COMBINATORICA 27 (5) (2007) 551–585
DOI: 10.1007/s00493-007-2183-y

**COMBINATORICA**
Bolyai Society – Springer-Verlag

# THE INTRINSIC DIMENSIONALITY OF GRAPHS

ROBERT KRAUTHGAMER, JAMES R. LEE*

Received January 9, 2005

We resolve the following conjecture raised by Levin together with Linial, London, and Rabinovich [Combinatorica, 1995]. For a graph $G$, let $\dim(G)$ be the smallest $d$ such that $G$ occurs as a (not necessarily induced) subgraph of $\mathbb{Z}_\infty^d$, the infinite graph with vertex set $\mathbb{Z}^d$ and an edge $(u,v)$ whenever $\|u-v\|_\infty = 1$. The *growth rate* of $G$, denoted $\rho_G$, is the minimum $\rho$ such that every ball of radius $r>1$ in $G$ contains at most $r^\rho$ vertices. By simple volume arguments, $\dim(G) = \Omega(\rho_G)$. Levin conjectured that this lower bound is tight, i.e., that $\dim(G) = O(\rho_G)$ for every graph $G$.

Previously, it was unknown whether $\dim(G)$ could be bounded above by any function of $\rho_G$. We show that a weaker form of Levin's conjecture holds by proving that $\dim(G) = O(\rho_G \log \rho_G)$ for any graph $G$. We disprove, however, the specific bound of the conjecture and show that our upper bound is tight by exhibiting graphs for which $\dim(G) = \Omega(\rho_G \log \rho_G)$. For several special families of graphs (e.g., planar graphs), we salvage the strong form, showing that $\dim(G) = O(\rho_G)$. Our results extend to a variant of the conjecture for finite-dimensional Euclidean spaces posed by Linial and independently by Benjamini and Schramm.

## 1. Introduction

The geometry of graphs, a fascinating area of combinatorics concerned with the geometric representation of graphs, has found many algorithmic applications in recent years. A very fruitful and actively studied line of research involves embedding the metric of a weighted graph into some finite-dimensional real-normed space (see, for instance, the surveys [13] and [19,

*Mathematics Subject Classification (2000):* 05C62; 51F99

* Supported by NSF grant CCR-0121555 and by an NSF Graduate Research Fellowship.

ch. 15]). In their seminal paper [16], Linial, London, and Rabinovich were the first to fully realize the algorithmic importance of low-distortion metric embeddings. However, their initial motivation was to understand the relationship between the *dimensionality* of a graph and its combinatorial properties.

The notion of dimensionality for a graph is usually based on a particular way of embedding the graph into some space that possesses an intrinsic dimension (e.g., a finite-dimensional Euclidean space). One then defines the dimension of a graph to be the least dimension into which it can be embedded. Several such notions have been extensively studied; see [18]. In [16], the authors wished to express the concept that graphs of "everywhere large diameter" should have low dimensionality. With the help of Leonid Levin, this concept was formalized as follows.

Let $\mathbb{Z}^d_\infty$ be the infinite graph with vertex set $\mathbb{Z}^d$ and an edge $(u,v)$ for two vertices $u$ and $v$ whenever $\|u-v\|_\infty = 1$. For a graph $G=(V,E)$, define $\dim(G)$ to be the smallest $d$ such that $G$ occurs as a (not necessarily induced) subgraph of $\mathbb{Z}^d_\infty$.

For a pair of vertices $u,v \in V$, we define $d_G(u,v)$ to be the distance between $u$ and $v$ in the shortest path metric on $G$. We denote by

$$B(v,r) = \{u \in V : d_G(u,v) \leq r\}$$

the closed ball of radius $r$ in $G$ centered at $v$, and define the *growth rate* of $G$ to be

$$\rho_G = \inf\{\rho : |B(v,r)| \leq r^\rho \text{ for all } v \in V \text{ and } r > 1\}.$$

Equivalently, $\rho_G = \sup\left\{\frac{\log|B(v,r)|}{\log r} : v \in V, r > 1\right\}$. Notice that $\rho_{\mathbb{Z}^d_\infty} = \Theta(d)$, so by a simple counting argument, we must have $\dim(G) = \Omega(\rho_G)$. Levin, together with Linial, London, and Rabinovich [16], conjectured that $O(\rho_G)$ dimensions suffice.

**Conjecture 1.** *For any graph $G$ with growth rate $\rho_G$, $G$ occurs as a (not necessarily induced) subgraph of $\mathbb{Z}_\infty^{O(\rho_G)}$. In other words, $\dim(G) = \Theta(\rho_G)$.*

In [16], it was shown that Conjecture 1 holds for the $k$-dimensional hypercube and the complete binary tree, but nothing beyond these two special cases was known. Indeed, it was not known whether $\dim(G)$ could be upper bounded by *any* function of $\rho_G$, even in the seemingly simpler case when the graph is a tree. Linial [15] asked about a Euclidean analogue to this notion of dimensionality.

**Question 2.** *For a graph $G=(V,E)$, what is the minimum dimension $d$, denoted $\dim_2(G)$, such that there exists a mapping $\gamma : V \to \mathbb{R}^d$ with the following properties?*

1. $\|\gamma(u)-\gamma(v)\|_2 \geq 1$ *for all* $u \neq v \in V$, *and*
2. $\|\gamma(u)-\gamma(v)\|_2 \leq 2$ *for all* $(u,v) \in E$.

Itai Benjamini and Oded Schramm [personal communication, 2003] independently asked a similar question: Is it the case that, for every infinite graph $G$ with $\rho(G)<\infty$, we have $\dim_2(G)<\infty$? In what follows, we resolve all these questions and give tight quantitative bounds. Linial remarked that the condition $\|\gamma(u)-\gamma(v)\|_2 \leq 2$ is somewhat arbitrary; indeed, we will see that it can be replaced by $\|\gamma(u)-\gamma(v)\|_2 \leq C$ for any fixed $C>1$ while affecting the value of $\dim_2(G)$ by only a constant factor (that depends on $C$), see Section 6.

## 1.1. Results and techniques

In Section 2, we provide a self-contained proof of Levin's conjecture for trees. We first perform a recursive decomposition of the tree. Each "level" of the decomposition is responsible for pairs of vertices whose distance (in the tree) falls into a certain range of the form $[r, r^2]$. For any single level, we use the probabilistic method to construct an embedding that handles all the respective pairs of vertices. We can embed each level separately and concatenate the $O(\log\log|V|)$ resulting mappings, but the dimension of the host lattice becomes too large. The key technique, which completes the proof, is a method of handling all the levels using the same coordinates. The solution involves "conservation" of randomness between levels – our hierarchical decomposition of the tree induces a partition of a corresponding probability space, and the probabilistic construction at later levels is restricted to randomizing over only the remaining "untapped" randomness (an idea that reappears throughout the paper). This result extends to graphs whose induced simple cycles are of bounded length by known low-distortion embeddings of such graphs into trees, due to [5,6].

In Section 3, we give a lower bound on the dimension necessary to embed low-degree expander graphs which shows that the strong form of Levin's conjecture does not hold in general. In particular, we show that for a $\log n$-degree expander $G$, $\dim(G)=\Omega(\rho_G \log \rho_G)$.

In Section 4, using different techniques, but many ideas from our proof for trees, we give a general upper bound on $\dim(G)$ in terms of certain graph decompositions. In this setting, choosing a good embedding for a level with high probability is more difficult; the approach we use is inspired by a technique of [21] (which was used there to embed planar graphs into Euclidean space with low distortion). Again, we must discover a delicate way of handling all the levels simultaneously. In Section 4.4, we employ the decomposition of [14] (see also [10]), combined with the results of Section 4, to prove the conjecture for any family of graphs which excludes a fixed minor (this includes planar graphs, for instance).

In Section 5, we modify a probabilistic decomposition of [17,3] for use with growth-restricted graphs. Our modifications are two-fold. First, the parameters of our decomposition depend on $\rho$ (and not on $n=|V|$ as in [17, 3]). This is essential to our application. Secondly, our decomposition is local in the sense that events which are far apart (in $G$) are mutually independent. As a result, we are able to apply the Lovász Local Lemma, yielding decompositions which, when combined with the results of Section 4, give $\dim(G)=O(\rho_G^3)$ for any graph $G$.

To obtain a tight upper bound of $\dim(G)=O(\rho_G \log \rho_G)$, we observe in Section 5.3 that the many steps of our embedding can be essentially performed "at once," removing the need to amplify the individual probabilities at each step. Here, it is essential that every step of the embedding process is "local" with respect to its "scale."

Finally, in Section 6, it is shown that all our results for $\dim(G)$ hold also for Linial's variant $\dim_2(G)$ [15], yielding a conclusive answer to Question 2, and positively resolving the question of Benjamini and Schramm. This follows from a standard application of Chernoff-type tail bounds to the random processes employed in previous sections.

### 1.2. Related work

Notions of dimensionality for graphs were perhaps first considered by Erdős, Harary, and Tutte [8]. The geometric representations of graphs have been extensively studied in other settings; see, for instance, the survey of Lovász and Vesztergombi [18]. As mentioned before, the related study of low-distortion metric embeddings has received increasing attention in recent years (see the surveys [13,19]).

Conjecture 1 is actually a dual of the bandwidth problem for graphs. The bandwidth problem asks for the minimum stretch of any edge in an embedding of the graph into $\mathbb{Z}=\mathbb{Z}_\infty^1$. Conjecture 1 asks for the minimum dimension needed to achieve a stretch of one (no stretch). Interestingly, the density bound $D=\max\{\frac{|B(v,r)|}{2r}\}$ (a one-dimensional analogue of the growth rate), which is a straightforward lower bound on the bandwidth, is conjectured to be within a $\log|V|$ factor of the bandwidth (this gap is met, for instance, by expanders). From [9], we know that the bandwidth and density differ by only a polylog($|V|$) factor. However, the techniques employed for the bandwidth problem do not seem to help in resolving the dual question.

Imposing a growth restriction like $|B(v,r)| \leq r^\rho$ on a graph has many analogs in the analysis of metric spaces, and in Riemannian geometry. For metric spaces, there are the notions of a *doubling metric* and *Ahlfors Q-regularity* (see [12]). A metric space $(X,d)$ is called doubling if there exists a constant $L > 0$ such that every ball in $X$ can be covered by $L$ balls of half the radius. Assouad [2] showed that, for any fixed $0<\epsilon<1$, the metric space $(X,d^\epsilon)$ (with distances raised to the power $\epsilon$) embeds into $\mathbb{R}^k$ with distortion $D$, where $k$ and $D$ depend only on $L$. This theorem is similar in spirit to a number of the results presented here. Although Assouad's methods are significantly different and do not apply to Levin's problem, the relationship is not entirely superficial; the techniques of Section 5 were used in [11] to obtain a new (quantitatively almost optimal) proof of Assouad's theorem.

### 1.3. Preliminaries

For a point $x=(x_1,\dots,x_d)\in\mathbb{R}^d$, we write $\|x\|_2=(\sum_{i=1}^d |x_i|^2)^{\frac{1}{2}}$ and $\|x\|_\infty = \max_{i=1}^d |x_i|$. For a graph $G$, we write $V(G)$ and $E(G)$ for the vertex and edge set of $G$, respectively.

**Definition 1.1.** For a graph $G=(V,E)$, a map $\varphi: V\to\mathbb{Z}^d$ is a *contraction* (or *contractive*) if $(u,v)\in E$ implies $\|\varphi(u)-\varphi(v)\|_\infty\leq 1$. Furthermore, if $\{\varphi_i\}$ is a finite set of mappings, we define the *direct sum*, $\varphi=\bigoplus_i \varphi_i$ to be the mapping $\varphi(u)=(\varphi_1(u),\varphi_2(u),\dots)$ (i.e., where coordinates are concatenated).

Notice that if a map $\varphi: V\to\mathbb{Z}^d$ is both contractive and injective, then $G$ occurs as a subgraph of $\mathbb{Z}^d_\infty$, and in particular, $\dim(G)\leq d$. We can think of any such embedding $\varphi$ as consisting of $d$ separate one-dimensional maps $\varphi_1,\dots,\varphi_d$ such that $\varphi=\bigoplus_{i=1}^n \varphi_i$. The following simple lemma will serve as our guide.

**Lemma 1.2.** *Let $G=(V,E)$ and $\varphi=\bigoplus_{i=1}^d \varphi_i$ where $\varphi_i: V\to\mathbb{Z}$, then the following are true.*

1. *$\varphi$ is a contraction $\iff$ every $\varphi_i$ is a contraction.*
2. *$\varphi$ is injective $\iff$ for every pair $u,v\in V$, there exists some $\varphi_i$ such that $\varphi_i(u)\neq\varphi_i(v)$.*

Unless otherwise stated, $\|\cdot\|=\|\cdot\|_\infty$ and all logarithms are to the base 2.

## 2. Trees

In this section we show that every tree $T$ with growth rate $\rho$ occurs as a subgraph of $\mathbb{Z}_\infty^d$ with $d=O(\rho)$ by exhibiting a map $\varphi: T \to \mathbb{Z}^d$ that is both contractive and injective.

### 2.1. Embedding trees by random walks

In light of Lemma 1.2, it is natural to define a distribution over random contractions and then argue that some such map must be injective.

