on ℓm vertices, and let $\pi : V_1 \to V_\ell$ be a bijection. There then exists a set $\mathcal{P}$ of m pairwise vertex-disjoint paths of order ℓ in G, such that for each $v \in V_1$ there is a path in $\mathcal{P}$ with ends v and $\pi(v)$.*

Let X and Y be disjoint sets of vertices in a digraph G. Observe that the underlying graph of $G[X \to Y]$ is then a bipartite graph with vertex classes X and Y. We say that $G[X \to Y]$ is (d, ε)-*regular* (respectively $(d_{\geq}, \varepsilon)$-*regular* or (d, ε)-*superregular*) to mean that this underlying graph is

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

(d,ε)-regular (respectively $(d_{\geq},\varepsilon)$-regular or (d,ε)-superregular). In this way we may apply the previous results of this subsection to digraphs.

The celebrated Regularity Lemma of Szemerédi [21, 22] states that every sufficiently large graph admits a partition such that almost all pairs of parts are regular in the sense we discuss here. Alon and Shapira [1] gave the following analogous result for directed graphs.

Lemma 20 (Regularity Lemma for digraphs [1]). *For all positive ε, K' there exist $K,\ n_0$ such that if G is a digraph of order $n \geq n_0$ and $d \in [0,1]$, then there exist a partition $V_0,\dots,V_k$ of $V(G)$ and a spanning subgraph G' of G such that*

(i) $K' \leq k \leq K$;

(ii) $|V_1| = \cdots = |V_k|$ *and* $|V_0| < \varepsilon n$;

(iii) $deg^+_{G'}(x) \geq deg^+_G(x) - (d+\varepsilon)n$ *for all* $x \in V(G)$;

(iv) $deg^-_{G'}(x) \geq deg^-_G(x) - (d+\varepsilon)n$ *for all* $x \in V(G)$;

(v) for all $i \in [k]$ *the digraph* $G'[V_i]$ *has no edges* ;

(vi) for all distinct i,j *with* $1 \leq i,j \leq k$, *either* $G'[V_i \to V_j]$ *is empty or* $G'[V_i \to V_j]$ *is* $(d_{\geq},\varepsilon)$*-regular.*

We refer to the sets $V_1,\dots,V_k$ as the *clusters* of G. For $d \in [0,1]$, the *reduced graph* R with parameters $\varepsilon,\ d$ and K' of G is a digraph we obtain by applying Lemma 20 to G with parameters ε, d and K'; the digraph R has vertex set $[k]$ and edges $i \to j$ precisely when $G'[V_i \to V_j]$ has density at least d. The following lemma will be used to obtain the reduced graph required in the allocation stage of the proof of Theorem 4.

Lemma 21. *Suppose that* $\frac{1}{n} \ll \frac{1}{K} \ll \frac{1}{K'} \ll \varepsilon \ll d \ll \eta \ll \alpha$. *If G is a digraph of order n with* $\delta^0(G) \geq \left(\frac{1}{2}+\alpha\right)n$, *then there exist an integer k with* $K' \leq k \leq K$, *a partition $V_0, V_1, \cdots, V_k$ of $V(G)$ and a digraph $R^\star$ with* $V(R^\star) = V_0 \dot\cup [k]$ *with the following properties.*

(a) $|V_0| < \varepsilon n$ *and* $m := |V_1| = \cdots = |V_k|$.

(b) The pairs $G[V_1 \to V_2], G[V_2 \to V_3], \dots, G[V_{k-1} \to V_k], G[V_k \to V_1]$, *are* (d,ε)*-super regular.*

(c) For all $i,j \in [k]$ *we have* $i \to j \in E(R^\star)$ *precisely when* $G[V_i \to V_j]$ *is* $(d_{\geq},\varepsilon)$*-regular.*

(d) For all $v \in V_0$ *and all* $i \in [k]$ *we have* $v \leftarrow i \in E(R^\star)$ *precisely when* $deg^-(v,V_i) \geq (1/2+\eta)m$, *and* $v \to i \in E(R^\star)$ *precisely when* $deg^+(v,V_i) \geq (1/2+\eta)m$.

(e) For all $i \in [k]$ *we have* $deg^0_{R^\star}(i,[k]) \geq (1/2+\eta)k$.

(f) For all $v \in V_0$ *we have* $deg^0_{R^\star}(v,[k]) > \alpha k/2$.

Proof. Introduce constants ε', d' with $\frac{1}{K'} \ll \varepsilon' \ll \varepsilon \ll d \ll d' \ll \eta$ and apply the digraph version of the Regularity Lemma (Lemma 20) to G with ε' and d' in place of ε' and d. This yields an integer k with $K' \leq k \leq K$, a partition $V(G) = V'_0 \dot\cup V'_1 \dot\cup \cdots \dot\cup V'_k$ and a spanning subgraph G' of G satisfying properties (i)–(vi) of Lemma 20. Define a graph R with vertex set $[k]$ in which $i \to j$ is an edge of R precisely if $G'[V'_i \to V'_j]$ is $(d'_{\geq},\varepsilon')$-regular. For each $i \in [k]$, if we choose a vertex $x \in V'_i$ then x has at least one outneighbour in each of at least $(deg^+_{G'}(x) - |V'_0|)/|V'_1| \geq ((1/2+\alpha)n - (d+\varepsilon)n - \varepsilon n)/(n/k) \geq (1/2+\alpha/2)k$ clusters V'_j with $j \in [k] \setminus \{i\}$. It follows that $\delta^+(R) \geq (1/2+\alpha/2)k$, and the same argument for inneighbours shows that $\delta^-(R) \geq (1/2+\alpha/2)k$, so $\delta^0(R) \geq (1/2+\alpha/2)k$. By Theorem 5 (specifically the case of a directed cycle due to Ghouila-Houri [10]) we find that R contains a directed Hamilton cycle. By relabelling the clusters if necessary, we assume without loss of generality that the edges of this cycle are $1 \to 2, 2 \to 3, \dots, k \to 1$.

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

By definition of $(d_{\geq}', \varepsilon')$-regularity, for each $i \in [k]$ the set V_i' contains at most $\varepsilon' n/k$ vertices which have fewer than $(d' - \varepsilon')|V_{i+1}'|$ outneighbours in V_{i+1}' and at most $\varepsilon' n/k$ vertices which have fewer than $(d' - \varepsilon')|V_{i-1}'|$ inneighbours in V_{i-1}'. By moving $2\varepsilon' n/k$ vertices (including all vertices with atypical degrees) from each V_i' to V_0' to obtain new sets $V_0, V_1, \dots, V_k$, we obtain a partition $V_0, V_1, \cdots, V_k$ of $V(G)$ for which (a) and (b) hold.

Let $R^\star$ be the digraph with vertex set $V_0 \dot\cup [k]$ and with edges defined as follows. For each distinct $i, j \in [k]$ we have $i \to j \in E(R^\star)$ if $G[V_i \to V_j]$ is $(d_{\geq}, \varepsilon)$-regular, and for each $v \in V_0$ and each $i \in [k]$ we have $v \to i \in E(R^\star)$ if $\deg_G^+(v, V_i) \geq (1/2 + \eta)m$ and $v \leftarrow i \in E(R^\star)$ if $\deg_G^-(v, V_i) \geq (1/2 + \eta)m$. This definition ensures Properties (c) and (d). Observe also that if $G[V_i' \to V_j']$ is $(d_{\geq}', \varepsilon')$-regular then $G[V_i \to V_j]$ is $(d_{\geq}, \varepsilon)$-regular, and so for each $i \in [k]$ we have $\deg^0_{R^\star}(i, [k]) \geq \deg^0_R(i) \geq (1/2 + \eta)k$, giving (e). Finally, consider any $v \in V_0$, and let $X \subseteq [k]$ be the set of all $i \in [k]$ for which $\deg_G^+(v, V_i) \geq (1/2 + \eta)m$. Then v has at most $|V_0| < \varepsilon n$ outneighbours in V_0, and at most $(1/2 + \eta)mk$ outneighbours in clusters V_i with $i \notin X$, so v has at least $\delta^0(G) - \varepsilon n - (1/2 + \eta)mk \geq \alpha n/2$ outneighbours in clusters V_i with $i \in X$; since each cluster contains m vertices it follows that $|X| \geq \alpha n/2m \geq \alpha k/2$, and so we have $\deg^+_{R^\star}(v, [k]) > \alpha k/2$. An identical argument for inneighbours shows that $\deg^-_{R^\star}(v, [k]) > \alpha k/2$, giving (f). □

3.4. Homomorphisms

As described in Section 2, a key idea for the proof of Theorem 4 is to find an allocation of the vertices of a subforest F of Q to the clusters of a reduced graph R of the host graph G. This allocation is a homomorphism from $F \to R$ with properties that will enable us to later embed each vertex of F into the cluster to which it is embedded. In preparation for this, we make the following definitions.

Let H and G be digraphs. A homomorphism $\varphi : H \to G$ is an edge-preserving map from $V(H)$ to $V(G)$, meaning that every edge $u \to v \in E(H)$ is mapped to an edge $\varphi(u) \to \varphi(v) \in E(G)$. The *$\varphi$-indegree* $\deg_\varphi^-(v)$ of $v \in V(H)$ in G is $\big|\varphi\big(N_H^-(v)\big)\big|$; the *$\varphi$-outdegree* $\deg_\varphi^+(v)$ of v is defined similarly, and the *φ-degree* of v is $\deg_\varphi(v) := \deg_\varphi^-(v) + \deg_\varphi^+(v)$. Note that vertices in $\varphi\big(N_H^-(v)\big) \cap \varphi\big(N_H^+(v)\big)$ are counted twice towards $\deg_\varphi(v)$. The *maximum degree* of φ is $\Delta(\varphi) := \max_{v \in V(H)} \deg_\varphi(v)$.

The embedding algorithm used in the proof of Theorem 4 relies on a property of large dense digraphs stated in Lemma 23. The form in which it is stated here is a generalisation of that used by Kühn, Mycroft, and Osthus [16, Lemma 2.5]. We begin with a definition.

Definition 22. *Let $\beta, \gamma, m > 0$, let G and R be digraphs and let S be an oriented star with centre c. Also let φ be a homomorphism from S to R, and let $J^- := \varphi\big(N_S^-(c)\big)$ and $J^+ := \varphi\big(N_S^+(c)\big)$, so $|J^-| + |J^+| = \Delta(\varphi)$. Let $\mathcal{V} = \{ V_i : i \in V(R) \}$ be a partition of $V(G)$ such that $\beta m \leq |V| \leq m$ for each $V \in \mathcal{V}$, fix a subset $U_j^\circ \subseteq V_j$ for each $\circ \in \{-, +\}$ and $j \in J^\circ$, and write $\mathcal{U}^\circ = \{ U_j^\circ \subseteq V_j : j \in J^\circ \}$ for each $\circ \in \{-, +\}$. A subset $X \subseteq V_{\varphi(c)}$ is $(\mathcal{U}^+, \mathcal{U}^-, \beta, \gamma, \varphi, m)$-good for S if for every possible choice of subsets $W_j^\circ \subseteq U_j^\circ$ of size $|W_j^\circ| \geq \beta m$ for each $\circ \in \{-, +\}$ and $j \in J^\circ$, there are at least $\gamma\sqrt{m}$ vertices $v \in X$ such that*

- *for all $j \in J^-$ we have $\deg^-(v, W_j^-) \geq \gamma m$, and*
- *for all $j \in J^+$ we have $\deg^+(v, W_j^+) \geq \gamma m$.*

Here is some motivation for Definition 22. We will embed the vertices of the tree one by one, and, after embedding a vertex $x \in V(T)$, we reserve sets of vertices for the children of x. If the reserved sets are always good, then this greedy embedding strategy will succeed. The next lemma states one sufficient condition for this to occur.

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

Lemma 23. *Suppose that* $\frac{1}{m} \ll \varepsilon \ll \gamma \ll \frac{1}{q}, \beta, d$. *Let G and R be digraphs and let S be an oriented star with centre c. Let φ be a homomorphism from S to R with $\Delta(\varphi) \le q$ and let $J^\circ := \varphi\big(N_S^\circ(c)\big)$ for each $\circ \in \{-,+\}$. Let $\mathcal{V} = \{ V_i : i \in V(R) \}$ be a partition of $V(G)$ such that $\beta m \le |V| \le m$ for all $V \in \mathcal{V}$, and for each $\circ \in \{-,+\}$ and $j \in J^\circ$ let $U_j^\circ \subseteq V_j$ be such that $G[V_{\varphi(c)} \leftarrow U^-_{\varphi(j^-)}]$ and $G[V_{\varphi(c)} \to U^+_{\varphi(j^+)}]$ are $(d_\ge, \varepsilon)$-regular for all $j^- \in J^-$ and $j^+ \in J^+$. Finally, write $\mathcal{U}^\circ = \{ U_j^\circ \subseteq V_j : j \in J^\circ \}$ for each $\circ \in \{-,+\}$. Then every subset $Y \subseteq V_{\varphi(c)}$ of size $|Y| \ge \gamma m/2$ contains a subset V with $|V| = \sqrt{m}$ which is $(\mathcal{U}^+, \mathcal{U}^-, \beta, \gamma, \varphi, m)$-good for S.*

Proof. Introduce new constants ε' and d' with $\varepsilon \ll \varepsilon' \ll \gamma \ll d' \ll \frac{1}{q}, \beta, d$. By removing leaves if necessary, we may assume that $\varphi(x) \ne \varphi(y)$ whenever x, y are both inleaves or both outleaves of S. Indeed, this does not change the statement we are trying to prove, since the definition of $(\mathcal{U}^+, \mathcal{U}^-, \beta, \gamma, \varphi, m)$-good is not affected by whether the number of in-leaves (respectively out-leaves) of S mapped to a fixed $x \in V(R)$ is precisely one or another positive integer. So we may proceed assuming that S has precisely q leaves $x_1, \dots, x_q$. For each $i \in [q]$ let $U_i := U^\bullet_{\varphi(x_i)}$, where $\bullet \in \{+,-\}$ is such that $x_i \in N_S^\bullet(c)$. For each $i \in [q]$, let $S_i := S\big[\{c, x_1, \dots, x_i\}\big]$, so $S_q = S$.

Fix $Y \subseteq V_{\varphi(c)}$ with $|Y| \ge \gamma m/2$. For each $t \in [q]$, each t-tuple $\mathbf{t} = (v_1, \dots, v_t)$ of distinct vertices with $v_j \in U_j$ for each $j \in [t]$, and each subset $Z \subseteq Y$, we write $N^{S_t}(\mathbf{t}, Z)$ for the set of vertices $v \in Z$ such that mapping $c \mapsto v$ and $x_j \mapsto v_j$ for each $j \in [t]$ gives a homomorphism from S_t to G. Also let T be the set of tuples $(v_1, \dots, v_q)$ of distinct vertices with $v_i \in U_i$ for each $i \in [q]$; call $\mathbf{t} \in T$ *bad* if $\big|N^S(\mathbf{t}, Y)\big| < d'|Y|$ and *good* otherwise.

Let B be the set of all bad tuples in T; our first goal is to show that B is small. To do this, set $V^0 := Y$, and suppose that we choose $v_i \in U_i$ for each $i \in [q]$ in turn, at each step setting $V^i := N^{S_i}\big((v_1, \dots, v_i), V^{i-1}\big)$. This yields a tuple $\mathbf{t} = (v_1, \dots, v_q) \in T$. For each $i \in [q]$ say that the choice of v_i is *bad* if $|V^i| < (d/2)|V^{i-1}|$. So if we do not make any bad choices, then for each $i \in [q]$ we have $|V^i| \ge (d/2)|V^{i-1}|$, and in particular $|V^q| \ge (d/2)^q |V^0| \ge d'|Y|$, so $\mathbf{t}$ is good. It follows that if the outcome $\mathbf{t}$ is not good, then for some $i \in [q]$ we must have made a bad choice. By definition of V^i this implies that $\deg^\circ(v_i, V^{i-1}) < |V^{i-1}|d/2$, where $\circ \in \{+,-\}$ is such that $c \in N_S^\circ(x_i)$. Now consider the smallest i for which we made a bad choice of v_i; this minimality property implies that $|V^{i-1}| \ge (d/2)^{i-1}|V^0| \ge (\gamma d^{i-1}/2^i)|V_{\varphi(c)}|$, and so $G[U_i \to V^{i-1}]$ (if $\circ = +$) or $G[V^{i-1} \to U_i]$ (if $\circ = -$) is $(d_\ge, \varepsilon')$-regular by Lemma 16. However, since $d/2 < d - \varepsilon'$, it follows that at most $\varepsilon'|U_i| \le \varepsilon' m$ vertices $b_i \in U_i$ have $\deg^\circ(b_i, V^{i-1}) < |V^{i-1}|d/2$, so there are at most $\varepsilon' m$ bad choices for v_i. We conclude that if $\mathbf{t}$ is bad then we must have made the first bad choice when choosing v_i for some $i \in [k]$, at which point there were at most $\varepsilon' m$ options for the bad choice, so in total we have $|B| \le k\varepsilon' m^q$ (since we have m options for each other choice).

Now choose a set $X \subseteq Y$ of size $|X| = \sqrt{m}$ uniformly at random. For each $\mathbf{t} \in T$ which is good, $\big|N^S(\mathbf{t}, X)\big|$ has hypergeometric distribution with expectation $\big|N^S(\mathbf{t}, Y)\big|(|X|/|Y|) \ge d'|X| = d'\sqrt{m}$. So by Theorem 14 the probability that $\big|N^S(\mathbf{t}, X)\big| < d'\sqrt{m}/2$ decreases exponentially with m. Since $|T \setminus B| \le |T| \le (\beta m)^q$, by taking a union bound we conclude that with high probability every $\mathbf{t} \in T \setminus B$ has $\big|N^S(\mathbf{t}, X)\big| \ge d'\sqrt{m}/2$. Fix a choice of X for which this event occurs.

