Ramsey numbers of degenerate graphs

By Choongbum Lee

Abstract

A graph is d-degenerate if all its subgraphs have a vertex of degree at most d. We prove that there exists a constant c such that for all natural numbers d and r, every d-degenerate graph H of chromatic number r with $|V(H)| \geq 2^{d^2 2^{cr}}$ has Ramsey number at most $2^{d 2^{cr}} |V(H)|$. This solves a conjecture of Burr and Erdős from 1973.

1. Introduction

Ramsey theory studies problems that can be grouped under the common theme that 'every large system contains a highly organized subsystem.' A classical example is the celebrated van der Waerden theorem [35] asserting that in every coloring of the natural numbers with a finite number of colors, one can find monochromatic arithmetic progressions of arbitrary finite length. This has motivated further results such as the Hales-Jewett theorem [21] and Szemerédi's theorem [34] and has had a tremendous influence on Combinatorics and related fields. See [20] for a comprehensive overview of Ramsey theory.

For a graph H, the *Ramsey number* of H, denoted $r(H)$, is defined as the minimum integer n such that in every edge two-coloring of K_n, the complete graph on n vertices, there exists a monochromatic copy of H. The name of the field has its origins in a 1930 paper of Frank P. Ramsey [30], who proved that $r(K_t)$ are finite for all natural numbers t. In 1935, Erdős and Szekeres [14] rediscovered Ramsey's theorem and brought it to more widespread attention by finding interesting geometric applications of its hypergraph extensions. Plenty of variants and applications have been found since then, and now it is considered to be one of the most important results in combinatorics, lying at the point of intersection of many fields.

There are many fascinating problems concerning bounds on Ramsey numbers of various graphs. Erdős and Szekeres, in the paper mentioned above, established a recurrence relation on the Ramsey numbers of complete graphs that implies $r(K_t) \leq \binom{2t-2}{t-1} = 2^{(2+o(1))t}$ for all natural numbers t. Later, in

1947, Erdős [13], in one of the earliest applications of the probabilistic method, proved $r(K_t) \geq 2^{(1/2+o(1))t}$. These two bounds together show that $r(K_t)$ is exponential in terms of the number of vertices t. There have been some interesting improvements on these bounds [8], [32], but despite a great amount of effort, the constants in the exponents remain unchanged. See the recent survey paper of Conlon, Fox, and Sudakov [10] for further information on graph Ramsey theory.

In 1973, Burr and Erdős [4] initiated the study of Ramsey numbers of sparse graphs and conjectured that the behavior of Ramsey numbers of sparse graphs should be dramatically different from that of complete graphs. A graph is *d-degenerate* if all its subgraphs contain a vertex of degree at most d. Degeneracy is a natural measure of sparseness of graphs as it implies that for all subsets of vertices X, there are fewer than $d|X|$ edges with both endpoints in X. Burr and Erdős conjectured that for every natural number d, there exists a constant $c = c(d)$ such that every d-degenerate graph H on n vertices satisfies $r(H) \leq cn$. This is in striking contrast with the case of complete graphs where the dependence on the number of vertices is exponential.

This conjecture has received much attention and motivated several important developments over the past 40 years. For example, the work of Chvátal, Rödl, Szemerédi, and Trotter [6] from 1983 that we will discuss below is one of the earliest applications of the regularity lemma, and it fostered further developments such as the blow-up lemma of Komlós, Sarközy, and Szemerédi [23]. Furthermore, progress on the conjecture is intertwined with the development of a powerful new tool in probabilistic combinatorics now known as dependent random choice. Kostochka and Rödl [24] used its primitive version to study a special case of the conjecture, and later, using a different method [25], established the first polynomial bound $r(H) \leq c_d n^2$ for all d-degenerate graphs H on n vertices (where c_d is a constant depending on d). The general framework of applying the dependent random choice technique to embed degenerate graphs was pioneered by Kostochka and Sudakov [26], who improved the bound of Kostochka and Rödl to a nearly linear bound $r(H) \leq 2^{c_d (\log n)^{2d/(2d+1)}} n$. Later, Fox and Sudakov [15], [16] refined the method to prove that $r(H) \leq 2^{c_d \sqrt{\log n}} n$. See the survey paper of Fox and Sudakov [17] for an overview of the history and applications of this fascinating method.

Linear bounds were established for special cases such as subdivisions of graphs by Alon [1], random graphs by Fox and Sudakov [16], and graphs with small bandwidth by the author [27]. The conjecture has also been examined for weaker notions of sparseness. Chvátal, Rödl, Szemerédi, and Trotter [6] proved that if H is a graph on n vertices of maximum degree at most Δ, then $r(H) \leq c(\Delta)n$, where $c(\Delta)$ is a constant depending only on Δ. Their proof used the regularity lemma, and thus the dependence of $c(\Delta)$ on Δ was of

tower type. This bound has been improved since then, by Eaton [12], Graham, Rödl, and Ruciński [18], [19], and then by Conlon, Fox, and Sudakov [9] to $r(H) \leq c^{\Delta \log \Delta} n$. For bipartite graphs, Conlon [7] and independently Fox and Sudakov [15] proved that a stronger bound $r(H) \leq c^{\Delta} n$ holds. These results are close to best possible since Graham, Rödl, and Ruciński [19] constructed bipartite graphs H on n vertices of maximum degree Δ satisfying $r(H) \geq c^{\Delta} n$ (for a different constant c).

Chen and Schelp [5] considered another measure of sparseness. They defined a graph to be *p-arrangeable* if there is an ordering $v_1, \ldots, v_n$ of the vertices such that for any vertex v_i, its neighbors to the right of v_i have together at most p neighbors to the left of v_i (including v_i). This is a measure of sparseness that lies strictly between degeneracy and bounded maximum degree. They showed that graphs with bounded arrangeability have Ramsey number linear in the number of vertices, implying, in particular, that the Burr-Erdős conjecture holds for planar graphs. Furthermore, Rödl and Thomas [31] showed that graphs with no K_p-subdivision have arrangeability less than p^8 and therefore have Ramsey number linear in the number of vertices.

In this paper, we build upon these developments and settle the conjecture of Burr and Erdős. We say that a graph G is *universal* for a family $\mathcal{F}$ of graphs if it contains all graphs $F \in \mathcal{F}$ as subgraphs. For an edge coloring of a graph, we say that a color is *universal* for a family $\mathcal{F}$ if the subgraph induced by the edges of that color is universal for $\mathcal{F}$.

THEOREM 1.1. *There exists a constant c such that the following holds for every natural number d, r, and n satisfying $n \geq 2^{d^2 2^{cr}}$. For every edge two-coloring of the complete graph on at least $2^{d 2^{cr}} n$ vertices, one of the colors is universal for the family of d-degenerate r-colorable graphs on at most n vertices.*

This settles the conjecture of Burr and Erdős since all d-degenerate graphs have chromatic number at most $d + 1$. Moreover, for fixed values of r, Theorem 1.1 is best possible up to the constant in the exponent. To see this, consider a random graph G on $(1 - \varepsilon) 2^d n$ vertices of density $\frac{1}{2}$ and let H be the complete bipartite graph $K_{d,n-d}$ with d vertices in one part and $n - d$ vertices in the other part. It is well known that in G, with high probability, every d-tuple of (distinct) vertices has fewer than $(1 - \frac{\varepsilon}{2})n$ common neighbors. Therefore, G does not contain a copy of H. Since the complement of G can also be seen as a random graph of density $\frac{1}{2}$, we see that the complement of G also does not contain a copy of H. Another way to see the tightness of Theorem 1.1 is by considering the construction of Graham, Rödl, and Ruciński [19] mentioned above.

The *density* of a graph is defined as the fraction of pairs of distinct vertices that form an edge. Most of the previous results mentioned above in fact provide

a density-embedding result for bipartite graphs, saying that every dense enough graph contains a copy. Note that the Ramsey number result follows from such a density-embedding result since in every edge two-coloring of a complete graph, one of the colors must have density at least $\frac{1}{2}$. In this context, the following theorem generalizes Theorem 1.1 to a density-embedding result.

THEOREM 1.2. *There exists a constant c such that the following holds for every natural number d and real number $\alpha \leq \frac{1}{2}$. For every natural number $n \geq \alpha^{-cd^2}$, if G is a graph with at least $\alpha^{-cd}n$ vertices and density at least α, then it is universal for the family of d-degenerate bipartite graphs on n vertices.*

Note that the complete bipartite graph $K_{d,n-d}$ mentioned above has a specific structure. Namely, every vertex on one side has bounded degree. Until now, even this special case, bipartite graphs with one side of bounded degree, of the conjecture was open. A result of Alon [1] implies the case when the degrees are bounded by two. The corresponding density-embedding result was proved by Fox and Sudakov [15], who suggested the general case as an interesting problem to examine. The theorem below addresses this special case of the conjecture and shows that we can improve Theorem 1.2 to get nearly best possible embedding results.

THEOREM 1.3. *Let d be a natural number and α, ε be positive real numbers satisfying $\alpha^{d(d-2)} \leq \varepsilon < 1$. Let G be a graph on $(1+\varepsilon)\alpha^{-d}n$ vertices of density at least α. Then G is universal for the family of bipartite graphs H on n vertices with a vertex partition $W_1 \cup W_2$ where all vertices in W_1 have at most d neighbors in W_2, and $\frac{|W_2|^d}{|W_2|(|W_2|-1)\cdots(|W_2|-d+1)} \leq 1 + \varepsilon$.*

This theorem strengthens previous density-embedding results for bipartite graphs with bounded maximum degree [7], [15] and is close to being best possible as can be seen by the example where G is a random graph and H is the complete bipartite graph $K_{d,n-d}$ discussed above. Let Q_n be the *hypercube*, which is the graph with vertex set $\{0,1\}^n$ where two vertices are adjacent if and only if they differ in exactly one coordinate. Theorem 1.3 with $\varepsilon = \frac{n^2}{2^n}$ and $\alpha = \frac{1}{2}$ shows that $r(Q_n) \leq 2^{2n} + n^2 2^n$ holds for all sufficiently large n. This bound improves by a constant factor, the current best known bound $r(Q_n) \leq 2^{2n+6}$ of Conlon, Fox, and Sudakov [11], proved using a different approach based on the local lemma. It is conjectured [4] that there exists a constant c such that $r(Q_n) \leq c2^n$ for all n.

The proofs of the three theorems above are based on dependent random choice and build upon several ideas developed for earlier applications. While the proofs of Theorems 1.1 and 1.2 are technically involved, Theorem 1.3 has a short proof that highlights one of the main differences in our usage of this technique as compared to the previous ones. We thus start by presenting the

proof of Theorem 1.3 in Section 2. In Section 3, we introduce some central concepts and notation and give a brief outline of the proofs of our main theorems. In Section 4, we develop the embedding strategy that will be used. The proofs of Theorems 1.1 and 1.2 will be given in Section 5 except for one key lemma, which will be proved in Section 6. We conclude with some remarks in Section 7.

Notation. For an integer m, define $[m] := \{1, 2, \ldots, m\}$, and for a pair of integers m_1, m_2, define $[m_1, m_2] := \{m_1, \ldots, m_2\}$, $[m_1, m_2) := \{m_1, \ldots, m_2-1\}$, $(m_1, m_2] := \{m_1+1, \ldots, m_2\}$, and $(m_1, m_2) := \{m_1+1, \ldots, m_2-1\}$. For a set of elements X, we define $X^t = X \times \cdots \times X$ as the set of all t-tuples in X. Let Q be a d-tuple and Q' be a d'-tuple of elements in some set. We use $Q \cup Q'$ to denote the $(d+d')$-tuple obtained by concatenating Q' to the end of Q.

Let $G = (V, E)$ be an n-vertex graph. For a vertex x and a set T, we define $\deg(x; T)$ as the number of neighbors of x in T and define $\deg(x) := \deg(x; V)$. For a set or an ordered tuple of vertices Q, define

$$N(Q; T) := \{x \in T : \forall y \in Q, \{x, y\} \in E\}$$

as the set of common neighbors in T of vertices in Q. Define $N(Q) := N(Q; V)$. For two sets X and Y, define $E(X, Y) = \{(x, y) \in X \times Y : \{x, y\} \in E\}$ and $e(X, Y) = |E(X, Y)|$. Furthermore, define $e(X) = \frac{1}{2}e(X, X)$ as the number of edges in X. Let $d(X, Y) = \frac{e(X,Y)}{|X||Y|}$ be the density of edges between X and Y. For a graph H, an *embedding* of H in G is an injective map $f : V(H) \to V(G)$ for which $\{f(v), f(w)\} \in E(G)$ whenever $\{v, w\} \in E(H)$. A *partial embedding of H in G defined on* $V' \subseteq V(H)$ is an embedding of $H[V']$ into G. For $V'' \subseteq V(H)$, an *extension to V''* of a partial embedding f on V' is an embedding $g : H[V' \cup V''] \to G$ such that $g|_{V'} = f$. We often abuse notation and denote the extended map using the same notation f.

Throughout the paper, we will be using subscripts such as in $x_{1.1}$ to indicate that x is the constant coming from Theorem/Lemma/Proposition 1.1.

2. One-side bounded bipartite graphs

Fox and Sudakov [15] asked if the Burr-Erdős conjecture holds for bipartite graphs where all vertices in one part have their degree bounded by d. In this section, we answer their question by establishing an embedding theorem for such graphs.

Definition 2.1. Let G be a bipartite graph with vertex partition $V_1 \cup V_2$. For a positive real number θ and an ordered d-tuple $Q \in V_2^d$, we define the *θ-defect of Q* as

$$\omega_\theta(Q) = \begin{cases} 0 & \text{if } |N(Q)| \geq \theta, \\ \frac{\theta}{|N(Q)|} & \text{otherwise.} \end{cases}$$

We simply write $\omega(Q)$ when θ is clear from the context.

For a given ordered d-tuple of vertices Q, the θ-defect $\omega_\theta(Q)$ defined above captures how the number of common neighbors of Q compares to some pre-scribed threshold θ. We give a penalty to the d-tuples having only a small number of common neighbors. Informally, if a set has small average θ-defect over d-tuples, then most d-tuples in it will have at least θ common neighbors. This weight function has been considered before by Alon, Krivelevich, and Su-dakov [2] with a fixed value of the threshold θ. The following lemma, based on dependent random choice, shows that we can find a set in which the average defect is small.

LEMMA 2.2. *Let d, t be natural numbers, s be a nonnegative integer satis-fying $t \geq s$, and η, ε be positive real numbers satisfying $\varepsilon < 1$. Let G be a bipar-tite graph of density at least α with vertex partition $V_1 \cup V_2$. Then there exists a set $A \subseteq V_2$ of size at least $|A| \geq \varepsilon^{1/d} \alpha^t |V_2|$ such that $\frac{1}{|A|^d} \sum_{Q \in A^d} \omega_\theta(Q)^s \leq \frac{\eta^t}{1-\varepsilon}$ for all $\theta \leq \eta \alpha^d |V_1|$.*

Proof. Since ω_θ is increasing in θ, it suffices to prove the lemma when $\theta = \eta \alpha^d |V_1|$. Throughout the proof we will use this fixed value of θ and use the notation ω_θ without the subscript. Let $\mathbf{X} \in V_1^t$ be a t-tuple of vertices chosen uniformly at random (where we allow repetition of vertices), and define $\mathbf{A} = N(\mathbf{X})$. Note that

$$\mathbb{E}\left[|\mathbf{A}|\right] = \sum_{v \in V_2} \mathbf{P}(v \in N(\mathbf{X}))$$

$$= \sum_{v \in V_2} \left(\frac{\deg(v)}{|V_1|}\right)^t \geq |V_2| \left(\frac{1}{|V_2|} \sum_{v \in V_2} \frac{\deg(v)}{|V_1|}\right)^t \geq \alpha^t |V_2|.$$

By convexity, we have $\mathbb{E}\left[|\mathbf{A}|^d\right] \geq \alpha^{dt} |V_2|^d$. Fix a d-tuple $Q \in V_2^d$, and let $|N(Q)| = \gamma \theta$. Since

$$\mathbf{P}\left(Q \in \mathbf{A}^d\right) = \left(\frac{|N(Q)|}{|V_1|}\right)^t = \left(\frac{\gamma \eta \alpha^d |V_1|}{|V_1|}\right)^t = \gamma^t \eta^t \alpha^{dt},$$

if $\gamma < 1$, then $\omega(Q)^s \cdot \mathbf{P}\left(Q \in \mathbf{A}^d\right) = \gamma^{t-s} \eta^t \alpha^{dt} \leq \eta^t \alpha^{dt}$. (We used the fact that $t \geq s$.) On the other hand, if $\gamma \geq 1$, then $\omega(Q) = 0$ by definition. Hence,

$$\mathbb{E}\left[\sum_{Q \in \mathbf{A}^d} \omega(Q)^s\right] = \sum_{Q \in V_2^d} \omega(Q)^s \cdot \mathbf{P}\left(Q \in \mathbf{A}^d\right) \leq |V_2|^d \cdot \eta^t \alpha^{dt}.$$

Therefore,

$$\mathbb{E}\left[|\mathbf{A}|^d - \frac{1-\varepsilon}{\eta^t} \sum_{Q \in \mathbf{A}^d} \omega(Q)^s\right] \geq \varepsilon \alpha^{dt} |V_2|^d.$$

Choose $\mathbf{X}$ so that the random variable on the left-hand side of the inequality above becomes at least as large as its expectation, and let A be $\mathbf{A}$ for this choice of vertices. Then $|A|^d \geq \varepsilon \alpha^{dt} |V_2|^d$, and thus $|A| \geq \varepsilon^{1/d} \alpha^t |V_2|$. The other claim follows from $|A|^d - \frac{1-\varepsilon}{\eta^t} \sum_{Q \in A^d} \omega(Q)^s \geq 0$. □

We now prove a density-embedding theorem for bipartite graphs with one part having bounded maximum degree. Theorem 1.3 follows from the theorem below by taking $\varepsilon \geq \alpha^{d(d-2)}$ since it implies $(1+\varepsilon)\alpha^{-d} \geq (\frac{1+\varepsilon}{\varepsilon})^{1/d} \alpha^{-2}$, and in every graph on $n = n_1 + n_2$ vertices of density α, we can find a bipartite subgraph with n_1 and n_2 vertices in each part, having density at least α.