Let $T=(V,E)$ be a tree whose growth rate is at most $\rho$. Fix an arbitrary root $r$ of $T$, and let $h$ be the height of $T$ (so that $V \subseteq B(r,h)$). Define $d_T$ to be the shortest path metric on $T$ and let $c>0$ be a sufficiently large constant to be determined later. Let $T_1, T_2, \ldots, T_{c\rho}$ be $c\rho$ weighted copies of $T$, where the weight of every edge in $T_i$ is chosen independently and uniformly at random from the set $\{-1,+1\}$. For each $v \in V$, let $v_i$ be the sum of the edge weights on the unique path from $r$ to $v$ in $T_i$. Finally, define the image of $v$ in $\mathbb{Z}^{c\rho}$ by $\varphi(v)=(v_1, v_2, \ldots, v_{c\rho})$.

Clearly $\varphi$ is a contraction. Now consider any two vertices $u,v \in V$ for which $d_T(u,v) \geq \sqrt{h}$. The probability that the images of $u$ and $v$ agree in any single coordinate, i.e. that $u_i = v_i$, is the probability that a random walk with $+1/{-1}$ steps and length $\sqrt{h}$ ends at its starting point, namely $O(h)^{-1/4}$. Hence the probability that $\varphi(u)=\varphi(v)$ is at most $O(h)^{-c\rho/4}$. Observe that since $T$ is contained in a ball of radius $h$ centered at $r$, it contains at most $h^\rho$ vertices. Taking a union bound over at most $h^{2\rho}$ pairs $\{u,v\} \in V^2$,

$$\Pr\left[\exists u,v \in V,\ d_T(u,v) \geq \sqrt{h} \text{ and } \varphi(u)=\varphi(v)\right] \leq h^{2\rho}\, O(h)^{-c\rho/4}.$$

It follows that there exists universal constants $h_0 > 0$, $c > 9$ such that for every tree $T$ of height $h \geq h_0$, there exists a map $\varphi: V \to \mathbb{Z}^{c\rho}$ for which $d_T(u,v) \geq \sqrt{h} \implies \varphi(u) \neq \varphi(v)$. In what follows, we carefully utilize this simple but powerful observation to show that $\dim(T)=O(\rho)$ for any tree $T$, thus proving Conjecture 1 for the special case of trees.

### 2.2. Relative embeddings and rooted subtrees

Consider a tree $T=(V,E)$ with $\rho=\rho_T$ and fix a root $r_T$ of $T$. Define a *rooted subtree* of $T$ to be a connected vertex-induced subgraph $X$ with a distinguished root $r_X$. Let $W=\{-1,0,+1\}$ be the set of edge weights. For a rooted subtree $X=(V_X,E_X)$, we define a *d-dimensional relative embedding* of $X$ to be a map $\mu_X : E_X \to W^d$. Finally, we will denote by $\mu_X^* : V_X \to \mathbb{Z}^d$ the *(absolute) embedding* induced by the relative embedding $\mu_X$ (with respect to the root $r_X$), which is the map obtained as follows: For a vertex $v \in V_X$, define its image $\mu_X^*(v) \in \mathbb{Z}^d$ as the sum of the edge weights $\mu_X(e)$ along the unique path from $r_X$ to $v$ in $X$. By our choice of $W$, induced embeddings are always contractions. Furthermore, given any contractive embedding $\varphi : V_T \to \mathbb{Z}^d$, there exists a unique $d$-dimensional relative embedding $\mu$ such that $\varphi = \mu^*$ (with respect to $r_X$). Let us define $\mathbf{0} = (0, 0, \ldots, 0)$ to be the all-zero vector. Notice that the construction of Section 2.1, when applied to a subtree $X$, yields a relative embedding of $X$ with the following desirable property.

**Lemma 2.1 (Relative embeddings).** *There exist constants $h_0$ and $c$ such that for every rooted subtree $X = (V_X, E_X)$ of $T$ with height at most $h$ where $h \geq h_0$, there is a relative embedding $\mu_X : E_X \to W^{c\rho}$ such that, for all $u, v \in V_X$, $\mu_X^*(u) \neq \mu_X^*(v)$ whenever $d_T(u, v) \geq \sqrt{h}$.*

In essence, $\mu_X^*$ "separates" points in $X$ which are far apart relative to the height of $T$. Notice that the above lemma only works for $h$ sufficiently large. When $h$ is bounded, i.e., $h = O(1)$, a brute force embedding suffices.

**Lemma 2.2 (Small subtrees).** *For any rooted subtree $X = (V_X, E_X)$ of $T$ with height at most $h_1$, there exists a relative embedding $\mu : E_X \to W^{\rho \log(h_1+2)}$ such that $\mu^*$ is injective.*

**Proof.** Let $V_X = \{v_1, v_2, \ldots, v_m\}$ and define $\varphi(v_i) = B(i)$ where $B(i)$ is the binary representation of $i-1$ written as a $\lceil \log m \rceil$-dimensional vector. Now $\varphi$ is clearly injective and also a contraction since $\|B(i) - B(j)\|_\infty \leq 1$ for all $i, j$. Finally, notice that since $X$ is of height $h = O(1)$, we have $m \leq h^\rho$ and hence $\lceil \log m \rceil \leq \rho \log(h_1 + 2)$. Let $\mu$ be the unique relative embedding such that $\varphi = \mu^*$. It follows that $\mu : E_X \to W^{\rho \log(h_1+2)}$ is a relative embedding with $\mu^*$ injective. ∎

Suppose $\{X_1, X_2, \ldots, X_k\}$ is a collection of vertex-disjoint rooted subtrees of $T$, and let each $X_i = (V_i, E_i)$ have root $r_i$. Furthermore, suppose that for each $X_i$, we have a relative embedding $\mu_i : E_i \to W^d$. Then we can define a relative embedding $\mu$ on all of $V$ by

$$\mu(e) = \begin{cases} \mu_i(e) & \text{if } e \in E_i \text{ for some } i \\ \mathbf{0} & \text{otherwise.} \end{cases}$$

Note that $\mu$ has the desirable property $\|\mu^*(u) - \mu^*(v)\| = \|\mu_i^*(u) - \mu_i^*(v)\|$ whenever $u, v \in V_i$. We will say that $\mu$ is obtained by *glueing* the relative embeddings $\{\mu_i\}$ together.

In the sequel, we will construct, for a given $\hat{h}>0$ (which need not be the height of the tree $T$), an embedding $\varphi: V \to \mathbb{Z}^{O(\rho)}$ that satisfies $\varphi(u) \neq \varphi(v)$ whenever $\hat{h}^{1/2} \le d(u,v) \le \hat{h}$. (This is essentially the "single scale" version of the conjecture.) To do this, we will partition $T$ into subtrees of height $O(\hat{h})$, find for each subtree a relative embedding that satisfies the desired property, and then glue all these relative embeddings into an embedding for $T$. There is the slight problem that for pairs $u,v$ with $\hat{h}^{1/2} \le d(u,v) \le \hat{h}$ that end up in different subtrees, we have no guarantee that their images (under $\mu^*$) will be distinct. To handle this, we will actually use two sets of disjoint subtrees which are "staggered" so that every pair $u,v$ with $d(u,v) \le h$ ends up in the same subtree in at least one of the sets. A far more challenging problem is that this embedding is guaranteed to handle only one value of $h$.

## 2.3. The leveled decomposition

Let $\operatorname{diam}(T)$ be the diameter of $T$. Set $k = \lceil \log\log \operatorname{diam}(T) \rceil$ and $\Delta = 2^{2^k}$, hence $\operatorname{diam}(T) \le \Delta \le \operatorname{diam}(T)^2$. We define $k$ *levels* $L_0, L_1, \ldots, L_{k-1}$ as follows. Level $i$ consists of two partitions of $T$ into rooted subtrees; denote these two partitions $A_i$ and $B_i$ and let $L_i = A_i \cup B_i$. The subtrees in $L_i$ will cover $T$ (in a sense that will be defined soon) and will each have height at most $3h(i)$, where $h(i) = \Delta^{1/2^i}$. (For convenience, we define $h(k)=1$.) To form $A_i$, let $O_i^A$ be the set of edges in $T$ whose *depth* (i.e., distance from the root $r_T$) is a multiple of $3h(i)$. Removing $O_i^A$ from $T$ results in a collection of disjoint subtrees; let $A_i$ consist of these subtrees, each rooted at its (unique) closest vertex to $r_T$. $B_i$ is formed similarly, except that $O_i^B$ is defined as the set of edges in $T$ whose depth modulo $3h(i)$ is equal to $h(i)$ (rather than 0). The edges in $O_i^A$ and $O_i^B$ are called the *open edges* of level $L_i$. The next lemma is easy to verify. In particular, property (3) follows from the "staggering" of the two sets of subtrees $A_i$ and $B_i$. Property (4) follows from the specifics of the decomposition; it provides a nesting that will turn out to be useful in Section 2.5.

**Lemma 2.3 (The leveled decomposition).** *For every tree $T=(V,E)$, the above construction satisfies the following properties.*

1. *Each $A_i$ and each $B_i$ is a partition of $V$.*
2. *The height of any subtree $X \in L_i$ is at most $3h(i)$.*
3. *For any pair $u,v \in V$ with $d(u,v) \le h(i)$, there is some $X \in L_i$ containing both $u$ and $v$.*

4. *For every $i$, $O^A_{i+1} \subseteq O^A_i$; hence, each subtree in $A_i$ is entirely contained in some subtree in $A_{i+1}$. The same holds for the subtrees in $B_i$. In this sense, each level is a refinement of the previous level.*

**Definition 2.4 (A separating map).** We say that $\varphi : V \to \mathbb{Z}^d$ *separates* $A_i$ (and similarly $B_i$) if, for every $X \in A_i$ and for every pair $u,v \in V(X)$ with $h(i+1) \leq d(u,v) \leq h(i)$, we have $\varphi(u) \neq \varphi(v)$.

Notice that if $\varphi$ separates $A_i$ and $B_i$ for all $i \in \{0,1,\dots,k-1\}$, then $\varphi$ is injective (by the properties in Lemma 2.3).

## 2.4. A first attempt

Consider the partition $A_i$ of $T$. Applying the embedding of Lemmas 2.1 and 2.2 to each $X \in A_i$ and glueing the relative embeddings together yields an induced embedding $\varphi^A_i$ which separates $A_i$. Let $\varphi^B_i$ be a similar embedding obtained from the partition $B_i$, and let $\varphi_i : V \to \mathbb{Z}^{O(\rho)}$ be defined as $\varphi_i = \varphi^A_i \oplus \varphi^B_i$. Then $\varphi_i$ separates $A_i$ and $B_i$. Finally, the map $\varphi = \varphi_0 \oplus \cdots \oplus \varphi_{k-1}$ separates every $A_i$ and every $B_i$, and is hence injective (by Lemma 1.2). Since $\varphi$ is also a contraction, it yields $\dim(T) = O(\rho k) = O(\rho \log\log \operatorname{diam}(T))$. Unfortunately, this bound depends on the diameter of $T$ and is therefore insufficient for our purposes.

## 2.5. Conserving randomness or "Not using all your ammo at once."

In the preceding section, we used too many dimensions because we needed a distinct set of coordinates for every level. In essence, determining the embedding for level $L_i$ leaves no randomness for "higher" levels $L_{i-1}, L_{i-2}, \dots, L_0$ (since all the edge weights in the relative embedding for $L_i$ are determined).

Now consider the open edges of level $L_i$, namely, $O^A_i$ and $O^B_i$, which run between disjoint subtrees. In Section 2.2, when the relative embeddings for subtrees are glued together, the open edges are assigned a weight of $\mathbf{0}$. But they might as well have been assigned any other weight in $W^d$. Clearly the resulting embedding would still be a contraction. Thus even after fixing a relative embedding for $L_i$, there is still some freedom left to us in deciding how to choose weights for the edges in $O^A_i$ and $O^B_i$. It turns out that the randomness stored in the unassigned open edges is sufficient.

We will now show that, after finding a relative embedding for the subtrees in $L_{i+2}$, there is still enough randomness left to embed the subtrees in $L_i$

simply by assigning random weights to the open edges of $L_{i+2}$. Notice that this process goes up two levels at a time, from $L_{i+2}$ to $L_i$, so we will need to do it twice, once for "even" levels and once for "odd" levels. This will increase the number of coordinates used by only a factor of 2.

**Theorem 2.5 (Embedding trees).** *Any tree $T$ with growth rate $\rho$ occurs as a (not necessarily induced) subgraph of $\mathbb{Z}_\infty^{O(\rho)}$, thus $\dim(T)=O(\rho)$.*

**Proof.** We will construct four contractions, $\varphi^A_{\text{even}}, \varphi^A_{\text{odd}}, \varphi^B_{\text{even}}, \varphi^B_{\text{odd}} : V \to \mathbb{Z}^{O(\rho)}$. Let $L_0, L_1, \dots, L_{k-1}$ be the levels of the decomposition for $T$, and assume for simplicity that $k$ is odd. Then $\varphi^A_{\text{even}}$ will separate $A_{k-1}, A_{k-3}, \dots, A_0$, $\varphi^A_{\text{odd}}$ will separate $A_{k-2}, A_{k-4}, \dots, A_1$, and $\varphi^B_{\text{even}}$ and $\varphi^B_{\text{odd}}$ will satisfy similar properties for the $B_i$. It will follow from the discussion in Section 2.4, that $\varphi = \varphi^A_{\text{even}} \oplus \varphi^A_{\text{odd}} \oplus \varphi^B_{\text{even}} \oplus \varphi^B_{\text{odd}}$ is a contractive injection, providing the desired embedding of $T$ into $\mathbb{Z}_\infty^{O(\rho)}$.