It remains to show that X is $(\mathcal{U}^+, \mathcal{U}^-, \beta, \gamma, \varphi, m)$-good for S. So fix subsets $W_j^\circ \subseteq U_j^\circ$ of size $|W_j^\circ| \ge \beta m$ for each $\circ \in \{+,-\}$ and each $j \in J^\circ$. Let T' be the set of tuples $(v_1, \dots, v_q)$ of distinct vertices with $v_i \in W^-_{\varphi(x_i)}$ for each $i \in [q]$ with $x_i \in N_S^-(c)$ and $v_i \in W^+_{\varphi(x_i)}$ for each $i \in [q]$ with $x_i \in N_S^-(c)$. So $|T'| \ge (\beta m)^q - qm^{q-1}$, and therefore at least $|T'| - |B| \ge (\beta m)^q - q^2 m^{q-1} - k\varepsilon' m^q \ge \beta^q m^q/2$ tuples $\mathbf{t} \in T'$ are good and so have $\big|N^S(\mathbf{t}, X)\big| \ge d'\sqrt{m}/2$. Let $\mathcal{P} := \big\{ (v, \mathbf{t}) : \mathbf{t} \in T',\ v \in N^S(\mathbf{t}, X) \big\}$, so

$$|\mathcal{P}| \ge \frac{\beta^q m^q}{2} \cdot \frac{d'\sqrt{m}}{2} \ge \frac{d'\beta^q}{4} m^{q+1/2}. \tag{1}$$

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

In particular, at least $\gamma\sqrt{m}$ vertices $v^\star \in X$ must have $v^* \in N^S(\mathbf{t}, X)$ for at least γm^q tuples $\mathbf{t} \in T'$. Indeed, we would otherwise contradict (1), since we would then have

$$|\mathcal{P}| < |T'| \cdot \gamma\sqrt{m} + |X| \cdot \gamma m^q \leq m^q \cdot \gamma\sqrt{m} + \sqrt{m} \cdot \gamma m^q \leq 2\gamma m^{q+1/2}.$$

Each such $v^\star \in X$ must then have at least $\gamma m^q / m^{q-1} = \gamma m$ inneighbours in W_j^- for each $j \in J^-$ and at least γm outneighbours in W_j^+ for each $j \in J^+$, as required. □

3.5. Matching vertices

We use two simple results about matchings, whose straightforward proofs we omit.

Lemma 24. *Let G be a bipartite graph with vertex classes V and W, and suppose every vertex in V has degree at least $\varepsilon|W|$. Then there exists a subgraph $H \subseteq G$ such that*

(i) $\deg_H(v) = 1$ *for each* $v \in V$,

(ii) $\deg_H(w) \leq 1 + \frac{|V|}{\varepsilon|W|}$ *for each* $w \in W$.

Fact 25 (3, Exercise 16.1.6). *Let M and N be edge-disjoint matchings in a graph G. If $|M| > |N|$, then there exist edge-disjoint matchings M' and N' in G such that $|M'| = |M| - 1$, $|N'| = |N| + 1$ and $M' \cup N' = M \cup N$.* □

3.6. Patterns and diamonds

We define a *pattern* $\hat{P}$ to be a rooted oriented path whose root is a leaf; this means that for a fixed path of length ℓ, a pattern is determined by whether each edge along the path is directed towards the root ('ascending') or away from it ('descending'), so there are 2^ℓ distinct patterns with path length ℓ. Now let T be an oriented tree with a fixed root r. Then we can consistently define the pattern of each bare path P in T which does not contain r. Indeed, let r' be the unique vertex of P which is closest in T to r. Because P is a bare path, r' must be an endvertex of P, and then the pattern of P is simply P (with the orientation inherited from T) with root r'.

Let P be a pattern with three vertices a, b and c, where a is the root and b is adjacent to both a and c. A *P-diamond* is an orientation of a cycle on four vertices u, v, v', w in which both $\{u, v, w\}$ and $\{u, v', w\}$ induce patterns which are isomorphic to P (with root u in each case). For example, if P is $a \to b \to c$ then the digraph H with $V(H) = \{u, v, v', w\}$ and $E(H) = \{u \to v,\ u \to v',\ v \to w,\ v' \to w\}$ is a P-diamond (see Figure 1). We call the paths uvw and $uv'w$ the *branches* of the diamond, and say that the P-diamond H has *prefix* u, *middle* $\{v, v'\}$ and *suffix* w; we denote the P-diamond with this prefix, middle, and suffix by $u \preceq_{v'}^{v} \succeq w$.. If P is clear from the context, we write *diamond* instead of P-diamond. A *P-diamond path* in a digraph D is a sequence of P-diamonds $\left(u_i \preceq_{v_i'}^{v_i} \succeq w_i\right)_{i=0}^{t}$ such that $v_i = v'_{i-1}$ for each $i \in [t]$; we say that this path *connects* v_0 and v'_t. Finally, let $\mathcal{D}$ be a set of P-diamonds in a graph G. We say that $\mathcal{D}$ is *connecting* in G if for each pair $u, v \in V(G)$ there exists a P-diamond path, only using diamonds in $\mathcal{D}$, which connects u and v.

Due to the nature of Theorem 4, most patterns in our proof will have order two or seven. For notational convenience we make the following definitions, which allows us to usefully speak of P-diamond graphs for these patterns also. For a pattern P of order seven, let $v_1, \ldots, v_7$ be the vertices of P in order as they appear in the path, with v_1 being the root, and let P' be the pattern induced in P by $\{v_3, v_4, v_5\}$, with v_3 being the root of P'. Then we write P-diamond to mean P'-diamond. Similarly, for a pattern P of order two, let u be the root and v the other vertex of P. Form a pattern P' by adding a third vertex w as an outneighbour of v (we keep u as the root of P'). Again we then write P-diamond to mean P'-diamond.

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press



Figure 1. Left: a (∘ → • ← •)-diamond. Right: a (∘ ← • ← •)-diamond (∘ is the root of the path).

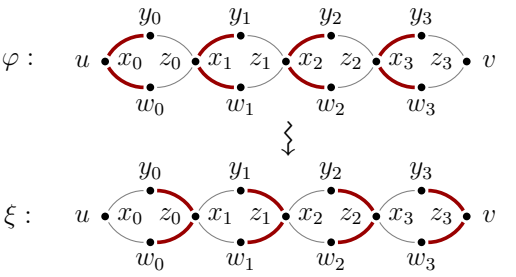


Figure 2. Remapping paths to the other branch of each diamond reduces the number of vertices mapped to u by one and increases the number of vertices mapped to v by one, with the number for each other vertex unchanged.

Lemma 26 (P-connected subgraphs). *Let G be a digraph of order n with $\delta^0(G) \geq \left(\frac{1}{2}+\alpha\right)n$ for some positive $\alpha \leq \frac{1}{2}$. If P is a pattern of order two, three or seven, then there exists a connecting set $\mathcal{D}$ of P-diamonds in G with $|\mathcal{D}| = n-1$ and so that no vertex of G appears in more than $4/\alpha$ diamonds in $\mathcal{D}$.*

Proof. By the definition of P-diamond for patterns of order two or seven, it suffices to prove the lemma in the case when P is a pattern of order three. Also, we may take $V(G) = [n]$. For each $i \in [n-1]$, let $\Diamond_i$ be the set of all P-diamonds with middle $\{i, i+1\}$. Let B_{pref} be a bipartite graph with vertex classes $\Diamond := \{\Diamond_1, \ldots \Diamond_{n-1}\}$ and $[n]$, with an edge between $\Diamond_i$ and $x \in [n]$ if x is a prefix of a P-diamond in $\Diamond_i$. Since i and $i+1$ each have at least $n/2 + \alpha n$ in- and out-neighbours, we have $\deg_{B_{\text{pref}}}(\Diamond_i) \geq 2\alpha n$. Therefore, by Lemma 24, there exists $H_{\text{pref}} \subseteq B_{\text{pref}}$ such that each vertex of $\Diamond$ is covered by precisely one edge of H_{pref} and each vertex of $[n]$ is covered by at most $1 + (n-1)/2\alpha n \leq 2/\alpha$ edges of H_{pref}. We define B_{suff} similarly for suffixes and obtain the corresponding graph H_{suff}. For each $i \in [n-1]$, let p_i be the neighbour of $\Diamond_i$ in H_{pref} and let s_i be the neighbour of $\Diamond_i$ in H_{suff}, so that the set $\mathcal{D} := \{p_i \prec_{i+1}^{i} \succ s_i : i \in [n-1]\}$ is a connecting set of P-diamonds in G with $|\mathcal{D}| = n-1$. Moreover, each vertex of G appears in at most $2/\alpha + 2 \leq 4/\alpha$ diamonds in $\mathcal{D}$: at most $1/\alpha$ times as a prefix, $1/\alpha$ times as a suffix, and twice in the middle. □

The next lemma is our main tool for adjusting vertex allocations (see Figure 2).

Lemma 27. *Let $m, k \in \mathbb{N}$, let R be a digraph of order k, let P be a pattern of order two, three or seven, and let $\mathcal{D}$ be a connecting set of P-diamonds in R. Also fix integers δ_v with $|\delta_v| < m$ for each $v \in V(R)$ such that $\sum_{v \in V(R)} \delta_v = 0$. Let $\mathcal{P}$ be a collection of pairwise vertex-disjoint oriented paths with pattern P, and let Q be the oriented graph which is the disjoint union of the paths in $\mathcal{P}$. If P has order two or three, then let M be the set consisting of the second vertex of each path in $\mathcal{P}$; if instead P has order seven, then let M be the set consisting of the fourth vertex of each path in $\mathcal{P}$. If there exists a homomorphism $\varphi: Q \to R$ such that for each diamond $D = x_i \prec_{w_i}^{y_i} \succ z_i$ in $\mathcal{D}$ at least km paths in $\mathcal{P}$ are mapped to $x_i y_i z_i$ and at least km paths are mapped to $x_i w_i z_i$, then there exists a homomorphism $\varrho: Q \to R$ such that $|\varrho^{-1}(v)| = |\varphi^{-1}(v)| + \delta_v$ for all $v \in V(R)$, and with the property that $\varrho(v) = \varphi(v)$ for every $v \in V(Q) \setminus M$.*

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

Proof. As in the previous lemma, it suffices to prove the lemma in the case when P is a pattern of order three. We proceed iteratively as follows. Let $u, v \in V(R)$ be such that $\delta_v < 0 < \delta_u$, and consider the P-diamond path from u to v. Let $\left(x_i \preceq^{y_i}_{w_i} \succeq z_i\right)_{i=1}^{t}$ be the sequence of diamonds in this path, so $u = y_1$ and $v = w_t$. For each $i \in [t]$, select a path in Q which is mapped to $x_i y_i w_i$, and modify the mapping so that it is now mapped to $x_i w_i z_i$ (see Figure 2). These changes result in a homomorphism $\xi : Q \to R$ such that $|\xi^{-1}(u)| = |\varphi^{-1}(u)| - 1$ and $|\xi^{-1}(v)| = |\varphi^{-1}(v)| + 1$, whereas $|\xi^{-1}(x)| = |\varphi^{-1}(x)|$ for all $x \in V(R) \setminus \{u, v\}$. So the number of paths mapped to any diamond branch is reduced by at most one; moreover we have $\xi(v) = \varphi(v)$ for every $v \in V(Q) \setminus M$. By iterating this procedure at most $\sum_v |\delta_v| \leq km$ times we achieve the desired mapping ϱ; since each diamond initially had at least km paths mapped to each branch, we never run out of paths to edit. □

3.7. Regular expander subdigraphs

We call a digraph D an *expander* if $\left|N^-(S)\right| > |S|$ and $\left|N^+(S)\right| > |S|$ for all nonempty proper subsets $S \subsetneq V(D)$. The next lemma shows that every digraph G which satisfies our minimum degree condition contains a spanning subgraph H which is a regular expander and which is not too dense. Moreover we may insist that H contains a given subgraph of G of bounded maximum degree.

Lemma 28. *Suppose that* $\frac{1}{n} \ll \frac{1}{f}, \alpha \leq 1$. *Let G be a digraph of order n with* $\delta^0(G) \geq (\frac{1}{2} + \alpha)n$. *If* $F \subseteq G$ *and* $\Delta^0(F) \leq f$, *then G contains a spanning d-regular subdigraph H such that*

(i) $F \subseteq H$,
(ii) $d \leq 25n^{2/3}/\alpha$, *and*
(iii) H is an expander.

Proof. Form a subgraph $H_p \subseteq G$ with $V(H_p) = V(G)$ by selecting each edge of G for inclusion in H_p with probability $p := n^{-1/3}$, with the choices for each edge being independent of all other choices.

Claim 29. *With high probability,*

(a) every vertex of H_p has in- and outdegree at most $2n^{2/3}$, *and*
(b) if S is a nonempty proper subset of $V(G)$, then $\left|N^-_{H_p}(S)\right|, \left|N^+_{H_p}(S)\right| > |S|$.

Proof of claim. Let $x \in V(H_p)$. Note that $\deg^-_{H_p}(x)$ and $\deg^+_{H_p}(x)$ are binomial random variables with expectation between $(\frac{1}{2} + \alpha)n^{2/3}$ and $n^{2/3}$. By Theorem 14 (applied with $a = 1$) we have

$$\mathbb{P}\left(\deg^-_{H_p}(x) > 2n^{2/3}\right) \leq 2\exp\left(-n^{2/3}/6\right) \quad \text{and}$$
$$\mathbb{P}\left(\deg^+_{H_p}(x) > 2n^{2/3}\right) \leq 2\exp\left(-n^{2/3}/6\right). \tag{2}$$

Taking a union bound over all n vertices we find that (a) holds with high probability. To prove (b), we will show that with high probability each nonempty proper subset $S \subseteq V(G)$ has $|N^+_{H_p}(S)| > |S|$; the same argument with directions reversed shows that with high probability each nonempty proper subset $S \subseteq V(G)$ has $|N^-_{H_p}(S)| > |S|$, giving the desired conclusion. So fix a nonempty proper subset $S \subseteq H_p$, and let Z_S denote the event that $|N^+_{H_p}(S)| \leq |S|$. We consider four cases.

If $|S| < n^{1/2}$, then choose any $x \in S$ and recall that $\deg^+_{H_p}(x)$ is a binomial random variable with expectation at least $(\frac{1}{2} + \alpha)n^{2/3}$. Applying Theorem 14 with $a = 1/2$ we find that

$$\mathbb{P}(Z_S) \leq \mathbb{P}\left(\deg^+_{H_p}(x) < n^{1/2}\right) \leq 2\exp\left(-\frac{n^{2/3}}{24}\right) \leq \exp\left(-\frac{n^{2/3}}{25}\right). \tag{3}$$

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

If $n^{1/2} \le |S| < n/2$, then for Z_S to occur there must exist a set $T \subseteq V(G)$ with $|T| \ge n/2$ such that $H_p[S \to T]$ is empty. So fix $T \subseteq V(G)$ with $|T| \ge n/2$, and observe that each $x \in S$ has $\deg^+_G(x, T) \ge \alpha n$. It follows that $e(G[S \to T]) \ge |S|\alpha n/2 \ge \alpha n^{3/2}/2$, and so $e(H_p[S \to T])$ has a binomial distribution with expectation at least $\alpha n^{7/6}/2$; by Theorem 14 applied with $a = 1$ we find the probability that $H_p[S \to T]$ is empty to be at most $2\exp(-\alpha n^{7/6}/6)$. Taking a union bound over all T we obtain

$$\mathbb{P}(Z_S) \le 2^{n+1} \exp\Big(-\frac{\alpha n^{7/6}}{6}\Big) \le \exp\Big(-\frac{\alpha n^{7/6}}{10}\Big). \tag{4}$$

Similarly, if $n/2 \le |S| \le n - n^{1/2}$, then for Z_S to occur there must exist a set $T \subseteq V(G)$ with $|T| \ge n^{1/2}$ such that $H_p[S \to T]$ is empty. So fix $T \subseteq V(G)$ with $T \ge n^{1/2}$, and observe that each $x \in T$ has $\deg^-_G(x, S) \ge \alpha n$. So $e(G[S \to T]) \ge |T|\alpha n/2 \ge \alpha n^{3/2}/2$, and so we obtain (4) exactly as in the previous case.

Finally, if $|S| > n - n^{1/2}$, then for each $x \in V(G)$ we have $\deg^-_G(x, S) \ge \deg^-_G(x) - n^{1/2} \ge n/2$, so $\deg^-_{H_p}(x, S)$ is a binomial random variable with expectation at least $n^{2/3}/2$. Applying Theorem 14 with $a = 1$ we deduce that $\mathbb{P}(\deg^-_{H_p}(x, S) = 0) \le 2\exp(-n^{2/3}/6)$, and so taking a union bound over all vertices we obtain

$$\mathbb{P}(Z_S) \le \sum_{x \in V(G)} \mathbb{P}\Big(\deg^-_{H_p}(x, S) = 0\Big) \le 2n \exp\Big(-\frac{n^{2/3}}{6}\Big) \le \exp\Big(-\frac{n^{2/3}}{25}\Big). \tag{5}$$

Taking a union bound, we obtain the desired bound of

$$\begin{aligned}\mathbb{P}\Big(\bigcup_S Z_S\Big) &\le \sum_{|S|<\sqrt{n}} \mathbb{P}(Z_S) \quad + \sum_{\sqrt{n}\le|S|\le n-\sqrt{n}} \mathbb{P}(Z_S) \quad + \sum_{n-\sqrt{n}<|S|} \mathbb{P}(Z_S) \\ &\le 2\sqrt{n}\binom{n}{\sqrt{n}} \underbrace{\exp\left(-\frac{n^{2/3}}{25}\right)}_{(3)\text{ and }(5)} + 2^n \underbrace{\exp\left(-\frac{\alpha n^{7/6}}{10}\right)}_{(4)} = o(1),\end{aligned}$$

where the sums and union each range over all proper nonempty subsets $S \subseteq V(G)$ with the specified sizes, and we use the bounds $\binom{n}{\sqrt{n}} \le (\sqrt{n}\, e)^{\sqrt{n}} \le \exp\big(\sqrt{n}\, \ln n\big)$. ◇

Returning to the proof of the lemma, fix an outcome of H_p such that (a) and (b) hold and let $H' := H_p \cup F$. Clearly each H with $H' \subseteq H \subseteq G$ satisfies both (i) and (iii). So to conclude the proof it suffices to find such an H satisfying (ii). By (a), we have $\Delta^0(H') \le 2n^{2/3} + f \le 3n^{2/3} - 1$. Hence, by Vizing's theorem, the edges of H' can be partitioned into at most $3n^{2/3}$ matchings. It follows by repeatedly applying Fact 25 that $E(H')$ admits a partition $\mathcal{M}$ into matchings such that $|\mathcal{M}| \le 3n^{2/3}$ and $|M| - |N| \le 1$ for all $M, N \in \mathcal{M}$ (more precisely, one may repeatedly apply Fact 25 to the largest and smallest matching in the partition, until all pairs of matchings have sizes differing by at most 1). By partitioning each matching in $\mathcal{M}$ into $\lceil \frac{7}{\alpha} \rceil$ pairwise disjoint submatchings with sizes as close as possible, we obtain a partition $M_1, \dots, M_d$ of $E(H')$ into matchings with $d \le 3n^{2/3}\lceil \frac{7}{\alpha} \rceil \le 25n^{2/3}/\alpha$, such that $|M_i| \le \alpha n/7 + 1 \le \alpha n/6$ for each $i \in [d]$, and $\big||M_i| - |M_j|\big| \le 1$ for all $i, j \in [d]$. The following procedure builds the desired H.