THEOREM 2.3. *Let ε be a positive real number. Let H be a bipartite graph on n vertices with a vertex partition $W_1 \cup W_2$ where all vertices in W_1 have at most d neighbors in W_2. Let G be a bipartite graph of density at least α with vertex partition $V_1 \cup V_2$ where $|V_1| \geq (1 + \varepsilon)\alpha^{-d}|W_1|$ and $|V_2| \geq (\frac{1+\varepsilon}{\varepsilon})^{1/d} \alpha^{-2}|W_2|$. If $\frac{|W_2|^d}{|W_2|(|W_2|-1)\cdots(|W_2|-d+1)} \leq 1+\varepsilon$, then G contains a copy of H.*

Proof. Let $\theta = |W_1|$, and note that $\theta \leq \frac{1}{1+\varepsilon}\alpha^d|V_1|$. Throughout the proof we will use ω without subscript with the understanding that $\omega = \omega_\theta$. Apply Lemma 2.2 with $t_{2.2} = 2$, $s_{2.2} = 1$, $d_{2.2} = d$, $\eta_{2.2} = \frac{1}{1+\varepsilon}$, and $\varepsilon_{2.2} = \frac{\varepsilon}{1+\varepsilon}$ to find a set $A \subseteq V_2$ of size $|A| \geq (\frac{\varepsilon}{1+\varepsilon})^{1/d}\alpha^2|V_2| \geq |W_2|$ for which $\sum_{Q \in A^d} \omega(Q) \leq (\frac{1}{1+\varepsilon})^2 \frac{1}{1-\varepsilon/(1+\varepsilon)}|A|^d = \frac{1}{1+\varepsilon}|A|^d$. By adding edges if necessary, we may assume that all vertices in W_1 have exactly d neighbors in W_2. Let ϕ be an injective map from W_2 to A chosen uniformly at random. For each vertex $v \in W_1$, let e_v be an ordered d-tuple of vertices obtained from $N(v)$ by arbitrarily ordering the vertices. Note that

$$\mathbb{E}\left[\sum_{v \in W_1} \omega(\phi(e_v))\right] = \sum_{v \in W_1} \mathbb{E}\left[\omega(\phi(e_v))\right] = \sum_{v \in W_1} \sum_{Q \in A^d} \omega(Q)\mathbf{P}(\phi(e_v) = Q)$$

$$\leq |W_1| \cdot \sum_{Q \in A^d} \frac{\omega(Q)}{|A|(|A|-1)\cdots(|A|-d+1)}.$$

Since $\sum_{Q \in A^d} \omega(Q) \leq \frac{1}{1+\varepsilon}|A|^d$ and $|A| \geq |W_2|$,

$$\mathbb{E}\left[\sum_{v \in W_1} \omega(\phi(e_v))\right] \leq |W_1| \cdot \frac{1}{1+\varepsilon} \frac{|W_2|^d}{|W_2|(|W_2|-1)\cdots(|W_2|-d+1)} \leq |W_1|.$$

Therefore, there exists a particular choice of ϕ such that $\sum_{v \in W_1} \omega(\phi(e_v)) \leq |W_1|$. Note that ϕ is trivially a partial embedding of H in G defined on W_2. We will now extend ϕ to W_1. Order the vertices in W_1 as $v_1, \ldots, v_{n_1}$ in decreasing order of $\omega(\phi(e_{v_i}))$ (where ties are broken arbitrarily). We will greedily embed the vertices of W_1 following this order. Suppose that we extended

ϕ to $\{v_1, \ldots, v_{i-1}\}$, and for simplicity define $e_i = e_{v_i}$. First assume that $\omega(\phi(e_i)) \neq 0$. Note that $i \cdot \omega(\phi(e_i)) \leq \sum_{v \in W_1} \omega(\phi(e_v)) \leq |W_1|$ and therefore $\omega(\phi(e_i)) \leq \frac{|W_1|}{i}$. Hence,

$$|N(\phi(e_i))| = \frac{\theta}{\omega(\phi(e_i))} \geq |W_1| \cdot \frac{i}{|W_1|} = i.$$

Since we have so far embedded at most $i-1$ vertices of W_1, we can define $\phi(v_i)$ as a vertex in $N(\phi(e_i)) \setminus \{\phi(v_1), \ldots, \phi(v_{i-1})\}$. On the other hand, if $\omega(\phi(e_i)) = 0$, then $|N(\phi(e_i))| \geq \theta \geq |W_1|$, and therefore we trivially have $N(\phi(e_i)) \setminus \{\phi(v_1), \ldots, \phi(v_{i-1})\} \neq \emptyset$ and can define $\phi(v_i)$ as a vertex in this set. Thus, we can find an embedding of H in G. $\square$

3. Preliminaries

3.1. *Decomposing H.* We start with the following simple lemma that decomposes a given degenerate graph into manageable pieces.

LEMMA 3.1. *Let H be an r-colorable d-degenerate graph on n vertices. Then there exist a natural number k and disjoint subsets $\{W_i^{(j)}\}_{i \in [k], j \in [r]}$ with the following properties*:

  (i) $k \leq \log_2 n$;
 (ii) *for all $(i,j) \in [k] \times [r]$, $|W_i^{(j)}| \leq 2^{-i+1} n$*;
(iii) *for all $j \in [r]$, the set $\bigcup_{i \in [k]} W_i^{(j)}$ is an independent set*; *and*
 (iv) *for all $(i,j) \in [k] \times [r]$, each vertex in $W_i^{(j)}$ has at most $4d$ neighbors in $\bigcup_{i' \geq i, j' \in [r]} W_{i'}^{(j')}$.*

Proof. Let $U_1 = V(H)$. For each $i \geq 1$, define $U_{i+1} \subseteq U_i$ as the set of vertices having degree at least $4d$ in the subgraph $H[U_i]$. Since H is d-degenerate, there are at most $d|U_i|$ edges in the subgraph $H[U_i]$. Therefore, we have $4d|U_{i+1}| \leq 2d|U_i|$, forcing $|U_{i+1}| \leq \frac{1}{2}|U_i|$. Since H has n vertices, we have $|U_i| \leq 2^{-(i-1)} n$ for each $i \geq 1$, and the process continues for $k \leq \log_2 n$ steps. Define $W_i = U_i \setminus U_{i+1}$ for all $i \in [k]$. Consider a proper r-coloring of H using $[r]$, and for each $j \in [r]$, define $W_i^{(j)}$ as the set of vertices of color j in W_i. One can easily check that all the conditions hold. $\square$

3.2. *Defect and average defect.* We will work with the decomposition given by Lemma 3.1. The main objective is to find a partition $\{V_i^{(j)}\}$ of the vertex set of the host graph G so that we can embed $W_i^{(j)}$ to $V_i^{(j)}$ piece by piece. As in the previous section, the existence of such an embedding depends on the 'average defect' of this partition. In this subsection, we introduce the central concepts and notation related to the defect function. We slightly generalize Definition 2.1 so that we consider the number of common neighbors in a specific set.

Definition 3.2. Let G be a graph. For a positive real number θ, a set $T \subseteq V(G)$, and an ordered tuple Q of vertices, we define the *θ-defect of Q in T* as

$$\omega_\theta(Q; T) = \begin{cases} 0 & \text{if } |N(Q; T)| \geq \theta, \\ \frac{\theta}{|N(Q; T)|} & \text{otherwise.} \end{cases}$$

We may simply write $\omega(Q; T)$ when θ is clear from the context.

One can easily check that monotonicity $\omega_\theta(Q; T) \leq \omega_{\theta'}(Q'; T')$ holds for all $\theta \leq \theta'$, $Q \subseteq Q'$ (as sets), and $T \supseteq T'$. Note that in the previous subsection, we utilized Lemma 2.2 only for the case $s = 1$, even though Lemma 2.2 provided a variety of bounds. From now on, it will be crucial to consider other values of s.

Definition 3.3. Let G be a graph and d, s be natural numbers. For a positive real number θ, a set $T \subseteq V(G)$, and a set $\mathcal{Q} \subseteq V(G)^d$, we define the *$s$-th moment of the θ-defect of $\mathcal{Q}$ in T* as

$$\mu_{s,\theta}(\mathcal{Q}; T) = \frac{1}{|\mathcal{Q}|} \sum_{Q \in \mathcal{Q}} \omega_\theta(Q; T)^s.$$

We may simply write $\mu_s(\mathcal{Q}; T)$ when θ is clear from the context. We extend the definition to $s = 0$ by defining $\omega_\theta(Q; T)^0$ to be 0 when $\omega_\theta(Q; T) = 0$, and 1 otherwise.

The monotonicity of the defect function ω and the fact that it equals either 0 or a real number at least 1 implies $\mu_{s,\theta}(\mathcal{Q}; T) \leq \mu_{s',\theta'}(\mathcal{Q}; T')$ for $s \leq s'$, $\theta \leq \theta'$, and $T \supseteq T'$. Note that the conclusion of Lemma 2.2 can be re-written as an upper bound on $\mu_{s,\theta}(A^d; V_1)$. Most previous applications of dependent random choice were proved by controlling the 0-th moment of the θ-defect function, which equals the probability that a random uniform d-tuple in $\mathcal{Q}$ has fewer than θ common neighbors, whereas here we will be considering higher moments. We will focus on the cases when $\mathcal{Q} = A_1 \times \cdots \times A_d$ for some (not necessarily distinct) sets A_i. The following proposition provides a useful relation between average defects of such product sets.

PROPOSITION 3.4. *If $A_1, \ldots, A_d, T$ are vertex subsets, then*

$$\mu_{s,\theta}\left(\prod_{i=1}^{d-1} A_i; T\right) \leq \mu_{s,\theta}\left(\prod_{i=1}^{d} A_i; T\right).$$

Proof. Since $\omega_\theta(Q; T) \leq \omega_\theta(Q'; T)$ for all $Q \subseteq Q'$,

$$\sum_{Q \in \prod_{i=1}^{d-1} A_i} \omega_\theta(Q; T)^s \leq \sum_{Q \in \prod_{i=1}^{d-1} A_i} \left(\frac{1}{|A_d|} \sum_{v_d \in A_d} \omega_\theta(Q \cup \{v_d\}; T)^s\right)$$

$$= \frac{1}{|A_d|} \sum_{Q \in \prod_{i=1}^{d} A_i} \omega_\theta(Q; T)^s.$$

Divide both sides of the inequality by $\prod_{i=1}^{d-1} |A_i|$ to obtain $\mu_{s,\theta}(\prod_{i=1}^{d-1} A_i; T) \leq \mu_{s,\theta}(\prod_{i=1}^{d} A_i; T)$. □

The following proposition shows that the contribution of d-tuples with repeated vertices towards the average defect is of small magnitude.

PROPOSITION 3.5. *Let $A_1, \ldots, A_d$ be vertex subsets all of sizes at least m, and let T be a vertex subset. Let $\partial \mathcal{Q} \subseteq \prod_{i \in [d]} A_i$ be the set of d-tuples $(v_1, \ldots, v_d)$ for which $v_i = v_j$ for some distinct $i, j \in [d]$. Then*

$$\sum_{Q \in \partial \mathcal{Q}} \omega_\theta(Q; T)^s \leq \frac{d(d-1)}{2m} \sum_{Q \in \prod_{i=1}^{d} A_i} \omega_\theta(Q; T)^s.$$

Proof. For each distinct $a, b \in [d]$, define $\mathcal{Q}_{a,b}$ as the set of d-tuples in $\partial \mathcal{Q}$ whose a-th and b-th coordinates coincide. Note that if $Q = (v_1, \ldots, v_d)$ and $v_{d-1} = v_d$, then $\omega_\theta(Q; T) = \omega_\theta((v_1, \ldots, v_{d-1}); T)$. Therefore, by Proposition 3.4,

$$\sum_{Q \in \mathcal{Q}_{d-1,d}} \omega_\theta(Q; T)^s = \sum_{Q \in \prod_{i=1}^{d-1} A_i} \omega_\theta(Q; T)^s \leq \frac{1}{|A_d|} \sum_{Q \in \prod_{i=1}^{d} A_i} \omega_\theta(Q; T)^s.$$

Since $|A_i| \geq m$ for all $i \in [d]$, the conclusion follows by summing the inequalities over all pairs a, b. □

The next proposition asserts that for a pair of sets V_1 and V_2, if the s-th moment of the average defect of V_1^d into V_2 is small, then all d-tuples in V_1^d have many common neighbors in V_2.

PROPOSITION 3.6. *Let d, s, and θ be natural numbers. For all sets of vertices V_1 and V_2, the number of d-tuples $Q \in V_1^d$ satisfying $|N(Q; V_2)| < \frac{\theta}{|V_1|^{d/s}}$ is less than $\mu_{s,\theta}(V_1^d; V_2)$.*

Proof. If a d-tuple $Q \in V_1^d$ satisfies $|N(Q; V_2)| < \frac{\theta}{|V_1|^{d/s}}$, then by definition, we have $\omega_\theta(Q; V_2)^s > |V_1|^d$. Therefore, the number of such d-tuples is less than $\mu_{s,\theta}(V_1^d; V_2)$. □

We will use Proposition 3.6 mostly in the case when $\mu_{s,\theta}(V_1^d; V_2) < 1$. For such cases, all d-tuples $Q \in V_1^d$ satisfy $|N(Q; V_2)| \geq \frac{\theta}{|V_1|^{d/s}}$.

3.3. *Outline of proof.* Let H be a d-degenerate graph and G be a host graph that we would like to embed H into. Let $\{W_i\}_{i \in [k]}$ be the partition of the vertex set of H obtained from Lemma 3.1, and suppose that we have disjoint vertex subsets $\{V_i\}_{i \in [k]}$ of G. We will embed W_i in V_i one vertex at a time, starting at $i = k$ and proceeding in decreasing order of index. Suppose that we have successfully embedded W_{i+1} in V_{i+1} (and all previous steps) and are about to embed W_i in V_i. Let ϕ denote the partial embedding that we have

at that moment. For each vertex $v \in W_i$, define $N^+(v) = N(v) \cap \bigcup_{j>i} W_j$. For simplicity, we assume that $|N^+(v)| = d$ for all $v \in V(H)$. We can force the extension of ϕ to W_i to be a homomorphism by embedding each vertex $v \in W_i$ to a common neighbor of vertices in $\phi(N^+(v))$. To obtain an embedding, we must additionally guarantee that ϕ is injective. Let e_v be an arbitrary ordering of the tuple $N^+(v)$. If the map constructed at the previous steps was well chosen so that the sum

$$(1) \qquad \sum_{v \in W_i} \omega_\theta(\phi(e_v); V_i)$$

is small, then we can proceed as in the proof of Theorem 2.3 to extend the map to W_i.

To briefly describe how to obtain control on (1), we consider the case when all vertices in $v \in W_i$ satisfy $N^+(v) \subseteq W_{i+1}$. In this case, if the embedding $\phi|_{W_{i+1}}$ was chosen as a uniform map from W_{i+1} to V_{i+1}, then we would expect the sum above to be $|W_i|\mu_{1,\theta}(V_{i+1}^d; V_i)$. Hence, we can make (1) small by finding sets $\{V_i\}_{i \in [k]}$ having small average defects in an appropriately defined way. However, even if we are given such subsets $\{V_i\}_{i \in [k]}$, our assumption that $\phi|_{W_{i+1}}$ is a uniform random map cannot be true. Nevertheless, we show that (1) is not too far from $\mu_{1,\theta}(V_{i+1}^d; V_i)$ by showing that our map is not too far from the random uniform map. For example, consider the simple case discussed above where all vertices in $v \in W_i$ satisfy $N^+(v) \subseteq W_{i+1}$. For a fixed vertex $v \in W_i$, let $w_1, \ldots, w_d$ be its neighbors in $N^+(v)$ and let $e_v = (w_1, \ldots, w_d)$. If the embedding was chosen uniformly at random, then the image of w_j would be chosen in the set V_{i+1}, but our algorithm chooses its image in $N(\phi(e_{w_j}); V_{i+1})$ instead. Hence, compared to the random uniform map, the expected value of $\omega_\theta(\phi(e_v); V_{i+1})$ could be larger by at most a factor $\prod_{j \in [d]} \frac{|V_{i+1}|}{|N(\phi(e_{w_j}); V_{i+1})|}$. Note that either $|N(\phi(e_{w_j}); V_{i+1})| \geq \theta$ or $\frac{|V_{i+1}|}{|N(\phi(e_{w_j}); V_{i+1})|} = \frac{|V_{i+1}|}{\theta} \omega_\theta(\phi(e_{w_j}); V_{i+1})$. Therefore, by the inequality of arithmetic and geometric means,

$$\prod_{j \in [d]} \frac{|V_{i+1}|}{|N(\phi(e_{w_j}); V_{i+1})|} \leq \frac{1}{d} \sum_{j \in [d]} \left(\frac{|V_{i+1}|}{|N(\phi(e_{w_j}); V_{i+1})|} \right)^d$$

$$\leq \frac{1}{d} \sum_{j \in [d]} \frac{|V_{i+1}|^d}{\theta^d} \left(1 + \omega_\theta(\phi(e_{w_j}); V_{i+1})^d \right).$$

The second term in the summand in expectation has value $\mu_{d,\theta}(V_{i+2}^d; V_{i+1})$, and thus if this quantity is less than 1 and everything worked as planned, then the right-hand side would be at most $2 \left(\frac{|V_{i+1}|}{\theta} \right)^d$. This means that in expectation, the sum (1) would be at most $2 \left(\frac{|V_{i+1}|}{\theta} \right)^d \cdot \mu_{1,\theta}(V_{i+1}^d; V_i)$. In general, if $N^+(v) \nsubseteq W_{i+1}$, then there will be more dependencies between the random variables and it gets trickier to control the events. However, the underlying

idea, comparing our algorithm with the uniform random map and measuring its deviation using higher moments of the average defect function, remains the same. We will formalize this idea in Section 4.