We will construct the map $\varphi^A_{\text{even}}$ inductively. The other maps are constructed similarly. Let $h_0$ and $c$ be the constant from Lemma 2.1 and let $k_0$ be the largest even integer such that $h(k_0) \geq h_0$. Since the trees in $A_{k_0}$ have height at most $3h(k_0)$, and since we may assume without loss of generality that $c \geq \log(3h_0+2)$, Lemma 2.2 yields a relative embedding $\mu_{k_0} : E \to W^{c\rho}$ for which $\mu^*_{k_0}$ separates $A_{k-1}, A_{k-3}, \dots, A_{k_0}$.

**The inductive step.** Now assume that we have a relative embedding $\mu_{i+2}: E \to W^{c\rho}$ for which $\mu^*_{i+2}$ separates $A_{k-1}, A_{k-3}, \dots, A_{i+2}$. We will show the existence of a relative embedding $\mu_i : E \to W^{c\rho}$ which satisfies

1. For all $T \in A_{i+2}$ and all $u, v \in V(T)$, $\|\mu^*_{i+2}(u) - \mu^*_{i+2}(v)\| = \|\mu^*_i(u) - \mu^*_i(v)\|$;
2. $\mu^*_i$ separates $A_i$.

Since the subtrees of $A_j$ for $j \geq i+2$ are all completely nested within the subtrees of $A_{i+2}$ (recall property (4) of Lemma 2.3), the first condition guarantees that $\mu^*_i$ separates $A_{k-1}, A_{k-3}, \dots, A_{i+2}$, since $\mu^*_{i+2}$ does.

To obtain $\mu_i$ from $\mu_{i+2}$, we will only change the edge weights in $O^A_{i+2}$, i.e. those running between disjoint subtrees of $A_{i+2}$. Condition (1) then follows immediately. The construction of the relative embedding $\mu_i$ is actually probabilistic; we shall randomly change edge weights in $O^A_{i+2}$, and show that (2) is satisfied with positive probability. For every $e \in O^A_{i+2}$, choose $\mu_i(e)$ uniformly at random from $\{-1,+1\}^{c\rho}$. For all other edges $e$, define $\mu_i(e) = \mu_{i+2}(e)$.

Let us now show that with positive probability, $\mu^*_i$ separates $A_i$. Fix some $X \in A_i$ and consider two points $u, v \in V(X)$ such that $d(u,v) \geq h(i+1) = h(i)^{1/2}$. Let $P_{uv}$ be the unique path from $u$ to $v$ in $X$. Since $P_{uv}$ has length at least $h(i)^{1/2}$ and each subtree of $A_{i+2}$ has height at most $3h(i+2) = 3h(i)^{1/4}$,

$P_{uv}$ must pass through at least $\frac{1}{3}h(i)^{1/4}$ such subtrees. In particular, the path includes at least $\frac{1}{3}h(i)^{1/4} - 2$ edges from $O^A_{i+2}$.

Now consider the part of $P_{uv}$ which is composed of edges whose weights are already fixed (i.e., edges not in $O^A_{i+2}$). The sum of their weights is fixed, and the probability that a random walk of length at least $\frac{1}{3}h(i)^{1/4} - 2$ (along open edges) is equal to the negation of any fixed amount is at most $O(h(i))^{-1/8}$. This also upper bounds the probability that the images of $u$ and $v$ (under $\mu_i^*$) agree in any single coordinate. So the probability of this occurring in $c\rho$ independent coordinates is

$$\Pr[\mu_i^*(u) = \mu_i^*(v)] = O(h(i))^{-c\rho/8}.$$

Finally, notice that $X$ has height at most $3h(i)$, and thus contains at most $(3h(i))^{2\rho}$ pairs of vertices. Since $h(i) \geq h_0$ and we may assume that $h_0$ and $c$ are sufficiently large constants (which are *independent of* $i$), the union bound $O(h(i))^{-c\rho/8}O(h(i))^{2\rho} < 1/2$ shows the existence of a map $\mu_i$ with the desired property in $X$. Continuing in this way for each disjoint subtree $X \in A_i$, we see that with positive probability, $\mu_i$ satisfies condition (2).

By induction, $\mu_0^*$ separates each of $A_{k-1}, \dots, A_0$. Setting $\varphi^A_{\text{even}} = \mu_0^*$ completes the proof. ∎

## 2.6. Alteration

We offer two simple lemmas on fixing maps which are sufficiently close to good embeddings.

**Lemma 2.6 (Almost injective embeddings).** *For a graph $G = (V, E)$, if there is a contractive map $\varphi : V \to \mathbb{Z}^d$ which is $k$-to-1, then there is a mapping $\varphi' : V \to \mathbb{Z}^{d + \lceil \log k \rceil}$ which is contractive and injective.*

**Proof.** To see this, suppose $z \in \mathrm{Im}(\varphi)$ and let $\varphi^{-1}(z)_1, \dots, \varphi^{-1}(z)_k$ be the $k$ possible preimages of $z$. Now define $\varphi'(\varphi^{-1}(z)_i) = (\varphi(z), B(i))$ where $B(i)$ is the $\lceil \log k \rceil$-digit binary representation of $i - 1$. ∎

**Lemma 2.7 (Almost contractive embeddings).** *For a graph $G = (V, E)$, if there is an injective mapping $\varphi : V \to \mathbb{Z}^d$ which satisfies $(u, v) \in E \Longrightarrow \|\varphi(u) - \varphi(v)\|_\infty \leq k$, then there is a mapping $\varphi' : V \to \mathbb{Z}^{d(2 + \log k)}$ which is contractive and injective.*

**Proof.** To get $\varphi'$, split $\mathbb{Z}^d$ up into cubes of side length $k$. Now contract every such cube to its center, and then scale all the coordinates by a factor of $1/k$. It is easy to see that the resulting map is a contraction, but since each cube contained at most $k^d$ points, the resulting map is only $k^d$-to-1. Applying Lemma 2.6, we can get a mapping which is injective and has $d+\lceil\log(k^d)\rceil$ dimensions. ∎

## 2.7. Graphs without long induced simple cycles

For a graph $G$, let $\lambda(G)$ be the length of the longest induced simple cycle in $G$. We will use Lemmas 2.6 and 2.7 and the following theorem of Brandstädt, Chepoi, and Dragan [5,6] to prove a result on graphs which have no long induced simple cycles.

**Theorem 2.8** **([6]).** *For any graph $G=(V,E)$, there exists a tree $T=(V,F)$ such that*

$$|d_G(u,v)-d_T(u,v)| = O(\lambda(G)). \tag{1}$$

**Theorem 2.9.** *For any graph $G$, let $\rho=\rho_G, \lambda=\lambda(G)$, then $G$ occurs as a subgraph of $\mathbb{Z}_\infty^{O(\rho\log^2[\lambda+2])}$.*

**Proof.** Let $\rho=\rho_G$, $\lambda=\lambda(G)$, and let $T$ be the corresponding tree of Theorem 2.8. First, setting $d_G(u,v)=1$ in (1), we see that edges are stretched in $T$ by at most $O(\lambda)$. Secondly, setting $d_T(u,v)=1$ in (1), we see that edges in $T$ correspond to paths in $G$ of length at most $O(\lambda)$. It follows that

$$\rho_T = \sup_{x,r}\frac{\log|B_T(x,r)|}{\log r} \le \sup_{x,r}\frac{\log|B_G(x,\lambda r)|}{\log r} \le \sup_{x,r}\frac{\log(\lambda r)^\rho}{\log r} \le \rho(1+\log\lambda).$$

So we can embed $T$ into $O(\rho\log(\lambda+2))$ dimensions by Theorem 2.5. The same mapping is also an injective embedding for $G$ that expands edges by at most $O(\lambda)$. Applying Lemma 2.7, we arrive at a contractive, injective embedding of $G$ into $O(\rho\log^2[\lambda+2])$ dimensions. ∎

**Corollary 2.10.** *Conjecture 1 is true for any class of graphs in which $\lambda(G)$ is bounded, yielding $\dim(G)=O(\rho)$. This class includes trees and chordal graphs.*

# 3. Expanders

Before we jump into the proof of our main theorem, let us take a moment to prove a lower bound for expander graphs.

**Definition 3.1 (Expander graphs).** An $n$-vertex graph $G=(V,E)$ will be called a *$\Theta(k)$-degree expander* for $k=k(n)\geq 2$ if it has the following properties:

(a) The degree of every vertex is $\Theta(k)$.
(b) The diameter of $G$ is $O(\log_k n)$.
(c) Every two disjoint subsets of $n/\log n$ vertices are connected by a path of length $O(\log_k \log n)$.

Observe that properties (b) and (c) follow from vertex expansion. Indeed, let $\Gamma(S)$ stand for the set of vertices with at least one neighbor in $S$, and suppose that for every $S\subset V$, we have $|\Gamma(S)|\geq\min\{\Omega(k|S|),\frac{2}{3}|V|\}$. Hence, there exists $t\leq O(\log_k n)$, such that every $u\in V$ satisfies $|\Gamma^t(u)|\geq\frac{2}{3}|V|$, and it follows that the diameter of $G$ is at most $2t=O(\log_k n)$. Property (c) follows similarly.

Consequently, for every $3\leq k(n)\leq\log n$ and every sufficiently large $n$, there exists a graph $G$ satisfying properties (a)–(c). An even simpler way to obtain such a graph $G$ is to take a (standard) 3-regular $n$-vertex expander $H$, and create $G$ on the same vertex set by a connecting two vertices $u,v$ whenever $d_H(u,v)\leq\log_3 k$. It is easy to satisfy property (a) by iteratively connecting by an edge the two vertices of lowest degree. A simple argument as above shows that the diameter of $G$ is $O(\log n)$ and that every two sets of size $n/\log n$ are connected by a path of length $O(\log\log n)$. Properties (b) and (c) now follow from the fact that for very two vertices $u,v$ we have $d_G(u,v)\leq O(d_H(u,v)/\log k)$.

## 3.1. A dimension lower bound

**Theorem 3.2.** *Let $G=(V,E)$ be a $\Theta(\log|V|)$-degree expander, then $\dim(G)=\Omega(\rho_G\log\rho_G)$. In particular, Conjecture 1 is not true (for general graphs).*

**Proof.** Let $G=(V,E)$ be a $\Theta(k)$-degree expander graph on $n$ vertices (see Definition 3.1), with $1\leq k\leq\log n$. It follows from Properties (a) and (b) that $\rho_G=\Theta\big(\frac{\log n}{\log\log_k n}\big)$. We shall show that if $G$ occurs as a subgraph of $\mathbb{Z}_\infty^d$ then $d\geq\Omega\big(\frac{\log n}{1+\log\log_k\log n}\big)$. Note that for $k=\log n$, this implies that $\dim(G)=\Omega(\log n)=\Omega(\rho_G\log\rho_G)$ and this lower bound is tight, up to constant factors, since the trivial upper bound $d=O(\log n)$ holds for any $n$-vertex graph (by a bijection into $\{0,1\}^{\lceil\log n\rceil}$).

Assume for contradiction that $G$ occurs as a subgraph of $\mathbb{Z}_\infty^d$ with $d=o\big(\frac{\log n}{1+\log\log_k\log n}\big)$. Let $\varphi$ be the corresponding embedding of $G$ into $\mathbb{Z}_\infty^d$, and let $\varphi_i$ be the projection of $\varphi$ on the coordinate $i = 1, \dots, d$. Let the set $S_i$ consist of the $n/\log n$ vertices $v \in V$ with smallest $\varphi_i(v)$, and let the set $L_i$ consist of the $n/\log n$ vertices $v \in V$ with largest $\varphi_i(v)$, breaking ties arbitrarily.

We claim that $\varphi_i(V \setminus (S_i \cup L_i))$ is contained in an interval of size $O(\log_k \log n)$. Indeed, by property (c) above $G$ contains a path of length $O(\log_k \log n)$ that connects some vertex $s \in S_i$ with some vertex $l \in L_i$, and since $\varphi$ is contractive, $\varphi_i(l) - \varphi_i(s) \le O(\log_k \log n)$. By the definition of $S_i$ and $L_i$, for every $v \in V \setminus (S_i \cup L_i)$ we have $\varphi_i(s) \le \varphi_i(v) \le \varphi_i(l)$, which proves the claim.