> **Procedure:** Let $G_0 := G - E(H')$. For each $i \in [d]$, in order, greedily choose a directed cycle C_i in $G_{i-1} \cup M_i$, such that C_i has length $3|M_i|$ and covers all edges in M_i; let C'_i be a directed Hamilton cycle in $G_{i-1} \setminus V(C_i)$ and let $G_i := G_{i-1} \setminus \big(E(C_i) \cup E(C'_i)\big)$. We set $H := \bigcup_{i \in [d]} (C_i \cup C'_i)$.

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

Let us check that these steps may be carried out. Fix $i \in [d]$; it suffices to show that C_i and C_i' exist. Let $M_i = \{u_1 \to v_1, \dots, u_r \to v_r\}$. Since $\delta^0(G_i) \geq \delta^0(G_0) - d \geq (1/2 + 3\alpha/4)n$, for each pair of vertices $x, y \in V(G_i)$ there exists at least $3\alpha n/4 - 2|M_i| \geq |M_i|$ vertices z such that $z \in (N^+_{G_i}(x) \cap N^-_{G_i}(y)) \setminus V(M_i)$. We may therefore choose distinct vertices $z_1, \dots, z_{|M_i|}$ with $z_j \in (N^+_{G_i}(v_j) \cap N^-_{G_i}(u_{j+1})) \setminus V(M_i)$ for each $j \in [r]$ (with addition taken modulo r). Let C_i be the cycle $u_1 \to v_1 \to z_1 \to u_2 \to \cdots \to u_r \to v_r \to z_r \to u_1$. Since $|C_i| = 3|M_i| \leq \alpha n/2$, it follows that $\delta^0\big(G_i \setminus V(C_i)\big) \geq (1/2 + \alpha/4)n$, so $G_i \setminus V(C_i)$ contains a directed Hamilton cycle C_i'.

To complete the proof, observe that $H' \subseteq H \subseteq G$ and that H is the edge-disjoint union of spanning subdigraphs $C_i \cup C_i'$ of G. Since each vertex of G has precisely one inneighbour and one outneighbour in $C_i \cup C_i'$ for each $i \in [d]$, we conclude that H is a spanning subgraph of G with $\deg^-_H(x) = \deg^+_H(x) = d \leq 25n^{2/3}/\alpha$ for each $x \in V(H)$, so (ii) holds.

We next establish a property of regular expander digraphs D which is crucial for our random allocation strategy, namely that if X is a random vertex of D (with some unspecified probability distribution on the vertices of D), and Y is a uniformly random outneighbour (or inneighbour) of X, then the distribution of Y is more uniform than the distribution of X, except in the case where X is uniformly-distributed, in which case the same is true of Y. This property is established in Lemma 31, using the notation we now introduce.

Let D be a digraph of order k, and let X be a random vertex of D. Formally speaking this means that X is a random variable with codomain $V(D)$ in some suitable probability space, so, writing $V(D) = \{x_1, \dots, x_k\}$ we have a probability distribution $\mathbb{P}(X = x_i) = p_i$ on the vertices of D. We then define the variation of X by

$$\mathrm{Var}(X) := \sum_{x \in V(D)} \left(\mathbb{P}(X = x) - \frac{1}{k} \right)^2.$$

So $\mathrm{Var}(X)$ is a measure of how uniform the distribution of X is, and in particular $\mathrm{Var}(X) = 0$ if and only if X is a uniformly-random vertex of D.

Consider the set $S \subseteq \mathbb{R}^k$ defined by $S = \{(p_1, \dots, p_k) : p_i \geq 0, \sum_{i=1}^k p_i = 1\}$, so S is the set of all possible probability distributions for a random vertex X of D. Observe that S is a convex polytope whose vertices are the unit vectors along each axis of the coordinate system. Moreover, the function $\mathrm{Var} : \mathbb{R}^k \to \mathbb{R}$ given by $\mathrm{Var}((p_1, \dots, p_k)) = \sum_{i=1}^k (p_i - 1/k)^2$ is convex, since it is a constant translation of the function which squares each coordinate. It follows that $\mathrm{Var}(\,\cdot\,)$ obtains its maximum value on S at a vertex of S, and therefore that for any distribution of a random vertex X in D we have

$$0 \leq \mathrm{Var}(X) \leq \left(1 - \frac{1}{k}\right)^2 + \frac{k-1}{k^2} = 1 - \frac{1}{k} < 1. \tag{6}$$

Our next lemma states that for every expander digraph and every assignment of weights to vertices, there exist vertices with somewhat distinct weights which share a common inneighbour (and the same is true for outneighbours).

Lemma 30. *Let D be an expander digraph of order n, let $f : V(D) \to \mathbb{R}$ and let $M := \max_{x,\, y \in V(D)} f(x) - f(y)$. If $n := |D| \geq 3$ and $M > 0$, then there exist u, x, $y \in V(D)$ such that x, $y \in N^-(u)$ and $f(y) - f(x) \geq M/(n-1)$ and, similarly, there exist v, w, $z \in V(D)$ such that w, $z \in N^+(v)$ and $f(w) - f(z) \geq M/(n-1)$.*

Proof. It suffices to prove the existence of u, x, y; the statement for v, w, z follows by an identical argument with the roles of inneighbours and outneighbours switched. Let $S_1, \dots, S_r$ be a partition of $V(D)$ such that for all x, $y \in V(D)$ we have $f(x) = f(y)$ if and only if x, $y \in S_i$ for some $i \in [r]$. Clearly, $1 < r \leq n$. Since f is constant in each set of this partition, we write $f(i)$ for the common value of f over all $x \in S_i$. We can assume that the sets are labelled so that $f(i) < f(j)$ whenever $i < j$. Note that $M = f(r) - f(1)$, and therefore $f(j+1) - f(j) \geq M/(r-1) \geq M/(n-1)$ for some

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

$j \in [r-1]$. Let $X := S_1 \cup \cdots \cup S_j$ and $Y := S_{j+1} \cup \cdots \cup S_n$. Since D is an expander, $\left|N^+(X)\right| > |X|$ and $\left|N^+(Y)\right| > |Y|$. Because $|X| + |Y| = n$ there must be a vertex $u \in N^+(X) \cap N^+(Y)$. Let $x \in X$ and $y \in Y$ be inneighbours of u. Then $f(y) - f(x) \geq f(j+1) - f(j) \geq M/(n-1)$ as desired. □

We are now ready to establish the key property of regular expanders we need for our allocation strategy. Note here that when we say that Y is a uniformly-random outneighbour of X, we mean that the distribution of the random vertex Y can be obtained by first choosing a vertex according to the distribution of X, then selecting a uniformly-random outneighbour of the chosen vertex.

Lemma 31. *Let D be an d-regular expander digraph of order k. Let X be a random vertex of D, and let Y be a uniformly-random outneighbour of X. Then*

$$Var(Y) \leq \left(1 - \frac{1}{2k^5}\right) Var(X). \tag{7}$$

The same bound holds if y is a uniformly-random inneighbour of x.

Proof. For each $x \in V = V(D)$ write $f(x) := \frac{1}{d}\left(\mathbb{P}(X = x) - \frac{1}{k}\right)$. We then have

$$\mathrm{Var}(X) = \sum_{x \in V} \left(\mathbb{P}(X = x) - \frac{1}{k}\right)^2 = \sum_{x \in V} (df(x))^2 = \sum_{y \in V} \sum_{x \in N_D^-(y)} df(x)^2,$$

where the final equality holds since D is d-regular so each vertex x appears in $N_D^-(y)$ for precisely d vertices y. Similarly, since each $x \in V$ has precisely d outneighbours, for each $y \in V$ we have $\mathbb{P}(Y = y) = \sum_{x \in N_D^-(y)} \frac{1}{d} \cdot \mathbb{P}(X = x)$, so

$$\mathrm{Var}(Y) = \sum_{y \in V} \left(\mathbb{P}(Y = y) - \frac{1}{k}\right)^2 = \sum_{y \in V} \left(\sum_{x \in N_D^-(y)} \frac{1}{d}\left(\mathbb{P}(X = x) - \frac{1}{k}\right)\right)^2 = \sum_{y \in V} \left(\sum_{x \in N_D^-(y)} f(x)\right)^2.$$

Combining these expressions we obtain

$$\begin{aligned}
\mathrm{Var}(X) - \mathrm{Var}(Y) &= \sum_{y \in V} \left(\sum_{x \in N_D^-(y)} df(x)^2 - \left(\sum_{x \in N_D^-(y)} f(x) \right)^2 \right) \\
&= \sum_{y \in V} \left(\left(\sum_{z,w \in N_D^-(y)} \frac{1}{2} f(z)^2 + \frac{1}{2} f(w)^2 \right) - \left(\sum_{z,w \in N_D^-(y)} f(z) f(w) \right) \right), \\
&= \sum_{y \in V} \left(\sum_{z,w \in N_D^-(y)} \frac{1}{2} \big(f(z) - f(w)\big)^2 \right) \geq \frac{1}{2} \max_{\substack{y \in V \\ z,w \in N_D^-(y)}} \big(f(z) - f(w)\big)^2,
\end{aligned} \tag{8}$$

where the second equality holds since $|N_D^-(y)| = d$ for each $y \in V$, meaning that the term $\frac{1}{2} f(x)^2$ is counted precisely $2d$ times in the latter expression, whilst the final inequality holds simply because a sum of non-negative terms is at least as large as its maximum term. Write $L := \max_{x \in V} \left|f(x)\right|$. We then have

$$\begin{aligned}
\mathrm{Var}(X) - \mathrm{Var}(Y) &\geq \frac{1}{2} \max_{\substack{y \in V \\ z,w \in N_D^-(y)}} \big(f(z) - f(w)\big)^2 \\
&\geq \frac{\max_{u,v \in V} \big(f(u) - f(v)\big)^2}{2(k-1)^2} \geq \frac{L^2}{2(k-1)^2} \geq \frac{\mathrm{Var}(X)}{2k(k-1)^2 d^2} \geq \frac{\mathrm{Var}(X)}{2k^5}.
\end{aligned}$$

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

Algorithm 32 (Vertex Allocation Algorithm)

Input: an oriented forest F, an ancestral order $t_1, \dots, t_n$ of $V(F)$, a digraph D with $\delta^0(D) \geq 1$, a set $Z \subseteq V(F)$ which contains the root of each component of F, a set $\mathcal{E} \subseteq E(F)$, and for each $e \in \mathcal{E}$ a permutation π_e of $V(F)$.

Output: a map $\varphi : V(F) \to V(D)$.

for $\tau = 1$ **to** n **do**

  if $t_\tau \in Z$ **then** choose $\varphi(t_\tau)$ arbitrarily.

  else

    Let t_σ be the parent of t_τ, let $x_\sigma = \varphi(t_\sigma)$ and let e be the edge of F between t_σ and t_τ.

    if $e \notin \mathcal{E}$ **then** set $\varphi(t_\tau) := \begin{cases} c_\sigma^+ & \text{if } t_\tau \in N_F^+(t_\sigma), \\ c_\sigma^- & \text{if } t_\tau \in N_F^-(t_\sigma) \end{cases}$.

    if $e \in \mathcal{E}$ **then** set $\varphi(t_\tau) := \begin{cases} \pi_e(x_\sigma) & \text{if } t_\tau \in N_F^+(t_\sigma), \\ \pi_e^{-1}(x_\sigma) & \text{if } t_\tau \in N_F^-(t_\sigma). \end{cases}$

  Pick $c_\tau^+ \in N_D^+(x_\tau)$ and $c_\tau^- \in N_D^-(x_\tau)$ uniformly at random independently of all previous choices.

Indeed, the first inequality is (8), whilst the second holds by Lemma 30. The third holds since $\sum_{x \in V} f(x) = 0$ and so $\max_{u,v \in V} (f(u) - f(v)) \geq L$, whilst the fourth holds since $\mathrm{Var}(X) = \sum_{x \in X} \big(df(x)\big)^2 \leq kd^2L^2$.

This completes the proof in the case where Y is a uniformly-random outneighbour of X; the argument for Y being a uniformly-random inneighbour of X is identical with the roles of inneighbours and outneighbours switched. □

3.8. General allocation algorithm

For the following algorithm, recall that if a tree T is a component of a forest F, then the root of T appears prior to each other vertex of T in any ancestral order on the vertices of F; together with the stipulation that the root of each component of F lies in Z, this ensures that we may always take t_σ to be the parent of t_τ when required to do so. Note also that the algorithm makes arbitrary choices of $\varphi(t_\tau)$ for vertices t_τ in Z. When we apply the algorithm later we will specify how these arbitrary choices should be made; the point of not specifying this here is that the results we prove in this section about Algorithm 32 hold no matter how we subsequently do this.

Recall the proof outline of Theorem 4 in Section 2: we aim to allocate the vertices of a forest F to the vertices of a reduced graph R so that vertices in F are allocated to a uniformly-random in- or out-neighbour of their parent (according to the direction of the corresponding edge), whilst some edges (those in chosen bare paths) should be allocated along a Hamilton cycle $1 \to 2 \to \cdots \to k \to 1$ in R. These edges will be contained in the set $\mathcal{E}$ given as input to Algorithm 32, and indeed moving one step on the Hamilton cycle is a permutation on the vertex set, as required. Also, the allocation of some vertices of F will be constrained by the existence of neighbours outside F which have already been allocated; these vertices form the set Z given as input to Algorithm 32, and can be allocated appropriately at the point the algorithm arrives at them. Our aim at this point, then, is to show that so long as Z is not too large, and there are no long paths entirely in $\mathcal{E}$, Algorithm 32 will yield an approximately-uniform allocation of vertices of F among the vertices of R (corresponding to clusters of G). This is asserted by the main result of this section, Lemma 34. Before that, we give a preliminary result describing how, for a long path P in F with no vertex from Z (except possibly the initial vertex of P) which contains many edges not in $\mathcal{E}$, the allocation of the final vertex in P by Algorithm 32 is essentially independent of the allocation of the initial vertex of P.

Let F be a forest and let $\mathcal{E}$ be a set of edges of F. For vertices u, v in the same component of F, we define the $\mathcal{E}$-distance from u to v, denoted $\mathrm{dist}_{\mathcal{E}}(u, v)$, to be the number of edges of $\mathcal{E}$ in the

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

(unique) path in F from u to v (actually we will mainly work with $\mathrm{dist}_{\overline{\mathcal{E}}}(u,v)$ where $\overline{\mathcal{E}}$ denotes the complement $E(F)\setminus\mathcal{E}$).

Proposition 33. *Let R be a digraph with vertex set $[k]$ which is a d-regular expander. Let F be an oriented forest with a fixed ancestral order, let $Z\subseteq V(F)$ be a set which contains the root of each component of F, let $\mathcal{E}$ be a set of edges of F, and for each edge $e\in\mathcal{E}$ let $\pi_e:[k]\to[k]$ be a permutation. Let $u_0,u_1,\dots,u_\ell$ be the vertices of a path in F for which u_0 is an ancestor of u_ℓ, $\mathrm{dist}_{\overline{\mathcal{E}}}(u_0,u_\ell)\geq 4k^5\log m$, and $u_i\notin Z$ for each $1\leq i\leq\ell$. If we apply Algorithm 32 to obtain a map $\varphi:V(F)\to V(R)$ then for all $x,y\in V(D)$,*

$$\mathbb{P}\left(\varphi(u_\ell)=y \mid \varphi(u_0)=x\right)=\frac{1}{k}\pm\frac{1}{m}.$$

Proof. For each $0\leq i\leq\ell$ let X_i be the value of $\varphi(u_i)$ conditioned on the event that $\varphi(u_0)=x$, so X_i is a random vertex of D. In particular, X_0 then takes value x with probability 1, so $\mathrm{Var}(X_0)=1-1/k<1$. Now consider some $i\in[\ell]$. If the edge e of F between u_{i-1} and u_i is in $\mathcal{E}$, then $X_i=\pi_e(X_{i-1})$ or $X_i=\pi_e^{-1}(X_{i-1})$ according to the direction of this edge; in either case we have $\mathrm{Var}(X_i)=\mathrm{Var}(X_{i-1})$. On the other hand, if the edge of F between u_{i-1} and u_i is not in $\mathcal{E}$, then X_i is a uniformly-random outneighbour or inneighbour of X_{i-1}, again according to the direction of this edge. In either case we have $\mathrm{Var}(X_i)\leq(1-1/2k^5)\mathrm{Var}(X_{i-1})$ by Lemma 31. We conclude that

$$\mathrm{Var}(X_\ell)\leq\left(1-\frac{1}{2k^5}\right)^{\mathrm{dist}_{\overline{\mathcal{E}}}(u_0,u_\ell)}\mathrm{Var}(X_0)<\left(1-\frac{1}{2k^5}\right)^{\mathrm{dist}_{\overline{\mathcal{E}}}(u_0,u_\ell)}.$$

By definition of Var we have $\mathbb{P}(X_\ell=y\mid\varphi(u_0)=x)=\frac{1}{k}\pm\sqrt{\mathrm{Var}(X_\ell)}$ for each $y\in V(D)$, so

$$\mathbb{P}\left(\varphi(u_\ell)=y \mid \varphi(u_0)=x\right)=\frac{1}{k}\pm\left(1-\frac{1}{2k^5}\right)^{\mathrm{dist}_{\overline{\mathcal{E}}}(u_0,u_\ell)/2}=\frac{1}{k}\pm\left(1-\frac{1}{2k^5}\right)^{2k^5\log m}=\frac{1}{k}\pm\frac{1}{m},$$

where the final inequality holds by the standard inequality $(1-x)\leq\mathrm{e}^{-x}$. □

Lemma 34. *Suppose that $\frac{1}{n}\ll\frac{1}{k}$. Let R be a digraph with vertex set $[k]$ which is a d-regular expander. Let F be an oriented forest on n vertices with a fixed ancestral order and let $Z\subseteq V(F)$ be a set which contains the root of each component of F. Let $\mathcal{E}$ be a set of edges of F which does not contain a path of length 7, and for each edge $e\in\mathcal{E}$ let $\pi_e:[k]\to[k]$ be a permutation. If we apply Algorithm 32 to obtain a map $\varphi:V(F)\to V(R)$, then for each set $S\subseteq V(F)$, with high probability we have for each $i\in[k]$ that*

$$|\varphi^{-1}(i)\cap S|=\frac{|S|}{k}\pm\left(\frac{n}{\log n}+(|Z|+3n^{1/3})\Delta(F)^{56k^5\log\log n}\right).$$

Proof. We may assume that $|S|\geq n/\log n$ as otherwise there is nothing to prove.