4. Embedding scheme

In this section, we develop the embedding scheme. The main new ingredient of our embedding algorithm compared to previous ones lies in the randomness. We recall the proof of Theorem 2.3 to illustrate the essence of this difference. There, we embedded a graph H with bipartition $W_1 \cup W_2$ into a graph V with bipartition $V_1 \cup V_2$. We started by finding a set $A \subseteq V_2$ with small defect into V_1 using dependent random choice. This step is identical to that from previous applications (except that we imposed a weaker bound on the defect function). In the second step, we found a map ϕ from W_2 to V_2. The main difference lies in this step. In previous applications, it was done greedily so that for all $v \in W_1$, the set $\phi(N(v))$ has many common neighbors. For our purpose, the weaker bound from the first step makes this strategy not viable. Therefore, we took a random map from W_2 to V_2 and used the first moment to find a map for which the number of common neighbors is controlled by $\mu(A^d; V_1)$. Finally, we embedded W_1 in V_1 so that each vertex $v \in W_1$ gets mapped to a vertex in the common neighbor of $\phi(N(v))$. The difficulty in extending this algorithm to nonbipartite graphs lies in the fact that there is no clear distinction between the second and third steps. That is, when mapping a vertex $v \in W_1$ to the common neighbor of $\phi(N(v))$, we must retain randomness so that the process can be continued.

Let G be a graph with disjoint vertex subsets $\{V_i\}_{i \in [k]}$ and H be a graph with a vertex partition $\{W_i\}_{i \in [k]}$ into independent sets where each vertex in W_i has at most d neighbors in $W_{i+1} \cup \cdots \cup W_k$ for all $i \in [k-1]$. Let $\{\theta_i\}_{i \in [k-1]}$ be natural numbers. For $i \in [k-1]$ and a vertex $x \in W_i$, define $N^+(x) = N(x) \cap \bigcup_{j \in [i+1, k]} W_j$. Add edges to H if necessary so that $|N^+(x)| = d$ for all vertices $x \in V(H) \setminus W_k$. For $i \in [k]$ and $x \in W_i$, let e_x be an (arbitrary) ordered d-tuple of vertices formed from $N^+(x)$. Define a random map $\psi : V(H) \to V(G)$ using the following 'random greedy' process:

1. Take an injection from W_k to V_k uniformly at random.
2. For $i \in [k-1]$, given a map ψ defined on $W_{i+1} \cup \cdots \cup W_k$, we then extend ψ to W_i. Let $x_1, x_2, \ldots, x_m$ be the vertices in W_i ordered so that $\omega_{\theta_i}(\psi(e_{x_j}); V_i)$ is decreasing in j. (Ties are broken arbitrarily.)
3. After embedding $x_1, \ldots, x_{j-1}$, define $e_j = e_{x_j}$ and $L_j = N(\psi(e_j); V_i) \setminus \{\psi(x_1), \ldots, \psi(x_{j-1})\}$ as the set of available vertices for x_j. Embed x_j using the first valid option among the following:
   3.1. if $N(\psi(e_j); V_i) = \emptyset$, then let $\psi(x_j)$ be a vertex in V_i chosen uniformly at random;

3.2. if $|L_j| < \frac{1}{2}|N(\psi(e_j); V_i)|$, then let $\psi(x_j)$ be a vertex in $N(\psi(e_j); V_i)$ chosen uniformly at random;

3.3. if $|L_j| \geq \frac{1}{2}|N(\psi(e_j); V_i)|$, then let $\psi(x_j)$ be a vertex in L_j chosen uniformly at random.

If we only run Steps 3.2 and 3.3 (but never Step 3.1) throughout the embedding process, then the resulting map is a homomorphism from H to G. As we will later see, Step 3.1 can easily be ruled out if we have control on the defect function, so one can safely assume that ψ is always an homomorphism from H to G. Given that ψ is a homomorphism, it suffices to show that ψ is injective, in order to show that it is an embedding. Thus, it will be crucial to understand when we will run Step 3.3 instead of Step 3.2.

Throughout the proof we consider the defect function $\omega_\theta(Q; T)$ for various different choices of θ, Q, and T. Since these parameters will always be a function of the vertex $x \in V(H)$ that we are about to embed, we make the following definitions to remove redundant information and simplify notation.

Definition 4.1. Suppose we are given $\{V_i\}_{i \in [k]}$, $\{W_i\}_{i \in [k]}$, and $\{\theta_i\}_{i \in [k]}$. For $i \in [k-1]$ and $x \in W_i$, make the following definitions:

(i) $\theta_x := \theta_i$ and $V_x := V_i$;

(ii) $\omega(x; \psi) := \omega_{\theta_x}(\psi(e_x); V_x)$;

(iii) $\mathcal{Q}_x := V_{i_1} \times \cdots \times V_{i_d}$, where $W_{i_1} \times \cdots \times W_{i_d}$ is the unique product space containing e_x;

(iv) $\mu_s(x) := \mu_{s,\theta_x}(\mathcal{Q}_x; V_x)$ and $\mu_s := \max_{x \in V(H)} \mu_s(x)$;

(v) $\gamma := \max\left\{1, \max_{i \in [k-1]} \frac{|V_i|}{\theta_i}\right\}$.

For $x \in V(H)$, if $\mu_s(x)$ is finite, then there are no d-tuples $Q \in \mathcal{Q}_x$ having $N(Q; V_x) = \emptyset$. Hence, we will never run Step 3.1 in such case. For a vertex $x \in W_i$, the parameter $\theta_x = \theta_i$ is the defect that is of interest when embedding x in V_i, as whether we run Step 3.2 or 3.3 depends on $|N(\psi(e_x); V_x)|$, which is closely related to $\omega(x; \psi) = \omega_{\theta_x}(\psi(e_x); V_x)$. Note that the enumeration of vertices in Step 2 is determined by the values $\omega(x; \psi)$. Further note that if the embedding algorithm chose the image of each vertex $y \in e_x$ as a random uniform vertex in V_y, then $\mu_s(x)$ would be the expected value of $\omega(x; \psi)^s$. This is an important observation, since we will later see that even though ψ does not always choose a random uniform vertex, γ can be used to bound the difference between the expected value of $\omega(x; \psi)^s$ defined by our random process and by the random uniform case. The following theorem establishes a sufficient condition for ψ to be an embedding.

THEOREM 4.2. *If $|V_k| \geq 2|W_k|$ and $\theta_i \geq 2|W_i|$ for all $i \in [k-1]$, then the probability that ψ does not induce an embedding of H in G is at most $2^{2d+2}\gamma^{2d}\mu_{4d} \sum_{i \in [k-1]} \frac{|W_i|}{\theta_i}$.*

The rest of this section focuses on proving Theorem 4.2. The proof of Theorem 1.1 then follows by finding subsets of vertices of G satisfying the condition of Theorem 4.2.

4.1. *Proof of Theorem* 4.2. For technical reasons, we consider auxiliary graphs. Add a set W_{k+1} of d vertices to H, make the bipartite graph between W_k and W_{k+1} complete, and between W_i and W_{k+1} empty for all $i \in [k-1]$. Denote the resulting graph as H'. Define $N^+(x) := N(x) \cap W_{k+1}$ for each $x \in W_k$. Note that now, for all $x \in V(H)$, we have $|N^+(x)| = d$. Let G' be a graph obtained from G by adding a set V_{k+1} of d vertices adjacent to all other vertices. Define $\theta_k = |V_k|$, and note that $\omega_{\theta_k}(Q; V_k) = 0$ for all $Q \in V_{k+1}^d$. Further note that for all $x \in W_k$, we have $\omega_{\theta_k}(x; \psi) = 0$ and $\mu_s(x) = 0$ regardless of the choice of ψ and s since $e_x \in W_{k+1}^d$. Note that Definition 4.1 now extends to $i = k$ as well. The parameters for $i \leq k-1$ do not change depending on whether we consider the graphs H, G, or the graphs H', G'. Furthermore, γ is identical when defined for the pair G or in G', since $\frac{|V_k|}{\theta_k} = 1$. Thus we abuse notation and use the parameters defined in Definition 4.1 as if they were defined for H', G'.

Throughout this section, we will apply the embedding scheme defined above to embed H' to G', but with a slightly modified first step:

1′. Take a map from W_{k+1} to V_{k+1} uniformly at random (instead of a random injection).

We then embed the rest of the graph using the same algorithm (where we also consider $i = k$ in Steps 2 and 3). Note that ψ restricted to the vertices $V(H)$ has the same distribution as the map previously defined without the sets W_{k+1} and V_{k+1}.

The following lemma gives a sufficient condition for ψ to be an embedding.

LEMMA 4.3. *Let* $G, H, G', H', \{V_i\}_{i \in [k+1]}, \{W_i\}_{i \in [k+1]}$, *and* ψ *be as described above. Suppose that* $\theta_i \geq 2|W_i|$ *for all* $i \in [k]$. *For a natural number* s, *let* $\phi : V(H') \to V(G')$ *be a map satisfying* $\mathbf{P}(\psi = \phi) > 0$ *and* $\sum_{x \in W_i} \omega(x; \phi)^s \leq \frac{1}{2}\theta_i$ *for all* $i \in [k]$. *Then* $\phi|_{V(H)}$ *is an embedding of* H *in* G.

Proof. As discussed in Section 3, we have $\sum_{x \in W_i} \omega(x; \phi) \leq \sum_{x \in W_i} \omega(x; \phi)^s$, and therefore it suffices to consider the case $s = 1$. Let $\phi : V(H') \to V(G')$ be a map satisfying $\sum_{x \in W_i} \omega(x; \phi) \leq \frac{1}{2}\theta_i$ for all $i \in [k]$. Fix $i \in [k]$ and condition on the event that $\psi = \phi$ after mapping the vertices $W_{i+1}, \ldots, W_k$. Let $x_1, x_2, \ldots, x_m$ be the vertices in W_i in decreasing order of $\omega(x_j; \phi)$, and define $e_j = e_{x_j}$ for each $j \in [m]$. Note that this is the order that the vertices in W_i will be mapped to V_i. Consider the j-th step. Since $\sum_{x \in W_i} \omega(x; \phi)$ is finite, we have $N(\phi(e_j); V_i) \neq \emptyset$. Hence, we will not run Step 3.1 when mapping v_j.

If $\omega(x_j; \phi) = 0$, then $|N(\phi(e_j); V_i)| \geq \theta_i \geq 2|W_i|$, and thus we determine $\psi(x_j)$ according to Step 3.3. If $\omega(x_j; \phi) \neq 0$, then by how we ordered the vertices, we have $j \cdot \omega(x_j; \phi) \leq \frac{1}{2}\theta_i$, and thus

$$\frac{\theta_i}{|N(\phi(e_j); V_i)|} = \omega(x_j; \phi) \leq \frac{\theta_i}{2j},$$

from which it follows that $|N(\phi(e_j); V_i)| \geq 2j$. Since we embedded at most $j - 1$ vertices of W_i prior to x_j, we determine $\psi(x_j)$ according to Step 3.3 when embedding x_j. Since $\mathbf{P}(\psi(x_j) = \phi(x_j)) > 0$ (conditioned on $\psi = \phi$ for all previous vertices), we see that $\phi(x_j)$ is in $N(\phi(e_j); V_i)$ and is distinct from all previously mapped vertices. Since the analysis applies to all steps of the embedding, we can conclude that $\phi|_{V(H)}$ is injective, and thus an embedding. $\square$

Lemma 4.3 shows that it is crucial to control the quantity $\sum_{x \in W_i} \omega(x; \psi)^s$. Note that the expected value of $\omega(x; \psi)^s$ is $\mu_s(x)$ if $\psi(e_x)$ were uniformly distributed in $\mathcal{Q}_x$. This was the case in the proof of Theorem 2.3, and there, thus we were able to conclude that $\sum_{x \in W_i} \omega(x; \psi)^s$ is small quite straightforwardly. Now our situation is more complicated, since $\psi(e_x)$ is no longer uniformly distributed in $\mathcal{Q}_x$. We gain control on the sum by comparing our distribution with the uniform distribution.

Let ν be the distribution on the set of maps $\phi : V(H') \to V(G')$ obtained by the random greedy embedding algorithm defined above. For a set of vertices $I \subseteq V(H)$, we say that $\tilde{\nu}$ is a probability distribution obtained from ν by *neutralizing* I if in the random greedy process above, every time we embed a vertex $x \in I$, instead of following Steps 3.1, 3.2, or 3.3, we choose the image of x as a uniform random vertex in V_x. For example, if we neutralize all vertices, then the resulting distribution is a uniform distribution over all maps $\phi : V(H') \to V(G')$ satisfying $\phi(W_i) \subseteq V_i$ for all $i \in [k + 1]$.

From now on, we use ψ to denote a random map from $V(H')$ to $V(G')$ whose distribution is determined by the probability measure under consideration. In contrast, we use ϕ to denote fixed (nonrandom) maps. Recall that $\gamma = \max\left\{1, \max_{i \in [k]} \frac{|V_i|}{\theta_i}\right\}$.

LEMMA 4.4. *Let $I_1 \subseteq V(H)$ be a subset of vertices, and let ν_1 be the distribution obtained from ν by neutralizing I_1. Let X be a nonnegative random variable depending only on the images of vertices in J for some $J \subseteq V(H)$. Define $I_2 = I_1 \cup J$ and define ν_2 as the distribution obtained from ν by neutralizing I_2. Suppose that $t = |J \setminus I_1| \geq 1$. Then*

$$\mathbb{E}_{\nu_1}[X] \leq 2^t \gamma^t \mathbb{E}_{\nu_2}[X] + 2^{2t-1} \gamma^{2t} \mathbb{E}_{\nu_2}[X^2] + \frac{1}{2t} \sum_{y \in J \setminus I_1} \mathbb{E}_{\nu_2}\left[\omega(y; \psi)^{2t}\right].$$

Therefore, if $X^2 \geq X$ holds with probability 1, *then*

$$\mathbb{E}_{\nu_1}[X] \leq 2^{2t}\gamma^{2t}\mathbb{E}_{\nu_2}[X^2] + \frac{1}{2t}\sum_{y \in J \setminus I_1} \mathbb{E}_{\nu_2}\left[\omega(y;\psi)^{2t}\right].$$

Note that the first term on the right-hand side, $\mathbb{E}_{\nu_2}[X^2]$, under the probability measure ν_2, is determined completely by a set of vertices whose images are chosen uniformly at random. Hence, this lemma allows us to compare the distribution on maps defined by our random greedy algorithm with the random uniform map.

Proof. For each vertex $y \in I_2 \setminus I_1$, define $C_y(\phi) = \frac{2|V_y|}{|N(\phi(e_y);V_y)|}$. For a set $W \subseteq V(H')$ of size $|W| = t$, fix a map $\phi : W \to V(G')$. We use the notation $\psi_t = \phi$ to indicate the event that the random map ψ obtained after embedding the first t vertices is ϕ. Define $I_W = I_2 \setminus (I_1 \cup W)$. Let ν_W be the distribution obtained from ν_1 by neutralizing $I_1 \cup I_W$. We prove that for all W, t and ϕ as above,

$$(2) \qquad \mathbb{E}_{\nu_1}\left[X \,\middle|\, \psi_t = \phi\right] \leq \mathbb{E}_{\nu_W}\left[X \cdot \prod_{y \in I_W} C_y(\psi) \,\middle|\, \psi_t = \phi\right].$$

We prove (2) by (downward) induction on t. If $t = |V(H)|$, then $I_W = \emptyset$ and so the random variables on both sides of the inequality equal X. Therefore, (2) trivially holds.

Let us now investigate the value t while assuming that (2) is true for all larger values. Let W be a set of size t and ϕ be a map defined on W. Conditioned on $\psi_t = \phi$, we know which vertex will be embedded next. Say that it is $x_{t+1} \in W_i$ for some $i \in [k]$. Define $W' = W \cup \{x_{t+1}\}$. For each $z \in V_i$, let ϕ_z be the extension of ϕ obtained by defining $\phi(x_{t+1}) = z$. Then

$$\mathbb{E}_{\nu_1}\left[X \,\middle|\, \psi_t = \phi\right] = \sum_{z \in V_i} \mathbf{P}_{\nu_1}\left(\psi_{t+1} = \phi_z \,\middle|\, \psi_t = \phi\right) \mathbb{E}_{\nu_1}\left[X \,\middle|\, \psi_{t+1} = \phi_z\right].$$

Therefore, by the inductive hypothesis, we have

$$\mathbb{E}_{\nu_1}\left[X \,\middle|\, \psi_t = \phi\right]$$

$$\leq \sum_{z \in V_i} \mathbf{P}_{\nu_1}\left(\psi_{t+1} = \phi_z \,\middle|\, \psi_t = \phi\right) \mathbb{E}_{\nu_{W'}}\left[X \cdot \prod_{y \in I_{W'}} C_y(\psi) \,\middle|\, \psi_{t+1} = \phi_z\right]$$

$$= \sum_{z \in V_i} \mathbf{P}_{\nu_1}\left(\psi_{t+1} = \phi_z \,\middle|\, \psi_t = \phi\right) \mathbb{E}_{\nu_W}\left[X \cdot \prod_{y \in I_{W'}} C_y(\psi) \,\middle|\, \psi_{t+1} = \phi_z\right],$$

where the second equality follows since the distributions of $\nu_{W'}$ and ν_W differ only on the image of x_{t+1}, which is fixed once we condition on $\psi_{t+1} = \phi_z$.