Finally, the set of vertices $V' = V \setminus (\cup_{i=1}^d (S_i \cup L_i))$ contains at least $n - dn/\log n > n/2$ vertices. By the above claim, $\varphi(V')$ is contained in a subset of the lattice $\mathbb{Z}^d$ formed by the Cartesian product of $d$ intervals of size $O(\log_k \log n)$. However, this subset of $\mathbb{Z}^d$ contains at most $(O(\log_k \log n))^d \le n/2$ points, which contradicts the assumption that $\varphi$ is injective. ∎

### 3.2. A distortion lower bound

It is well-known that the bandwidth of an expander is $\Omega(n)$. The bandwidth may be seen as the maximum stretch of any edge in an injective embedding of the graph into $\mathbb{Z}^1_\infty$, thus embeddings into $\mathbb{Z}^d_\infty$ provide a generalization of the bandwidth. We show now that even in much higher dimensions, the edges of the expander must be stretched by a large factor.

**Corollary 3.3.** *For $1 \le k \le \log n$, any embedding of an $n$-vertex $\Theta(k)$-degree expander $G$ into $\mathbb{Z}_\infty^{O(\rho_G)}$ stretches at least one edge to length $(\log n)^{\Omega(1)}$.*

**Proof.** Proceed similar to the proof of Theorem 3.2, with $d = O(\rho_G) = O\left(\frac{\log n}{\log \log n}\right)$. Now if every edge is stretched by at most $\alpha$, then two sets of $n/\log n$ vertices have their lattice images within distance $O(\alpha \log_k \log n)$ of each other. It follows that $O(\alpha \log_k \log n)^{O(\rho_G)} \le n/2$, and for $\alpha = (\log n)^{o(1)}$ we derive a contradiction. ∎

## 4. Divide & conquer: Upper bounds from graph decomposition

In this section, we will use the ideas of Section 2 to prove a result on general graphs in terms of their decompositions (Theorem 4.2). As an example application of this general result, we will show that Conjecture 1 holds for graphs excluding a fixed minor (Section 4.4).

**Outline.** At a conceptual level, our embedding method for a general graph $G$ bears some similarity to the case of trees, although several technical aspects differ significantly. We cannot, for example, form a coordinate by assigning edge weights independently at random (the weights along every cycle would have to sum to 0). Here is an outline of the proof, with simplified notation and constants. Our first step is to focus on a single "scale" $r$, and show a contraction $\varphi_r$ such that $\varphi_r(u) \neq \varphi_r(v)$ for every two vertices $u, v$ with $r^{1/2} \leq d(u,v) \leq r$. The construction of $\varphi_r$ follows a divide and conquer approach – we decompose the graph into (overlapping) "clusters", embed each cluster separately, and then "glue" these embeddings together. The decomposition guarantees that for every pair $u, v$ as above there exists at least one cluster $C$ that contains them both, hence it suffices that the embedding $\varphi_C$ of this cluster satisfies $\varphi_C(u) \neq \varphi_C(v)$. The decomposition further guarantees that each cluster $C$ has diameter at most $r^{O(1)}$, and then the growth rate bound implies $|C| \leq r^{O(\rho_G)}$.

One key difference between trees and general graphs is in the "divide" stage. For trees, we were able to design such a decomposition directly (Section 2.3), using only two "layers", i.e. two partitions of $V$. When embedding general graphs, we consider decompositions with more layers, but we must ensure that the number of layers is bounded by, say, $O(1)$ or $O(\rho_G)$, since it affects the dimension of the resulting embedding. Furthermore, for trees our decompositions were nested (in the sense that finer partitions were refinements of coarser ones); for general decompositions we will have to force this nesting property to hold.

Another key difference is in the "conquer" (or combining) stage. In trees, it is quite easy to glue embeddings of disjoint subtrees, using the concept of relative embeddings (Section 2.2). In general graphs, we ensure that embeddings of disjoint clusters can be glued together by restricting ourselves to cluster embeddings in which the "boundary" of the cluster is mapped to the all-zeros vector. One side effect of this restriction is that for $u, v \in C$ lying on or "close" to the boundary of $C$, we cannot require that $\varphi_C(u) \neq \varphi_C(v)$. In addition, to embed a single cluster $C$, we have to resort to a more sophisticated method, inspired by Rao [21]; roughly speaking, we employ another decomposition that breaks $C$ into subclusters and map each subcluster independently at random. This inner decomposition of $C$ into subclusters has the same requirements as the outer decomposition, but it is applied with a different parameter, namely – each subcluster's diameter is less than $r^{1/2}$. To show that this embedding of $C$ succeeds with high probability we use a union bound over the at most $|C|^2 \leq r^{O(\rho_G)}$ pairs $u, v \in C$. Notice that the inner decomposition guarantees, for pairs $u, v \in C$ as above, that $\varphi_C(u)$ and

$\varphi_C(v)$ are independent, while the outer decomposition limits the size of the subproblem $C$, enabling the use of a union bound. On top of this, we have to adapt the technique of conserving randomness (Section 2.5) to this new embedding method.

**Preliminaries.** In what follows, let $G=(V,E)$ be a simple graph with growth rate $\rho=\rho_G$. A *cluster* of $G$ is a simply subset $S\subseteq V$, though we will usually use this terminology in the context of a partition which contains $S$.

Define the *boundary of a cluster* $S$ as $\partial S=\{u\in S:\exists(u,v)\in E, v\notin S\}$. The boundary of a collection $\mathcal{C}$ of clusters is defined as $\partial\mathcal{C}=\cup_{S\in\mathcal{C}}\partial S$. For a cluster $S\subseteq V$ and a partition $P$ of $V$, the *induced partition* (on the cluster $S$) is defined as $Q=\{C\cap S:C\in P\}\setminus\{\emptyset\}$. As before, let $\mathbf{0}=(0,0,\dots,0)$ be the all-zero vector.

For $u,v\in V$ let $d(u,v)$ be the distance between $u$ and $v$ in the shortest path metric of $G$. We stress that even when a particular cluster $S$ is considered, $d(u,v)$ denotes the distance in $G$ and not in $S$. In particular, define the *diameter* (sometimes called weak diameter) of $S\subseteq V$ to be $\operatorname{diam}(S)=\sup_{u,v\in S}d(u,v)$. As usual, for $S\subseteq V$ we define $d(u,S)=\inf_{v\in S}d(u,v)$.

**The padded decomposition.** We first discuss our method of choosing clusters. For the rest of this section, fix an arbitrary constant $\alpha>1$. We will not explicitly state the dependence of other constants on $\alpha$, but it will be clear that $\alpha=O(1)$ suffices for our purposes.

**Definition 4.1 (The padded decomposition).** A set $\{P_1,P_2,\dots,P_m\}$ of $m$ partitions of $V$ is called an *$r$-padded decomposition of $G$ with $m$ layers* if the following properties are satisfied.

1. If $C\in\cup_{i=1}^m P_i$, then $\operatorname{diam}(C)\le r^\alpha$.
2. For every $u\in V$ there exists some $C\in\cup_{i=1}^m P_i$ such that $B(u,3r)\subseteq C$.

We can now state our main result about embeddings obtained from graph decompositions. Its proof appears in Section 4.3.

**Theorem 4.2 (Embedding via decomposition).** *Let $G$ be a graph with $\rho=\rho_G$. If for every $4\le r\le\operatorname{diam}(G)$ there exists an $r$-padded decomposition of $G$ with $m$ layers, then $\dim(G)=O(m^2\rho)$.*

## 4.1. Relative embeddings

Suppose we are given a cluster $S\subseteq V$. Define a *$d$-dimensional relative embedding* of $S$ to be a contraction $\varphi:S\to\mathbb{Z}^d$ such that $\varphi(\partial S)=\mathbf{0}$, i.e. the boundary is mapped to $\mathbf{0}$. Suppose further that we would like to find a relative embedding of $S$ with the following property (parameterized by $r>0$): For every

$u,v \in S$ with $d(u,v) > r$ and such that $B(u, 3r^{1/2}) \subseteq S$, we have $\varphi(u) \neq \varphi(v)$. In other words, since we are imposing the rather stringent condition that $\varphi(\partial S) = \mathbf{0}$, we only make requirements on vertices that are far enough from the boundary.

We will produce such an embedding using a technique inspired by the methods of Rao [21]. Each coordinate is formed by partitioning $S$ into clusters of diameter at most $r$, so $u, v$ as above must end up in different clusters. We then define the image of a vertex to be the distance from that vertex to the boundary of its cluster. To achieve injectiveness with high probability, we "perturb" the images by randomly contracting each cluster's boundary inward.

For ease of notation, we define an *$r$-inner decomposition* to be an $r^{1/\alpha}$-padded decomposition; in this case, clusters have $\mathrm{diam}(C) \leq r$ and vertices have "padding" of the form $B(u, 3r^{1/\alpha})$. We now show how to use an $r$-inner decomposition to produce a good relative embedding.

**Lemma 4.3 (Relative embeddings).** *Suppose that $G$ has an $r$-inner decomposition with $m$ layers, and let $S \subseteq V$ be a cluster with $|S| \leq r^{O(\rho)}$. Then there exists a relative embedding $\varphi : S \to \mathbb{Z}^{O(m\rho)}$ such that for every $u, v \in S$ with $d(u,v) > r$ and $B(u, 3r^{1/\alpha}) \subseteq S$, we have $\varphi(u) \neq \varphi(v)$.*

**Proof.** The $m$ partitions produced by the $r$-inner decomposition induce $m$ partitions $Q_1, \ldots, Q_m$ of $S$ (recall the definition of an induced partition). For each $Q_j$ we will construct a map $\varphi_j : S \to \mathbb{Z}^{c\alpha\rho}$, where $c > 0$ is a constant to be determined later.

Fix some partition $Q_j$ and form a single coordinate $\varphi_j^0 : S \to \mathbb{Z}$ as follows: For every $C \in Q_j$, choose some $r_C \in \{0, 1, \ldots, r^{1/\alpha}\}$ uniformly at random and let $\partial_C^* = \{v \in C : d(v, \partial C) \leq r_C\}$ (this is the boundary of $C$ randomly contracted inward). Now for each $u \in S$ let $C_u \in Q_j$ be the cluster containing $u$ and define $\varphi_j^0(u) = d(u, \partial_{C_u}^*)$. Recall that $d(\cdot,\cdot)$ denotes distance in $G$, and thus $\varphi_j^0(u) = \max\{0, d(u, \partial C_u) - r_{C_u}\}$.

Clearly $\varphi_j^0(u)$ is a contraction, since $(u,v) \in E$ implies that either $u$ and $v$ are in the same cluster $C \in Q_j$ and then $|d(u, \partial_C^*) - d(v, \partial_C^*)| \leq 1$, or each of them belongs to the boundary of its cluster and then $\varphi_j^0(u) = \varphi_j^0(v) = 0$. It is also clear that $u \in \partial S$ implies $u \in \partial C$ for some $C \in Q_j$ and hence $\varphi_j^0(u) = 0$. Thus $\varphi_j^0(\partial S) = 0$.

Now independently form $c\alpha\rho$ such coordinates (each time picking fresh values for the $r_C$) and let $\varphi_j$ be the direct sum of the resulting maps, where $c > 0$ is a sufficiently large constant to be determined later. Finally, set $\varphi = \varphi_1 \oplus \cdots \oplus \varphi_m$. From the properties of $\varphi_j$, we conclude that $\varphi : S \to \mathbb{Z}^{cm\rho}$ is a contraction which maps $\partial S$ to $\mathbf{0}$, i.e., a relative embedding.

Consider a pair $u,v$ with $d(u,v) > r$ and such that $B(u,3r^{1/\alpha}) \subseteq S$. It follows from property (2) of Definition 4.1 that there exists a partition $Q_j$ of $S$ and a subset $C \in Q_j$ for which $B(u,3r^{1/\alpha}) \subseteq C$. Since $d(u,v) > r$, the two vertices $u$ and $v$ must lie in different subsets of $Q_j$. It follows that, in any single coordinate $\varphi_j^0$ of the map $\varphi_j$, the value of $\varphi_j^0(u)$ is distributed uniformly over an interval of size $r^{1/\alpha}$ independently of the value $\varphi_j^0(v)$, hence $\Pr[\varphi_j^0(u) = \varphi_j^0(v)] \leq r^{-1/\alpha}$. Thus the probability that $u$ and $v$ collide in all $c\alpha\rho$ coordinates of $\varphi_j$ is $\Pr[\varphi_j(u) = \varphi_j(v)] \leq r^{-c\rho}$. Since $|S| \leq r^{O(\rho)}$, there are at most $r^{O(2\rho)}$ such pairs $u,v$, and hence the probability that there exists a pair that collides is at most $r^{O(2\rho)} r^{-c\rho} < 1/2$, if the constant $c$ is chosen to be sufficiently large. The existence of a map $\varphi$ satisfying the lemma follows. ∎

## 4.2. A first attempt

Here is a simple approach which will fail in the end, but will give some intuition as to how the padded decomposition will be used. We will use the padded decomposition (Definition 4.1) to decompose $G$ into layers of disjoint clusters. Using Section 4.1 we can then find a relative embedding for each cluster; glueing all these embedding together, we shall arrive at a good embedding for $G$. Note that the padded decomposition is being first to decompose the graph $G$ into clusters which will be separately embedded, and then *inside* each cluster to compute a good relative embedding for that cluster.