By Corollary 11 we may choose a set $Y\subseteq V(F)$ with $|Y|\leq 3n^{1/3}$ such that every component of $F-Y$ has size at most $n^{2/3}$. Write $Z^*:=Y\cup Z$, so $|Z^*|\leq|Z|+3n^{1/3}$. Also let $T_1,\dots,T_s$ be the components of $F-Z^*$, so $|T_i|\leq n^{2/3}$ for each $i\in[s]$, and for each $i\in[s]$ let z_i be the nearest ancestor in Z^* of vertices in T_i; note that z_i exists since the root of each component of F is in Z, and z_i is well-defined since each vertex in T_i has the same nearest ancestor in Z^*. Moreover, choose the indices of the T_i and z_i so that if z_i is an ancestor of z_j then $i<j$ (this can be achieved by having the order $z_1,z_2,\dots,z_s$ be the restriction of the ancestral order on F to the vertices z_i). Let B be the set of all vertices $z\in V(F)$ for which there exists $x\in Z^*$ which is an ancestor of z with $\mathrm{dist}_F(x,z)\leq 56k^5\log\log n$, so $|B|\leq|Z^*|\Delta(F)^{56k^5\log\log n}$. For each $i\in[s]$ set $F_i:=V(T_i)\setminus B$ for each $i\in[s]$. So the sets $F_1,\dots,F_s$ are pairwise disjoint subsets of $V(F)$ with the following properties.

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

(i) $\left|\bigcup_{i\in[s]} F_i\right| \geq n - (|Z| + 3n^{1/3})\Delta(F)^{56k^5 \log\log n}$, since every vertex of F is in some F_i except for the vertices in B.

(ii) $|F_i| \leq |T_i| \leq n^{2/3}$ for each $i \in [s]$.

(iii) For each $i \in [s]$, each $x \in \bigcup_{j<i} V(F_j)$, and each $y \in F_i$, either there is no path from x to y in F, or the path from x to y in F includes z_i.

(iv) For any $i \in [s]$ and $y \in F_i$ we have $\mathrm{dist}_F(z_i, y) \geq 56k^5 \log\log n$.

Define random variables X_i^j for each $i \in [s]$ and $j \in [k]$ by

$$X_i^j := \frac{|\varphi^{-1}(j) \cap F_i \cap S|}{n^{2/3}}$$

so X_i^j is the number of vertices of $F_i \cap S$ allocated to cluster j, normalised by $n^{2/3}$ so that, by (ii), each X_i^j lies in the range $[0, 1]$. Observe, crucially, that (iii) implies that for each $i \in [s]$ the allocation of vertices in F_i conditioned on the value of $\varphi(z_i)$ is independent of $\{\varphi(w){:}w \in \bigcup_{j<i} F_j\}$. Indeed, for any fixed value of $\varphi(z_i)$ the value of $\varphi(x)$ for $x \in F_i$ depends only on the random choices made by Algorithm 32 for descendants of z_i on the path between z_i and x, none of which are in Z^* by choice of z_i, and each of these choices are independent of all other choices made by the algorithm. Hence, for each $q \in [k]$, we have $\mathbb{E}(X_i^j \mid X_{i-1}^j, \dots, X_1^j, \varphi(z_i) = q) = \mathbb{E}(X_i^j \mid \varphi(z_i) = q)$. It follows that for every $i \in [s]$ and $j \in [k]$ we have

$$\begin{aligned}\mathbb{E}(X_i^j \mid X_{i-1}^j, \dots, X_1^j) &\leq \max_{q\in[k]} \mathbb{E}(X_i^j \mid X_{i-1}^j, \dots, X_1^j, \varphi(z_i) = q) = \max_{q\in[k]} \mathbb{E}(X_i^j \mid \varphi(z_i) = q) \\ &= \max_{q\in[k]} \frac{\sum_{x\in F_i\cap S} \mathbb{P}(x \in V_j \mid \varphi(z_i) = q)}{n^{2/3}} \leq \frac{1}{k}\left(1 + \frac{1}{2\log n}\right)\frac{|F_i \cap S|}{n^{2/3}}.\end{aligned}$$

To see that the final inequality holds, note that by (iv) we have $\mathrm{dist}_F(z_i, y) \geq 56k^5 \log\log n$; since $\mathcal{E}$ does not contain a path of length seven, at least a seventh of the edges on the path between z_i and y are not in $\mathcal{E}$, and it follows that $\mathrm{dist}_{\overline{\mathcal{E}}}(z_i, y) \geq 8k^5 \log\log n \geq 4k^5 \log(2k \log n)$. So we may apply Proposition 33 with $2k \log n$ in place of m, giving the desired inequality.

We apply Lemma 15 with

$$\mu := \frac{1}{k}\left(1 + \frac{1}{2\log n}\right)\frac{|S|}{n^{2/3}} \geq \frac{1}{k}\left(1 + \frac{1}{2\log n}\right)\sum_{i\in[s]} \frac{|F_i \cap S|}{n^{2/3}},$$

which (since $\frac{1}{n} \ll \frac{1}{k}$) yields

$$\mathbb{P}\Big(\sum_{i\in[s]} X_i^j > \Big(1 + \frac{1}{3\log n}\Big)\mu\Big) \leq \exp\left(\frac{-\mu}{27(\log n)^2}\right) \leq \exp\left(-\frac{|S|}{27kn^{2/3}(\log n)^2}\right) \leq \exp\big(-n^{1/4}\big).$$

Taking a union bound we find that with high probability, for every $i \in [s]$ and $j \in [k]$ the event described does not occur, implying that for every $j \in [k]$ we have

$$n^{2/3}\sum_{i\in[s]} X_i^j \leq n^{2/3}\left(1 + \frac{1}{3\log n}\right)\mu \leq \frac{|S|}{k}\left(1 + \frac{1}{\log n}\right).$$

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

In other words, for each $j \in [k]$ we have $\left|\varphi^{-1}(j) \cap S \cap \bigcup_{i\in[s]} F_i\right| \le \frac{|S|}{k}\left(1+\frac{1}{\log n}\right)$. Since every vertex is in $\varphi^{-1}(j)$ for precisely one $j \in [k]$, it follows that for each $j \in [k]$ we have

$$\begin{aligned}\frac{|S|}{k}\left(1+\frac{1}{\log n}\right)+\left|S\setminus\bigcup_{i\in[s]}F_i\right| &\ge \left|\varphi^{-1}(j)\cap S\right| \ge \left|S\cap\bigcup_{i\in[s]}F_i\right| - \frac{(k-1)|S|}{k}\left(1+\frac{1}{\log n}\right)\\ &\ge \frac{|S|}{k}-\frac{|S|}{\log n}-\left|S\setminus\bigcup_{i\in[s]}F_i\right|,\end{aligned}$$

and together with (i) this gives the desired conclusion. □

4. Proof of Theorem 4

Note that Theorem 4(1) follows from Theorem 4(2) by appending to Q a directed path of order $|G| - n \ge \alpha n$ joined to Q by a single edge (yielding a graph Q' of order $|V(G)|$), and adjusting the remaining constants accordingly. Hence, it suffices to prove Theorem 4(2). We do this by establishing the following more general result (to see that this implies Theorem 4(2), observe that the fact that every edge of Q_0 is subdivided at least once implies that Q is 2-degenerate, and deleting the vertices of Q_0 from Q yields a 1-degenerate subgraph).

Theorem 35. *Suppose $\frac{1}{n} \ll \lambda \ll \alpha$. Let Q be a 2-degenerate graph of order n with maximum degree $\Delta(Q) \le \exp(\sqrt{\log n})$, and suppose that Q can be made 1-degenerate by deleting from it at most $n^{0.99}$ vertices. If Q contains either λn pairwise vertex-disjoint bare paths of order 7 or λn pairwise disjoint edges incident to leaves, then every orientation of Q is contained in every directed graph G of order n with $\delta^0(G) \ge (1/2+\alpha)n$.*

To prove Theorem 35, introduce new constants $K, K', \varepsilon, \varepsilon', \gamma, \beta, d, \eta$ with

$$\frac{1}{n} \ll \frac{1}{K} \ll \frac{1}{K'} \ll \varepsilon \ll \varepsilon' \ll \gamma \ll \beta \ll d \ll \lambda \ll \eta \ll \alpha.$$

Let $\mathcal{P}_{\text{undir}}$ be a collection of λn pairwise vertex-disjoint paths in Q, where either each $P \in \mathcal{P}_{\text{undir}}$ is a bare path of order 7, or each $P \in \mathcal{P}_{\text{undir}}$ is a single edge (i.e., a path of length 1) one of whose endvertices is a leaf. Fix an arbitrary orientation of Q; our goal is then to construct an embedding ϱ of Q into G.

4.1. Anatomy of the treelike structure

We assume without loss of generality that at most one component of Q is a tree. We may do this because if more than one component of Q is a tree, then by Proposition 13 we may add edges to Q to connect all these trees into a single tree without affecting the conditions on Q in the statement of the theorem.

Let Q_0 be a set of at most $n^{0.99}$ vertices of Q whose deletion turns Q into a 1-degenerate graph. In other words, every component of $Q - Q_0$ is a tree; for each such tree T, let A_T be the set of attachments of T in Q (these are vertices of T with neighbours in Q_0). Observe that the total number of attachments, over all components, is then $|\bigcup_T A_T| \le |Q_0|\Delta(Q)$, where the union is taken over all components T of $Q - Q_0$. Also, our previous assumption implies that at most one tree has no attachments.

We apply Proposition 12 to each component T of $Q - Q_0$, with A_T in place of X; this yields a set $Y_T \subseteq V(T)$ with $|Y_T| \le \max(6|A_T|, 1)$ such that $A_T \subseteq Y_T$, so that each component of $T - Y_T$ contains at most $|T|/2 \le n/2$ vertices, and such that each component T' of $T - Y_T$ has either one or two attachments in T (these are vertices with neighbours in Y_T), each of which has only

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

one neighbour in Y_T (and hence in $V(Q) \setminus V(T)$), with the additional property that if T' has two attachments in T then these are not adjacent.

We set $V_{\text{ground}} := Q_0 \cup \bigcup_T Y_T$, where the union is taken over all components T of $Q - Q_0$. So

$$|V_{\text{ground}}| \leq |Q_0| + \Big|\bigcup_T Y_T\Big| \leq |Q_0| + 6\Big|\bigcup_T A_T\Big| + 1 \leq |Q_0|(1 + 6\Delta(Q)) + 1 < 7n^{0.99} \exp\left(\sqrt{\log n}\right) \leq n^{0.995}. \tag{9}$$

We also set $F = Q - V_{\text{ground}}$, so F is a forest, and let $\mathcal{T}$ be the set of components of F. So each $T \in \mathcal{T}$ is an oriented tree with $|T| \leq n/2$, and our choice of V_{ground} ensures that each $T \in \mathcal{T}$ has either one or two attachments in Q (these are vertices with a neighbour in V_{ground}), that each attachment has at most one neighbour in V_{ground}, and that if T has two attachments, then these vertices are not adjacent in T. Moreover, by (9) we have

$$|\mathcal{T}| \leq |V_{\text{ground}}|\Delta(Q) < n^{0.995} \exp\left(\sqrt{\log n}\right) \leq n^{0.999}. \tag{10}$$

For each $T \in \mathcal{T}$ fix r^T as the root of T. We may then fix a tidy ancestral order $\prec$ of the forest F by Lemma 8. So in particular, r^T appears before any other vertex of T in the order $\prec$. If T has two attachments in Q then we refer to the other attachment as the *secondary attachment* s^T of T. Let $\hat{r}^T$ (respectively, $\hat{s}^T$) denote the sole neighbour of r^T (respectively, s^T) which lies outside of T, and let p^T denote the parent of s^T in T (where we consider T rooted at r^T). So the vertices $r^T, p^T, s^T, \hat{r}^T$ and $\hat{s}^T$ are all distinct. We partition $\mathcal{T}$ into sets $\mathcal{T}_1$ and $\mathcal{T}_2$ such that

$$n/3 \leq \Big|\bigcup_{T \in \mathcal{T}_1} V(T)\Big|, \Big|\bigcup_{T \in \mathcal{T}_2} V(T)\Big| \leq 2n/3,$$

which is possible since $|T| \leq n/2$ for every $T \in \mathcal{T}$. Call a path $P \in \mathcal{P}_{\text{undir}}$ *unfit* if P either

(i) contains a vertex of V_{ground}, or

(ii) contains either p^T, r^T or s^T for some $T \in \mathcal{T}$.

We call the remaining paths in $\mathcal{P}_{\text{undir}}$ *fit*. Observe that $|\mathcal{P}_{\text{undir}}| = \lambda n$, and $\mathcal{P}_{\text{undir}}$ contains at most $|V_{\text{ground}}| + 3|\mathcal{T}| \leq \lambda n/3$ unfit paths. So by relabelling $\mathcal{T}_1$ and $\mathcal{T}_2$ if necessary, we may assume that there are at least $\lambda n/3$ fit paths in $\mathcal{P}_{\text{undir}}$ which are entirely contained in trees in $\mathcal{T}_1$. Furthermore, considering the fixed roots r^T chosen for each $T \in \mathcal{T}$, each such path P has a well-defined pattern (as defined in Section 3.6, with respect to the root r^T of the tree T which contains P). Since there are at most 2^6 possible patterns for the orientation of a rooted path $P \in \mathcal{P}_{\text{undir}}$, we may choose a set $\mathcal{P} \subseteq \mathcal{P}_{\text{undir}}$ with $|\mathcal{P}| = \lambda n/2^8$ such that every $P \in \mathcal{P}$ is fit, lies in some $T \in \mathcal{T}_1$, and has the same pattern $\hat{P}$. For each $P \in \mathcal{P}$ let $v_1^P, \ldots, v_7^P$ be the vertices of P, labelled with $v_1^P \prec v_2^P \prec \ldots \prec v_7^P$.

4.2. Reduced graph

We next construct a regular partition of G, which plays a crucial role in the allocation and embedding phases. We apply Lemma 21 to G to obtain an integer k with $K' \leq k \leq K$, a partition $V_0 \dot\cup V_1 \dot\cup \cdots \dot\cup V_k$ of $V(G)$ and a digraph $R^\star$ with $V(R^\star) = V_0 \dot\cup [k]$ satisfying properties (a)–(f) (in particular, m is defined to be the common size of the clusters $V_1, \ldots, V_k$, and satisfies $(1-\varepsilon)n/k \leq m \leq n/k$). Note that since $|V_0| \leq \varepsilon n$ we have

$$\delta^0(G - V_0) \geq (1/2 + \alpha)n - |V_0| \geq (1/2 + \alpha/2)n. \tag{11}$$

By Lemma 26 applied to $R^\star[[k]]$, there is a connecting set $\mathcal{D}$ of $\hat{P}$-diamonds in $R^\star[[k]]$ with $|\mathcal{D}| = k - 1$ such that each $i \in [k]$ lies in at most $4/\eta$ diamonds in $\mathcal{D}$. Let $H^\diamond = \bigcup_{\diamondsuit \in \mathcal{D}} \diamondsuit$, and let $H \subseteq$

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

$R^\star[[k]]$ be the Hamilton cycle $1 \to 2 \to \cdots \to k \to 1$ (recall that Lemma 21 guarantees that this is indeed a cycle in $R^\star[[k]]$). By Lemma 28, $R^\star[[k]]$ contains a spanning d_J-regular expander J, with $H \subseteq J$.

4.3. Allocation

The next step in the proof is to allocate vertices to clusters (vertices of V_{ground} will in fact be embedded at this step). To do this, we fix an embedding ϱ of the vertices in V_{ground} and let φ_0 be the associated allocation. We then choose an allocation φ_{root} for the roots of each $T \in \mathcal{T}$, and fix a set $I(T) \subseteq [k]$ of candidate allocations for each secondary attachment s^T. Finally the allocation of Q is completed using Claim 36 (proved in Section 4.5).