If $I_{W'} = I_W$, then $x_{t+1} \notin I_2$, and thus $\mathbf{P}_{\nu_1}(\psi_{t+1} = \phi_z | \psi_t = \phi) = \mathbf{P}_{\nu_W}(\psi_{t+1} = \phi_z | \psi_t = \phi)$. Therefore, the right-hand side above is

$$\sum_{z \in V_i} \mathbf{P}_{\nu_W}(\psi_{t+1} = \phi_z | \psi_t = \phi) \, \mathbb{E}_{\nu_W}\left[X \cdot \prod_{y \in I_W} C_y(\psi) \,\middle|\, \psi_{t+1} = \phi_z \right]$$

$$= \mathbb{E}_{\nu_W}\left[X \cdot \prod_{y \in I_W} C_y(\psi) \,\middle|\, \psi_t = \phi \right],$$

proving our claim. On the other hand if $I_{W'} \neq I_W$, then $I_{W'} = I_W \cup \{x_{t+1}\}$ and $x_{t+1} \in J \backslash I_1$. Note that the image of x_{t+1}, under the probability distribution ν_1, will be chosen uniformly in a set of size $|V_i|$ (Case 3.1), $|N(\phi(e_{x_{t+1}}); V_i)|$ (Case 3.2), or at least $\frac{|N(\phi(e_{x_{t+1}}); V_i)|}{2}$ (Case 3.3). Since $\mathbf{P}_{\nu_W}(\psi_{t+1} = \phi_z | \psi_t = \phi) = \frac{1}{|V_i|}$, it follows that

$$\mathbf{P}_{\nu_1}(\psi_{t+1} = \phi_z | \psi_t = \phi) \leq \frac{2}{|N(\phi(e_{x_{t+1}}); V_i)|}$$

$$= C_{x_{t+1}}(\phi) \cdot \mathbf{P}_{\nu_W}(\psi_{t+1} = \phi_z | \psi_t = \phi).$$

Therefore,

$$\mathbb{E}_{\nu_1}\left[X \,\middle|\, \psi_t = \phi \right]$$

$$\leq \sum_{z \in V_i} \mathbf{P}_{\nu_1}(\psi_{t+1} = \phi_z | \psi_t = \phi) \, \mathbb{E}_{\nu_W}\left[X \cdot \prod_{y \in I_{W'}} C_y(\psi) \,\middle|\, \psi_{t+1} = \phi_z \right]$$

$$\leq \sum_{z \in V_i} C_{x_{t+1}}(\phi) \cdot \mathbf{P}_{\nu_W}(\psi_{t+1} = \phi_z | \psi_t = \phi) \, \mathbb{E}_{\nu_W}\left[X \cdot \prod_{y \in I_{W'}} C_y(\psi) \,\middle|\, \psi_{t+1} = \phi_z \right].$$

Since $C_{x_{t+1}}(\psi) = \frac{2|V_{x_{t+1}}|}{|N(\psi(e_{x_{t+1}}); V_{x_{t+1}})|}$ and $e_{x_{t+1}} \in \left(\bigcup_{j \geq i+1} W_j \right)^d$, the value of $C_{x_{t+1}}(\psi)$ is determined once we condition on $\psi_{t+1} = \phi_z$. Since $I_W = I_{W'} \cup \{v_{t+1}\}$, we have

$$\mathbb{E}_{\nu_1}\left[X \,\middle|\, \psi_t = \phi \right]$$

$$\leq \sum_{z \in V_i} \mathbf{P}_{\nu_W}(\psi_{t+1} = \phi_z | \psi_t = \phi) \, \mathbb{E}_{\nu_W}\left[X \cdot \prod_{y \in I_W} C_y(\psi) \,\middle|\, \psi_{t+1} = \phi_z \right]$$

$$= \mathbb{E}_{\nu_W}\left[X \cdot \prod_{y \in I_W} C_y(\psi) \,\middle|\, \psi_t = \phi \right].$$

This proves the claim.

The inequality (2) for $t = 0$ gives $\mathbb{E}_{\nu_1}[X] \leq \mathbb{E}_{\nu_2}\left[X \cdot \prod_{y \in I_2 \backslash I_1} C_y(\psi) \right]$. Note that $C_y(\psi) = \frac{2|V_y|}{|N(\psi(e_y); V_y)|} \leq 2|V_y| \cdot \max\left\{ \frac{1}{\theta_y}, \frac{\omega(y; \psi)}{\theta_y} \right\}$. Therefore, if $|I_2 \backslash I_1| = t$,

then

$$\prod_{y \in I_2 \setminus I_1} C_y(\psi) \leq 2^t \prod_{y \in I_2 \setminus I_1} |V_y| \left(\frac{\max\{1, \omega(y; \psi)\}}{\theta_y} \right)$$

$$\leq \left(2^t \prod_{y \in I_2 \setminus I_1} \frac{|V_y|}{\theta_y} \right) \frac{1}{t} \sum_{y \in I_2 \setminus I_1} (\max\{1, \omega(y; \psi)\})^t$$

$$\leq 2^t \gamma^t \frac{1}{t} \sum_{y \in I_2 \setminus I_1} (1 + \omega(y; \psi)^t).$$

Hence, for $C = 2^t \gamma^t$,

$$\mathbb{E}_{\nu_1}[X] \leq \mathbb{E}_{\nu_2}[CX] + \frac{1}{t} \sum_{y \in I_2 \setminus I_1} \mathbb{E}_{\nu_2}\left[CX \cdot \omega(y; \psi)^t \right]$$

$$\leq \mathbb{E}_{\nu_2}[CX] + \frac{1}{t} \sum_{y \in I_2 \setminus I_1} \frac{1}{2} \left(\mathbb{E}_{\nu_2}\left[C^2 X^2 + \omega(y; \psi)^{2t} \right] \right)$$

$$\leq \mathbb{E}_{\nu_2}[CX] + \frac{1}{2} \mathbb{E}_{\nu_2}\left[C^2 X^2 \right] + \frac{1}{2t} \sum_{y \in I_2 \setminus I_1} \mathbb{E}_{\nu_2}\left[\omega(y; \psi)^{2t} \right]. \qquad \square$$

We plan to gain control on the defects $\omega(x; \psi)$ by using Lemma 4.4. As explained above, the first term in the right-hand side of Lemma 4.4 gives a direct comparison between our process and the random uniform map. However, the second term in the right-hand side of Lemma 4.4 is still problematic since it is in general determined by vertices that are not yet neutralized. We repeatedly apply Lemma 4.4 to further gain control on these terms. As we proceed, the set of neutralized vertices further propagates and eventually there will be no vertices left to be neutralized.

LEMMA 4.5. *For all $x \in V(H)$, $\mathbb{E}\left[\omega(x; \psi)^{2d} \right] \leq 2^{2d+1} \gamma^{2d} \mu_{4d}$.*

Proof. For a vertex $x \in W_i$, let T_0 be the vertex-weighted rooted tree labelled by vertices in $V(H)$ with a single root vertex labelled x having weight 1. For $j \geq 0$, suppose that we constructed a weighted rooted labelled tree T_j, and let $L_j = V(T_j) \setminus V(T_{j-1})$. For a node $a \in L_j$, define $y_a \in V(H)$ as the label of a, and $P(a)$ as the set of vertices on the path from the root to the parent node of a in the tree T_j. Let $I(a) = W_{k+1} \cup \bigcup_{b \in P(a)} N^+(y_b)$, and let $F(a) = N^+(y_a) \setminus I(a)$. For each vertex $y \in F(a)$, add a child to a labelled y and let its weight be $\frac{\sigma(a)}{2|F(a)|}$, where $\sigma(a)$ is the weight of a in T_j. Let T_{j+1} be the tree obtained by doing this process for all $a \in L_j$. Note that the process eventually stops since each path from a root to a leaf has labels of the form $(y_{i_1}, y_{i_2}, \ldots, y_{i_s})$ for vertices $y_{i_j} \in W_{i_j}$ satisfying $i_1 < i_2 < \cdots < i_s$. Let T_t be the final tree; i.e., t is the minimum integer for which $T_t = T_{t+1}$. We may assume that $T_1 \subseteq T_2 \subseteq \cdots \subseteq T_t$.

Define $\gamma' = 2^{2d}\gamma^{2d}$. Let ν be the probability measure on maps $\psi : V(H') \to V(G')$ induced by our random embedding algorithm. For a tree T_j and its node a, let ν_a denote the probability measure obtained from ν by neutralizing $I(a)$. We claim that for each $j \geq 0$,

$$\mathbb{E}_\nu\left[\omega(x;\psi)^{2d}\right] \leq \sum_{a \in V(T_{j-1})} \sigma(a)\gamma'\mu_{4d} + \sum_{a \in L_j} \sigma(a)\mathbb{E}_{\nu_a}[\omega(y_a;\psi)^{2d}],$$

where for $j = 0$ we let T_{-1} be the empty graph. The root node r is the unique node in T_0 and has weight $\sigma(r) = 1$. Moreover, ν_r and ν have identical distribution since $I(r) = W_{k+1}$ and we defined the map from W_{k+1} to V_{k+1} to be a random uniform map. Therefore, the inequality above holds for $j = 0$. (In fact, equality holds.)

Suppose that the claim holds for some $j \geq 0$. For a vertex $a \in L_j$, let $\Gamma(a)$ be the set of children of a in T_{j+1}, and let ν_0 be the probability measure obtained from ν_a by neutralizing the vertices in $N^+(a)$. If $\Gamma(a) = \emptyset$, then $N^+(y_a) \subseteq I(a)$ and thus $\mathbb{E}_{\nu_a}[\omega(y_a;\psi)^{2d}] \leq \mu_{2d} \leq \mu_{4d}$. Otherwise if $\Gamma(a) \neq \emptyset$, then for all $b \in \Gamma(a)$, we have $\nu_0 = \nu_b$. Therefore, by Lemma 4.4, we have

$$(3) \quad \mathbb{E}_{\nu_a}[\omega(y_a;\psi)^{2d}] \leq \gamma' \cdot \mathbb{E}_{\nu_0}[\omega(y_a;\psi)^{4d}] + \frac{1}{2|\Gamma(a)|}\sum_{b \in \Gamma(a)} \mathbb{E}_{\nu_b}[\omega(y_b;\psi)^{2|\Gamma(a)|}].$$

Since the images of $N^+(a)$ are neutralized in the measure ν_0, by definition we have $\mathbb{E}_{\nu_0}[\omega(y_a;\psi)^{4d}] \leq \mu_{4d}$. Since $\gamma' \geq 1$, we will use the same bound (3) for nodes $a \in L_j$ having $\Gamma(a) = \emptyset$ with the understanding that the second term equals zero for such nodes. Therefore, by the inductive hypothesis and the fact $|\Gamma(a)| = |F(a)| \leq d$ for all $a \in L_j$, we have

$$\mathbb{E}_\nu\left[\omega(x;\psi)^{2d}\right] \leq \sum_{a \in V(T_{j-1})} \sigma(a)\gamma'\mu_{4d} + \sum_{a \in L_j} \sigma(a)\mathbb{E}_{\nu_a}[\omega(y_a;\psi)^{2d}]$$

$$\leq \sum_{a \in V(T_{j-1})} \sigma(a)\gamma'\mu_{4d}$$

$$+ \sum_{a \in L_j} \sigma(a)\left(\gamma' \cdot \mu_{4d} + \frac{1}{2|F(a)|}\sum_{b \in \Gamma(a)} \mathbb{E}_{\nu_b}[\omega(y_b;\psi)^{2d}]\right)$$

$$= \sum_{a \in V(T_j)} \sigma(a)\gamma'\mu_{4d} + \sum_{a \in L_{j+1}} \sigma(a)\mathbb{E}_{\nu_a}[\omega(y_a;\psi)^{2d}],$$

thus proving the claim.

Since $T_t = T_{t+1}$, in the end we see that $\mathbb{E}_\nu\left[\omega(x;\psi)^{2d}\right] \leq \sum_{a \in V(T_t)} \sigma(a)\gamma'\mu_{4d}$. For each $j \geq 0$, let $U_j \subseteq V(T_t)$ be the set of nodes of T_t that are at distance j from the root node. Since $\sigma(a) = \sum_{b \in \Gamma(a)} 2\sigma(b)$ holds for every non-leaf node $a \in V(T_t)$, a simple recursive argument implies $\sum_{a \in U_j} 2^j\sigma(a) \leq 1$, or

equivalently $\sum_{a \in U_j} \sigma(a) \leq 2^{-j}$ for all $j \geq 0$. Since T_t is a finite tree, this implies

$$\sum_{a \in V(T_t)} \sigma(a) \leq \sum_{j \geq 0} \sum_{a \in U_j} \sigma(A) \leq \sum_{j \geq 0} 2^{-j} = 2.$$

Therefore,

$$\mathbb{E}_{\nu} \left[\omega(x; \psi)^{2d} \right] \leq \sum_{a \in V(T_t)} \sigma(a) \gamma' \mu_{4d} \leq 2^{2d+1} \gamma^{2d} \mu_{4d}. \qquad \square$$

Lemma 4.5 can be used to control sums of the form $\sum_{x \in W_i} \omega(x; \psi)^s$ that appear in Lemma 4.3. Using it, we can prove Theorem 4.2, which gives a quantifiable condition that will guarantee the existence of an embedding. It is equivalent to the following form since we defined $\theta_k = |V_k|$ when adding the artificial set W_{k+1} at the beginning of this subsection.

THEOREM 4.2. *If $\theta_i \geq 2|W_i|$ for all $i \in [k]$, then the probability that ψ does not induce an embedding of H in G is at most $2^{2d+2} \gamma^{2d} \mu_{4d} \sum_{i \in [k-1]} \frac{|W_i|}{\theta_i}$.*

Proof. Define $\lambda_i = \sum_{x \in W_i} \omega(x; \psi)^{2d}$ for each $i \in [k]$. If $\sum_{i \in [k]} \frac{\lambda_i}{\theta_i} < \frac{1}{2}$, then we have $\lambda_i < \frac{1}{2} \theta_i$ for all $i \in [k]$, and therefore by Lemma 4.3, ψ induces an embedding of H in G. Note that $\lambda_k = 0$ since $\omega(x; \psi) = 0$ for all $x \in W_k$. Thus, ψ may not induce an embedding of H in G only if $\sum_{i \in [k]} \frac{\lambda_i}{\theta_i} = \sum_{i \in [k-1]} \frac{\lambda_i}{\theta_i} \geq \frac{1}{2}$. By Markov's inequality, the probability of this event is at most $2\mathbb{E} \left[\sum_{i \in [k-1]} \frac{\lambda_i}{\theta_i} \right]$. By Lemma 4.5, for all $x \in V(H)$, we have $\mathbb{E} \left[\omega(x; \psi)^{2d} \right] \leq 2^{2d+1} \gamma^{2d} \mu_{4d}$. Therefore, $\mathbb{E} [\lambda_i] \leq |W_i| \cdot 2^{2d+1} \gamma^{2d} \mu_{4d}$ and

$$2\mathbb{E} \left[\sum_{i \in [k-1]} \frac{\lambda_i}{\theta_i} \right] \leq 2 \cdot \sum_{i \in [k-1]} 2^{2d+1} \frac{|W_i|}{\theta_i} \gamma^{2d} \mu_{4d} \leq 2^{2d+2} \gamma^{2d} \mu_{4d} \sum_{i \in [k-1]} \frac{|W_i|}{\theta_i}. \quad \square$$

5. Proof of the main theorem

In this section, we prove Theorems 1.1 and 1.2. We first prove the bipartite case, Theorem 1.2, for which a density-type embedding result holds and the proof is slightly more simple. We then prove our main theorem, Theorem 1.1, in Section 5.2.

The following lemma is a slight variant of Lemma 2.2 for nonbipartite graphs. Since we need to add further conditions to the conclusion, instead of stating the outcome of dependent random choice in terms of a particular set, we state it in terms of the expected value of the random variables of interest. We omit its proof since it follows from the proof of Lemma 2.2 after making straightforward modifications.

LEMMA 5.1. *Let d, s and t be natural numbers satisfying $t \geq s$, and let η, α be positive real numbers. Let G be a graph with two sets $V_1, V_2 \subseteq V(G)$*

satisfying $e(V_1, V_2) \geq \alpha |V_1||V_2|$. Let $\mathbf{X}$ be a t-tuple in V_1^t chosen uniformly at random, and let $\mathbf{A} = N(\mathbf{X}; V_2)$. Then for all $\theta \leq \eta \alpha^d |V_1|$,

$$\mathbb{E}[\mathbf{A}^d] \geq \alpha^{dt} |V_2|^d \quad and \quad \mathbb{E}\left[\sum_{Q \in \mathbf{A}^d} \omega_\theta(Q; V_1)^s\right] \leq |V_2|^d \cdot \eta^t \alpha^{dt}.$$

In particular, this implies the existence of a set A such that $|A| \geq \frac{1}{2}\alpha^t |V_2|$ and $\mu_s(A_2^d; V_1) \leq 2\eta^t$.

The outline of the proof in both cases, bipartite and nonbipartite, are the same. We first prepare the host graph by repeatedly applying Lemma 5.1 to find a collection of r sets $\{A_j\}_{j \in [r]}$ that have small average defect towards each other. This framework was first developed by Kostochka and Sudakov in [26] and its variants have been used in much subsequent work [15], [16], [27]. Although the statement we need easily follows from the proof of Fox and Sudakov's application [16], we cannot use their work as a blackbox since we must control higher moments of the defect function. The following proposition summarizes the main observation underlying this framework.