Let $k = \lceil \log\log \operatorname{diam}(G) \rceil$, and set $r_i = 2^{2^i}$ for $i \in \{1,\ldots,k\}$, and $r_0 = 0$. We apply an $r$-padded decomposition with $r = r_1,\ldots,r_k$. For each value of $r$, this decomposition will break the graph into clusters of diameter at most $r^\alpha$ such that every two vertices within a distance $r$ are contained in some such cluster $S$ and are "far" from the boundary of $S$ (as otherwise we cannot ensure that they are "separated" by the relative embedding for $S$).

**An embedding for one level.** Assume that $i > 1$. Let $\{P_1, P_2, \ldots, P_m\}$ be the partitions produced by the $r_i$-padded decomposition. We will show how to construct a contraction $\varphi_i : V \to \mathbb{Z}^{O(m^2\rho)}$ that satisfies: For every pair $u,v \in V$ with $r_{i-1} < d(u,v) \leq r_i$, we have $\varphi_i(u) \neq \varphi_i(v)$.

Fix a partition $P_j$ of $V$. For every cluster $S \in P_j$, compute a relative embedding $\psi_S : S \to \mathbb{Z}^{O(m\rho)}$ by applying Lemma 4.3 with the parameter $r$ set to $r_{i-1} = r_i^{1/2} \geq 2$. Note that the lemma is applicable since $\operatorname{diam}(S) \leq r_i^\alpha$

implies $|S| \le r_i^{\alpha\rho} = r_{i-1}^{O(\rho)}$. Now for every $u \in V$, set $\varphi_{ij}(u) = \psi_S(u)$ where $S \in P_j$ is the cluster containing $u$. Notice that this map is well-defined since $P_j$ is a partition of $V$. Also, notice that it is a contraction, for suppose $(u,v) \in E$. If $u$ and $v$ are in the same cluster $S$, then $\|\varphi_{ij}(u) - \varphi_{ij}(v)\| = \|\psi_S(u) - \psi_S(v)\| \le 1$ since $\psi_S$ is a relative embedding, and hence a contraction. If $u$ and $v$ are in different clusters, then $\varphi_{ij}(u) = \varphi_{ij}(v) = \mathbf{0}$ since both of $u$ and $v$ are on the boundary of their cluster. Finally, set $\varphi_i = \varphi_{i1} \oplus \cdots \oplus \varphi_{im}$.

Now consider some $u,v \in V$ with $r_{i-1} < d(u,v) \le r_i$. By property (2) of Definition 4.1, there exists some partition $P_j$ and a cluster $S \in P_j$ such that $B(u, 3r_i) \subseteq S$. Thus $u,v \in S$, and certainly $B(u, 3r_{i-1}^{1/\alpha}) \subseteq S$, so by Lemma 4.3, $\psi_S(u) \ne \psi_S(v)$. It follows that $\varphi_{ij}(u) \ne \varphi_{ij}(v)$, and hence $\varphi_i(u) \ne \varphi_i(v)$.

**The base case.** For $r = r_1 = O(1)$, we will construct $\varphi_1$ in a special way so that for all $u,v \in V$ with $0 < d(u,v) \le r_1$ we have $\varphi_1(u) \ne \varphi_1(v)$. We break $G$ into clusters using an $r_1$-padded decomposition as above, but then revert to a much simpler relative embedding technique: Given a cluster $S = \{v_1, v_2, \ldots, v_s\}$ with $\operatorname{diam}(S) \le r_1^{\alpha}$, define the relative embedding $\psi_S(v_i) = B(i)$ if $v_i \notin \partial S$ and $\psi_S(v_i) = \mathbf{0}$ otherwise, where $B(i)$ is the binary representation of $i$ as an $O(\rho)$-dimensional vector. Notice that the number of coordinates meets our needs, since $s = |S| \le r_1^{\alpha\rho} \le 2^{O(\rho)}$. This map is a contraction and satisfies $\psi_S(u) \ne \psi_S(v)$ whenever $u,v \notin \partial S$. Using this technique in the above argument (instead of Lemma 4.3) yields the desired map $\varphi_1$. In fact, this map uses only $O(m\rho)$ coordinates, so we append 0's to every image and extend it to $O(m^2\rho)$ coordinates.

**Putting it all together.** If we let $\varphi = \varphi_1 \oplus \varphi_2 \oplus \cdots \oplus \varphi_k$, we see that $\varphi$ is a contractive, injective embedding since for any distinct $u,v \in V$, the distance $d(u,v)$ falls into some range $r_{j-1} < d(u,v) \le r_i$ and thus $\varphi_j(u) \ne \varphi_j(v)$.

Assuming that for every $r \ge 4$ we can construct $r$-padded decompositions with $m$ layers, then each $\varphi_{ij}$ (which we obtained by applying Lemma 4.3) uses $O(m\rho)$ coordinates, and thus each $\varphi_i$ uses $O(m^2\rho)$ coordinates. It follows that the final embedding $\varphi$ uses $O(m^2\rho \log\log \operatorname{diam}(G))$ coordinates in all. It turns out that this bound is of the right form, except for the dependence on $\operatorname{diam}(G)$, so our next goal will be to eliminate this term. In the case of trees, we achieved this goal by exploiting some "untapped randomness", namely, after fixing a relative embedding for a level, we were still free to assign arbitrary weights to the open edges of that level. In the next section, we exploit a similar observation, namely that the boundary of a cluster need not be mapped to $\mathbf{0}$, because the edges running between clusters are still "open."

### 4.3. Forced nesting, contracted clusters, and untapped randomness

We improve over the preceding failed attempt by reusing the coordinates when proceeding inductively from finer partitions to coarser ones. Informally, the main idea is to consider every cluster of the finer partition as a single entity whose embedding is "rigid", achieved by contracting each such cluster into a single vertex. For this approach to work, we need the padded decompositions to be nested, achieved by a "forced nesting" technique. We introduce these two notions and then prove Theorem 4.2.

**Contracted clusters.** Suppose we have a relative embedding $\psi_C$ for each cluster $C$ in a partition $P$. Previously, we "glued" these embeddings by setting $\Psi(u)=\psi_C(u)$ where $C\in P$ is the cluster containing $u$. This yielded a contraction $\Psi$ defined on all of $V$ with the property that whenever $u,v$ belong to the same cluster $C\in P$, we have $\|\Psi(u)-\Psi(v)\|=\|\psi_C(u)-\psi_C(v)\|$. The following gives a simple way of maintaining this property, while allowing some freedom in choosing $\Psi$.

**Definition 4.4 (Contracted graph).** Let $P$ be a partition of $V$. The *contracted graph* (with respect to $P$) is the graph $\hat{G}=(\hat{V},\hat{E})$ obtained from $G$ by contracting, in the graph-theoretic sense, each cluster $C\in P$ to a single vertex.

Each cluster $C\in P$ corresponds to a vertex in $\hat{G}$, and vice versa. Hence, we may identify $\hat{V}=P$ and then $\hat{E}=\{(C_1,C_2):\exists(u_1,u_2)\in E, u_1\in C_1\in P, u_2\in C_2\in P\}$. We let $d_{\hat{G}}(\cdot,\cdot)$ be the shortest path metric in the contracted graph $\hat{G}$, and $B_{\hat{G}}(\cdot,\cdot)$ be a closed ball in $\hat{G}$. We shall keep using $d(\cdot,\cdot)$ and $B(\cdot,\cdot)$ when referring to the corresponding notions in the given graph $G$. A cluster $\hat{S}$ in $\hat{G}$ and its boundary $\partial\hat{S}$ are defined similarly to those in $G$.

**Lemma 4.5.** *Let $\hat{G}=(\hat{V},\hat{E})$ be the contracted graph with respect to a partition $P$ of $V$. Suppose that for each cluster $C\in P$ we have a relative embedding $\psi_C:C\to\mathbb{Z}^d$, and suppose that we have a contraction $\hat{\psi}:\hat{V}\to\mathbb{Z}^d$. For every $u\in V$, define the map $\Psi:V\to\mathbb{Z}^d$ as follows: $\Psi(u)=\hat{\psi}(C_u)+\psi_{C_u}(u)$, where $C_u\in P$ is the cluster containing $u$. Then $\Psi$ is a contraction and for all $u,v$ in the same cluster $C\in P$, we have $\|\Psi(u)-\Psi(v)\|=\|\psi_C(u)-\psi_C(v)\|$.*

**Proof.** By definition, for all $u,v\in C$ and $C\in P$, we have $\|\psi(u)-\psi(v)\|=\|\psi_C(u)-\psi_C(v)\|$. In particular, $\psi$ contracts every edge whose endpoints are in the same cluster $C$. For $(u,v)\in E$ where $u,v$ are in different clusters of $P$, we have $\|\psi(u)-\psi(v)\|=\|\hat{\psi}(C_u)-\hat{\psi}(C_v)\|\le 1$ because $\psi_P$ is a contraction and $(u,v)\in E$ implies $(C_u,C_v)\in\hat{E}$. ∎

**Forced nesting.** Given two partitions $P$ and $Q$ of $V$, we say that $P$ is a *refinement* of $Q$ if for all $C \in P$ and there exists $S \in Q$ such that $C \subseteq S$. It turns out that we can force such a nesting and remove only a negligible amount of padding.

**Lemma 4.6.** *Let $\mathcal{P} = \{P_1, \ldots, P_m\}$ be an $r$-padded decomposition of $G$, and $\tilde{\mathcal{P}} = \{\tilde{P}_1, \ldots, \tilde{P}_m\}$ be an $\tilde{r}$-padded decomposition of $G$, where $r \geq \tilde{r}^\alpha$. Then we can modify $\mathcal{P}$ so that, for every $j \in \{1, \ldots, m\}$, the partition $P_j$ is a refinement of $\tilde{P}_j$, with only a constant factor loss in the padding guarantee of Definition 4.1, namely condition (2) is replaced with:*

*(2') For every $u \in V$ there exists some $C \in \cup_{i=1}^m P_i$ such that $B(u, 2r) \subseteq C$.*

**Proof.** Whenever a cluster $C \in P_j$ contains only a portion of a cluster $C' \in \tilde{P}_j$ (i.e. $\emptyset \neq C \cap C' \neq C'$), modify $P_j$ by breaking $C$ into the cluster $C \setminus C'$ and a singleton cluster $\{u\}$ for every $u \in C \cap C'$. This process is continued (in an arbitrary order) until there is no such cluster $C$. When this happens, for every $j \in \{1, \ldots, m\}$, the partition $\tilde{P}_j$ is a refinement of $P_j$, as desired.

Notice that condition (1) of Definition 4.1 remains satisfied since each modification only decreases the diameter of clusters in $P_j$. Condition (2) is still satisfied if we replace $B(u, 3r)$ by $B(u, 3r - \tilde{r}^\alpha) \supseteq B(u, 2r)$, which completes the proof. ∎

We are now ready to prove Theorem 4.2. Here is a brief outline: We shall construct relative embeddings from the bottom up, i.e., proceeding inductively from finer partitions to coarser partitions. Once we find an embedding that is "good" for level $j$, we modify it so that it becomes good also for level $j - t$, for some constant $t$. This modification will involve constructing a random embedding of the contracted graph (essentially via Lemma 4.3) and combining it with the existing embedding (using Lemma 4.5). Repeating this $t = O(1)$ times, starting at levels $1, 2, \ldots, t$, respectively, will yield an embedding that uses only $O(tm^2\rho) = O(m^2\rho)$ coordinates.

**Proof of Theorem 4.2 (Embedding via decomposition).** Suppose that for every $4 \leq r \leq \mathrm{diam}(G)$, and every cluster $S \subseteq V$, there exists an $r$-padded decomposition of $S$ with $m$ layers. Let $\alpha > 1$ be the constant from Definition 4.1. As before, define $k = \lceil \log \log \mathrm{diam}(G) \rceil$, $r_i = 2^{2^i}$ for $i \in \{1, \ldots, k\}$, and $r_0 = 0$. For every $i \in \{1, \ldots, k\}$ let $\mathcal{P}_i = \{P_1^i, \ldots, P_m^i\}$ be the $r_i$-padded decomposition of $G$. Let $t = t(\alpha)$ be a positive integer to be chosen later (e.g., $t \geq \alpha^3$ suffices assuming $\alpha \geq 2$).

We first make sure that this series of decompositions $\mathcal{P}_1, \ldots, \mathcal{P}_k$ is nested. Iteratively, for $i = t+1, t+2, \ldots, k$, apply Lemma 4.6 to the padded decompositions $\mathcal{P}_i$ and $\mathcal{P}_{i-t}$. Thereafter, for every $j \in \{1, \ldots, m\}$, the partition $P_j^{i-t}$ is a refinement of $P_j^i$ with only a constant factor loss in the padding guarantee (which will be sufficient for what follows).

**Definition 4.7 (A separating map).** We say that $\varphi : V \to \mathbb{Z}^d$ *separates* a partition $P_j^i$ if, for all $u,v \in V$ such that $r_{i-1} < d(u,v) \le r_i$ and such that there is a cluster $S \in P_j^i$ with $B(u, 2r_i) \subseteq S$, we have $\varphi(u) \neq \varphi(v)$.