Since Q is 2-degenerate, the same is true of $Q[V_{\text{ground}}]$, so the small size of V_{ground} allows us to greedily embed its vertices to $G - V_0$ as follows. Fix an ordering $r_1, r_2, \ldots, r_t$ of the vertices in V_{ground} such that for each $i \in [t]$ the vertex r_i has at most two neighbours in $\{r_1, \ldots, r_{i-1}\}$, and fix the image $\varrho(r_1)$ as some arbitrary vertex in $V(G) \setminus V_0$. By (9) and (11), for all $\bullet, \diamond \in \{-, +\}$ and all $x, y \in V(G)$, we have

$$\left|(N^\bullet_G(x) \cap N^\diamond(y)) \setminus V_0\right| \geq \alpha n - \varepsilon n > |V_{\text{ground}}|,$$

and hence for each $1 < i \leq t$ there is an appropriate choice for the image $\varrho(r_i)$ among the vertices of $V(G) \setminus V_0$ which are not yet in the image of ϱ. For each $x \in V_{\text{ground}}$, having fixed $\varrho(x)$, set $\varphi_0(x)$ so that $\varrho(x) \in V_{\varphi_0(x)}$; this means that each vertex of V_{ground} is allocated to the cluster to which it is embedded.

Let us now allocate the roots r^T of each $T \in \mathcal{T}$. For each $v \in V_{\text{ground}}$, and each $\bullet \in \{-, +\}$, by (11) we may choose $i(v, \bullet) \in [k]$ such that

$$\deg^\bullet_G(\varrho(v), V_{i(v,\bullet)}) \geq (1 + \alpha)m/2. \tag{12}$$

For each $T \in \mathcal{T}$, set $\varphi_{\text{root}}(r^T) = i(\hat{r}^T, \bullet)$ where $r^T \in N^\bullet(\hat{r}^T)$. Moreover, for each $T \in \mathcal{T}$ with a secondary attachment $s^T \in N^\bullet(\hat{s}^T)$, recall that $\hat{s}^T$ is the unique neighbour of s^T in V_{ground}, so $\varrho(\hat{s}^T)$ has been defined. Let $I(T)$ be the set of $j \in [k]$ such that

$$\deg^\bullet_G(\varrho(\hat{s}^T), V_j) \geq \eta m. \tag{13}$$

Since each cluster has at most $m \leq n/k$ vertices, we then have

$$|I(T)| \geq (1/2 + \eta)k, \tag{14}$$

since otherwise $\deg^\bullet_{G - V_0}(\varrho(\hat{s}^T)) < \eta m \cdot k + m \cdot (1/2 + \eta)k \leq (1/2 + 2\eta)n$, contradicting (11). We now allocate all remaining vertices of Q through the following claim (proved in Section 4.5), for which we define

$$g := \left\lceil \frac{\lambda m}{2^{10}} \right\rceil$$

Claim 36. *There exist disjoint $\mathcal{P}^0, \mathcal{P}^H \subseteq \mathcal{P}$ and an extension of φ_0 to a map $\varphi : V(Q) \to V(R^\star)$ such that*

- *(i) for each $T \in \mathcal{T}$ the restriction φ^T of φ to T is a homomorphism from T to $R^\star$ with $\Delta(\varphi^T) \leq 5$;*
- *(ii) $|\mathcal{P}^0| = |V_0|$, and the restriction of φ to either the centres v^P_4 of paths $P \in \mathcal{P}^0$ (if these paths have order 7) or the non-root vertices v^P_2 of paths $P \in \mathcal{P}^0$ (if these paths have order 2, in which case v^P_2 is a leaf of Q) is a bijection from that set to V_0;*
- *(iii) for each $i \in [k]$ we have $|\varphi^{-1}(i) \cap N| \leq 6\varepsilon m/\alpha$, where $N := \bigcup_{x \in V(Q):\varphi(x) \in V_0} N^-_Q(x) \cup N^+_Q(x)$;*
- *(iv) φ maps precisely m vertices to each $i \in [k]$;*

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

(v) *for each* $P \in \mathcal{P}^H$, *the restriction of* φ *to* P *is a homomorphism from* P *to* H;

(vi) φ *maps precisely* g *roots* v_1^P *of paths in* $\mathcal{P}^H$ *to each* $i \in [k]$;

(vii) *For each* $T \in \mathcal{T}$ *we have* $\varphi(r^T) = \varphi_{\text{root}}(r^T)$ *and, if* T *has a secondary attachment* s^T, *then* $\varphi(s^T) \in I(T)$.

4.4. Embedding

Fix $\mathcal{P}^0$, $\mathcal{P}^H$, and an allocation φ extending φ_0 as obtained from Claim 36, and let M be the set of pendant vertices of paths in $\mathcal{P}^H$, so $M := \{v_2^P : P \in \mathcal{P}^H\}$ if paths in $\mathcal{P}$ have order 2, and $M := \{v_2^P, v_3^P, v_4^P, v_5^P, v_6^P : P \in \mathcal{P}^H\}$ if the paths in $\mathcal{P}$ have order 7. Also say that $v \in V(Q)$ is a *distinguished vertex* if $\varphi(v) \in V_0$. For each distinguished vertex $v \in V(Q)$ set $\varrho(v) = \varphi(v)$. So ϱ now embeds all vertices in V_{ground} and all distinguished vertices. The bulk of the embedding is achieved by the following claim (proved in Section 4.6).

Claim 37. *There exists an extension of* ϱ *to an embedding of* $Q - M$ *in* G *such that properties (i)–(iv) below hold. For each* $i \in [k]$, *write*

$$\begin{aligned} U_i &:= \{\, \varrho(u) \in V_i : u = v_1^P \text{ for some } P \in \mathcal{P}^H \,\}, \\ W_i &:= \{\, \varrho(w) \in V_i : w = v_P^7 \text{ for some } P \in \mathcal{P}^H \,\}, \text{ and} \\ V_i^\star &:= V_i \setminus \varrho(Q - M). \end{aligned}$$

(i) *The embedding* ϱ *respects the allocation, meaning that* $\varrho(x) = \varphi(x)$ *if* $\varphi(x) \in V_0$ *and* $\varrho(x) \in V_{\varphi(x)}$ *for all other* $x \in V(Q) \setminus M$.

(ii) *If the paths in* $\mathcal{P}$ *have order two then for each* $i \in [k]$ *we have* $|U_i| = |V_i^*| = g$ *and* $W_i = \emptyset$.

(iii) *If the paths in* $\mathcal{P}$ *have order seven then for each* $i \in [k]$ *we have* $|U_i| = |W_i| = g$ *and* $|V_i^*| = 5g$.

(iv) *For each* $i \in [k]$ *the graphs* $G[V_{i-1}^\star \to U_i]$, $G[U_i \to V_{i+1}^\star]$, $G[V_{i-1}^\star \to W_i]$, $G[W_i \to V_{i+1}^\star]$ *and* $G[V_{i-1}^\star \to V_i^\star]$ *are each* (β, ε')*-superregular.*

Fix such an (extended) embedding ϱ of $Q - M$ in G. We complete ϱ to an embedding of Q in G by defining the images of vertices in M, which are the pendant vertices of paths in $\mathcal{P}^H$, as follows.

Suppose first that the paths in $\mathcal{P}$ each have order 2, in which case they are pairwise-disjoint edges incident to leaves. So each path $P \in \mathcal{P}^H$ has a root vertex v_1^P and another vertex v_2^P which is a leaf vertex of Q, and either v_2^P is an outneighbour of v_1^P for every $P \in \mathcal{P}^H$ or v_2^P is an inneighbour of v_1^P for every $P \in \mathcal{P}^H$. We assume the former; the argument for the latter case is similar. For each $i \in [k]$ we have $|U_i| = |V_{i+1}^\star| = g$ by Claim 37 (ii); since $G[U_i \to V_{i+1}^\star]$ is (β, ε')-superregular by Claim 37 (iv), it follows that $G[U_i \to V_{i+1}^\star]$ contains a perfect matching M_i by Lemma 18. For each $P \in \mathcal{P}^H$ with $v_1(P) \in U_i$ set $\varrho(v_2^P)$ to be the vertex in $V_{i+1}^\star$ which M_i matches to $\varrho(v_1^P)$; doing this for each $i \in [k]$ gives the desired embedding ϱ of Q in G.

Now suppose instead that the paths in $\mathcal{P}$ each have order 7. For each $i \in [k]$, let $Z_i^1 = U_i$ and $Z_i^7 = W_i$, and choose uniformly at random an equipartition of $V_i^\star$ into five sets Z_i^j for $j \in [2, 3, 4, 5, 6]$; the choice for each $i \in [k]$ is independent of all others. So by Claim 37 (iii) we have $|Z_i^j| = g$ for each $i \in [k]$ and $j \in [7]$. Moreover, for each $j \in \{2, 3, 4, 5, 6\}$ the set Z_i^j is a uniformly-random subset of $V_i^\star$ of size $|V_i^\star|/5$, whilst tautologically Z_i^1 is a uniformly-random subset of U_i of size $|U_i|$ and Z_i^7 is a uniformly-random subset of W_i of size $|W_i|$. Using Claim 37 (iv) we may apply Lemma 17 and take a union bound over both events for each $i \in [k]$ and $\ell \in [6]$ to find that, with positive probability, for each $i \in [k]$ and each $\ell \in [6]$ the graphs

$$G[Z_{i-1}^{\ell+1} \to Z_i^\ell] \text{ and } G[Z_i^\ell \to Z_{i+1}^{\ell+1}] \text{ are each } (\beta, 5\varepsilon') - \text{superregular}, \tag{15}$$

with addition on the indices taken modulo k. Fix an outcome for which each of these events occurs.

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

Recall that all paths in $\mathcal{P}$ have the same pattern $\hat{P}$. Let $u_1, \dots, u_7$ be the vertices of $\hat{P}$, ordered as they appear in $\hat{P}$ with root u_1. For each $\ell \in [7]$, let $\Sigma(\ell)$ denote the difference between the number of forward and backward edges in the subpath of P from u_1^P up to u_ℓ^P. In other words, we define

$$\Sigma(\ell) := \big|\{j \in [\ell-1] : u_j \to u_{j+1} \in E(\hat{P})\}\big| - \big|\{j \in [\ell-1] : u_j \leftarrow x_{j+1} \in E(\hat{P})\}\big|,$$

and remark that $\Sigma(1) = 0$. With this definition, (15) implies that for each $\ell \in [6]$ and $i \in [k]$ the sets $Z^{\ell}_{i+\Sigma(\ell)}$ and $Z^{\ell+1}_{i+\Sigma(\ell+1)}$ form a $(\beta, 5\varepsilon')$-superregular pair in the direction of the edge between u_ℓ and $u_{\ell+1}$.

Fix $i \in [k]$ and let $\mathcal{P}_i^H := \{P \in \mathcal{P}^H : \varrho(v_1^P) \in V_i\}$. So each $P \in \mathcal{P}_i^H$ has $\varrho(v_1^P) \in U_i = Z_i^1$ and also $\varrho(v_7^P) \in W_{i+\Sigma(7)} = Z^7_{i+\Sigma(7)}$ by Claim 36 (v) and Claim 37 (i), while $v_2^P, \dots, v_6^P$ remain to be embedded. Let $\pi : Z_i^1 \to Z^7_{i+\Sigma(7)}$ be the bijection with $\pi(\varrho(v_1^P)) = \varrho(v_7^P)$ for each $P \in \mathcal{P}_i^H$, and let L_i be the 7-layer graph with vertex classes $Z^j_{i+\Sigma(j)}$ for $j \in [7]$. By applying Lemma 19 to L_i, we obtain a collection $\mathcal{P}_i'$ of g pairwise vertex-disjoint paths of order 7 in L_i such that for each $P \in \mathcal{P}_i^H$ there is a path $P' \in \mathcal{P}_i'$ with ends $\varrho(v_1^P)$ and $\varrho(v_7^P)$. Let the vertices of this path be $x_1^P, x_2^P, \dots, x_7^P$ in that order, so $x_1^P = \varrho(v_1^P)$, $x_7^P = \varrho(v_7^P)$ and $x_j^P \in Z^j_{i+\Sigma(j)}$ for each $j \in [7]$. Set $\varrho(v_j^P) = x_j^P$ for each $P \in \mathcal{P}_i^H$ and $j \in \{2,3,4,5,6\}$; doing this for each $i \in [k]$ gives the desired embedding ϱ of Q in G.

This completes the proof of Theorem 4 except for the proofs of Claims 36 and 37, which are contained in the next two sections.

4.5. Proof of Claim 36

Let F_1 and F_2 be the forests formed by the trees in $\mathcal{T}_1$ and $\mathcal{T}_2$ respectively. The allocation proceeds in two phases. Roughly speaking, in the first phase we allocate F_1 using Algorithm 32, enforcing that the root of each $T \in \mathcal{T}$ is mapped in accordance with φ_{root} and that each secondary attachment is mapped within a set V_i with $i \in I(T)$; we also ensure that paths in $\mathcal{P}$ are mapped along H. To conclude this phase, we modify the allocation of some paths in $\mathcal{P}$, re-routing them so as to go through every vertex in V_0 and through the connecting set $\mathcal{D}$ of diamonds in $R^\star[[k]]$. The resulting allocation φ_1 will map precisely one vertex of F_1 to each vertex of V_0, and will map the remaining vertices of F_1 approximately uniformly among $i \in [k]$, so that $\big|\,|\varphi_1^{-1}(i)| - |\varphi_1^{-1}(j)|\,\big| \le 10\varepsilon n/(\eta k)$ for all $i, j \in [k]$.

In the second phase, we build an allocation φ_2 of F_2 using a biased allocation algorithm (Algorithm 32 is not really modified, but we apply it to an auxiliary digraph, which will produce the desired bias). This ensures that the combination φ_{join} of φ_0, φ_1 and φ_2 will map $V(Q)$ much more uniformly over $[k]$. In particular, for all $i, j \in [k]$ we shall have $\big|\,|\varphi_{\mathrm{join}}^{-1}(i)| - |\varphi_{\mathrm{join}}^{-1}(j)|\,\big| < 4n \log\log n / \log n$. Finally we complete the proof by modifying the allocation of paths routed through diamonds to obtain a perfectly uniform allocation φ with the desired properties.

For each path $P \in \mathcal{P}$, given an allocation π of the root v_1^P of P, a *canonical* allocation of P is a homomorphism π from P into H which extends π.

The first phase. Define permutations $\pi^+ : [k] \to [k]$ and $\pi^- : [k] \to [k]$ such that $\pi^+(i) = i+1$ and $\pi^-(i) = i-1$ for each $i \in [k]$, with addition and subtraction taken modulo k. So π^+ maps each $i \in [k]$ to its outneighbour in H, and likewise π^- maps each $i \in [k]$ to its inneighbour in H. Let $\mathcal{E} := \bigcup_{P \in \mathcal{P}} E(P)$, and for each edge $e \in \mathcal{E}$ joining v_ℓ^P and $v_{\ell+1}^P$, define π_e by

$$\pi_e := \begin{cases} \pi^- & \text{if } v_{\ell+1}^P \in N^-_{F_1}(v_\ell^P), \\ \pi^+ & \text{otherwise.} \end{cases}$$

We apply Algorithm 32 to F_1 and J to obtain a map $\psi : V(F_1) \to [k]$. We do with π_e as defined above for each edge in $\mathcal{E}$, and with $Z = \bigcup_{T \in \mathcal{T}_1} \{r^T, s^T\}$. Our choice of allocation for vertices

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

of Z is as follows: for each $T \in \mathcal{T}_1$ we set $\psi(r^T) = \varphi_{\mathrm{root}}(r^T)$, whilst for each $T \in \mathcal{T}_1$ with a secondary attachment s^T we set $\psi(s^T)$ to be some $j \in [k]$ which lies in both $I(T)$ and the appropriate neighbourhood of the image of p^T (recall that p^T is the parent of s^T, so $\psi(p^T)$ will already have been defined when the algorithm comes to choose $\psi(s^T)$). More precisely, fix $\bullet \in \{-,+\}$ so that $s^T \in N_T^\bullet(p^T)$, and choose $\psi(s^T) \in N^\bullet_{R^\star[[k]]}\big(\psi(p^T)\big) \cap I(T)$ (this is possible since $|I(T)| > k/2$ by (14) and $\delta^0\big(R^\star[[k]]\big) > k/2$ by Lemma 21 (e)). Observe that, with these definitions, Algorithm 32 ensures that for each edge $u \to v$ of F_1 we have that $\psi(u) \to \psi(v)$ is an edge of J, *except* possibly when $\{u,v\} = \{p^T, s^T\}$, in which case the fact that $\psi(s^T)$ was chosen in $N^\bullet_{R^\star[[k]]}\big(\psi(p^T)\big)$ ensures that $\psi(u) \to \psi(v)$ is an edge of $R^\star[[k]]$. Since $J \subseteq R^\star[[k]]$, it follows that ψ is an allocation of F_1 to $R^\star[[k]]$ (that is, a homomorphism from F_1 to $R^\star[[k]]$). Moreover, ψ has the following crucial properties. First, our choices of $\mathcal{E}$ and π_e ensure that the allocation of each path in $\mathcal{P}$ by ψ is canonical. Second, for each $T \in \mathcal{T}_1$ we have $\psi(r^T) = \varphi_{\mathrm{root}}(r^T)$ and, if T has a secondary attachment, then our choice of $\psi(s^T)$ ensures that $\psi(s^T) \in I(T)$. Third, by Lemma 34 (with $S = V(F_1)$), with high probability we have for each $i \in [k]$ that

$$|\psi^{-1}(i)| = \frac{v(F_1)}{k} \pm \left(\frac{n}{\log n} + (|Z| + 3n^{1/3})\Delta(F_1)^{56k^5 \log\log n}\right) = \frac{v(F_1)}{k} \pm \frac{2n}{\log n}, \tag{16}$$

using that $|Z| \le 2n^{0.999}$ by (10) and that $3n^{0.999}\Delta(F_1)^{56k^5\log\log n} \le n/\log n$. Fourth, set $\Lambda := \{v_1^P : P \in \mathcal{P}\}$, so Λ is the set of roots of paths in $\mathcal{P}$, and for each $i \in [k]$ let $\mathcal{P}_i \subseteq \mathcal{P}$ consist of all paths $P \in \mathcal{P}$ with $\psi(v_1^P) = i$, that is, whose root is allocated to V_i. Then $|\Lambda| = |\mathcal{P}| = \lambda n/2^8$, and a similar application of Lemma 34 (with $S = \Lambda$) shows that with high probability we have for each $i \in [k]$ that

$$|\mathcal{P}_i| = |\psi^{-1}(i) \cap \Lambda| = \frac{\lambda n}{2^8 k} \pm \frac{2n}{\log n} \ge \frac{\lambda m}{2^9} \tag{17}$$

Fix an outcome of Algorithm 32 for which ψ has each of these properties. We now modify ψ to obtain the allocation φ_1 which is our goal in this phase; the procedure for this varies according to whether the paths in $\mathcal{P}$ have order 2 or 7.