PROPOSITION 5.2. *Let G be a graph with sets $V_1, A_2 \subseteq V(G)$. Let $\mathbf{X}$ be a t-tuple in A_2^t chosen uniformly at random, and let $\mathbf{A}_1 = N(\mathbf{X}; V_1)$. Then*

$$\mathbb{E}\left[\mu_{s,\theta}(A_2^d; \mathbf{A}_1)\right] = \mu_{s,\theta}(A_2^{d+t}; V_1).$$

Proof. For all $Q \in A_2^d$, we have $N(Q; \mathbf{A}_1) = N(Q \cup \mathbf{X}; V_1)$. Therefore, $\omega_\theta(Q; \mathbf{A}_1)^s = \omega_\theta(Q \cup \mathbf{X}; V_1)^s$, where we consider $Q \cup \mathbf{X}$ as a $(d+t)$-tuple. Hence,

$$\mathbb{E}[\omega_\theta(Q; \mathbf{A}_1)^s] = \mathbb{E}[\omega_\theta(Q \cup \mathbf{X}; V_1)^s] = \sum_{Y \in A_2^t} \omega_\theta(Q \cup Y; V_1)^s \cdot \mathbf{P}(\mathbf{X} = Y).$$

Since $\mathbf{P}(\mathbf{X} = Y) = \frac{1}{|A_2|^t}$, it follows that

$$\mathbb{E}\left[\mu_{s,\theta}(A_2^d; \mathbf{A}_1)\right] = \mathbb{E}\left[\frac{1}{|A_2|^d} \sum_{Q \in A_2^d} \omega_\theta(Q; \mathbf{A}_1)^s\right]$$

$$= \frac{1}{|A_2|^{d+t}} \sum_{Q \in A_2^d} \sum_{Y \in A_2^t} \omega_\theta(Q \cup Y; V_1)^s.$$

The last expression equals $\mu_{s,\theta}(A_2^{d+t}; V_1)$ by definition. $\quad\square$

By repeatedly applying Proposition 5.2 following techniques developed in previous works, we can obtain a collection of r sets $\{A_j\}_{j \in [r]}$ that have small average defect towards each other. Once we are given such a collection, we use the following lemma producing sets $\{V_i^{(j)}\}$ that are needed for the embedding scheme of Section 4. Its proof follows from standard probabilistic methods but is rather technical and thus will be given separately in another section. For an r-tuple of sets $\{A_j\}_{j \in [r]}$, we define $A_{-j} = \bigcup_{j' \in [r] \setminus \{j\}} A_{j'}$ for each $j \in [r]$.

LEMMA 5.3. *Let k, d, s, r be fixed natural numbers satisfying $r \geq 2$, $s \geq 4d$, and let $\varepsilon, \varepsilon'$ be fixed positive real numbers. Let m be a natural number satisfying $m \geq 2^{200}(\varepsilon')^{-80}\varepsilon^{-80d}r^{160d}d^{40}k^{40}s^8$, and let θ be a natural number satisfying $\theta \geq \varepsilon m$. Let p_i for $i \in [k]$ be positive real numbers satisfying $\sum_{i \in [k]} p_i \leq 1$ and $p_i \geq m^{-1/(10d)}$ for all $i \in [k]$. Suppose that $\{A_j\}_{j \in [r]}$ are vertex subsets with $\varepsilon m \leq |A_j| \leq m$, satisfying $\mu_{s,\theta}(A^d_{-j}; A_j) < \frac{1}{2}$ for all $j \in [r]$. Define $\theta_i = \frac{1}{2r}p_i\theta$ for all $i \in [k]$. Then there exist sets $\{V_i^{(j)}\}_{(i,j) \in [k] \times [r]}$ satisfying the following conditions:*

(i) *For all $i \in [k]$ and $j \in [r]$, we have $V_i^{(j)} \subseteq A_j$ and $|V_i^{(j)}| \leq p_i|A_j|$.*
(ii) *For all $((i,j), (i_1, j_1), \ldots, (i_d, j_d)) \in ([k] \times [r])^{d+1}$ satisfying $j_1, \ldots, j_d \neq j$,*

$$\mu_{s,\theta_i}\left(\prod_{a \in [d]} V_{i_a}^{(j_a)}; V_i^{(j)}\right) \leq \max\left\{\varepsilon', 8r^d\varepsilon^{-d}\mu_{s,\theta}(A^d_{-j}; A_j)\right\}.$$

Moreover, if $r = 2$, then the factor $r^d\varepsilon^{-d}$ can be replaced by 1.
(iii) *For all $(i,j) \neq (i',j')$, the sets $V_i^{(j)}$ and $V_{i'}^{(j')}$ are disjoint.*

We would like to note that the improvement for $r = 2$ in item (ii) plays an important role in the proof of Theorem 5.5, as it allows us to prove a single exponential bound for bipartite graphs (as opposed to the double exponential bound for general graphs).

5.1. *Bipartite graphs.* We first find a pair of sets (A_1, A_2) for which the d-tuples in A_1^d have small average defect into A_2, and vice versa for d-tuples in A_2. This will be achieved by applying Lemma 5.1 twice. The first application will give a set A_2 such that the d-tuples in A_2^d have small average defect into $V(G)$. We then apply the lemma again by choosing a random set $\mathbf{X} \subseteq A_2$ to obtain a set A_1 whose d-tuples have small average defect into A_2. Proposition 5.2 can be used to show that this pair of sets has the claimed properties.

LEMMA 5.4. *Let m, d, s, t be natural numbers and α, η be positive real numbers satisfying $t \geq s$ and $\eta \leq \frac{1}{16}\alpha^{2d}$. Let G be a bipartite graph of minimum degree at least αm with vertex partition $V_1 \cup V_2$, where $|V_1| = |V_2| = m$. Then there exist sets $A_1 \subseteq V_1$ and $A_2 \subseteq V_2$ satisfying the following properties:*

(i) *$|A_i| \geq \frac{1}{4}\alpha^t m$ for both $i = 1, 2$;*
(ii) *for all $\theta \leq \frac{1}{2}\eta\alpha^{d+t}m$, $\mu_{s,\theta}(A_1^d; A_2) \leq 2\eta^{t/2}$ and $\mu_{s,\theta}(A_2^d; A_1) \leq 2\eta^{t/2}$.*

Proof. Since $\omega_{\theta'} \leq \omega_{\theta}$ holds for all $\theta' \leq \theta$, it suffices to consider the case when $\theta = \frac{1}{2}\eta\alpha^{d+t}m$. Throughout the proof, we will consider ω and μ with this fixed value of θ and hence will omit θ from the subscripts. Since $\theta \leq \eta\alpha^{d+t}|V_1|$, we can apply Lemma 5.1 with $d_{5.1} = d + t$ to find a set $A_2 \subseteq V_2$ of size $|A_2| \geq \frac{1}{2}\alpha^t|V_2|$ such that $\mu_s(A_2^{d+t}; V_1) \leq 2\eta^t$.

By the minimum degree condition on G, we know that the subgraph of G induced on $V_1 \cup A_2$ has density at least α. Let $\mathbf{X}$ be a t-tuple in A_2^t chosen uniformly at random, and let $\mathbf{A}_1 = N(\mathbf{X})$. Since $\theta \leq \eta \alpha^d |A_2|$, Lemma 5.1 with $(V_1)_{5.1} = A_2$ and $(V_2)_{5.1} = V_1$ implies

$$\mathbb{E}\left[|\mathbf{A}_1|^d - \frac{1}{2\eta^t} \sum_{Q \in \mathbf{A}_1^d} \omega(Q; A_2)^s \right] \geq \frac{1}{2} \alpha^{dt} |V_1|^d.$$

By Proposition 5.2, it follows that $\mathbb{E}[\mu_s(A_2^d; \mathbf{A}_1)] = \mu_s(A_2^{d+t}; V_1) \leq 2\eta^t$. Since $\eta^{t/2} \leq \frac{1}{4} \alpha^{dt}$,

$$\mathbb{E}\left[|\mathbf{A}_1|^d - \frac{1}{2\eta^t} \sum_{Q \in \mathbf{A}_1^d} \omega(Q; A_2)^s - \frac{|V_1|^d}{2\eta^{t/2}} \mu_s(A_2^d; \mathbf{A}_1) \right] \geq \frac{1}{4} \alpha^{dt} |V_1|^d.$$

Let X be a particular choice of $\mathbf{X}$ for which the random variable on the left-hand side becomes at least as large as its expected value, and let A_1 be $\mathbf{A}_1$ for this choice of $\mathbf{X}$. First, $|A_1|^d \geq \frac{1}{4} \alpha^{dt} |V_1|^d$ implies $|A_1| \geq \frac{1}{4} \alpha^t |V_1|$. Second, $|A_1|^d - \frac{1}{2\eta^t} \sum_{Q \in A_1^d} \omega(Q; A_2)^s \geq 0$ implies $\mu_s(A_1^d; A_2) \leq 2\eta^t$. Third, $|A_1|^d - \frac{|V_1|^d}{2\eta^{t/2}} \mu_s(A_2^d; A_1) \geq 0$ implies $\mu_s(A_2^d; A_1) \leq 2\eta^{t/2}$. $\square$

The proof of the bipartite case of the Burr–Erdős conjecture follows by combining Lemma 5.4 with Lemma 5.3 and then using the embedding scheme from Section 4. Even though the theorem below is stated for graphs of large minimum degree, the density-embedding theorem result follows since every graph on n vertices of density at least α contains a subgraph of minimum degree at least $\frac{1}{2} \alpha n$.

THEOREM 5.5. *There exists a positive real number c such that for every natural number d and positive real number α, the following holds for all m satisfying $m \geq \alpha^{-cd^2}$. If G is a graph on m vertices with minimum degree at least αm, then it is universal for the family of d-degenerate bipartite graphs on at most $d^{-1} 2^{-21} \alpha^{72d} m$ vertices.*

Proof. Define $n = d^{-1} 2^{-21} \alpha^{72d} m$. Let H be a d-degenerate bipartite graph on at most n vertices. By Lemma 3.1, there exists a vertex partition $V(H) = \bigcup_{i \in [k]} (W_i^{(1)} \cup W_i^{(2)})$ satisfying the following properties:

(i) $k \leq \log_2 n$;

(ii) for all $(i, j) \in [k] \times [2]$, we have $|W_i^{(j)}| \leq 2^{-i+1} n$;

(iii) both $\bigcup_{i \in [k]} W_i^{(1)}$ and $\bigcup_{i \in [k]} W_i^{(2)}$ are independent sets; and

(iv) for all $(i, j) \in [k] \times [2]$, each vertex $v \in W_i^{(j)}$ has at most $4d$ neighbors in $\bigcup_{i' \geq i} W_{i'}^{(1)} \cup W_{i'}^{(2)}$.

Define $t = s = 32d$ and $\eta = \frac{1}{128}\alpha^{36d}$. Define $\theta = \frac{1}{8}\eta\alpha^{4d+t}m$. We can use the concentration of hypergeometric distribution (see, for example, [22]) to find a partition $V_1 \cup V_2$ of $V(G)$ for which $|V_1| = |V_2| = \frac{m}{2}$ and the bipartite subgraph induced on $V_1 \cup V_2$ has minimum degree at least $2^{-1/(36d)}\alpha m$. Since $\eta \leq \frac{1}{16}(2^{-1/(36d)}\alpha)^{8d}$ and $\theta = \frac{1}{2}\eta \cdot (2^{-1/(36d)}\alpha)^{4d+t} \cdot \frac{m}{2}$, we can apply Lemma 5.4 with $d_{5.4} = 4d$ and $\alpha_{5.4} = 2^{-1/(36d)}\alpha$ to obtain sets $A \subseteq V_1$ and $B \subseteq V_2$ satisfying the following properties:

(i) $|A|, |B| \geq \frac{1}{4}2^{-t/(36d)}\alpha^t\frac{m}{2} \geq \frac{1}{16}\alpha^t m$, and
(ii) $\mu_{s,\theta}(A^{4d}; B) \leq 2\eta^{t/2}$ and $\mu_{s,\theta}(B^{4d}; A) \leq 2\eta^{t/2}$.

Define $p_i = c2^{-i/(80d)}$, where c is a positive constant defined so that $\sum_{i\in[k]} p_i = 1$. Then

$$c = \frac{1}{\sum_{i\in[k]} 2^{-i/(80d)}} \geq \frac{1}{\sum_{i=0}^{\infty} 2^{-i/(80d)}} = 1 - 2^{-1/(80d)}$$

$$\geq 1 - \left(1 - \frac{\ln 2}{80d} + \frac{(\ln 2)^2}{2(80d)^2}\right) \geq \frac{1}{160d}.$$

Thus, for all $i \in [k]$, we have $p_i \geq p_k = c2^{-k/(80d)} \geq c2^{-\log_2 n/(80d)} = cn^{-1/(80d)} \geq \frac{1}{160d}n^{-1/(80d)}$.

We can apply Lemma 5.3 to the pair of sets (A, B) with $r_{5.3} = 2$, $s_{5.3} = s$, $d_{5.3} = 4d$, $\varepsilon_{5.3} = \frac{\theta}{m} = \frac{1}{8}\eta\alpha^{4d+t}$, $(\varepsilon')_{5.3} = 16\eta^{t/2}$, and p_i as above since $s \geq 16d$, $p_i \geq (\frac{m}{2})^{-1/(40d)}$, and

$$|A|, |B| \geq \frac{1}{16}\alpha^t m \geq 2^{200}(16\eta^{t/2})^{-80}\left(\frac{1}{8}\eta\alpha^{4d+t}\right)^{-80d} r^{160d}(4d)^{40}k^{40}s^8,$$

given that c is large enough. (Recall that $m \geq \alpha^{-cd^2}$.) This provides partitions $A = \bigcup_{i\in[k]} A_i$ and $B = \bigcup_{i\in[k]} B_i$ satisfying the following conditions for $\theta_i = \frac{1}{4}p_i\theta$:

(i) for all $i \in [k]$, $|A_i| \leq p_i|A|$ and $|B_i| \leq p_i|B|$;
(ii) for all $(j, i_1, \ldots, i_{4d}) \in [k]^{4d+1}$,

$$\mu_{s,\theta_j}\left(\prod_{a\in[4d]} B_{i_a}; A_j\right) \leq \max\{16\eta^{t/2}, 8\mu_{s,\theta}(B^{4d}; A)\} \leq 16\eta^{t/2};$$

(iii) for all $(j, i_1, \ldots, i_{4d}) \in [k]^{4d+1}$,

$$\mu_{s,\theta_j}\left(\prod_{a\in[4d]} A_{i_a}; B_j\right) \leq \max\{16\eta^{t/2}, 8\mu_{s,\theta}(A^{4d}; B)\} \leq 16\eta^{t/2}.$$

We now apply the embedding scheme described in Section 4. For each $i \in [k]$, we will map $W_i^{(1)}$ to A_i and $W_i^{(2)}$ to B_i. We proceed by choosing one pair $(i, j) \in [k] \times [2]$ at a time, following the lexicographical ordering on $[k] \times [2]$ in

the reverse order. For each set $W_i^{(j)}$, its corresponding defect parameter used in the embedding scheme is $\theta_i = \frac{1}{4}p_i\theta$, and therefore

$$\gamma \leq \max_{i \in [k]} \frac{\max\{|A_i|, |B_i|\}}{\theta_i} \leq \max_{i \in [k]} \frac{p_i m/2}{(p_i\theta/4)} = \frac{2m}{\theta} \leq \frac{16}{\eta\alpha^{4d+t}}.$$

By the properties above, the maximum average s-th moment defect μ_s is at most $16\eta^{t/2}$. For each $(i, j) \in [k] \times [2]$, we have

$$\frac{|W_i^{(j)}|}{\theta_i} \leq \frac{2^{-i+1}n}{\frac{1}{4}c2^{-i/(80d)}\theta} = \frac{2^{-i+1}n}{\frac{1}{4}c2^{-i/(80d)}(\eta\alpha^{4d+t}m)/8}$$

$$\leq \frac{10240dn}{2^{i/2}\eta\alpha^{4d+t}m} \leq \frac{d2^{21}n}{2^{i/2}\alpha^{72d}m} = \frac{1}{2^{i/2}}.$$

Therefore,

$$2^{2(4d)+2} \sum_{i \in [k]} \sum_{j \in [2]} \frac{|W_i^{(j)}|}{\theta_i}\gamma^{2(4d)} \cdot \mu_{4(4d)} \leq 2^{8d+3} \sum_{i \in [k]} \frac{1}{2^{i/2}} \left(\frac{16}{\eta\alpha^{4d+t}}\right)^{8d} \cdot 16\eta^{t/2}$$

$$= 2^{8d+3} \sum_{i \in [k]} \frac{1}{2^{i/2}} \cdot 2^{32d+4}\eta^{t/2-8d}\alpha^{-8d(4d+t)}$$

$$\leq 2^{40d+9} \cdot \eta^{8d}\alpha^{-288d^2}$$

$$\leq 2^{40d+9} \cdot (\alpha^{36d}2^{-7})^{8d}\alpha^{-288d^2}$$

$$< 2^{-16d+9} < 1.$$

Hence, by Theorem 4.2, we can find a copy of H in G. $\qquad\square$

5.2. *General graphs.* For general graphs, we prove the lemma corresponding to Lemma 5.4 in two steps. In the first step we find sets $A_1, A_2, \ldots, A_r$ that have small average defect in one fixed direction.

LEMMA 5.6. *Let d, s, t, r be nonnegative integers satisfying $t \geq s$ and η be a positive real number. In every edge two-coloring of the complete graph K_n with red and blue, in the red graph or the blue graph, there exist sets $A_1 \subseteq \cdots \subseteq A_r$ satisfying the following conditions*:

(i) $|A_j| \geq 2^{-2(t+1)(r-1)}n$ *for all $j \in [r]$; and*

(ii) *for all $\theta \leq \eta2^{-d-2(t+1)(r-1)}n$, $\mu_{s,\theta}(A_j^d; A_{j+1}) \leq 2\eta^t$ for all $j \leq r - 1$.*

Proof. Define $A_{2(r-1)} = V(G)$, and arbitrarily color it with one of the colors. For $i \in [2(r-1)]$, suppose that we constructed a set A_i of size at least $2^{-(t+1)(2r-2-i)}n$. There exists a color, say c_i, of density at least $\frac{1}{2}$ in the set A_i. Since $\theta \leq \eta2^{-d}|A_i|$, we may apply Lemma 5.1 to the subgraph induced on A_i with the edges of color c_i to find a set $A_{i-1} \subseteq A_i$ of size $|A_{i-1}| \geq 2^{-t-1}|A_i|$ such that $\mu_{s,\theta}(A_{i-1}^d; A_i) \leq 2\eta^t$ in the graph consisting of the edges of color c_i. Color the set A_{i-1} with the color that we used.