To prove the theorem it suffices to construct, for each $j \in \{1,\dots,m\}$ and each $i_0 \in \{1,\dots,t\}$, a contractive map $\varphi_{i_0,j} : V \to \mathbb{Z}^{O(m\rho)}$ that separates every partition $P_j^i$ for which $i \equiv i_0 \pmod t$. Indeed, the direct sum $\varphi = \bigoplus_{i_0,j} \varphi_{i_0,j}$ is a contractive embedding of $G$ into the $O(tm^2\rho)$-dimensional lattice, which separates every $P_i^j$ and is thus injective. Since $t$ is a constant, it would follow that $\dim(G) = O(m^2\rho)$.

Fixing $j \in \{1,\dots,m\}$ and $i_0 \in \{1,\dots,t\}$, it remains to construct only the map $\varphi_{i_0,j}$ satisfying the stated properties. We shall actually prove, by induction on $i$, a more general assertion: If $c = c(t,\alpha)$ is a sufficiently large constant, then for every $i \in \{1,\dots,k\}$ with $i \equiv i_0 \pmod t$, there exists a contractive map $\psi^i : V \to \mathbb{Z}^{cm\rho}$ such that

i. $\psi^i$ separates each of $P_j^{i_0}, P_j^{i_0+t}, \dots, P_j^i$, and
ii. the restriction of $\psi^i$ to any cluster $C \in P_j^i$ is a relative embedding of $C$.

**The base case.** For $i = i_0$ we have $r_{i_0} \le r_t = O(1)$, and thus we can use the base case from Section 4.2 to generate a map $\psi^{i_0}$ that separates $P_j^{i_0}$. This map need only use $O(\rho)$ coordinates, since we have only a single partition $P_j^{i_0}$ at hand rather than $m$ partitions, but we can extend the map to use $cm\rho$ coordinates by appending 0's to every image.

**The inductive step.** Suppose we have a contractive map $\psi^i$ which satisfies properties (i) and (ii), and let us construct $\psi^{i+t}$ using Lemma 4.5. Define $\hat{G} = (\hat{V}, \hat{E})$ as the contracted graph of $G$ with respect to the partition $P_j^i$. For every cluster $C \in P_j^i$, let $\psi_C : C \to \mathbb{Z}^{cm\rho}$ be the restriction of $\psi_i$ to $C$; by the induction hypothesis, $\psi_C$ is a relative embedding of $C$. It remains to define an embedding $\hat{\psi} : \hat{V} \to \mathbb{Z}^{cm\rho}$, and then we can let $\psi^{i+t} : V \to \mathbb{Z}^{cm\rho}$ be the map yielded by Lemma 4.5.

Since $\hat{V} = P_j^i$ is a refinement of $P_j^{i+t}$, the latter naturally yields a partition $\hat{P}_j^{i+t}$ of $\hat{V}$, as follows. With every cluster $S \in P_j^{i+t}$ we associate a cluster $\hat{S} = \{C \in \hat{V} : C \subseteq S\}$ in $\hat{G}$. It is then easy to verify that $\hat{P}_j^{i+t} = \{\hat{S} : S \in P_j^{i+t}\}$ is a partition of $\hat{V}$.

Below, we shall define the mapping $\hat{\psi}$ on a single cluster $\hat{S} \in \hat{P}_j^{i+t}$ in $\hat{G}$. By construction, $\hat{\psi}$ will be a relative embedding of that cluster $\hat{S}$ (with respect to $\hat{G}$) i.e. $\hat{\psi}(\partial\hat{S})=\mathbf{0}$, and thus the resulting embedding will necessarily be a contractive embedding of the entire $\hat{V}$. In fact, we shall define $\hat{\psi}$ on each $\hat{S}$ randomly.

Fix a cluster $S\in P_j^{i+t}$ and the corresponding $\hat{S}\in\hat{P}_j^{i+t}$. Now construct an "inner decomposition" for $\hat{S}$, as follows: Take an $m$-layer $r_{i+t}^{1/\alpha}$-inner decomposition of $G$ (note that $r_{i+t}^{1/\alpha^2}\geq 4$), and force $P_j^i$ to be a refinement of each layer of this inner decomposition using Lemma 4.6. The forced nesting procedure is applicable because $r_{i+t}^{1/\alpha^2}\geq r_i^{\alpha}$. This procedure modifies only the inner decomposition (and not $P_j^i$), changing its padding guarantee to $2r_{i+t}^{1/\alpha^2}$. Each layer of the inner decomposition is a partition of $V$, and hence induces a partition of $S$. Denote these $m$ partitions of $S$ by $Q_1,\dots,Q_m$. Since $P_j^i$ is a refinement of the inner decomposition, each partition $Q_l$ of $S$ naturally yields a partition of $\hat{S}$ which we shall denote by $\hat{Q}_l$. (This is similar to the way $\hat{P}_j^{i+t}$ was defined.)

We generate the map $\hat{\psi}:\hat{S}\to\mathbb{Z}^{c\alpha m\rho}$ in a random fashion, using an argument similar to Lemma 4.3[1]. Fix a partition $Q_l$ and form a single coordinate $f:\hat{S}\to\mathbb{Z}$ randomly as follows: For every $C\in\hat{Q}_l$ choose a value $r_C\in\{0,1,\dots,r_i\}$ uniformly at random and let $f(w)=\max\{0,d_{\hat{G}}(w,\partial\hat{Q}_l)\}$. Now independently form $c\alpha\rho$ such coordinates (each time picking fresh values for the $r_C$), and let $\hat{\psi}_l:\hat{S}\to\mathbb{Z}^{c\alpha\rho}$ be the direct sum of the resulting maps. We apply the above to every partition $Q_l$ and set $\hat{\psi}=\hat{\psi}_1\bigoplus\cdots\bigoplus\hat{\psi}_m$.

Notice that once $\hat{\psi}$ is defined on $\hat{S}$, the map $\psi^{i+t}$ yielded by Lemma 4.5 is defined on $S$. Since the former map is constructed in a random fashion, the latter mapping is randomized as well. The next two lemmas analyze this randomized embedding of $S$.

**Claim 4.8.** *Fix $S\in P_j^{i+t}$. Then $\psi^{i+t}:S\to\mathbb{Z}^{c\alpha m\rho}$ is a relative embedding of $S$.*

**Proof.** Let us first show that $\hat{\psi}:\hat{S}\to\mathbb{Z}^{c\alpha\rho}$ is a relative embedding of $\hat{S}$ (with respect to $\hat{G}$). Indeed, it is easy to see that every coordinate $f$ generated as above using some partition $\hat{Q}_l$ is a contraction, and that for every vertex $w\in\partial\hat{S}$, we have $w\in\partial\hat{Q}_l$ and hence $f(w)=0$. It follows that $\hat{\psi}$, which is a direct sum of such maps $f$, is a relative embedding of $\hat{S}$.

Now consider $\psi^{i+t}$; we know it is a contraction from Lemma 4.5, so we only need to show that $\psi^{i+t}(\partial S)=\mathbf{0}$. Fix $u\in\partial S$ and let $C_u\in P_j^i$ be the

[1] Applying this lemma directly to $\hat{S}$ would generate $\hat{\psi}$ such that $\hat{\psi}(C)\neq\hat{\psi}(C')$ for certain $C,C'\in\hat{V}$, but it does not guarantee that as a result $\psi^{i+t}(u)\neq\psi^{i+t}(v)$ for suitable $u\in C$, $v\in C'$.

cluster containing $u$. Then $u \in \partial C_u$ (recall that $C_u \subseteq S$) and thus $\psi_{C_u}(u) = \mathbf{0}$. It also follows that $C_u \in \partial \hat{S}$, and hence $\hat{\psi}(C_u) = \mathbf{0}$. By definition, $\psi^{i+t}(u) = \hat{\psi}(C_u) + \psi_{C_u}(u) = \mathbf{0}$, which proves the claim. ∎

**Claim 4.9.** *Fix $S \in P_j^{i+t}$. Then with probability at least $1/2$, for all $u, v \in S$ with $r_{i+t-1} < d(u,v) \le r_{i+t}$ and with $B(u, 2r_{i+t}) \subseteq S$, we have $\psi^{i+t}(u) \neq \psi^{i+t}(v)$.*

**Proof.** Fix $u, v \in S$ with $r_{i+t-1} < d(u,v) \le r_{i+t}$ and $B(u, 2r_{i+t}) \subseteq S$. Let $C_u \in P_j^i$ be the cluster containing $u$, and let $C_v \in P_j^i$ be similarly for $v$. Clearly, $u, v \in S \in P_j^{i+t}$; recalling that $P_j^i$ is a refinement of $P_j^{i+t}$ we get that $C_u, C_v \in \hat{S}$.

The ball $B = B(u, 2r_{i+t}^{1/\alpha^2})$ is contained in some cluster of the inner decomposition, even after the forced nesting with $P_j^i$. By the above, $B \subseteq S$ and thus $B$ is entirely contained also in some cluster of some partition $Q_l$ of $S$. Since clusters in $P_j^i$ have diameter at most $r^\alpha$ (in $G$), we have:

(a) $r_{i+t-1}/r_i^\alpha < d_{\hat{G}}(C_u, C_v) \le r_{i+t}$,

(b) $B_{\hat{G}}(C_u, 2r_{i+t}^{1/\alpha^2}/r_i^\alpha)$ is entirely contained in some cluster of $\hat{Q}_l$.

Now consider $\hat{\psi}_l(C_u)$ and $\hat{\psi}_l(C_v)$. Clusters in $\hat{Q}_l$ have diameter (in $\hat{G}$) at most $r_{i+t}^{1/\alpha} \le r_{i+t-1}/r_i^\alpha$, so (a) implies that $C_u$ and $C_v$ reside in different clusters of $\hat{Q}_l$. This, in conjunction with (b) (and because $2r_{i+t}^{1/\alpha^2}/r_i^\alpha \ge r_i$), implies that each coordinate of $\hat{\psi}_l(C_u)$ is distributed uniformly over an interval of size $r_i$, independently of the corresponding coordinate in $\hat{\psi}(C_v)$. Recalling that $\hat{\psi} = \hat{\psi}_1 \bigoplus \cdots \bigoplus \hat{\psi}_m$, $\psi^{i+t}(u) = \psi^i(u) + \hat{\psi}(C_u)$, and $\psi^{i+t}(v) = \psi^i(v) + \hat{\psi}(C_v)$, we get that each of at least $c\alpha\rho$ coordinates of $\psi^{i+t}(u)$ is distributed uniformly over an interval of size $r_i$, independently of the corresponding coordinate in $\psi^{i+t}(v)$. Therefore,

$$\Pr\left[\psi^{i+t}(u) = \psi^{i+t}(v)\right] \le r_i^{-c\alpha\rho}.$$

The number of pairs $u, v \in S$ is at most $|S|^2 \le (r_{i+t}^\alpha)^{2\rho} = r_i^{2^{t+1}\alpha\rho}$, the claim follows via a union bound if only we choose the constant $c \ge 2^{t+2}$. ∎

As mentioned before, we use the preceding claim to construct for every $S \in P_j^{i+t}$ a map $\hat{\psi} : \hat{S} \to \mathbb{Z}^{O(m\rho)}$. This collection of maps gives an embedding $\hat{\psi} : \hat{V} \to \mathbb{Z}^{O(m\rho)}$, and using Lemma 4.5 we produce from it an embedding $\psi^{i+t} : V \to \mathbb{Z}^{O(m\rho)}$ that satisfies the induction hypothesis for $i + t$, and this completes the proof of Theorem 4.2. ∎

### 4.4. Graphs excluding a fixed minor

Let $G$ be a graph that excludes a $K_{s,s}$ minor for some fixed $s$. By adapting a decomposition technique of Klein, Plotkin, and Rao [14] we construct, for any value $r \geq 1$, an $r$-padded decomposition of $G$ with only $O(2^s)$ layers. Applying Theorem 4.2, we then arrive at the main result of this section.

**Theorem 4.10 (Excluded minor families).** *Conjecture 1 is true for any family of graphs that excludes a fixed minor. For such graphs,* $\dim(G) = O(\rho_G)$.

**Proof.** By Theorem 4.2 it suffices to show how to produce an $r$-padded decomposition with $m = 2^s$ layers for any graph that excludes a $K_{s,s}$ minor, as this implies that $\dim(G) \leq O(4^s \rho_G)$.

To this end, consider such a graph $G = (V, E)$ and fix a value $r$. First, construct a Breadth-First-Search (BFS) tree from an arbitrary vertex $v \in V$, and compute for every vertex its BFS level (i.e., its distance from $v$). Then cut the tree every $12r$ BFS levels by removing any edge connecting a vertex of BFS-level $j$ to a vertex of BFS-level $j+1$ for any $j \equiv 0 \pmod{12r}$. Let $C_0$ denote the resulting set of connected components of $G$. Next, take another copy of $G$ and cut it similarly but starting from BFS level $6r$, i.e., remove any edge connecting a vertex of BFS-level $j$ to a vertex of BFS-level $j+1$ for any $j \equiv 6r \pmod{12r}$. Let $C_1$ denote the resulting set of connected components of $G$. Now for each $C_i$, apply the same procedure on all the connected components in $C_i$, namely, choose an arbitrary vertex, construct a BFS tree, and form two sets of connected components $C_{i0}$ and $C_{i1}$ by making the cuts as above (staggered, each at intervals of size $12r$). Repeat this process $s$ times and let $C_q$ for $q \in \{0,1\}^s$ denote the $2^s$ final sets of connected components. From [14] we know that every connected component in every final $C_q$ has diameter at most $O(r)$. In addition, for every vertex $u$ the entire ball $B(u, 3r)$ is uncut in at least one final set $C_q$, because each step performs two different sets of "staggered" cuts, at least one of which must avoid the entire ball $B(u, 3r)$. ∎

## 5. A general dimension upper bound

In this section we give a tight upper bound on the dimension of general graphs: $\dim(G) = O(\rho_G \log \rho_G)$ for any graph $G$. (In Section 3, we showed that this upper bound is met by expanders.) First, we devise a decomposition for growth-restricted metrics (Section 5.1) and use Theorem 4.2 to obtain a weaker upper bound of $O(\rho_G^3)$ (Section 5.2). Then, by combining the previous arguments more carefully and utilizing some Chernoff-type tail bounds, we obtain the aforementioned tight upper bound (Section 5.3). We shall use some terminology from Section 5.2.