Let us handle the former case first, that is, when each path in $\mathcal{P}$ is an edge incident to a leaf. Let $\bullet \in \{-,+\}$ be the sign such that $v_1^P \in N^\bullet(v_2^P)$ for all $P \in \mathcal{P}$. By Lemma 21 (f), we have $\big|N^\bullet_{R^\star}(x) \cap [k]\big| \ge \alpha k/2$ for each $x \in V_0$, so by Lemma 24 there exists a function $\mathfrak{g} : V_0 \to [k]$ such that $\mathfrak{g}(x) \in N^\bullet_{R^\star}(x)$ for each $x \in V_0$ and $\big|\mathfrak{g}^{-1}(i)\big| \le 1 + 2|V_0|/(\alpha k) \le 3\varepsilon m/\alpha$ for each $i \in [k]$. Next, for each $\diamondsuit \in \mathcal{D}$ let $\mathcal{D}_i$ be the set of diamonds in $\diamondsuit$ with prefix i, so for each $i \in [k]$ we have $|\mathcal{D}_i| \le 4/\eta$ by our choice of $\mathcal{D}$. For each $i \in [k]$ choose pairwise disjoint subsets $\mathcal{P}_i^0$, $\mathcal{P}_i^\diamond$, $\mathcal{P}_i^H \subseteq \mathcal{P}_i$ with sizes $|\mathcal{P}_i^0| = \big|\mathfrak{g}^{-1}(i)\big|$, $|\mathcal{P}_i^\diamond| = 2n|\mathcal{D}_i|/k^2$ and $|\mathcal{P}_i^H| = g$. This is possible by (17) since

$$|\mathcal{P}_i^0| + |\mathcal{P}_i^\diamond| + |\mathcal{P}_i^H| = \big|\mathfrak{g}^{-1}(i)\big| + \frac{2n|\mathcal{D}_i|}{k^2} + g \le \frac{3\varepsilon m}{\alpha} + \frac{2n}{k^2}\cdot\frac{4}{\eta} + \left\lceil\frac{\lambda m}{2^{10}}\right\rceil \le \frac{\lambda m}{2^9} \le |\mathcal{P}_i|.$$

Let $\mathcal{P}^0 := \bigcup_{i\in[k]} \mathcal{P}_i^0$, $\mathcal{P}^\diamond := \bigcup_{i\in[k]} \mathcal{P}_i^\diamond$ and $\mathcal{P}^H := \bigcup_{i\in[k]} \mathcal{P}_i^H$, so in particular $|\mathcal{P}^0| = \sum_{i\in[k]} \big|\mathfrak{g}^{-1}(i)\big| = |V_0|$.

Our choice of $\mathcal{P}^0$ allows us to choose, for each $P \in \mathcal{P}^0$, an image $\varphi_1(v_2^P) \in \mathfrak{g}^{-1}\big(\psi(v_1^P)\big)$ so that the chosen images $\varphi_1(v_2^P)$ for each $P \in \mathcal{P}^0$ are all distinct. We also set $\varphi_1(v_1^P) := \psi(v_1^P)$ for each $P \in \mathcal{P}^0$. So φ_1 is a homomorphism from the paths in $\mathcal{P}^0$ to $R^\star$ whose restriction to $\{v_2^P : P \in \mathcal{P}^0\}$ is a bijection from that set to V_0, which will ensure (ii). Moreover, the neighbours in Q of vertices mapped to V_0 are the vertices v_1^P for $P \in \mathcal{P}$, of which at most $\big|\mathfrak{g}^{-1}(i)\big| \le 3\varepsilon m/\alpha$ are mapped to each $i \in [k]$; this will ensure (iii).

Similarly, our choice of $\mathcal{P}^\diamond$ allows us to choose a map $\mathfrak{h}{:}\mathcal{P}^\diamond \to \mathcal{D}$ so that for each $i \in [k]$ each $P \in \mathcal{P}_i^\diamond$ has $\mathfrak{h}(P) \in \mathcal{D}_i$ and so that each $\diamondsuit \in \mathcal{D}$ has $|\mathfrak{h}^{-1}(\diamondsuit)| = 2n/k^2$. This means that for each $\diamondsuit \in \mathcal{D}$ we may do the following. Let u be the prefix of $\diamondsuit$ (so $\diamondsuit \in \mathcal{D}_u$), and let $\{v,v'\}$ be the middle

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

of $\Diamond$. Choose n/k^2 paths $P \in \mathfrak{h}^{-1}(\Diamond)$ and for each set $\varphi_1(v_2^P) := v$; for each of the remaining n/k^2 paths $P \in \mathfrak{h}^{-1}(\Diamond)$ set $\varphi_1(v_2^P) := v'$. Also set $\varphi_1(v_1^P) := u = \psi(v_1^P)$ for every $P \in \mathfrak{h}^{-1}(\Diamond)$. So φ_1 gives a homomorphism from the paths in $\mathcal{P}^\diamond$ to $R^\star[[k]]$ which maps n/k^2 paths to each branch of each diamond $\Diamond \in \mathcal{D}$.

Finally, for each vertex $u \in V(F_1)$ for which $\varphi_1(u)$ has not yet been defined, set $\varphi_1(u) := \psi(u)$ (so u remains allocated as in the outcome of Algorithm 32). Since we previously set $\varphi_1(v_1^P) = \psi(v_1^P)$ for each path $P \in \mathcal{P}^0 \cup \mathcal{P}^\diamond$, we then have that φ_1 is a homomorphism from F_1 to $R^\star$. Moreover, the only vertices $u \in V(F_1)$ which may have $\psi(u) \neq \varphi_1(u)$ are vertices v_2^P for paths $P \in \mathcal{P}^0 \cup \mathcal{P}^\diamond$. Consequently, for each $i \in [k]$ the number of vertices $u \in V(F_1)$ with $\psi(u) = i$ and $\varphi_1(u) \neq i$ is at most $|\mathcal{P}^0_{i^*}| + |\mathcal{P}^\diamond_{i^*}|$, where $i^* = i - 1$ if $v_1^P \to v_2^P$ and $i^* = i + 1$ if $v_1^P \leftarrow v_2^P$. Similarly, the number of vertices $v \in V(F_1)$ with $\varphi_1(v) = i$ and $\psi(v) \neq i$ is at most $(4/\eta)(n/k^2)$ since v is in at most $4/\eta$ diamonds in $\mathcal{D}$. We conclude that for each $i \in [k]$ we have

$$\big|\, |\varphi_1^{-1}(i)| - |\psi^{-1}(i)| \,\big| \leq \frac{3\varepsilon m}{\alpha} + \frac{2n}{k^2} \cdot \frac{4}{\eta} + \frac{4}{\eta} \cdot \frac{n}{k^2} \leq \frac{4\varepsilon m}{\alpha}. \tag{18}$$

Now consider instead the case where the paths in $\mathcal{P}$ are bare paths of order 7. In this case, for each $i \in [k]$ we choose pairwise disjoint subsets $\mathcal{P}_i^0$, $\mathcal{P}_i^\diamond$, $\mathcal{P}_i^H \subseteq \mathcal{P}_i$ with $|\mathcal{P}_i^0| = |V_0|/k \pm 1$, $|\mathcal{P}_i^\diamond| = 2n|\mathcal{D}|/k^3 \pm 1$ and $|\mathcal{P}_i^H| = g$; the precise values of $|\mathcal{P}_i^0|$ and $|\mathcal{P}_i^\diamond|$ for $i \in [k]$ are chosen so that $\sum_{i\in[k]} |\mathcal{P}_i^0| = |V_0|$ and $\sum_{i\in[k]} |\mathcal{P}_i^\diamond| = 2n|\mathcal{D}|/k^2$. By (17) it is possible to make these choices since

$$|\mathcal{P}_i^0| + |\mathcal{P}_i^\diamond| + |\mathcal{P}_i^H| \leq \frac{|V_0|}{k} + 1 + \frac{2n|\mathcal{D}|}{k^3} + 1 + g \leq 2 + \frac{\varepsilon n}{k} + \frac{2n}{k^2} + \left\lceil \frac{\lambda m}{2^{10}} \right\rceil \leq \frac{\lambda m}{2^9} \leq |\mathcal{P}_i|.$$

As before we let $\mathcal{P}^0 := \bigcup_{i\in[k]} \mathcal{P}_i^0$, $\mathcal{P}^\diamond := \bigcup_{i\in[k]} \mathcal{P}_i^\diamond$ and $\mathcal{P}^H := \bigcup_{i\in[k]} \mathcal{P}_i^H$, so in particular $|\mathcal{P}^0| = |V_0|$ and $|\mathcal{P}_i^\diamond| = 2n|\mathcal{D}|/k^2$.

Fix a bijection $p : \mathcal{P}^0 \to V_0$. For each $P \in \mathcal{P}^0$ set $\varphi_1(v_4^P) = p(P)$. Also, for each $P \in \mathcal{P}^0$ set $\varphi_1(v_1^P) = \psi(v_1^P)$ and $\varphi_1(v_7^P) = \psi(v_7^P)$. Next, let $\bullet, \circ \in \{-, +\}$ be such that $v_3^P \in N_P^\bullet(v_4^P)$ and $v_5^P \in N_P^\circ(v_4^P)$, and choose $\varphi_1(v_3^P) \in N_{R^\star}^\bullet(\varphi_1(v_4^P))$ and $\varphi_1(v_5^P) \in N_{R^\star}^\circ(\varphi_1(v_4^P))$. Since $R^\star$ satisfies property (f) of Lemma 21, for each $P \in \mathcal{P}^0$ there are at least $\alpha k/2$ options for the choices of $\varphi_1(v_3^P)$ and of $\varphi_1(v_5^P)$. So by Lemma 24 we may make these choices so that for each $i \in [k]$ at most $1 + |V_0|/(\alpha k/2) \leq 3\varepsilon m/\alpha$ paths $P \in \mathcal{P}^0$ have $\varphi_1(v_3^P) = i$ and at most $3\varepsilon m/\alpha$ paths $P \in \mathcal{P}^0$ have $\varphi_1(v_5^P) = i$. Next, for each $P \in \mathcal{P}^0$, choose $\varphi_1(v_2^P)$ to be an appropriate common neighbour of $\varphi_1(v_1^P)$ and $\varphi_1(v_3^P)$ in $R^\star[[k]]$ (here 'appropriate' means respecting the direction of the edges between v_1^P and v_2^P and between v_2^P and v_3^P), and likewise choose $\varphi_1(v_6^P)$ to be an appropriate common neighbour of $\varphi_1(v_5^P)$ and $\varphi_1(v_7^P)$ in $R^\star[[k]]$. Since $\delta^0(R^\star[[k]]) \geq (1/2 + \eta)k$, there are at least $2\eta k$ options for each image, so by Lemma 24 we may make these choices in such a way that for each $i \in [k]$ at most $1 + \varepsilon n/(2\eta k) \leq \varepsilon m/\eta$ paths $P \in \mathcal{P}^0$ have $\varphi_1(v_2^P) = i$ and at most $\varepsilon m/\eta$ paths $P \in \mathcal{P}^0$ have $\varphi_1(v_6^P) = i$. Then φ_1 is a homomorphism from the paths in $\mathcal{P}^0$ to $R^\star$ whose restriction to $\{v_4^P : P \in \mathcal{P}^0\}$ is a bijection from that set to V_0, which will ensure (ii). Moreover, the neighbours in Q of vertices mapped to V_0 are the vertices v_3^P and v_5^P for $P \in \mathcal{P}$, of which at most $6\varepsilon m/\alpha$ are mapped to each $i \in [k]$; this will ensure (iii).

Next choose a map $f : \mathcal{P}^\diamond \mapsto \mathcal{D}$ such that for each $\Diamond \in \mathcal{D}$ we have $|f^{-1}(\Diamond)| = 2n/k^2$ (this is possible since $|\mathcal{P}^\diamond| = 2n|\mathcal{D}|/k^2$). For each $\Diamond \in \mathcal{D}$ we do the following. Let u, $\{v, v'\}$ and w be the prefix, middle and suffix of $\Diamond$. Choose n/k^2 paths $P \in f^{-1}(\Diamond)$ and for each set $\varphi_1(v_3^P) = u$, $\varphi_1(v_4^P) = v$, $\varphi_1(v_5^P) = w$; for each of the remaining n/k^2 paths $P \in f^{-1}(\Diamond)$ set $\varphi_1(v_3^P) = u$, $\varphi_1(v_4^P) = v'$, $\varphi_1(v_5^P) = w$. In both cases also set $\varphi_1(v_1^P) = \psi(v_1^P)$ and $\varphi_1(v_7^P) = \psi(v_7^P)$. Then, choose $\varphi_1(v_2^P)$ and $\varphi_1(v_6^P)$ exactly as we did for paths $P \in \mathcal{P}^0$; as before there are at least $2\eta k$ options for each image, so we may make these choices so that for each $i \in [k]$ at most $(2n|\mathcal{D}|/k^2)/(2\eta k) + 1 \leq 2n/(\eta k^2)$

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

paths $P \in \mathcal{P}^\diamond$ have $\varphi_1(v_2^P) = i$ and at most $2n/(\eta k^2)$ paths $P \in \mathcal{P}^\diamond$ have $\varphi_1(v_6^P) = i$. Then φ_1 gives a homomorphism from the paths in $\mathcal{P}^\diamond$ to $R^\star[[k]]$ which maps n/k^2 paths to each branch of each diamond $\diamondsuit \in \mathcal{D}$.

As in the previous case our final step is to set $\varphi_1(u) := \psi(u)$ for each vertex $u \in V(F_1)$ for which $\varphi_1(u)$ has not yet been defined (so u remains allocated as in the outcome of Algorithm 32). Since we previously set $\varphi_1(v_1^P) = \psi(v_1^P)$ and $\varphi_1(v_7^P) = \psi(v_7^P)$ for each $P \in \mathcal{P}^0 \cup \mathcal{P}^\diamond$, we then we then have that φ_1 is a homomorphism from F_1 to $R^\star$. Moreover, the only vertices $u \in V(F_1)$ which may have $\psi(u) \neq \varphi_1(u)$ are the pendant vertices $v_2^P, \dots, v_6^P$ of paths $P \in \mathcal{P}^0 \cup \mathcal{P}^\diamond$. Consequently, for each $i \in [k]$ the number of vertices $u \in V(F_1)$ with $\psi(u) = i$ and $\varphi_1(u) \neq i$ is at most $5 \max_{i\in[k]} |\mathcal{P}_i^0| + 5 \max_{i\in[k]} |\mathcal{P}_i^\diamond| \leq 5(\varepsilon n/k + 1) + 10n|\mathcal{D}|/k^3 \leq 6\varepsilon m$. Similarly, the number of vertices $v \in V(F_1)$ with $\varphi_1(v) = i$ and $\psi(v) \neq i$ is at most $6\varepsilon m/\alpha + 2\varepsilon m/\eta + (2n/k^2) \cdot (4/\eta) + 4n/(\eta k^2) \leq 3\varepsilon m/\eta$ since v is in at most $4/\eta$ diamonds in $\mathcal{D}$. We conclude that for each $i \in [k]$ we have

$$\big|\, |\varphi_1^{-1}(i)| - |\psi^{-1}(i)| \,\big| \leq 6\varepsilon m + \frac{3\varepsilon m}{\eta} \leq \frac{4\varepsilon m}{\eta}. \tag{19}$$

In both cases we have obtained a homomorphism φ_1 from F_1 to $R^\star$ with desired properties for the images of vertices of paths in $\mathcal{P}^0$ and $\mathcal{P}^\diamond$. Moreover, combining (16) with (18) or (19) (according to the case), in both cases for each $i \in [k]$ we have

$$\big|\varphi_1^{-1}(i)\big| = \frac{v(F_1)}{k} \pm \left(\frac{2n}{\log n} + \frac{4\varepsilon m}{\eta}\right) = \frac{v(F_1)}{k} \pm \frac{5\varepsilon m}{\eta}. \tag{20}$$

The second phase. In our second allocation phase, we use an auxiliary graph which is a weighted blow-up of J to build an allocation φ_2 of F_2 in J. This will compensate for the non-uniform usage of clusters, drastically reducing the differences between the numbers of vertices mapped to each cluster. Let $n_2 := v(F_2)$, and for each $i \in [k]$ set

$$\alpha_i := \frac{1}{n_2}\left(\frac{n - |V_0|}{k} - |\varphi_0^{-1}(i)| - |\varphi_1^{-1}(i)|\right) \qquad \text{and} \qquad b_i := \alpha_i \log\log\log n_2. \tag{21}$$

(We remark that the $\log\log\log n_2$ term in the definition of b_i is chosen simply to grow very slowly as a function of n_2; any other sufficiently slowly growing function of n_2 would work equally well.) Since $\sum_{i\in[k]} |\varphi_0^{-1}(i)| = |V_{\text{ground}}|$, $\sum_{i\in[k]} |\varphi_1^{-1}(i)| = v(F_1) - |V_0|$ and $n = v(F_1) + v(F_2) + |V_{\text{ground}}|$, we have

$$\sum_{i\in[k]} \alpha_i = 1. \tag{22}$$

By (9), (20), (21) and the facts that $n_2 \geq n/3$ and $|\varphi_0^{-1}(i)| \leq |V_{\text{ground}}|$, for each $i \in [k]$ we have

$$\begin{aligned}\alpha_i &\geq \frac{1}{n_2}\left(\frac{n}{k} - \frac{|V_0|}{k} - |V_{\text{ground}}| - \frac{v(F_1)}{k} - \frac{5\varepsilon m}{\eta}\right)\\ &= \frac{1}{n_2}\left(\frac{n_2}{k} - \frac{|V_0|}{k} - \frac{k-1}{k}|V_{\text{ground}}| - \frac{5\varepsilon m}{\eta}\right) \geq \frac{1}{k} - \frac{6\varepsilon n}{\eta n_2 k} \geq \frac{1}{k}\left(1 - \frac{20\varepsilon}{\eta}\right).\end{aligned} \tag{23}$$