Repeat the process to find sets $A_{2r-2} \supseteq A_{2r-3} \supseteq \cdots \supseteq A_0$. Note that $|A_0| \geq 2^{-(t+1)(2r-2)}n$. By the pigeonhole principle, we can find r indices $i_1 < \cdots < i_r$ for which A_{i_j} are all colored by the same color, say red. These sets satisfy Property (i). Since $\mu_{s,\theta}(Q; X) \leq \mu_{s,\theta}(Q; Y)$ holds for all sets $X \supseteq Y$, we have $\mu_{s,\theta}(A^d_{i_{a-1}}; A_{i_a}) \leq \mu_{s,\theta}(A^d_{i_{a-1}}; A_{i_{a-1}+1}) \leq 2\eta^t$ in the red graph for all $a \in [2, r]$. Thus, the sets $A_{i_1}, A_{i_2}, \ldots, A_{i_r}$ satisfy the claimed properties. □

Given the sets $A_1, \ldots, A_r$ constructed in the previous lemma, we run r more rounds of dependent random choice to produce an r-tuple of sets that have small average defect towards each other. At the i-th round, we will choose a random t_i-tuple $\mathbf{X}_i \in A_i^{t_i}$ and update each set A_j for $j \neq i$ to $N(\mathbf{X}) \cap A_j$. This will ensure that the average defect of A^d_{-i} into A_i is small. The next lemma shows that all the conditions are maintained throughout this process.

LEMMA 5.7. *Let $d, r, s,$ and t be nonnegative integers satisfying $t \geq s$. Let ξ and θ be positive real numbers satisfying $\xi \leq 2^{-20(d+t)\cdot 8^{r+1}r}$ and $\theta \leq \xi^2 n$. In every edge two-coloring of the complete graph K_n with red and blue, in the red graph or the blue graph, there exist sets A_j for $j \in [r]$ satisfying the following properties*:

(i) *$|A_j| \geq \theta$ for all $j \in [r]$; and*
(ii) *$\mu_{s,\theta}(A^d_{-j}; A_j) \leq \xi^t$ for all $j \in [r]$.*

Proof. For $i = 0, 1, \ldots, r$, define $t_i = 8^{r+1-i}(d+t)$ and $d_i = d + \sum_{j=i+1}^{r} t_j$. Note that $\frac{1}{6} t_i \geq d_i \geq t_{i+1}$ for all $i \in [r-1]$. Define $\xi_0 = 2^{-20t_0 r}$, and note that $\xi \leq \xi_0$. Define $\theta_0 = \xi n$, $\theta_1 = \xi \theta_0 = \xi^2 n$, and note that it suffices to prove the statement for $\theta = \theta_1$.

We may apply Lemma 5.6 with $d_{5.6} = d_0$, $s_{5.6} = 0$, $t_{5.6} = t_0$, $r_{5.6} = r$, and $\eta_{5.6} = 2^{-16t_0 r}$ since $\theta_1 \leq \eta 2^{-d_0 - (t_0+1)(2r-2)}n$. This gives sets $B_1 \subseteq \cdots \subseteq B_r$ in, without loss of generality, the red graph, satisfying $|B_j| \geq 2^{-4t_0 r}n$ for all $j \in [r]$, and $\mu_{0,\theta_0}(B_j^{d_0}; B_{j+1}) \leq 2(2^{-16t_0 r})^{t_0} \leq \xi^{t_0/2}$ for all $j \in [r-1]$. Let $A_{0,j} = B_j$ for all $j \in [r]$. For each $i = 0, 1, \ldots, r$, we will iteratively construct sets $\{A_{i,j}\}_{j \in [r]}$ satisfying the following properties:

(a) $|A_{i,j}| \geq \theta_0$ for all $j \in [i, r]$;
(b) $|A_{i,j}| \geq \theta_1$ for all $j < i$;
(c) $\mu_{0,\theta_0}(A_{i,j}^{d_i}; A_{i,j+1}) \leq \xi^{t_i/2}$ for all $j \in [i, r-1]$;
(d) $\mu_{s,\theta_1}(A_{i,-j}^{d_i}; A_{i,j}) \leq \xi^{t_i/2}$ for all $j \leq i$;
(e) $A_{i,j} \subseteq A_{i,j+1}$ for all $j > i$.

Note that the properties hold for $i = 0$. (The only relevant properties are (a),(c) and (e).) In the end, the sets $A_j = A_{r,j}$ for $j \in [r]$ satisfy Properties (i) and (ii) by Properties (b) and (d) since $d_r = d$, $\theta_1 = \xi^2 n$ and $t_r = 8(d+t) \geq 2t$.

Suppose that for some $i \geq 1$, we have constructed sets $\{A_{i-1,j}\}_{j \in [r]}$ for which the properties hold. For simplicity, we abuse notation and write $B_j = $

$A_{i-1,j}$ for all $j \in [r]$. Let $\mathbf{X} \in B_i^{t_i}$ be a t_i-tuple chosen uniformly at random. For each $j \neq i$, define $\mathbf{A}_j = B_j \cap N(\mathbf{X})$ and define $\mathbf{A}_i = B_i$. There are several events that we consider.

Event 1. $|\mathbf{A}_j| \geq \theta_0$ for all $j \geq i$ and $|\mathbf{A}_j| \geq \theta_1$ for all $j < i$.

Note that the claim trivially holds for $\mathbf{A}_i$ since $|\mathbf{A}_i| = |B_i| \geq \theta_0$. For $j > i$, since $B_i \subseteq B_{j-1}$,

$$\mathbf{P}\left(|\mathbf{A}_j| < \theta_0\right) = \mu_{0,\theta_0}(B_i^{t_i}; B_j) \leq \left(\frac{|B_{j-1}|^{t_i}}{|B_i|^{t_i}}\right) \mu_{0,\theta_0}(B_{j-1}^{t_i}; B_j)$$

$$\leq \left(\frac{n}{\xi n}\right)^{t_i} \mu_{0,\theta_0}(B_{j-1}^{t_i}; B_j) = \xi^{-t_i}\mu_{0,\theta_0}(B_{j-1}^{t_i}; B_j).$$

Since $d_{i-1} \geq t_i$, by Proposition 3.4 and Property (c), we have $\mu_{0,\theta_0}(B_{j-1}^{t_i}; B_j) \leq \mu_{0,\theta_0}(B_{j-1}^{d_{i-1}}; B_j) \leq \xi^{t_{i-1}/2}$. Hence, the probability that $|\mathbf{A}_j| < \theta_0$ is at most $\xi^{-t_i+t_{i-1}/2} \leq \xi^{t_{i-1}/4}$. Similarly, for $j < i$, since $B_i \subseteq B_{-j}$,

$$\mathbf{P}\left(|\mathbf{A}_j| < \theta_1\right) = \mu_{0,\theta_1}(B_i^{t_i}; B_j) \leq \left(\frac{|B_{-j}|^{t_i}}{|B_i|^{t_i}}\right) \mu_{0,\theta_1}(B_{-j}^{t_i}; B_j)$$

$$\leq \left(\frac{n}{\xi^2 n}\right)^{t_i} \mu_{0,\theta_1}(B_{-j}^{t_i}; B_j)$$

$$\leq \xi^{-2t_i}\mu_{s,\theta_1}(B_{-j}^{d_{i-1}}; B_j).$$

By Property (d), this is at most $\xi^{-2t_i+t_{i-1}/2} \leq \xi^{t_{i-1}/4}$. Since there are r total events, the probability of Event 1 is at least $1 - r\xi^{t_{i-1}/4} > \frac{3}{4}$.

Event 2. $\mu_{0,\theta_0}(B_j^{d_i}; \mathbf{A}_{j+1}) \leq 4r\xi^{t_{i-1}/4}$ for all $j \in [i, r-1]$.

For $j \in [i, r-1]$, since $B_i \subseteq B_j$, we have

$$\mathbb{E}[\mu_{0,\theta_0}(B_j^{d_i}; \mathbf{A}_{j+1})] = \frac{1}{|B_i|^{t_i}} \sum_{Q \in B_i^{t_i}} \mu_{0,\theta_0}(B_j^{d_i}; N(Q) \cap B_{j+1})$$

$$\leq \left(\frac{|B_j|}{|B_i|}\right)^{t_i} \frac{1}{|B_j|^{t_i}} \sum_{Q \in B_j^{t_i}} \mu_{0,\theta_0}(B_j^{d_i}; N(Q) \cap B_{j+1})$$

$$\leq \xi^{-2t_i}\mathbb{E}_{\nu}[\mu_{0,\theta_0}(B_j^{d_i}; \mathbf{A}_{j+1})],$$

where ν is the probability distribution obtained by choosing $\mathbf{X}$ uniformly in the set $B_j^{t_i}$ instead of $B_i^{t_i}$. By Proposition 5.2 and Property (c),

$$\mathbb{E}_{\nu}[\mu_{0,\theta_0}(B_j^{d_i}; \mathbf{A}_{j+1})] \leq \mu_{0,\theta_0}(B_j^{d_{i-1}}; B_{j+1}) \leq \xi^{t_{i-1}/2}.$$

Since $\xi^{-2t_i} \cdot \xi^{t_{i-1}/2} \leq \xi^{t_{i-1}/4}$, by Markov's inequality, the probability of Event 2 for a fixed index j is less than $1 - \frac{1}{4r}$. Since there are at most r choices for the index j, the probability of Event 2 is greater than $\frac{3}{4}$.

Event 3. $\mu_{s,\theta_1}(B_{-j}^{d_i}; \mathbf{A}_j) \le 4r\xi^{t_{i-1}/4}$ for all $j < i$.

As in Event 2, we have $\mathbb{E}[\mu_{s,\theta_1}(B_{-j}^{d_i}; \mathbf{A}_j)] \le \xi^{-2t_i}\mu_{s,\theta_1}(B_{-j}^{d_{i-1}}; B_j) \le \xi^{t_{i-1}/4}$ (by Property (d)). Therefore, by Markov's inequality, the probability of Event 3 for a fixed j is less than $1 - \frac{1}{4r}$. Since there are at most r choices for j, the probability of Event 3 is greater than $\frac{3}{4}$.

Event 4. $\sum_{Q \in \mathbf{A}_{-i}^{d_i}} \omega_{\theta_1}(Q; B_i)^s \le 4|B_{-i}|^{d_i}\xi^{t_i}$.

Note that

$$\mathbb{E}\left[\sum_{Q \in \mathbf{A}_{-i}^{d_i}} \omega_{\theta_1}(Q; B_i)^s\right] \le \sum_{Q \in B_{-i}^{d_i}} \omega_{\theta_1}(Q; B_i)^s \cdot \mathbf{P}\left(Q \in \mathbf{A}_{-i}^{d_i}\right).$$

Since $\theta_1 = \xi\theta_0 \le \xi|B_i|$, if $\omega_{\theta_1}(Q; B_i) \ne 0$, then $|N(Q; B_i)| = \frac{\theta_1}{\omega_{\theta_1}(Q;B_i)} \le \frac{\xi|B_i|}{\omega_{\theta_1}(Q;B_i)}$ and thus $\mathbf{P}(Q \in \mathbf{A}_{-i}^{d_i}) = \left(\frac{|N(Q;B_i)|}{|B_i|}\right)^{t_i} \le \left(\frac{\xi}{\omega_{\theta_1}(Q;B_i)}\right)^{t_i} \le \frac{\xi^{t_i}}{\omega_{\theta_1}(Q;B_i)^s}$. Thus, $\mathbb{E}\left[\sum_{Q \in \mathbf{A}_{-i}^{d_i}} \omega_{\theta_1}(Q; B_i)^s\right] \le |B_{-i}|^{d_i}\xi^{t_i}$. Hence, with probability greater than $\frac{3}{4}$, we have $\sum_{Q \in \mathbf{A}_{-i}^{d_i}} \omega_{\theta_1}(Q; B_i)^s \le 4|B_{-i}|^{d_i}\xi^{t_i}$.

Therefore, with positive probability all four events hold. Let X be a particular choice of $\mathbf{X}$ for which all four events hold, and define $A_{i,j} = \mathbf{A}_j$ for this choice of X. Event 1 immediately implies Properties (a) and (b). Since $|A_{i,j}| \ge \theta_1 \ge \xi^2|B_j|$ holds for all $j \in [r]$, it follows by Event 2 that for all $j \in [i, r-1]$,

$$\mu_{0,\theta_0}(A_{i,j}^{d_i}; A_{i,j+1}) \le \xi^{-2d_i}\mu_{0,\theta_0}(B_j^{d_i}; A_{i,j+1}) \le \xi^{-2d_i} \cdot 4r\xi^{t_{i-1}/4} \le \xi^{t_i/2},$$

implying Property (c). Similarly, Event 3 implies $\mu_{s,\theta_1}(A_{i,-j}^{d_i}; A_j) \le \xi^{t_i/2}$ for all $j < i$. Finally, Event 4 implies

$$\frac{1}{|A_{i,-i}|^{d_i}} \sum_{Q \in A_{i,-i}^{d_i}} \omega_{\theta_1}(Q; A_{i,i})^s \le \frac{\xi^{-2d_i}}{|B_{-i}|^{d_i}} \sum_{Q \in B_{-i}^{d_i}} \omega_{\theta_1}(Q; B_i)^s \le \xi^{-2d_i} \cdot 4\xi^{t_i} \le \xi^{t_i/2},$$

proving Property (d). Property (e) follows from the same property the previous step and the definition $A_{i,j} = N(X) \cap A_{i-1,j}$ for $j \ne i$. $\square$

We conclude this section with the proof of our main theorem, Theorem 1.1.

THEOREM 1.1. *There exists a constant c such that the following holds for every natural number d, r, and n satisfying $n \ge 2^{d^2 2^{cr}}$. For every edge two-coloring of the complete graph on at least $2^{d2^{cr}}n$ vertices, one of the colors is universal for the family of d-degenerate r-colorable graphs on at most n vertices.*

Proof. Define $m = 2^{d2^{cr}}n$ for a large enough constant c. Suppose that we are given an edge coloring of K_m with two colors red and blue. Define

$t = s = 32d$ and $\xi = 2^{-d2^{(c/2)r}}$. Define $\theta = \xi^2 m$. Lemma 5.7 with $d_{5.7} = 4d$ implies that in the red graph or the blue graph, there exist sets $\{A_j\}_{j \in [r]}$ satisfying the following properties:

(i) $|A_j| \geq \theta$ for all $j \in [r]$; and

(ii) $\mu_{s,\theta}(A_{-j}^{4d}; A_j) \leq \xi^t$.

Define G as the subgraph of K_m consisting of the edges of the color realizing the two properties above.

Let H be a d-degenerate r-colorable graph on at most n vertices. By Lemma 3.1, there exists a vertex partition $V(H) = \bigcup_{(i,j) \in [k] \times [r]} W_i^{(j)}$ satisfying the following properties:

(i) $k \leq \log_2 n$;

(ii) for all $(i, j) \in [k] \times [r]$, we have $|W_i^{(j)}| \leq 2^{-i+1} n$;

(iii) for all $j \in [r]$, the set $\bigcup_{i \in [k]} W_i^{(j)}$ is an independent set; and

(iv) for all $(i, j) \in [k] \times [r]$, each vertex $v \in W_i^{(j)}$ has at most $4d$ neighbors in $\bigcup_{i' \geq i, j' \in [r]} W_{i'}^{(j')}$.

Define $p_i = c' 2^{-i/(80d)}$, where c' is a positive constant defined so that $\sum_{i \in [k]} p_i = 1$. Then

$$c' = \frac{1}{\sum_{i \in [k]} 2^{-i/(80d)}} \geq \frac{1}{\sum_{i=0}^{\infty} 2^{-i/(80d)}}$$

$$= 1 - 2^{-1/(80d)} \geq 1 - \left(1 - \frac{\ln 2}{80d} + \frac{(\ln 2)^2}{2(80d)^2}\right) \geq \frac{1}{160d}.$$

Thus, for all $i \in [k]$, we have $p_i \geq p_k = c' 2^{-k/(80d)} \geq c' 2^{-\log_2 n/(80d)} = c' n^{-1/(80d)} \geq \frac{1}{160d} n^{-1/(80d)}$. Since $p_i \geq n^{-1/(40d)}$, $s \geq 16d$, and $m > n \geq 2^{d^2 2^{cr}}$, if c is large enough, then we can apply Lemma 5.3 to G with $d_{5.3} = 4d$, $\varepsilon_{5.3} = \xi^2$, $(\varepsilon')_{5.3} = \xi^{t/2}$ and sets $\{A_j\}_{j \in [r]}$, to obtain sets $\{V_i^{(j)}\}_{(i,j) \in [k] \times [r]}$ satisfying the following conditions for $\theta_i = \frac{1}{2r} p_i \theta$:

(i) for all $(i, j) \in [k] \times [r]$, $|V_i^{(j)}| \leq p_i |A_j|$; and

(ii) for all $((i, j), (i_1, j_1), \ldots, (i_{4d}, j_{4d})) \in ([k] \times [r])^{4d+1}$ satisfying $j_1, \ldots, j_{4d} \neq j$,

$$\mu_{s,\theta_i}\left(\prod_{a \in [4d]} V_{i_a}^{(j_a)}; V_i^{(j)}\right) \leq \max\left\{\xi^{t/2}, 8r^{4d}\xi^{-4d}\mu_{s,\theta}(A_{-j}^{4d}; A_j)\right\} \leq \xi^{t/2}.$$

We now apply the embedding scheme described in Section 4. We choose one pair $(i, j) \in [k] \times [r]$ at a time, following the lexicographical ordering of $[k] \times [r]$ in the reverse order, and map $W_i^{(j)}$ to $V_i^{(j)}$. For each set $W_i^{(j)}$, its corresponding defect parameter used in the embedding scheme is $\theta_i = \frac{1}{2r} p_i \theta$, and therefore,

$$\gamma \leq \max_{i \in [k]} \frac{\max_{j \in [r]} |V_i^{(j)}|}{\theta_i} \leq \max_{i \in [k]} \frac{p_i m}{p_i \xi^2 m / 2r} = \frac{2r}{\xi^2}.$$

Moreover, by Property (ii) above, the maximum average s-th moment defect satisfies $\mu_s \leq \xi^{t/2}$. For all $(i, j) \in [k] \times [r]$, since $m \geq 2^{d2cr} n \geq 640 r d \xi^{-3} n$ for large enough c,

$$\frac{|W_i^{(j)}|}{\theta_i} = \frac{2^{-i+1}n}{\frac{1}{2r}c'2^{-i/(80d)}\theta} \leq \frac{640 r d n}{2^{i/2} \cdot \xi^2 m} \leq \frac{1}{2^{i/2}} \cdot \xi.$$

Therefore, for large enough c,

$$2^{8d+2}\gamma^{8d}\mu_{16d} \sum_{i \in [k]} \sum_{j \in [r]} \frac{|W_i^{(j)}|}{\theta_i} \leq 2^{8d+2} \left(\frac{2r}{\xi^2}\right)^{8d} \xi^{t/2} \sum_{i \in [k]} \frac{r\xi}{2^{i/2}}$$
$$\leq 2^{16d+4} r^{8d+1} \xi < 1.$$

Hence, by Theorem 4.2, we can find a monochromatic copy of H. $\square$

6. Random pruning

In this section, we prove Lemma 5.3, the final ingredient of the proof. One of our main tools is the following concentration inequality (see [29, Th. 3.1]).