### 5.1. Partitioning of growth-restricted graphs

Linial and Saks [17] and Bartal [3] show that for any graph $G=(V,E)$ and $1 \le r \le \mathrm{diam}(G)$, there exists a probabilistic partitioning of $G$ into disjoint clusters of diameter at most $O(r\ln|V|)$, such that for any pair of vertices $u,v\in V$, the probability that $u$ and $v$ end up in different clusters is at most $d(u,v)/r$. Let $\rho=\rho_G$. In this section, we give a similar decomposition, but we replace the diameter bound of $O(r\ln|V|)$ with a bound that is independent of $|V|$, namely $O(\rho r\ln r)$, for any $r\ge\rho$. Our partitioning method is similar to those of [17] and [3], but different in a subtle and crucial way: It is local. Events which are sufficiently far apart are mutually independent.

First, take the continuous exponential distribution with mean $r>0$, truncate it at some value $M>0$ and rescale the remaining density function. The resulting distribution, which we denote $\mathrm{Texp}(r,M)$, has density function $p(z)=\frac{e^{M/r}}{r(e^{M/r}-1)}e^{-z/r}$ for $z\in(0,M)$.

**The partitioning procedure.** Let $V=\{v_1,v_2,\dots,v_n\}$ and let $r\ge\rho$. For each $v_t\in V$, choose independently a radius $r_t$ according to the distribution $\mathrm{Texp}(r,8\rho r\ln r)$. Now define $S_t=B(v_t,r_t)\setminus\cup_{i=1}^{t-1}B(v_i,r_i)$ as the set of vertices $v$ for which $B(v_t,r_t)$ is the first ball containing $v$. Finally, define the set of clusters to be $\mathcal{C}=\{S_1,\dots,S_n\}$.

It is easy to see that $\mathcal{C}$ is a partition of $V$, and that the (weak) diameter of every cluster $C\in\mathcal{C}$ is bounded by $\mathrm{diam}(C)\le 16\rho r\ln r$. Further analysis will require the following simple facts. In particular, (3) shows that if $M\ge 2r$, the truncated exponential distribution is "almost" memoryless.

**Fact 5.1.** *Consider a random variable* $R\sim\mathrm{Texp}(r,M)$ *for* $M\ge 2r>0$. *Then:*

1. *For all* $\beta\ge0$, $\Pr[R\ge\beta]\le 2e^{-\beta/r}$.
2. *For all* $\beta\ge0$, $\Pr[R\le\beta]\le 2(1-e^{-\beta/r})\le 2\beta/r$.
3. *For all* $\beta\ge0$ *and* $R_0\le M/2$, $\Pr[R\le R_0+\beta\,|\,R\ge R_0]\le 2\beta/r$.

For a vertex $u\in V$ and $x\ge0$, let $\mathcal{E}_u^x$ be the event that $B(u,x)$ is split between multiple clusters, i.e., that no cluster $C\in\mathcal{C}$ fully contains $B(u,x)$.

**Lemma 5.2.** *Let* $u\in V$ *and* $r\ge16\rho$, *and* $x\ge0$. *Then* $\Pr[\mathcal{E}_u^x]\le 10x/r$.

**Proof.** Assume $x \le r$ (the theorem says nothing for larger $x$) and let $\mathcal{B} = B(u,x)$, $B_t = B(v_t, r_t)$. Let us say that the ball $\mathcal{B}$ is *cut by the ball* $B_t$ if $\emptyset \ne S_t \cap \mathcal{B} \ne \mathcal{B}$ while for all $i<t$, $S_i \cap \mathcal{B} = \emptyset$. $\mathcal{E}_u^x$ is precisely the event that there is a ball $B_t$ that cuts $\mathcal{B}$. Let us separate these balls $B_t$ into two classes, depending on the distance from $v_t$ to $u$. Define $\mathcal{E}_{\text{far}}$ to be the event that there exists $B_t$ that cuts $\mathcal{B}$ and $d(v_t,u) \ge 4\rho r \ln r$. Define $\mathcal{E}_{\text{near}}$ to be the event that there exists $B_t$ that cuts $\mathcal{B}$ and $d(v_t,u) < 4\rho r \ln r$.

Fix $v_t$ with $d(v_t,u) \ge 4\rho r \ln r$ and notice that by Fact 5.1,

$$\Pr[B_t \text{ cuts } \mathcal{B}] \le \Pr[r_t \ge 4\rho r \ln r - x] \le 2r^{-4\rho} e^{x/r} \le 6r^{-4\rho}.$$

But the number of such $v_t$ for which $B_t$ can possibly cut $\mathcal{B}$ is at most the number of vertices in a ball of radius $8\rho r \ln r + x \le r^3$ which is at most $r^{3\rho}$. Taking a union bound over all such possible $v_t$, we see that $\Pr[\mathcal{E}_{\text{far}}] \le 6r^{-4\rho} r^{3\rho} \le 6/r^\rho \le 6/r$. Thus we are left only to bound the probability of $\mathcal{E}_{\text{near}}$.

Let the random variable $T$ be the minimum $t$ such that $B_T \cap \mathcal{B} \ne \emptyset$ (note that possibly $v_T \in \mathcal{B}$). The ball $B_T$ can either cut $\mathcal{B}$ (in which case $\mathcal{E}_u^x$ occurs) or contain $\mathcal{B}$ (and then $\mathcal{B} \subseteq S_T$ is not cut by any ball $B_t$). By the principle of deferred decision it suffices to upper bound the conditional probability $\Pr[\mathcal{E}_{\text{near}} | T = t]$ for an arbitrary $t$. To this end, we may assume that $d(v_t,u) \le 4\rho r \ln r$ (as otherwise this conditional probability is 0) and then $\mathcal{E}_{\text{near}}$ happens if and only if $B_t$ cuts $\mathcal{B}$, which in turn happens only if $r_t \le d(v_t,u) + x$. Hence,

$$\Pr[\mathcal{E}_{\text{near}} \,|\, T = t] \le \Pr\left[r_t \le d(v_t,u) + x \,|\, r_t \ge d(v_t,u) - x\right] \le \frac{4x}{r},$$

where we have used Fact 5.1 in conjunction with $d(v_t,u) \le 4\rho r \ln r$. Thus, $\Pr[\mathcal{E}_{\text{near}}] = \sum_t \Pr[T = t] \cdot \Pr[\mathcal{E}_{\text{near}} | T = t] \le 4x/r$ and $\Pr[\mathcal{E}_u^x] \le \Pr[\mathcal{E}_{\text{near}}] + \Pr[\mathcal{E}_{\text{far}}] \le 10x/r$. ∎

### 5.2. Layered decomposition of growth-restricted graphs

Now we describe how to obtain an $r$-padded decomposition with $O(\rho_G)$ layers for general graphs $G$. Plugging these values into Theorem 4.2 yields an embedding into $O(\rho_G^3)$ dimensions. We will only be able to show the existence of such decompositions under the assumption that $r \ge \rho$. In the case where $r \le 16\rho$, clusters of diameter $r^{O(1)}$ have at most $\rho^{O(\rho)}$ points, so we will be able to embed these by brute force using only $O(\rho \log \rho)$ dimensions (similar to the base case of Section 4.2). The final result appears in Theorem 5.5.

**Theorem 5.3** **(Decomposition theorem).** *For every graph $G=(V,E)$ and every $r \geq 16\rho_G$, there exists an $r$-padded decomposition with $m=O(\rho_G)$ layers.*

**Proof.** Let $\rho=\rho_G$ and assume $r \geq 16\rho$. To produce a single layer of the decomposition (a partition of $V$ into clusters), we will use the procedure of Section 5.1, with the parameter $r$ (in that procedure and in Lemma 5.2) set to $r^2$. Notice that the clusters produced have diameter at most $32\rho r^2 \ln r \leq r^4$. For a vertex $v \in V$, let $\mathcal{E}_v$ be the event that the ball of radius $3r$ about $v$ is cut (i.e., split amongst two or more clusters). From Lemma 5.2, we know that $\Pr[\mathcal{E}_v] \leq 3/r$.

Now produce $m$ layers independently (with fresh random coins each time) and let $\mathcal{E}_v^m$ be the event that the ball of radius $3r$ about $v$ is cut in *every* layer. Clearly $\Pr[\mathcal{E}_v^m] \leq (3/r)^m$. We would like to say that $\Pr\left[\bigwedge_{v\in V} \overline{\mathcal{E}_v^m}\right] > 0$. If we could show this with $m=O(\rho)$, the theorem would follow. And indeed, this is our goal. We will employ the following symmetric form of the Lovász Local Lemma, see e.g. [1].

**Lemma 5.4** **(Lovász Local Lemma).** *Let $A_1,\ldots,A_n$ be events in an arbitrary probability space. Suppose that for each $A_i$ there is a set that contains all but at most $d$ of the other events $A_j$, such that $A_i$ is mutually independent of this set of events. If for all $i \in \{1,\ldots,n\}$ we have $\Pr[A_i] \leq p$, and $ep(d+1) \leq 1$, then $\Pr[\wedge_{i=1}^n \overline{A_i}] > 0$.*

Let $r_1 = 2r^3 \ln r + 6r$. An event $\mathcal{E}_u^m$ is mutually independent of all events $\mathcal{E}_v^m$ for which $d(u,v) > r_1$ because every ball in the partitioning of Section 5.1 has radius at most $r^3 \ln r$ and thus cannot intersect both $B(u,3r)$ and $B(v,3r)$. It follows that $\mathcal{E}_u^m$ is mutually independent of the set of all events $\mathcal{E}_v^m$ except those for which $v \in B(u,r_1)$, and there are at most $d=r_1^\rho$ such vertices $v$. Thus if $\Pr[\mathcal{E}_u^m] \leq 1/e(d+1)$, we can apply the local lemma and the theorem is proved. But this is easily accomplished by choosing say $m=\lceil 8\rho \rceil$. By applying Lemma 5.4 we conclude that there exists an $r$-padded decomposition for $V$ (with $\alpha=4$). ∎

**Theorem 5.5.** *For every graph $G$ with growth rate $\rho_G$, $\dim(G)=O(\rho_G^3)$.*

**Proof (sketch).** We provide only a sketch of the proof, since a better upper bound is given in the next section. We use the proof of Theorem 4.2, except that instead of the induction's base case being $r=O(1)$, we start with a level corresponding to $r=\rho^{O(1)}$. In this case, clusters have diameter $r^{O(1)} = \rho^{O(1)}$, so we can easily give a relative embedding for each cluster using only $O(\rho \log \rho)$ coordinates, similar to the base case of 4.2. The rest of the proof then proceeds unchanged, including the inductive step in which we use the decomposition provided by Theorem 5.3. ∎

**Remark: Algorithmiziation of the Local Lemma.** Although most of the techniques in this paper can easily be interpreted algorithmically, this is not immediately obvious for our use of the local lemma. Fortunately, it is not difficult to see that standard techniques suffice (see, e.g., [1, Chapter 5]).

### 5.3. A tight upper bound

As mentioned previously, Theorem 5.3, combined with Theorem 4.2, shows that $\dim(G)=O(\rho_G^3)$ for every graph $G$. By carefully combining the previous arguments and utilizing some Chernoff-type tail bounds, we are able to find a tight upper bound, $\dim(G)=O(\rho_G \log \rho_G)$; see Theorem 5.8 below.

We first strengthen the decomposition of Theorem 5.3. Given $m$ layers (partitions of $V$) $P_1, P_2, \ldots, P_m$, we say that a vertex $u \in V$ is *padded* in a layer $j$ if there exists a cluster $C \in P_j$ such that $B(u,3r) \subseteq C$ (otherwise, we say that $u$ is unpadded in layer $j$). We next show a decomposition in which every vertex is padded in most of the layers (rather than in one layer).

**Theorem 5.6 (Strengthened decomposition theorem).** *For every graph $G=(V,E)$ and every $r \geq 36\rho_G$, there exists an $r$-padded decomposition with $m=O(\rho_G)$ layers, in which:*

*(2") For every $u \in V$ there are $\frac{3}{4}m$ partitions $P_j$ in which there is $C \in P_j$ with $B(u,3r) \subseteq C$.*

**Proof.** Similar to the proof of Theorem 5.3, we construct $m=O(\rho_G)$ layers of randomized partitions that always satisfy requirement (1), and argue that, with positive probability, requirement (2") is satisfied. The probability that $u$ is unpadded in a single layer is at most $3/r$ (this followed from Lemma 5.2), so the expected number of layers in which $u$ is unpadded is at most $3m/r$. We now need the following Chernoff-type tail bound (see, e.g., [20, Chapter 4]).