Let B be a (blow-up) graph of $R^\star[[k]]$, obtained by replacing each $i \in [k]$ by a set B_i with precisely b_i vertices, with $x \to y \in E(B)$ if and only if $x \in B_i$, $y \in B_j$ and $i \to j \in E(R^\star)$. Note that $v(B) = \log\log\log n_2$ by (21) and (22). Also B contains a spanning d'_J-regular expander subdigraph

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

J^{blow} by Lemma 28, since

$$\begin{aligned}\delta^0(B) \geq \delta_0\big(R^\star[[k]]\big) \cdot \min_{i\in[k]} b_i &= \left(\frac{1}{2}+\eta\right) k \min_{i\in[k]} \alpha_i \nu(B)\\ &\overset{(23)}{\geq} \left(\frac{1}{2}+\eta\right) k \cdot \nu(B) \cdot \frac{1}{k}\left(1-\frac{20\varepsilon}{\eta}\right)\\ &\geq \left(\frac{1}{2}+\frac{\eta}{2}\right)\nu(B).\end{aligned}$$

We apply Algorithm 32 to F_2 and B to obtain a map $\psi_B{:}V(F_2)\to V(B)$. This application is simpler than the application of Algorithm 32 in Phase 1. Specifically, we apply Algorithm 32 with $\mathcal{E}=\emptyset$ and with $Z=\bigcup_{T\in\mathcal{T}_2}\{r^T,s^T\}$. Our choice of allocation for vertices of Z is essentially the same as in Phase 1: for each $T\in\mathcal{T}_2$ we set $\psi_B(r^T)$ to be an arbitrary vertex in $B_{\varphi_{\text{root}}(r^T)}$, and for each $T\in\mathcal{T}_2$ with a secondary attachment we choose $\psi_B(s^T)$ in both $\bigcup_{i\in I(T)} B_i$ and the appropriate neighbourhood of the image of p^T (recall that p^T is the parent of s^T in T, so $\psi_B(p^T)$ has already been defined when Algorithm 32 considers s^T). In particular, the latter condition ensures that the map ψ_B is a homomorphism from F_2 to B. Moreover, applying Lemma 34 with $S=V(F_2)$ we find that with high probability we have for each $x\in V(B)$ that

$$|\psi_B^{-1}(x)| = \frac{n_2}{\nu(B)} \pm \left(\frac{n_2}{\log n_2} + \big(|Z|+3n_2^{1/3}\big)\Delta(F_2)^{56\nu(B)^5\log\log n_2}\right) = \frac{n_2}{\nu(B)} \pm \frac{2n}{\log n}, \tag{24}$$

where we use the bounds $|Z|\leq 2n^{0.999}$ and $3n^{0.999}\Delta(F_2)^{56\nu(B)^5\log\log n_2}\leq n/\log n$.

We recover the desired allocation φ_2 of F_2 to $R^\star[[k]]$ by 'collapsing' the allocation of each blown-up vertex. More precisely, we define $\varphi_2: V(F_2)\to[k]$ by putting $\varphi_2(v)=i$ for each $v\in V(F_2)$ whenever $\psi_B(v)\in B_i$. So φ_2 is a homomorphism from F_2 to $R^\star[[k]]$ with $\varphi_2(r^T)=\varphi_{\text{root}}(r^T)$ for each $T\in\mathcal{T}_2$ and $\varphi_2(s^T)\in I(T)$ for each $T\in\mathcal{T}_2$ with a secondary attachment. From (24) it follows that for each $i\in[k]$ we have

$$|\varphi_2^{-1}(i)| = \sum_{x\in B_i}|\psi_B^{-1}(x)| = \frac{b_i n_2}{\nu(B)} \pm \frac{2nb_i}{\log n} = \alpha_i n_2 \pm \frac{2n\log\log n}{\log n}. \tag{25}$$

The conclusion. Let $\varphi_{\text{join}}: V(Q)\to R^\star$ be the map formed by combining φ_0, φ_1, and φ_2. More precisely, for each $x\in V(Q)$ we set $\varphi_{\text{join}}(x)=\varphi_0(x)$ if $x\in V_{\text{ground}}$, $\varphi_{\text{join}}(x)=\varphi_1(x)$ if $x\in V(F_1)$ and $\varphi_{\text{join}}(x)=\varphi_2(x)$ if $x\in V(F_2)$. For each $i\in[k]$ we then have

$$\begin{aligned}|\varphi_{\text{join}}^{-1}(i)| &= |\varphi_0^{-1}(i)| + |\varphi_1^{-1}(i)| + |\varphi_2^{-1}(i)|\\ &\overset{(25)}{=} |\varphi_0^{-1}(i)| + |\varphi_1^{-1}(i)| + \alpha_i n_2 \pm \frac{2n\log\log n}{\log n}\\ &\overset{(21)}{=} |\varphi_0^{-1}(i)| + |\varphi_1^{-1}(i)| + \frac{n-|V_0|}{k} - |\varphi_0^{-1}(i)| - |\varphi_1^{-1}(i)| \pm \frac{2n\log\log n}{\log n}\\ &= m \pm \frac{2n\log\log n}{\log n}.\end{aligned}$$

Finally, for each $i\in[k]$ set $\delta_i := m - |\varphi_{\text{join}}^{-1}(i)|$, so $\delta_i < n/k^3$ and $\sum_{i\in[k]}\delta_i = 0$. Recall that our choice of φ_1 ensured that for each $\Diamond\in\mathcal{D}$ at least n/k^2 paths $P\in\mathcal{P}^\diamond$ were mapped to each branch of $\mathcal{D}$. So we may apply Lemma 27. Writing $\mathcal{P}^\diamond$ also for the oriented graph which is the disjoint union of the paths in $\mathcal{P}^\diamond$, this yields a homomorphism φ from $\mathcal{P}^\diamond$ to $R^\star[[k]]$ with $|\varphi^{-1}(i)| = |\varphi_{\text{join}}^{-1}(i)| + \delta_i$ for each $i\in[k]$ and such that $\varphi(x)=\varphi_{\text{join}}(x)$ for all vertices of each $P\in\mathcal{P}^\diamond$ except for the central vertex v_4^P if P has order 7 or the leaf vertex v_2^P if P has order 2. Setting also $\varphi(x)=\varphi_{\text{join}}(x)$ for every other vertex x of Q we obtain a map $\varphi(x){:}V(Q)\to V(R^\star)$ extending φ_0 such that, for every $i\in[k]$,

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

we have

$$|\varphi^{-1}(i)| = |\varphi_{\text{join}}^{-1}(i)| + \delta_i = m,$$

so φ satisfies (iv).

To conclude the proof of the claim, let us argue that φ satisfies the remaining stated properties. Recall that φ_1 is a homomorphism from F_1 to $R^\star$ and φ_2 is a homomorphism from F_2 to $R^\star$; it follows that for each $T \in \mathcal{T}$ the restriction of φ_{join} to T is a homomorphism from T to $R^\star$, and so the same is true of φ. Next observe that for each vertex $u \in V(F)$, our applications of Algorithm 32 allocated all in-children of u to the same vertex of $R^\star$ and all out-children of u to the same vertex of $R^\star$, except in the cases where the edge between u and the child was an edge of $\mathcal{E}$ or if the child was a secondary attachment; each of these exceptions can occur for at most one child of u. Counting also the parent of u, we conclude that $\Delta(\psi), \Delta(\psi_B) \leq 5$; it follows that $\Delta(\varphi) \leq 5$, so we have (i).

Our choice of $\mathcal{P}^H$ immediately gives (vi). Moreover, each path $P \in \mathcal{P}^H$ was allocated canonically by ψ; since for each $P \in \mathcal{P}^H$ and $v \in V(P)$ we have $\varphi(v) = \varphi_{\text{join}}(v) = \varphi_1(v) = \psi(v)$, it follows that we have (v). Now recall that our choice of the sets $\mathcal{P}_i^0$ and the map φ_1 ensured that (ii) and (iii) held with φ_1 in place of φ. Since we have $\varphi(v) = \varphi_{\text{join}}(v) = \varphi_1(v)$ for all vertices v in paths in $\mathcal{P}^0$, we have (ii) and (iii). Finally, for (vii) recall for each $i \in \{1, 2\}$ that the choice of φ_i ensured that for every $T \in \mathcal{T}_i$ we had $\varphi_i(r^T) = \varphi_{\text{root}}(r^T)$, so $\varphi(r^T) = \varphi_{\text{root}}(r^T)$ also. Moreover, if T had a secondary attachment s^T, then we had $\varphi_i(s^T) \in I(T)$, and so $\varphi(s^T) \in I(T)$ also.

4.6. Proof of Claim 37

In this section we describe and analyse a greedy algorithm that extends ϱ to an embedding of $Q - M$ into G. This algorithm embeds each vertex of F in turn, in the order of $\prec$. Call a vertex $v \in V(F)$ *special* if v is an attachment of some $T \in \mathcal{T}$, if v is a neighbour of a distinguished vertex, or if $v \in \{v_1^P, v_7^P\}$ for some $P \in \mathcal{P}^H$. The algorithm treats these vertices of G specially, embedding them into fixed sets which are chosen beforehand for that purpose; another pre-chosen set will never be embedded to by the algorithm, and this will provide the necessary superregularity properties. Specifically, we reserve pairwise disjoint subsets A, B, Π_1, Π_7, and Π_{other} of $V(G) \setminus V_0$ obtained from the following claim.

Claim 38. *There exist pairwise disjoint subsets A, B, Π_1, Π_7, $\Pi_{\text{other}} \subseteq V(G) \setminus \big(V_0 \cup \varrho(V_{\text{ground}})\big)$ such that for all $i \in [k]$ and all $T \in \mathcal{T}$ the following statements hold.*

- *(i)* $|A \cap V_i| = |B \cap V_i| = |\Pi_{\text{other}} \cap V_i| = g/10$ *and* $|\Pi_1 \cap V_i| = |\Pi_7 \cap V_i| = g + 2\beta m$.
- *(ii) For all $x \in V_i$ and all $Y \in \{\Pi_1, \Pi_7, \Pi_{\text{other}}\}$ we have $\deg^-(x, Y \cap V_{i-1}),\ \deg^+(x, Y \cap V_{i+1}) \geq |Y|d/2k$.*
- *(iii) For all $x \in V_0$, all $\bullet \in \{-, +\}$ and all $j \in N^\bullet_{R^\star}(x)$, we have $\deg^\bullet_G(x, B \cap V_j) \geq |B|/2k$.*
- *(iv) For each attachment a of T, and $\bullet \in \{-, +\}$ if $u \in N^\bullet_Q(a)$ is the (unique) neighbour of a outside of T and $\circ$ is the opposite sign to $\bullet$, then*

$$\deg^\circ(\varrho(u), A \cap V_{\varphi(a)}) \geq |A|\eta/2k.$$

Proof. For each $i \in [k]$ let $V_i' = V_i \setminus \varrho(V_{\text{ground}})$, and choose a collection of pairwise disjoint subsets $A \cap V_i'$, $B \cap V_i'$, $\Pi_{\text{other}} \cap V_i'$, $\Pi_1 \cap V_i'$, $\Pi_7 \cap V_i'$ of V_i', with sizes dictated by (i), uniformly at random among all such collections and independently of the choices for each $i' \neq i$. In particular this means that each set is chosen uniformly at random among all subsets of V_i' of the specified size. Define each of A, B, Π_1, Π_7, and Π_{other} to be the union over all $i \in [k]$ of the corresponding intersections. So (i) holds by our choice of the sets; we now use concentration inequalities to show that with high probability we also have the properties (ii)–(iv).

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

Let $\mathcal{A} = \{A, B, \Pi_1, \Pi_7, \Pi_{\text{other}}\}$. Fix $i \in [k]$, and recall that $G[V_{i-1} \to V_i]$ and $G[V_i \to V_{i+1}]$ are both (d, ε)-superregular. Hence, for each $x \in V_{i-1}$, each $y \in V_{i+1}$ and all $X \in \mathcal{A}$, we have that $\deg^+(x, X \cap V_i')$ and $\deg^-(y, X \cap V_i')$ are random variables with hypergeometric distribution and with expectation at least $|X \cap V_i'|(d-\varepsilon) - |V_{\text{ground}}| \geq 2|X \cap V_i|d/3$. So, by Theorem 14, the probability that any one of these random variables has value strictly less than $|X \cap V_i|d/2$ decreases exponentially with n. By taking a union bound over all $i \in [k]$, all $x \in V_i$ and each $X \in \{\Pi_1, \Pi_7, \Pi_{\text{other}}\}$, it follows that with high probability all these random variables have value at least $|X \cap V_i|d/2 = |X|d/2k$ (for each choice of X), which implies that (ii) holds with high probability. Similar union bound arguments complete the proof of the claim, using Lemma 21 (d) for (iii), and using both (12) and (13) for (iv). □

In the case where the paths in $\mathcal{P}$ are bare paths of order 7, fix sets A, B, Π_1, Π_7, and Π_{other} as in Claim 38. We will embed attachments of trees in $\mathcal{T}$ into A, neighbours of distinguished vertices into B, vertices v_1^P of paths $P \in \mathcal{P}^H$ into Π_1 and vertices v_7^P of paths $P \in \mathcal{P}^H$ into Π_7, while Π_{other} will be used to ensure the superregularity properties of 37 (iv). In the other case, where the paths in $\mathcal{P}$ have order 2, we instead fix sets A, B, Π_1, and Π_{other} with the properties in Claim 38 but take Π_7 to be empty. We embed to these sets as in the previous case, except now there are no vertices v_7^P for $P \in \mathcal{P}^H$ to be embedded to Π_7. In both cases all other vertices of $Q - M$ will be embedded outside of these sets (recall that distinguished vertices have already been embedded to V_0). We also insist that each attachment of each tree in $\mathcal{T}$ is embedded in the appropriate neighbourhood of the image of its neighbour in V_{ground}, and each neighbour of a distinguished vertex is embedded within the appropriate neighbourhood of the image of the distinguished vertex. For an easy description of these rules for embedding, for each $v \in V(F)$ which is not a distinguished vertex we set

$$Z_v = \begin{cases} V_{\varphi(v)} \cap A \cap N_G^\bullet\big(\varrho(\hat{r}^T)\big) & \text{if } v = r^T \text{ for some } T \in \mathcal{T} \text{ and } r^T \in N_Q^\bullet(\hat{r}^T), \\ V_{\varphi(v)} \cap A \cap N_G^\bullet\big(\varrho(\hat{s}^T)\big) & \text{if } v = s^T \text{ for some } T \in \mathcal{T} \text{ and } s^T \in N_Q^\bullet(\hat{r}^T), \\ V_{\varphi(v)} \cap B \cap N_G^\bullet\big(\varrho(z)\big) & \text{if } v \text{ is a neighbour of a distinguished vertex } z \text{ and } v \in N_Q^\bullet(z), \\ V_{\varphi(v)} \cap \Pi_1 & \text{if } v = v_1^P \text{ for some } P \in \mathcal{P}^H, \\ V_{\varphi(v)} \cap \Pi_7 & \text{if } v = v_7^P \text{ for some } P \in \mathcal{P}^H, \text{ and} \\ V_{\varphi(v)} \setminus (A \cup B \cup \Pi_1 \cup \Pi_7 \cup \Pi_{\text{other}}) & \text{if } v \text{ is not special,} \end{cases}$$

and we will embed v within Z_v unless $v \in M$ (in which case we do not embed v at all). We do this by the following greedy algorithm.

► **Setup.** Let Ξ be the initial image of ϱ (that is, Ξ contains the images of vertices in V_{ground} and of distinguished vertices). Let $t_1 \prec t_2 \prec \ldots \prec t_{v(F)}$ be the vertices of F, ordered as in $\prec$. Beginning at time $\tau = 1$, we take the following steps.

► **1) Update available vertices.** Say that a vertex $u \in V(F)$ is *open* if $u \prec t_\tau$ but u has a child v with $t_\tau \prec v$. Let

$$\text{open}(\tau) := \{u \in V(F) : u \text{ is open}\},$$

and

$$\text{used}(\tau) := \big(\Xi \cup \{\varrho(t_\sigma) : \sigma \prec \tau, t_\sigma \notin M\}\big)$$

so $\text{used}(\tau)$ is the set of vertices of G which have already been occupied by the image of a vertex of Q. Let

$$\text{reserved}(\tau) = \bigcup_{u \in \text{open}(\tau)} \bigcup_{v \in C(u)} R(v)$$

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

be the set of vertices which are currently reserved for children of open vertices, and for each $w \in V(F)$ with $t_\tau \prec w$ let

$$Z_{w,\tau} = Z_w \setminus \big(\text{used}(\tau) \cup \text{reserved}(\tau)\big).$$

For vertices $w \in V(F)$ for which a reserved set $R(w)$ has not yet been selected, $Z_{w,\tau}$ is the set which w is presently permitted to reserve vertices from (we make no use of the sets $Z_{w,\tau}$ for those vertices w for which $R(w)$ has already been selected). Finally, for each $w \in \{t_\tau\} \cup C(t_\tau)$ set

$$\mathcal{U}^+(w) = \big\{Z_{x,\tau} : x \in C^+(w)\big\} \text{ and } \mathcal{U}^-(w) = \big\{Z_{x,\tau} : x \in C^-(w)\big\}$$

and let $S_w \subseteq F$ be the oriented star with centre w whose leaves are the children of w. Say that a vertex $v_\tau \in V_{\varphi(t_\tau)}$ is t_τ*-good* if for each $\circ \in \{-,+\}$ and each $w \in C^\circ(t_\tau)$ we have $\deg^\circ(v_\tau, Z_{w,\tau}) \geq \gamma m$.

► **2) Embed v_τ.** If t_τ is a distinguished vertex or a vertex of M, then do nothing at this step. Otherwise, embed t_τ as follows.

- If $t_\tau = r^T$ for some $T \in \mathcal{T}$, then choose a t_τ-good vertex $v_\tau \in Z_{t_\tau,w}$ and set $\varrho(t_\tau) = v_\tau$.
- Otherwise, choose a t_τ-good vertex $v_\tau \in R(v_\tau)$ and set $\varrho(t_\tau) = v_\tau$.