THEOREM 6.1. *Let* $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ *be a family of independent random variables with* X_i *taking values in a set* Ω_i *for each* i. *Suppose that the real-valued function* f *defined on* $\prod_{i \in [n]} \Omega_i$ *satisfies* $|f(\vec{x}) - f(\vec{y})| \leq c_i$ *whenever the vectors* $\vec{x}$ *and* $\vec{y}$ *differ only in the* i-*th coordinate. Then*

$$\mathbf{P}\left(\left|f(\mathbf{X}) - \mathbb{E}[f(\mathbf{X})]\right| \geq t\right) \leq 2e^{-2t^2/\sum_{i \in [n]} c_i^2}.$$

The next lemma shows that given the collection of sets obtained in Lemmas 5.4 and 5.6, we can further impose that the defect is not concentrated too much on individual vertices. This additional condition will help us later when taking a random partition.

LEMMA 6.2. *Let* ε *be a fixed positive real number, and let* d, s, r *be fixed natural numbers satisfying* $s \geq 4d$. *Let* m *be a natural number satisfying* $m \geq 2^{24}s^8d^8r^{18}\varepsilon^{-8}$, *and let* θ *be a natural number satisfying* $\theta \geq \varepsilon m$. *Let* $A_1, A_2, \ldots, A_r$ *be (not necessarily disjoint) vertex subsets satisfying* $\varepsilon m \leq |A_i| \leq m$ *and* $\mu_{s,\theta}(A_{-i}^d; A_i) < 1$ *for all* $i \in [r]$. *Then there exist subsets* $B_i \subseteq A_i$ *for all* $i \in [r]$ *satisfying the following properties*:

(i) $|B_i| \geq 2^{-1/(2d)}|A_i|$ *for all* $i \in [r]$;
(ii) $\mu_{s,\theta}(B_{-i}^d; B_i) \leq 2\mu_{s,\theta}(A_{-i}^d; A_i)$ *for all* $i \in [r]$; *and*
(iii) *for every* $i \in [r]$ *and* $v \in B_{-i}$, $\sum_{Q:v \in Q \in B_{-i}^d} \omega_\theta(Q; B_i)^s \leq 2|B_{-i}|^{d-5/8}$.

Proof. We will fix θ throughout the proof and thus for simplicity will use the notation $\omega_\theta = \omega$ and $\mu_{s,\theta} = \mu_s$. Let $R_i \subseteq A_{-i}$ be the set of vertices $v \in A_{-i}$

such that $\sum_{Q:v\in Q\in A^d_{-i}} \omega(Q; A_i)^s \geq |A_{-i}|^{d-5/8}$. Then

$$|R_i| \cdot |A_{-i}|^{d-5/8} \leq \sum_{v\in R_i} \sum_{Q:v\in Q\in A^d_{-i}} \omega(Q; A_i)^s \leq d \cdot |A_{-i}|^d \mu_s(A^d_{-i}; A_i) < d|A_{-i}|^d.$$

Therefore, $|R_i| < d|A_{-i}|^{5/8} \leq drm^{5/8}$. For each $i \in [r]$, let B_i be the set obtained from A_i by removing the vertices in $\bigcup_{j\in[r]} R_j$. Then $|B_i| \geq |A_i| - dr^2m^{5/8} \geq 2^{-1/(2d)}|A_i|$ and (i) holds.

Fix $i \in [r]$. Since $s \geq 4d$, by Proposition 3.6, all d-tuples in A^d_{-i} have at least $\frac{\theta}{|A_{-i}|^{1/4}} \geq \frac{\varepsilon m}{(rm)^{1/4}} \geq \varepsilon r^{-1/4}m^{3/4}$ common neighbors in A_i. Since B_i is obtained from A_i by removing at most $dr^2m^{5/8}$ vertices, all d-tuples $Q \in B^d_{-i}$ satisfy

$$|N(Q; B_i)| \geq \left(1 - \frac{dr^2m^{5/8}}{|N(Q; A_i)|}\right)|N(Q; A_i)| \geq \left(1 - \frac{dr^2m^{5/8}}{\varepsilon r^{-1/4}m^{3/4}}\right)|N(Q; A_i)|$$

$$\geq \left(1 - \frac{dr^{9/4}}{\varepsilon m^{1/8}}\right)|N(Q; A_i)| \geq 2^{-1/(2s)}|N(Q; A_i)|.$$

Hence, $\omega(Q; B_i) \leq 2^{1/(2s)}\omega(Q; A_i)$. Therefore,

$$\mu_s(B^d_{-i}; B_i) = \frac{1}{|B_{-i}|^d} \sum_{Q\in B^d_{-i}} \omega(Q; B_i)^s \leq \frac{1}{|B_{-i}|^d} \sum_{Q\in A^d_{-i}} 2^{1/2}\omega(Q; A_i)^s$$

$$= 2^{1/2}\frac{|A_{-i}|^d}{|B_{-i}|^d}\left(\frac{1}{|A_{-i}|^d} \sum_{Q\in A^d_{-i}} \omega(Q; A_i)^s\right) \leq 2\mu_s(A^d_{-i}; A_i).$$

This proves (ii). Moreover, since $B_i = A_i \setminus R_i$, each vertex $v \in B_i$ satisfies $\sum_{Q:v\in Q\in A^d_{-i}} \omega(Q; A_i)^s < |A_{-i}|^{d-5/8}$. Therefore,

$$\sum_{Q:v\in Q\in B^d_{-i}} \omega(Q; B_i)^s \leq \sum_{Q:v\in Q\in B^d_{-i}} 2^{1/2}\omega(Q; A_i)^s \leq 2^{1/2}|A_{-i}|^{d-5/8} \leq 2|B_{-i}|^{d-5/8},$$

proving (iii). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Lemma 6.2 prepares an r-tuple of sets with small average defect towards each other by further imposing that the defect is not concentrated too much on individual vertices. We now prove Lemma 5.3 by taking a random partition and showing that the defect is well-distributed.

LEMMA 5.3. *Let k, d, s, r be fixed natural numbers satisfying $r \geq 2$, $s \geq 4d$, and let $\varepsilon, \varepsilon'$ be fixed positive real numbers. Let m be a natural number satisfying $m \geq 2^{200}(\varepsilon')^{-80}\varepsilon^{-80d}r^{160d}d^{40}k^{40}s^8$, and let θ be a natural number satisfying $\theta \geq \varepsilon m$. Let p_i for $i \in [k]$ be positive real numbers satisfying $\sum_{i\in[k]} p_i \leq 1$ and $p_i \geq m^{-1/(10d)}$ for all $i \in [k]$. Suppose that $\{A_j\}_{j\in[r]}$ are vertex subsets with $\varepsilon m \leq |A_j| \leq m$, satisfying $\mu_{s,\theta}(A^d_{-j}; A_j) < \frac{1}{2}$ for all $j \in [r]$. Define $\theta_i = \frac{1}{2r}p_i\theta$*

for all $i \in [k]$. Then there exist sets $\{V_i^{(j)}\}_{(i,j) \in [k] \times [r]}$ satisfying the following conditions:

(i) *For all $i \in [k]$ and $j \in [r]$, we have $V_i^{(j)} \subseteq A_j$ and $|V_i^{(j)}| \leq p_i|A_j|$.*

(ii) *For all $((i,j),(i_1,j_1),\ldots,(i_d,j_d)) \in ([k] \times [r])^{d+1}$ satisfying $j_1,\ldots,j_d \neq j$,*

$$\mu_{s,\theta_i}\left(\prod_{a \in [d]} V_{i_a}^{(j_a)}; V_i^{(j)}\right) \leq \max\left\{\varepsilon', 8r^d\varepsilon^{-d}\mu_{s,\theta}(A_{-j}^d; A_j)\right\}.$$

Moreover, if $r = 2$, then the factor $r^d\varepsilon^{-d}$ can be replaced by 1.

(iii) *For all $(i,j) \neq (i',j')$, the sets $V_i^{(j)}$ and $V_{i'}^{(j')}$ are disjoint.*

Proof. Note that the given condition on $\{p_i\}_{i \in [k]}$ implies $k \leq m^{1/(10d)}$. By Lemma 6.2, we can find subsets $B_j \subseteq A_j$ for $j \in [r]$ satisfying the following conditions:

(a) $|B_j| \geq 2^{-1/(2d)}|A_j|$ for all $j \in [r]$;

(b) $\mu_{s,\theta}(B_{-j}^d; B_j) \leq 2\mu_{s,\theta}(A_{-j}^d; A_j) < 1$ for all $j \in [r]$; and

(c) for every $j \in [r]$ and $v \in B_{-j}$, we have

$$\sum_{Q:v \in Q \in B_{-j}^d} \omega_\theta(Q; B_j)^s \leq 2|B_{-j}|^{d-5/8}.$$

Define $q_i = \frac{1}{r}p_i$ for $i \in [k]$. Color the vertices with $[k] \times [r]$ randomly where each vertex receives color (i,j) with probability q_i and the outcome for each vertex is independent. For each $(i,j) \in [k] \times [r]$, let $V_i^{(j)} \subseteq B_j$ be the set of vertices of color (i,j). Let E_1 be the event that for all $(i,j) \in [k] \times [r]$ and all d-tuples $Q \in B_{-j}^d$, we have $|N(Q; V_i^{(j)})| \geq \frac{1}{2}q_i|N(Q; B_j)|$. Let E_2 be the event that for all $((i,j),(i_1,j_1),\ldots,(i_d,j_d)) \in ([k] \times [r])^{d+1}$ satisfying $j_1,\ldots,j_d \neq j$,

$$\mu_{s,\theta}\left(\prod_{a \in [d]} V_{i_a}^{(j_a)}; B_j\right) \leq \max\left\{\varepsilon', 4r^d\varepsilon^{-d}\mu_{s,\theta}\left(B_{-j}^d; B_j\right)\right\},$$

where if $r = 2$, then we replace the factor $r^d\varepsilon^{-d}$ by 1. Let E_3 be the event that for all $(i,j) \in [k] \times [r]$, we have $\frac{1}{2^{1/(2d)}}q_i|B_j| \leq |V_i^{(j)}| \leq 2q_i|B_j|$.

We first show that Properties (i), (ii), and (iii) follow conditioned on the events E_1, E_2 and E_3. Property (iii) immediately follows from the definition of the sets. Property (i) follows from E_3 and $|B_j| \leq |A_j|$ for $j \in [r]$. For some $j \in [r]$, fix a d-tuple $Q \in B_{-j}^d$. Since $|N(Q; V_i^{(j)})| \geq \frac{1}{2}q_i|N(Q; B_j)|$ holds by E_1, if $|N(Q; B_j)| \geq \theta$, then we have $|N(Q; V_i^{(j)})| \geq \frac{1}{2}q_i\theta$. Therefore, if $\omega_\theta(Q; B_j) = 0$, then $\omega_{\theta_i}(Q; V_i^{(j)}) = 0$. Otherwise, if $\omega_\theta(Q; B_j) \neq 0$, then $\omega_\theta(Q; B_j) = \frac{\theta}{|N(Q;B_j)|} \geq \frac{q_i\theta}{2|N(Q;V_i^{(j)})|} = \omega_{\theta_i}(Q; V_i^{(j)})$. Therefore, $\omega_{\theta_i}(Q; V_i^{(j)}) \leq \omega_\theta(Q; B_j)$ in both cases. This implies that for all $((i,j),(i_1,j_1),\ldots,(i_d,j_d)) \in$

$([k] \times [r])^{d+1}$, we have by E_2,

$$\mu_{s,\theta_i}\left(\prod_{a \in [d]} V_{i_a}^{(j_a)}; V_i^{(j)}\right) \leq \mu_{s,\theta}\left(\prod_{a \in [d]} V_{i_a}^{(j_a)}; B_j\right)$$

$$\leq \max\left\{\varepsilon', 4r^d\varepsilon^{-d}\mu_{s,\theta}\left(B_{-j}^d; B_j\right)\right\}.$$

Therefore, by (b), we have Property (ii). (Similarly Property (ii) holds for $r = 2$ as well.)

It thus suffices to show that the intersection of E_1, E_2, E_3 has nonzero probability. To compute the probability of E_1, note that for $(i,j) \in [k] \times [r]$, since $s \geq 4d$ and $\mu_{s,\theta}(B_{-j}; B_j) < 1$, by Proposition 3.6, all d-tuples $Q \in B_{-j}^d$ have at least $\frac{\theta}{|B_{-j}|^{1/4}} \geq \frac{\varepsilon m}{(rm)^{1/4}} = \varepsilon r^{-1/4}m^{3/4}$ common neighbors in B_j. Hence, for a fixed d-tuple $Q \in B_{-j}^d$,

$$\mathbb{E}\left[|N(Q; V_i^{(j)})|\right] = q_i|N(Q; B_j)| \geq q_i\varepsilon r^{-1/4}m^{3/4}.$$

Therefore, by Theorem 6.1, the probability that $|N(Q; V_i^{(j)})| < \frac{1}{2}q_i|N(Q; B_i)|$ for a fixed $Q \in B_{-j}^d$ and $(i,j) \in [k] \times [r]$ is at most $2e^{-2(q_i\varepsilon r^{-1/4}m^{3/4})^2/m} = 2e^{-2q_i^2\varepsilon^2 r^{-1/2}m^{1/2}}$. Since there are at most $(rm)^d$ choices for $Q \in B_{-j}^d$, there are at most $rk \cdot (rm)^d$ such events. By the union bound,

$$\mathbf{P}(E_1^c) \leq r^{d+1}km^d \cdot 2e^{-2q_i^2\varepsilon^2 r^{-1/2}m^{1/2}} \leq r^{d+1}km^d \cdot 2e^{-2m^{1/4}} < \frac{1}{4}.$$

Note that by Theorem 6.1,

$$\mathbf{P}\left(2^{-1/(2d)}q_i|B_j| \leq |V_i^{(j)}| \leq 2q_i|B_j|\right) \geq 1 - 2e^{-2(1-2^{-1/(2d)})^2q_i^2|B_j|}$$

$$\geq 1 - 2e^{-\varepsilon/(64d^2r^2)m^{1/2}} > 1 - \frac{1}{3(kr)^{d+1}}.$$

There are $(kr)^{d+1}$ choices for $((i,j),(i_1,j_1),\ldots,(i_d,j_d)) \in ([k] \times [r])^{d+1}$, and thus $P(E_3) > \frac{3}{4}$.