► **3) Declare reserved sets for children of t_τ.** For each $\bullet \in \{-,+\}$ and each child $w \in C^\bullet(t_\tau)$ which is not a distinguished vertex, choose a set $R(w) \subseteq Z_{w,\tau}$ with $|R(w)| = \sqrt{m}$ which is $\big(\mathcal{U}^+(w), \mathcal{U}^-(w), \beta, \gamma, \varphi, m\big)$-good for S_w, and make these choices so that the sets $R(w)$ for $w \in C(t_\tau)$ are pairwise disjoint. Moreover, if t_τ is neither a distinguished vertex nor an element of M then we insist additionally that $R(w) \subseteq N_G^\bullet(v_\tau)$.

► **4) Loop.** If $\tau = v(F)$ then terminate; otherwise increment τ and return to Step 1.

Proof that the algorithm runs successfully. Observe that the Setup and Steps 1 and 4 consist solely of definitions. We shall prove that (I) it is possible to make the choices required by the embedding algorithm at Steps 2 and 3, and (II) that if the choices of the algorithm can be made, then it produces an embedding ϱ of $Q - M$ in G with the properties stated in the claim.

We first prove (II). At the start of the algorithm, the images of V_{ground} and of the distinguished vertices are already fixed under ϱ, which is an embedding of the subgraph of Q induced by these vertices into G (in particular, ϱ bijectively maps the distinguished vertices to V_0). So to show that the outcome of the algorithm is a embedding ϱ from $Q - M$ to G, it suffices to show that ϱ is injective and also that for each vertex $t_\tau \in F$ we have

(a) $\varrho(t_\tau) \in N_G^+(\varrho(u))$ for each $u \in N_Q^-(t_\tau)$ which was already embedded at time τ, and

(b) $\varrho(t_\tau) \in N_G^-(\varrho(u))$ for each $u \in N_Q^+(t_\tau)$ which was already embedded at time τ.

To do this, observe first that since the algorithm considered the vertices of Q in an ancestral order, for each $t_\tau \in V(F)$ with a parent t_σ in F the parent t_σ had already been embedded at time τ. Furthermore, the image $\varrho(t_\tau)$ of t_τ was chosen within the set $R(t_\tau)$ of vertices reserved for the embedding of t_τ, which in turn was previously chosen to be a subset of the appropriate neighbourhood of t_σ at time σ. Observe also that the set $R(t_\tau)$ was chosen to avoid used(σ), and no vertices we embedded into $R(t_\tau)$ between times σ and ρ, so no vertex was embedded to $\varrho(t_\tau)$ before time τ. Moreover, for the $\bullet \in \{-,+\}$ with $t_\tau \in N_Q^\bullet(t_\sigma)$ we have $\varrho(t_\tau) \in N_G^\bullet(\varrho(t_\sigma))$, as desired. It remains to demonstrate the desired property in the case where a neighbour of t_τ in Q other than a parent of t_τ was already embedded at time τ. This can occur in the following three ways.

(1) $t_\tau = r^T$ for some $T \in \mathcal{T}$. In this case $\hat{r}^T$ is the only neighbour of t_τ which has previously been embedded, and the definition of Z_{t_τ} ensures that t_τ is embedded in the appropriate neighbourhood of $\varrho(\hat{r}^T)$. Moreover, the embedding of t_τ was chosen to avoid used(τ), so

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

no vertex was embedded to $\varrho(t_\tau)$ before time τ. Together with the previous observation that this holds for vertices t_τ with a parent in F, this shows that ϱ is injective.

(2) $t_\tau = s^T$ for some $T \in \mathcal{T}$. In this case $\hat{s}^T$ and the parent p^T of t_τ in T are the only neighbours of t_τ which have previously been embedded. The definition of Z_{t_τ} ensures that t_τ is embedded in the appropriate neighbourhood of $\hat{\varrho}(s^T)$.

(3) t_τ is the parent of a distinguished vertex v. In this case the parent of v_τ and v are the only neighbours of v_τ which have previously been embedded. Again the definition of Z_{t_τ} ensures that t_τ is embedded in the appropriate neighbourhood of $\varrho(v)$.

So the output of the algorithm is indeed an embedding ϱ of $Q - M$ in G. Also our choice of the sets Z_v ensures that we have (i). Now suppose that the paths in $\mathcal{P}$ have order two. In this case Claim 36 (i), (ii), and (iii) together imply that for each $i \in [k]$ precisely m vertices of Q, precisely g vertices of M, and precisely g vertices v_1^P of paths $P \in \mathcal{P}^H$ are allocated to V_i, and then (ii) follows from (i). Similarly, if the paths in $\mathcal{P}$ have order seven then Claim 36 (i), (ii), and (iii) together imply that for each $i \in [k]$ precisely m vertices of Q, precisely $5g$ vertices of M, precisely g vertices v_1^P of paths $P \in \mathcal{P}^H$ and precisely g vertices v_7^P of paths $P \in \mathcal{P}^H$ are allocated to V_i, so we have (iii).

For (iv), suppose first that the paths in $\mathcal{P}^H$ have order 2. This means that $\Pi_7, W_1, \dots, W_k$ are all empty. Since the embedding respects the allocation, for each $i \in [k]$ precisely g vertices are embedded to $\Pi_1 \cap V_i$, so $|U_i| = g$. Moreover, among the m vertices $u \in V(Q)$ with $\varphi(u) = i$, the vertices which have not been embedded in V_i are precisely the g vertices v_2^P for $P \in \mathcal{P}^H$ with $\varphi(v) = i$, so $|V_i^\star| = g$ as well. Since $g = \lceil \frac{\lambda m}{2^{10}} \rceil$, it follows that $G[V_{i-1}^\star \to U_i]$, $G[U_i \to V_{i+1}^\star]$ and $G[V_{i-1}^\star \to V_i^\star]$ are each (d, ε')-regular by Lemma 16. To check the degree condition required for superregularity in (iv), note that by Claim 38 (ii) and the fact that precisely g vertices v_1^P of paths $P \in \mathcal{P}^H$ are embedded to $\Pi_1 \cap V_i$, for each $i \in [k]$ and each $x \in V_{i+1}$ we have

$$\deg^-(x, U_i) \geq \frac{|\Pi_1| d}{2k} - \big(|\Pi_1 \cap V_i| - g\big) \geq \frac{kgd}{2k} - 2\beta m \geq \beta g = \beta |U_i|,$$

and similarly for each $i \in [k]$ and each $x \in V_{i-1}$ we have $\deg^+(x, U_i) \geq \beta |U_i|$. In the case where the paths in $\mathcal{P}^H$ have order 7, we obtain identical bounds with W_i and Π^7 in place of U_i and Π^1 respectively, as in this case for each $i \in [k]$ precisely g vertices v_7^P of paths $P \in \mathcal{P}^H$ are embedded to $\Pi_1 \cap V_i$. Moreover, since no vertex is ever embedded to Π_{other}, we have $\Pi_{\text{other}} \cap V_i \subseteq V_i^\star$ for each $i \in [k]$. This means that by Claim 38 (ii) for each $i \in [k]$ and each $x \in V_i$ we have

$$\deg^-(x, V_{i-1}^\star) \geq \deg^-(x, \Pi_{\text{other}} \cap V_{i-1}) \geq \frac{kg}{10} \cdot \frac{d}{2k} \geq \beta g = \beta |V_{i-1}^\star|,$$

and similarly $\deg^+(x, V_{i+1}^\star) \geq \beta |V_{i+1}^\star|$. It follows that each of the graphs in (iv) is (β, ε')-superregular, concluding the proof of (II).

We now turn to proving (I). The following invariant plays a crucial role in our analysis: for each $w \in V(F)$ which is not distinguished and each time τ in the execution of the algorithm we have

$$|Z_{w,\tau}| = \big| Z_w \setminus \big(\text{used}(\tau) \cup \text{reserved}(\tau)\big) \big| \geq \beta m. \tag{26}$$

To verify (26), note that since $\prec$ is a tidy ancestral order of F, for each τ we have $|\text{open}(\tau)| < \log_2 n$ and so $|\text{reserved}(\tau)| \leq (\log_2 n)\Delta(Q)\sqrt{m} \leq n^{2/3}$. Moreover, Claim 38 and the properties of φ guaranteed by Claim 36 imply that for each $i \in [k]$ we have

$$|\Pi_1 \cap \varphi(V(Q) \setminus M) \cap V_i| = g = |\Pi_7 \cap \varphi(V(Q) \setminus M) \cap V_i| = g,$$

except in the case where paths in $\mathcal{P}$ have order two, in which case we do not have the latter equality because Π_7 is empty. For each $i \in [k]$ we also have

$$\big|A \cap \varphi\big(V(Q) \setminus M\big) \cap V_i\big| 2|\mathcal{T}| \overset{(10)}{\leq} 2n^{0.999} \leq \beta m,$$

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

and, by Claim 36 (iii).

$$\left|B \cap \varphi\big(V(Q) \setminus M\big) \cap V_i\right| \leq \frac{6\varepsilon m}{\alpha} \leq \beta m,$$

Finally,

$$\left|\big(V_i \setminus (A \cup B \cup \Pi_1 \cup \Pi_7 \cup \Pi_{\text{other}})\big) \cap \varphi\big(V(Q) \setminus M\big)\right| \leq \left| V_i \cap \varphi\Big(V(Q) \setminus \bigcup_{P \in \mathcal{P}^H} P \Big) \right|,$$

and we can bound this quantity by $|V_i| - 2g$ if the paths in $\mathcal{P}$ each have order two and by $|V_i| - 7g$ if the paths in $\mathcal{P}$ each have order seven. Together with the sizes of the sets A, B, Π_1, Π_7, and Π_{other} and bounds on degrees given in Claim 38 (iii) and (iv) (recalling also that Π_7 is empty if the paths in $\mathcal{P}$ have order two), these bounds imply (26). In particular, it follows that when we define the families $\mathcal{U}^+(w)$ and $\mathcal{U}^-(w)$, each set within these families contains at least βm vertices.

In Step 2 we wish to choose a t_τ-good vertex v_τ. If $t_\tau = r^T$ for some $T \in \mathcal{T}$, then we need to do this with $v_\tau \in Z_{t_\tau,w}$. Since $|Z_{t_\tau,\tau}| \geq \beta m$ and $|Z_{w,\tau}| \geq \beta m$ for each $w \in C(t_\tau)$ by (26), and $\Delta(\varphi^T) \leq 5$ by Claim 36, we may apply Lemma 23 to obtain a $\big(\mathcal{U}^+(v_\tau), \mathcal{U}^-(v_\tau), \beta, \gamma, \varphi, m\big)$-good set for S_{t_τ} in Z_{t_τ}. By definition each vertex in this set is t_τ-good, so we can choose v_τ as required. On the other hand, if t_τ is not the root of a tree in $\mathcal{T}$, then t_τ has a parent t_σ in F, and we wish to choose v_τ in $R(t_\tau)$, the set previously reserved at time σ for the future embedding of t_τ. The set $R(t_\tau)$ was chosen at time σ to be a subset of $\subseteq Z_{t_\tau,\sigma}$ which was $\big(\mathcal{U}^+(v_\tau), \mathcal{U}^-(v_\tau), \beta, \gamma, \varphi, m\big)$-good for S_{t_τ}. Since no vertex was embedded in or reserved from $R(t_\tau)$ between times σ and τ we have $R(t_\tau) \subseteq Z_{t_\tau,\sigma}$, so by definition of $\big(\mathcal{U}^+(v_\tau), \mathcal{U}^-(v_\tau), \beta, \gamma, \varphi, m\big)$-good and (26) we can choose $v_\tau \in R(t_\tau)$ as required.

Finally, in Step 3 we wish to reserve a set $R(w)$ for each child $w \in C(t_\tau)$. If t_τ is neither a distinguished vertex nor a vertex of M, then we have just selected a t_τ-good vertex v_τ for the image of t_τ, and we require that $R(w) \subseteq N^\bullet_G(v_\tau) \cap Z_{w,\tau}$, where $\bullet \in \{+,-\}$ is such that $w \in C^\bullet(t_\tau)$. The t_τ-goodness of v_τ ensures that $|N^\bullet_G(v_\tau) \cap Z_{w,\tau}| \geq \gamma m$ for each $\bullet \in \{+,-\}$ and $w \in C^\bullet(t_\tau)$. Along with (26) and the fact that $\Delta(\varphi^T) \leq 5$ this enables us to apply Lemma 23 to obtain the desired $R(w)$ for each $w \in C(t_\tau)$ (since we only need $|N^\bullet_G(v_\tau) \cap Z_{w,\tau}| \geq \gamma m/2$ to apply Lemma 23, we may do this so that the sets $R(w)$ are also pairwise disjoint, as required). If instead t_τ is a distinguished vertex or a vertex of M, then we instead require just that $R(w) \subseteq Z_{w,\tau}$. By (26) we have $|Z_{w,\tau}| \geq \beta m$ for each $w \in C(t_\tau)$, and so again we may apply Lemma 23 to obtain the desired $R(w)$ for each $w \in C(t_\tau)$. This concludes the proof of Claim 37.

Remark 39. The embedding algorithm can be significantly simplified if the goal is to embed an almost spanning structure (say, with at most $(1-\alpha)n$ vertices). In particular the sets B, Π_1, Π_7, and Π_{other} would no longer be needed, and we can completely avoid embedding vertices to V_0.

Acknowledgements

We thank the anonymous reviewers for their thoughtful comments, and for suggesting a shift of the focus of the paper toward the more general result. We believe that our consequent restructuring of the content of this manuscript has improved the quality of the paper, with clearer and more concise arguments.

References

[1] Alon, N., Duke, R. A., Lefmann, H., Rödl, V. and Yuster, R. (1994) The algorithmic aspects of the regularity lemma. *J. Algorithm.* **16**(1) 80–109.

[2] Balogh, J., Csaba, B. and Samotij, W. (2011) Local resilience of almost spanning trees in random graphs. *Random Struct. Algor.* **38**(1–2) 121–139.

[3] Bondy, J. A. and Murty, U. S. R. (2008) *Graph Theory*, Vol. 244 of Graduate Texts in Mathematics. Springer-Verlag, New York.

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press

[4] Böttcher, J., Han, J., Kohayakawa, Y., Montgomery, R., Parczyk, O. and Person, Y. (2019) Universality for bounded degree spanning trees in randomly perturbed graphs. *Random Struct. Algor.* **55**(4) 854–864.
[5] Böttcher, J., Schacht, M. and Taraz, A. (2009) Proof of the bandwidth conjecture of Bollobás and Komlós. *Math. Ann.* **343**(1) 175–205.
[6] Clemens, D., Ferber, A., Glebov, R., Hefetz, D. and Liebenau, A. (2015) Building spanning trees quickly in maker-breaker games. *SIAM J. Discrete Math.* **29**(3) 1683–1705.
[7] DeBiasio, L., Kühn, D., Molla, T., Osthus, T. and Taylor, A. (2015) Arbitrary orientations of hamilton cycles in digraphs. *SIAM J. Discrete Math.* **29**(3) 1553–1584.
[8] DeBiasio, L. and Molla, T. (2015) Semi-degree threshold for anti-directed Hamiltonian cycles. *Electron. J. Comb.* **22**(4) P4.34. https://www.combinatorics.org/ojs/index.php/eljc/article/view/v22i4p34
[9] Diestel, R. (1997) *Graph Theory*, Vol. 173 of Graduate Texts in Mathematics. Springer-Verlag, New York. Translated from the 1996 German original.
[10] Ghouila-Houri, M. A. (1960) Une condition suffisante d'existence d'un circuit Hamiltonien. *Comptes Rendus de l'Académie des Sciences* **25** 495–497.
[11] Janson, S., Łuczak, T. and Ruciński, A. (2000) *Random Graphs*. Wiley-Interscience, New York.
[12] Kathapurkar, A. and Montgomery, R. (2022) Spanning trees in dense directed graphs. *J. Comb. Theory, Ser. B.* **156** 223–249.
[13] Komlós, J., Sárközy, G. N. and Szemerédi, E. (1995) Proof of a packing conjecture of Bollobás. *Comb. Prob. Comp.* **4**(3) 241–255.
[14] Komlós, J., Sárközy, G. N. and Szemerédi, E. (2001) Spanning trees in dense graphs. *Comb. Prob. Comp.* **10** 397–416.
[15] Krivelevich, M., Kwan, M. and Sudakov, B. (2017) Bounded-degree spanning trees in randomly perturbed graphs. *SIAM J. Discrete Math.* **31**(1) 155–171.
[16] Kühn, D., Mycroft, R. and Osthus, D. (2011) An approximate version of Sumner's universal tournament conjecture. *J. Combin.l Theory, Ser. B.* **101**(6) 415–447.
[17] McDiarmid, C. (1998) Concentration. In *Probabilistic Methods for Algorithmic Discrete Mathematics*, Vol. 16 of Algorithms Combin., Berlin: Springer, pp. 195–248.
[18] Moon, J. W. (1970) Counting labelled trees. Number 1 in Canadian mathematical monographs. *Canadian Mathematical Congress*. **1** 113.
[19] Mycroft, R. and Naia, T. (2018) Unavoidable trees in tournaments. *Random Struct. Algor.* **53**(2) 352–385.
[20] Sudakov, B. and Vondrák, J. (2010) A randomized embedding algorithm for trees. *Combinatorica* **30**(4) 445–470.
[21] Szemerédi, E. (1975) On sets of integers containing no k elements in arithmetic progression. *Acta Arith.* **27** 199–245, Collection of articles in memory of Juriĭ Vladimirovič Linnik.
[22] Szemerédi, E. (1978) Regular partitions of graphs. In *Problèmes combinatoires et théorie des graphes* (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976), Paris: CNRS, pp. 399–401.

Cite this article: Mycroft R and Naia T (2026). Trees and treelike structures in dense digraphs. *Combinatorics, Probability and Computing* **35**, 486–523. https://doi.org/10.1017/S0963548326100418

https://doi.org/10.1017/S0963548326100418 Published online by Cambridge University Press