To compute the probability of E_2, fix

$$((i,j),(i_1,j_1),\ldots,(i_d,j_d)) \in ([k] \times [r])^{d+1}.$$

To simplify notation, define $\mathcal{Q} = \prod_{a \in [d]} V_{i_a}^{(j_a)}$ and $\mathcal{B} = \prod_{a \in [d]} B_{j_a}$. Note that

$$(4) \qquad \mathbb{E}\left[\sum_{Q \in \mathcal{Q}} \omega_\theta(Q; B_j)^s\right] = \sum_{Q \in \mathcal{B}} \omega_\theta(Q; B_j)^s \cdot \mathbf{P}(Q \in \mathcal{Q}).$$

Fix $Q \in \mathcal{B}$. If all vertices in Q are distinct, then $\mathbf{P}(Q \in \mathcal{Q}) = \prod_{a \in [d]} q_{i_a}$. Otherwise, if $\partial\mathcal{B}$ is the set of d-tuples in $\mathcal{B}$ where not all vertices are distinct, then by Proposition 3.5, we have $\sum_{Q \in \partial\mathcal{B}} \omega_\theta(Q; B_j)^s \leq \frac{d(d-1)}{2\varepsilon m}\sum_{Q \in \mathcal{B}} \omega_\theta(Q; B_j)^s \leq$

$\frac{d(d-1)}{2\varepsilon m} \sum_{Q \in B_{-j}^d} \omega_\theta(Q; B_j)^s < \varepsilon^{-1} d^2 r^d m^{d-1}$, by Property (b) and $|B_{-j}| \leq rm$. Thus, in (4),

$$\mathbb{E}\left[\sum_{Q \in \mathcal{Q}} \omega_\theta(Q; B_j)^s\right] \leq \varepsilon^{-1} d^2 r^d m^{d-1} + \left(\prod_{a \in [d]} q_{i_a}\right) \cdot \sum_{Q \in \mathcal{B}} \omega_\theta(Q; B_j)^s$$

$$= \varepsilon^{-1} d^2 r^d m^{d-1} + \left(\prod_{a \in [d]} q_{i_a}\right) |\mathcal{B}| \mu_{s,\theta}(\mathcal{B}; B_j).$$

For each vertex $v \in B_{-j}$, by (c), we know that $\sum_{Q:v \in Q \in \mathcal{B}} \omega_\theta(Q; B_j)^s \leq 2|B_{-j}|^{d-5/8}$. Therefore, the random variable $\sum_{Q \in \mathcal{Q}} \omega_\theta(Q; B_j)^s$ can change by at most $2|B_{-j}|^{d-5/8} \leq 2r^d m^{d-5/8}$ if we change the outcome of a single vertex. Since $\prod_{a \in [d]} q_{i_a} \geq r^{-d} m^{-1/10}$, by Theorem 6.1, the probability that the random variable is greater than $\lambda := \left(\prod_{a \in [d]} q_{i_a}\right) |\mathcal{B}| \max\left\{2\mu_{s,\theta}(\mathcal{B}; B_j), \frac{\varepsilon'}{2}\right\}$ is at most $2e^{-2\frac{(\lambda')^2}{m \cdot (2r^d m^{d-5/8})^2}} = 2e^{-(\lambda')^2/(2r^{2d} m^{2d-1/4})}$, where

$$\lambda' := \lambda - \mathbb{E}\left[\sum_{Q \in \mathcal{Q}} \omega_\theta(Q; B_j)^s\right]$$

$$\geq \lambda - \max\left\{3\varepsilon^{-1} r^d d^2 m^{d-1}, \frac{3}{2}\left(\prod_{a \in [d]} q_{i_a}\right) |\mathcal{B}| \mu_{s,\theta}(\mathcal{B}; B_j)\right\} \geq \frac{1}{4}\lambda.$$

Since

$$\lambda = \left(\prod_{a \in [d]} q_{i_a}\right) |\mathcal{B}| \max\left\{2\mu_{s,\theta}(\mathcal{B}; B_j), \frac{\varepsilon'}{2}\right\}$$

$$\geq \left(r^{-1} m^{-1/(10d)}\right)^d \left(2^{-1/(2d)} \varepsilon m\right)^d \cdot \frac{\varepsilon'}{2} = \frac{\varepsilon' \varepsilon^d}{4r^d} m^{d-1/10},$$

the above mentioned probability is at most

$$2e^{-(\lambda')^2/(2r^{2d} m^{2d-1/4})} \leq 2e^{-\lambda^2/(32r^{2d} m^{2d-1/4})} \leq 2e^{-2^{-9}(\varepsilon')^2 \varepsilon^{2d} r^{-4d} m^{1/20}} < \frac{1}{4(kr)^{d+1}}.$$

Since there are at most $(kr)^{d+1}$ choices of indices, by the union bound, we see that $\sum_{Q \in \mathcal{Q}} \omega_\theta(Q; B_j)^s \leq \lambda$ holds for all choices, with probability greater than $\frac{3}{4}$. We show that E_2 holds if both this event and E_3 hold. For a fixed $((i,j),(i_1,j_1),\ldots,(i_d,j_d)) \in ([k] \times [r])^{d+1}$, following the notation above, event

E_3 implies that $\left(\prod_{a\in[d]} q_{i_a}\right)|\mathcal{B}| \leq 2^{1/2}|\mathcal{Q}|$. Therefore,

$$\sum_{Q\in\mathcal{Q}} \omega_\theta(Q;B_j)^s \leq \lambda = \left(\prod_{a\in[d]} q_{i_a}\right)|\mathcal{B}| \max\left\{2\mu_{s,\theta}(\mathcal{B};B_j), \frac{\varepsilon'}{2}\right\}$$

$$\leq 2^{1/2}|\mathcal{Q}| \max\left\{2\mu_{s,\theta}(\mathcal{B};B_j), \frac{\varepsilon'}{2}\right\}$$

$$\leq |\mathcal{Q}| \max\left\{2^{3/2}\mu_{s,\theta}(\mathcal{B};B_j), \varepsilon'\right\}.$$

Note that if $r = 2$, then $\mathcal{B} = B^d_{-j}$ and thus $\mu_{s,\theta}(\mathcal{B};B_j) = \mu_{s,\theta}(B^d_{-j};B_j)$. Therefore, we have E_2. If $r \neq 2$, then since all sets B_a for $a \in [r]$ have size between $2^{-1/(2d)}\varepsilon m$ and m, we see that $|B_{-j}| \leq 2^{1/(2d)}r\varepsilon^{-1}|B_a|$ for all $a \neq j$. Therefore, $|B^d_{-j}| \leq 2^{1/2}(r\varepsilon^{-1})^d|\mathcal{B}|$, and

$$\mu_{s,\theta}(\mathcal{B};B_j) = \frac{1}{|\mathcal{B}|} \sum_{Q\in\mathcal{B}} \omega_\theta(Q;B_j)^s$$

$$\leq \frac{2^{1/2}(r\varepsilon^{-1})^d}{|B_{-j}|^d} \sum_{Q\in B^d_{-j}} \omega_\theta(Q;B_j)^s = 2^{1/2}r^d\varepsilon^{-d}\mu_{s,\theta}(B^d_{-j};B_j).$$

Thus, E_2 holds with probability greater than $\frac{1}{2}$. We proved that there is nonzero probability that all E_1, E_2, E_3 hold, and as seen above, this proves the lemma. $\qquad\square$

7. Concluding remarks

Original form of the Burr-Erdős conjecture. The definition of Ramsey numbers can be extended to pairs of graphs. For a pair of graphs H_1 and H_2, the *Ramsey number* of the pair (H_1, H_2), denoted $r(H_1, H_2)$ is the minimum integer n such that in every edge coloring of K_n with two colors red and blue, there exists a red copy of H_1 or a blue copy of H_2. The *arboricity* of a graph is the minimum number of forests into which its edge set can be partitioned. The original conjecture of Burr and Erdős [4] can be stated as follows.

CONJECTURE 7.1. *For every natural number d, there exists a constant c such that for every pair of graphs H_1 and H_2, each having arboricity at most d, $r(H_1, H_2) \leq c(|V(H_1)| + |V(H_2)|)$.*

It is well known that arboricity and degeneracy are within a factor two of each other, and hence Theorem 1.1 implies this conjecture. Moreover, our proof can be extended to more than two colors.

Determining the constant. For graphs with fixed chromatic number, the constant c_d we found is exponential in d, which is best possible up to the constant in the exponent. For general degenerate graphs, the constant c_d we

obtained is double-exponential in d, and it still remains to understand the correct behavior of this constant. The corresponding question for bounded degree graphs is reasonably well understood since Conlon, Fox, and Sudakov [9], [7], [15] proved $r(G) \leq c^{\Delta \log \Delta} |V(G)|$ for all graphs of degree at most Δ and $r(G) \leq c^{r\Delta} |V(G)|$ for all r-colorable graphs of degree at most Δ. These bounds are close to being best possible since Graham, Rödl, and Ruciński proved that there are bipartite graphs of maximum degree Δ having $r(G) \geq c^{\Delta} |V(G)|$ (for some different constant c). Moreover, the author [28] proved that a transference principle holds for bounded degree graphs, and thus if G has a 'simple' structure, then the bound on the constant can be significantly improved. For example, if there exists a homomorphism f from G to a graph H having maximum degree at most d where $|f^{-1}(v)| = o(n)$ for each $v \in V(H)$, then $r(G) \leq c^{d \log d} |V(G)|$. Hence, in this case the constant does not grow together with the maximum degree of G.

Further applications of the technique. We used a random greedy embedding algorithm together with dependent random choice. Theorem 4.2 shows that our embedding algorithm succeeds with probability greater than $\frac{1}{2}$. Hence, a careful analysis will show that there are in fact many copies of the graph of interest. It would be interesting to find further applications of these methods. For instance, by using a variant of the proof of Lemma 5.7 as in [27] together with the embedding methods developed in this paper, one can show that a weak version of the blow-up lemma holds for degenerate graphs. Namely, for all d, δ and sufficiently large n, there exist ε and c such that if $\{V_i\}_{i \in [r]}$ are disjoint vertex subsets each having size at least n and (V_i, V_j) are (ε, δ)-dense for each distinct i, j, then it contains as subgraphs all d-degenerate r-colorable graphs with at most $c^d n$ vertices. One might be able to further develop this approach as in [27] and extend the bandwidth theorem of Böttcher, Schacht, and Taraz [3] to degenerate graphs of sublinear bandwidth.

Related problems. As observed by Burr and Erdős, graphs with at least $(1+\varepsilon)n \log n$ edges have Ramsey numbers superlinear in the number of vertices. On the other hand, they showed that there are graphs with $cn \log n$ edges for some constant c that have Ramsey numbers linear in the number of vertices. It would be interesting to further classify the graphs that have Ramsey number linear in terms of its number of vertices. An interesting test case is the class of hypercubes Q_n, for which we slightly improved the previous best known bound to $r(Q_n) = (1 + o_n(1))2^{2n}$. Burr and Erdős conjectured that there exists a constant c such that $r(Q_n) \leq c2^n$ holds for all natural numbers n.

In a similar direction, there has been much work aimed at understanding the Ramsey number of a graph in terms of its number of edges. The most notable result in this direction was proved by Sudakov [33], who confirmed a

conjecture of Erdős and Graham by showing that $r(H) \leq 2^{c\sqrt{m}}$ holds for all graphs H with m edges. Also, Conlon, Fox, and Sudakov have an interesting conjecture [10, Conj. 2.16], asking whether $\log(r(H)) = \Theta(d(H) + \log n)$ holds for all n-vertex graphs H, where $d(H)$ is the degeneracy of H.

References

[1] N. ALON, Subdivided graphs have linear Ramsey numbers, *J. Graph Theory* **18** (1994), 343–347. MR 1277513. Zbl 0811.05046. https://doi.org/10.1002/jgt.3190180406.

[2] N. ALON, M. KRIVELEVICH, and B. SUDAKOV, Turán numbers of bipartite graphs and related Ramsey-type questions, *Combin. Probab. Comput.* **12** (2003), 477–494. MR 2037065. Zbl 1060.05050.

[3] J. BÖTTCHER, M. SCHACHT, and A. TARAZ, Proof of the bandwidth conjecture of Bollobás and Komlós, *Math. Ann.* **343** (2009), 175–205. MR 2448444. Zbl 1229.05132. https://doi.org/10.1007/s00208-008-0268-6.

[4] S. A. BURR and P. ERDŐS, On the magnitude of generalized Ramsey numbers for graphs, in *Infinite and Finite Sets*, *Vol.* 1 (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), *Colloq. Math. Soc. János Bolyai* **10**, North-Holland, Amsterdam, 1975, pp. 215–240. MR 0371701. Zbl 0316.05110.

[5] G. CHEN and R. H. SCHELP, Graphs with linearly bounded Ramsey numbers, *J. Combin. Theory Ser. B* **57** (1993), 138–149. MR 1198403. Zbl 0725.05076. https://doi.org/10.1006/jctb.1993.1012.

[6] C. CHVATÁL, V. RÖDL, E. SZEMERÉDI, and W. T. TROTTER, JR., The Ramsey number of a graph with bounded maximum degree, *J. Combin. Theory Ser. B* **34** (1983), 239–243. MR 0714447. Zbl 0547.05044. https://doi.org/10.1016/0095-8956(83)90037-0.

[7] D. CONLON, Hypergraph packing and sparse bipartite Ramsey numbers, *Combin. Probab. Comput.* **18** (2009), 913–923. MR 2550376. Zbl 1191.05064. https://doi.org/10.1017/S0963548309990174.

[8] D. CONLON, A new upper bound for diagonal Ramsey numbers, *Ann. of Math.* **170** (2009), 941–960. MR 2552114. Zbl 1188.05087. https://doi.org/10.4007/annals.2009.170.941.

[9] D. CONLON, J. FOX, and B. SUDAKOV, On two problems in graph Ramsey theory, *Combinatorica* **32** (2012), 513–535. MR 3004807. Zbl 1289.05332. https://doi.org/10.1007/s00493-012-2710-3.

[10] D. CONLON, J. FOX, and B. SUDAKOV, Recent developments in graph Ramsey theory, in *Surveys in Combinatorics* 2015, *London Math. Soc. Lecture Note Ser.* **424**, Cambridge Univ. Press, Cambridge, 2015, pp. 49–118. MR 3497267. Zbl 1352.05123. https://doi.org/10.1017/CBO9781316106853.

[11] D. Conlon, J. Fox, and B. Sudakov, Short proofs of some extremal results II,
      J. Combin. Theory Ser. B **121** (2016), 173–196. MR 3548291. Zbl 1348.05209.
      https://doi.org/10.1016/j.jctb.2016.03.005.

[12] N. Eaton, Ramsey numbers for sparse graphs, *Discrete Math.* **185** (1998), 63–75.
      MR 1614289. Zbl 0955.05071. https://doi.org/10.1016/S0012-365X(97)00184-2.

[13] P. Erdös, Some remarks on the theory of graphs, *Bull. Amer. Math. Soc.*
      53 (1947), 292–294. MR 0019911. Zbl 0032.19203. https://doi.org/10.1090/
      S0002-9904-1947-08785-1.

[14] P. Erdös and G. Szekeres, A combinatorial problem in geometry, *Compositio
      Math.* **2** (1935), 463–470. MR 1556929. Zbl 0012.27010. Available at http://
      www.numdam.org/item?id=CM_1935__2__463_0.

[15] J. Fox and B. Sudakov, Density theorems for bipartite graphs and related
      Ramsey-type results, *Combinatorica* **29** (2009), 153–196. MR 2520279. Zbl 1212.
      05261. https://doi.org/10.1007/s00493-009-2475-5.

[16] J. Fox and B. Sudakov, Two remarks on the Burr-Erdős conjecture, *European
      J. Combin.* **30** (2009), 1630–1645. MR 2548655. Zbl 1223.05185. https://doi.
      org/10.1016/j.ejc.2009.03.004.

[17] J. Fox and B. Sudakov, Dependent random choice, *Random Structures Al-
      gorithms* **38** (2011), 68–99. MR 2768884. Zbl 1215.05159. https://doi.org/10.
      1002/rsa.20344.

[18] R. L. Graham, V. Rödl, and A. Ruciński, On graphs with linear Ramsey
      numbers, *J. Graph Theory* **35** (2000), 176–192. MR 1788033. Zbl 0965.05073.

[19] R. L. Graham, V. Rödl, and A. Ruciński, On bipartite graphs with linear
      Ramsey numbers, *Combinatorica* **21** (2001), 199–209. MR 1832445. Zbl 0989.
      05078. https://doi.org/10.1007/s004930100018.

[20] R. L. Graham, B. L. Rothschild, and J. H. Spencer, *Ramsey Theory*,
      second ed., *Wiley-Intersci. Ser. Discrete Math. Optim.*, John Wiley & Sons, New
      York, 1990. MR 1044995. Zbl 0705.05061.

[21] A. W. Hales and R. I. Jewett, Regularity and positional games, *Trans. Amer.
      Math. Soc.* **106** (1963), 222–229. MR 0143712. Zbl 0113.14802. https://doi.org/
      10.1090/S0002-9947-1963-0143712-1.

[22] S. Janson, T. Łuczak, and A. Rucinski, *Random Graphs*, *Wiley-Intersci.
      Ser. Discrete Math. Optim.*, John Wiley & Sons, New York, 2000. MR 1782847.
      Zbl 0968.05003. https://doi.org/10.1002/9781118032718.

[23] J. Komlós, G. N. Sárközy, and E. Szemerédi, Blow-up lemma, *Combinator-
      ica* **17** (1997), 109–123. MR 1466579. Zbl 0880.05049. https://doi.org/10.1007/
      BF01196135.

[24] A. V. Kostochka and V. Rödl, On graphs with small Ramsey numbers, *J.
      Graph Theory* **37** (2001), 198–204. MR 1834850. Zbl 0994.05094. https://doi.
      org/10.1002/jgt.1014.

[25] A. V. Kostochka and V. Rödl, On graphs with small Ramsey numbers. II,
      Combinatorica **24** (2004), 389–401. MR 2085363. Zbl 1063.05101. https://doi.
      org/10.1007/s00493-004-0024-9.

[26] A. Kostochka and B. Sudakov, On Ramsey numbers of sparse graphs, *Combin. Probab. Comput.* **12** (2003), 627–641. MR 2037075. Zbl 1062.05099.

[27] C. Lee, Embedding degenerate graphs of small bandwidth, 2015. arXiv 1501.05350.

[28] C. Lee, A transference principle for Ramsey numbers of bounded degree graphs, 2015. arXiv 1504.06285.

[29] C. McDiarmid, Concentration, in *Probabilistic Methods for Algorithmic Discrete Mathematics*, *Algorithms Combin.* **16**, Springer-Verlag, New York, 1998, pp. 195–248. MR 1678578. Zbl 0927.60027. https://doi.org/10.1007/978-3-662-12788-9_6.

[30] F. P. Ramsey, On a Problem of Formal Logic, *Proc. London Math. Soc.* **S2-30** (1928), 264. MR 1576401. Zbl 55.0032.04. https://doi.org/10.1112/plms/s2-30.1.264.

[31] V. Rödl and R. Thomas, Arrangeability and clique subdivisions, in *The Mathematics of Paul Erdős, II*, *Algorithms Combin.* **14**, Springer-Verlag, New York, 1997, pp. 236–239. MR 1425217. Zbl 0864.05073. https://doi.org/10.1007/978-3-642-60406-5_20.

[32] J. Spencer, Ramsey's theorem—a new lower bound, *J. Combinatorial Theory Ser. A* **18** (1975), 108–115. MR 0366726. Zbl 0296.05003. https://doi.org/10.1016/0097-3165(75)90071-0.

[33] B. Sudakov, A conjecture of Erdős on graph Ramsey numbers, *Adv. Math.* **227** (2011), 601–609. MR 2782204. Zbl 1213.05178. https://doi.org/10.1016/j.aim.2011.02.004.

[34] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Acta Arith.* **27** (1975), 199–245. MR 0369312. Zbl 0303.10056.

[35] B. Van der Waerden, Beweis einer baudetschen vermutung, *Nieuw Arch. Wisk* **15** (1927), 212–216. Zbl 53.0073.12.

Massachusetts Institute of Technology, Cambridge, MA
E-mail: cb_lee@math.mit.edu