**Lemma 5.7 (A tail bound).** *Let $X_1, X_2, \ldots, X_n$ be independent Poisson trials such that, for $1 \leq i \leq n$, $\Pr[X_i=1]=p_i$ and $0<p_i<1$. Then for $X=\sum_i X_i$, $\mu=E[X]$, and any $\delta>0$,*

$$\Pr[X>(1+\delta)\mu] < \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu < \left(\frac{e}{1+\delta}\right)^{(1+\delta)\mu}.$$

Let $m_u$ be the expected number of layers in which a vertex $u$ is unpadded, and let $\mathcal{E}_u$ be the event that $u$ is unpadded in more than $\frac{1}{4}m$ layers. Then, applying the above lemma,

$$\Pr[\mathcal{E}_u] = \Pr\left[Y_u > \frac{r}{12}\cdot\frac{3m}{r}\right] \le \frac{1}{(r/12e)^{m/4}}.$$

Applying Lemma 5.4 the same way we did in the proof of Theorem 5.3 for a suitable $m = O(\rho_G)$, we conclude that $\Pr\left[\bigwedge_{v\in V}\overline{\mathcal{E}_v^m}\right] > 0$. ∎

**Theorem 5.8 (Embedding of general graphs).** *For every graph $G = (V,E)$ with growth rate $\rho_G$, $\dim(G) = O(\rho_G \log\rho_G)$.*

**Proof.** We adapt the proof of Theorem 4.2, as follows. First, we use the strengthened decomposition from Theorem 5.6. Second, we employ a more careful analysis that exploits the independence of coordinates constructed from different layers together with the local lemma, instead of applying a union bound separately in each cluster.

In the sequel, we describe the modifications to that proof. In fact, we will only show that any one level (in the sense of Section 4.2) can be embedded into $O(\rho\log\rho)$ dimensions. Using the nesting techniques of Section 4.3, the existence of a contractive and injective embedding that uses only $O(\rho\log\rho)$ coordinates follows.

Let $\rho = \rho_G$. Fix $r = 2^{2^i}$ and suppose we are given an $r$-padded decomposition with $m = O(\rho)$ layers $P_1, P_2, \dots, P_m$ strengthened as per Theorem 5.6. We may assume further that $m \ge \rho$, because otherwise we can just duplicate every layer $\lceil\rho/m\rceil$ times. Let $c > 0$ be a constant to be determined later. We shall construct, for each cluster $S \in P_j$, a relative embedding $\psi_S : S \to \mathbb{Z}^{c\log\rho}$, such that for all $u,v \in V$ with $\sqrt{r} < d(u,v) \le r$ there exist a layer $P_j$ and a cluster $S \in P_j$ such that $u,v \in S$ and $\psi_S(u) \ne \psi_S(v)$. Letting $\varphi_j(u) = \psi_{S_u}(u)$ where $S_u \in P_j$ is the cluster containing $u$, and setting $\varphi = \oplus_{j=1}^m \varphi_j$, we will conclude that $\varphi(u) \ne \varphi(v)$ for all $u,v \in V$ with $\sqrt{r} < d(u,v) \le r$. As before, in the base case we shall replace the requirement $\sqrt{r} < d(u,v) \le r$ with $0 < d(u,v) \le r$.

**The base case.** Assume $16\rho \le r \le \rho^{O(1)}$ (note that this is where our decomposition breaks down). We produce, for every cluster $S \in P_j$, a relative embedding $\psi_S : S \to \mathbb{Z}^{c\log\rho}$ as follows: If $u \notin \partial S$, then let $\psi_S(u)$ be a $(c\log\rho)$-dimensional vector chosen uniformly at random from $\{0,1\}^{c\log\rho}$, and let $\psi_S(u) = \mathbf{0}$ otherwise. This $\psi_S$ is clearly a contraction.

Consider two vertices $u,v$ with $0 < d(u,v) \le r$ and let $P_j$ be a layer in which $u$ is padded. In this layer, $u$ and $v$ belong to the same cluster $S \in P_j$

and $u \notin \partial S$, so $\Pr[\psi_S(u) = \psi_S(v)] \le (\frac{1}{2})^{c \log \rho} = \rho^{-c}$. Let $\mathcal{E}_{u,v}$ be the event that, in the resulting embedding $\varphi = \oplus_{j=1}^m \varphi_j$, we have $\varphi(u) = \varphi(v)$. For this event to happen, it must be that in every layer in which $u$ is padded, we have $\psi_S(u) = \psi_S(v)$. It then follows that

$$\Pr[\mathcal{E}_{u,v}] = \Pr[\varphi(u) = \varphi(v)] \le \rho^{-cm/2} \le \rho^{-c\rho/2}.$$

There are $\Omega(|V|)$ events $\mathcal{E}_{u,v}$ and we would like to argue that with positive probability, none of them occur. Again, the local lemma comes to our rescue. It is not difficult to see that $\mathcal{E}_{u,v}$ is independent of all events $\mathcal{E}_{u',v'}$ for which $d(u,u') > r$ (because the image of $u$ is chosen independently of the images of all the corresponding $v$, $u'$, and $v'$). It follows that $\mathcal{E}_{u,v}$ is mutually independent of all but at most $d = r^{2\rho} \le \rho^{O(\rho)}$ other events. Choosing the constant $c$ to be a large enough relative to the constants in the bound $r \le \rho^{O(1)}$, we see that $\Pr[\mathcal{E}_{u,v}] \le \rho^{-c\rho/2} \le 1/e(d+1)$. Thus, applying Lemma 5.4 yields an embedding for which none of the events $\mathcal{E}_{u,v}$ occur.

**Higher levels.** Assume $r \ge (16\rho)^2$. Recall that in Theorem 4.2, for each layer $P_j$ and each cluster $S \in P_j$, we produced a random relative embedding $\psi_S : S \to \mathbb{Z}^{O(m \log \rho)}$ by constructing (using Lemma 4.3) $O(\log \rho)$ coordinates from each layer of an $m$-layer $r^{1/2}$-inner decomposition. Here, instead, for each $S \in P_j$ we shall produce a random relative embedding $\psi_S : S \to \mathbb{Z}^c$ (which is of course stronger than using $O(\log \rho)$ coordinates), by constructing only $c = O(1)$ coordinates from layer $j$ of the inner decomposition. More formally, take an $m$-layer $r^{1/2}$-inner decomposition of $G$. Now for each $P_j$ and each cluster $S \in P_j$, construct each of the $c$ coordinates of $\psi_S$ as follows (this is analogous to Lemma 4.3): Let $Q_j$ be the partition of $S$ induced by layer $j$ of the inner decomposition. For every cluster $C \in Q_j$, choose $r_C \in \{0, 1, \ldots, r^{1/2\alpha}\}$ uniformly at random. For every $u \in S$, let $C_u \in Q_j$ be the cluster containing $u$, and define the image of $u$ to be $\max\{0, d(u, \partial C_u) - r_{C_u}\}$. Clearly, $\psi_S$ is a relative embedding of $S$.

Consider two vertices $u, v$ with $\sqrt{r} < d(u,v) \le r$. Since $u$ is padded in at least $\frac{3}{4}m$ layers each decomposition (i.e., the layers $\{P_1, \ldots, P_m\}$ and in the inner decomposition), we see that for at least $\frac{1}{2}m$ values of $j$, the vertex $u$ is padded both in $P_j$ and in layer $j$ of the inner decomposition. Fix such $j$, and let $S \in P_j$ be the cluster containing $u$. It follows that $u, v$ belong to the same cluster $S \in P_j$, but to different clusters of the inner decomposition. Since $u$ is padded in layer $j$ of the inner decomposition, each coordinate of $\psi_S(u)$ is chosen at random from an interval of size at least $r^{-1/2\alpha}$, independently of $\psi_S(v)$, and thus the probability it collides with the corresponding coordinate of $\psi_S(v)$ is at most $r^{-1/2\alpha}$. Hence,

$$\Pr[\varphi_j(u) = \varphi_j(v)] = \Pr[\psi_S(u) = \psi_S(v)] \le r^{-c/2\alpha},$$

and the embedding $\varphi = \oplus_{j=1}^m \varphi_j$ satisfies $\Pr[\varphi(u) = \varphi(v)] \le r^{-cm/4\alpha} \le r^{-c\rho/4\alpha}$. Finally, we would like to apply Lemma 5.4 on the events $\mathcal{E}_{u,v} = \{\varphi(u) = \varphi(v)\}$ where $\sqrt{r} < d(u,v) \le r$. It can be seen that every event $\mathcal{E}_{u,v}$ is mutually independent of all the other events $\mathcal{E}_{u',v'}$ but the $r^{2\rho}$ events for which $d(u,u') \le 3r$ (because every cluster of the inner decomposition is mapped independently). Hence, for $c > 0$ a sufficiently large constant we can apply Lemma 5.4, which completes the embedding of a single level.

As mentioned before, the theorem follows by incorporating the nesting techniques of Section 4.3. It is straightforward to verify the details. ∎

## 6. Related notions of dimensionality

**Theorem 6.1 (Euclidean embeddings).** *The upper bounds for* $\dim(G)$ *in Theorems 5.8 and 4.10 hold also for* $\dim_2(G)$.

**Proof.** Consider a contractive, injective embedding $\varphi$ of $G = (V, E)$ into $\mathbb{Z}^d$ such that for some fixed $0 < \varepsilon < 1$, and every two distinct vertices $u, v$, their images $\varphi(u)$ and $\varphi(v)$ differ in at least $\varepsilon d$ coordinates. Our proof of Theorem 5.8 can be easily modified to yield such an embedding (for some universal constant $\varepsilon$) by applying appropriate Chernoff bounds when the coordinates are formed (see the application of Lemma 5.7 in Section 5.3, for instance). A similar modification to the proof of Theorem 4.10 results with $\varepsilon$ that depends only on the size $s$ of the excluded minor, namely, $\varepsilon \ge \Omega(1/4^s)$, because a modified proof of Theorem 4.2 guarantees that the images of every two distinct vertices are different in $\Omega(\rho)$, out of the $O(m^2\rho)$, coordinates.

After scaling $\varphi$ by $(\varepsilon d)^{-\frac{1}{2}}$, this embedding satisfies:

1. $\|\varphi(u) - \varphi(v)\|_2 \ge 1$ for all $u \ne v \in V$; and
2. $\|\varphi(u) - \varphi(v)\|_2 \le 1/\sqrt{\varepsilon}$ for all $(u,v) \in E$.

We now show that, for every fixed $0 < \nu < 1$, the constant $1/\sqrt{\varepsilon}$ can be reduced to $(1+\nu)^3$ for an arbitrarily small constant $\nu > 0$, while increasing the dimension $d$ only to $O(d)$, where the hidden constant depends only on $\varepsilon$ and $\nu$. Let $K = K(\varepsilon, \nu) \ge 1/\sqrt{\varepsilon}$ be determined later. For $x \in \mathbb{R}^d$, let $B_2(x, r) = \{y \in \mathbb{R}^d : \|x - y\|_2 < r\}$. Since, for every distinct $u, v \in V$, the balls $B_2(\varphi(u), \frac{1}{2})$ and $B_2(\varphi(v), \frac{1}{2})$ are disjoint, simple volume arguments show that, for all $u \in V$, $|\{v \in V : \|\varphi(v) - \varphi(u)\|_2 \le 2K\}| \le (c_1 K)^d$, where $c_1 > 0$ is a universal constant (constant independent of $\varepsilon$ and $\nu$).

We claim that there exists a map $\beta : V \to \mathbb{R}^{c_2 d}$, where $c_2 = c_2(K,\nu)$, satisfying

$$(*)\quad 1 \le \|\beta(u)-\beta(v)\|_2 \le 1+\nu \qquad \text{for } u \ne v \in V \text{ with } \|\varphi(u)-\varphi(v)\|_2 \le K.$$

Assuming the existence of such a map $\beta$, we arrive at our final map $\gamma : V \to \mathbb{R}^{O(d)}$ defined by the direct sum $\gamma = \frac{1}{K}\varphi \oplus \beta$, which satisfies

$$(1')\quad \|\gamma(u)-\gamma(v)\|_2 \ge 1 \qquad \text{for all } u \ne v \in V; \text{ and}$$
$$(2')\quad \|\gamma(u)-\gamma(v)\|_2 \le \sqrt{1/K^2\varepsilon + (1+\nu)^2} \qquad \text{for all } (u,v) \in E.$$

Choosing $K(\varepsilon,\nu)$ to be sufficiently large, say $K = 1/\sqrt{\varepsilon\nu}$, we see that the bound in (2') is at most $(1+\nu)^3$.

It remains now to prove the existence of such $\beta$. Let $c_2 = c_2(K,\nu)$ be determined later. We define $\beta(u)$ to be a random vector in $\mathbb{R}^{c_2 d}$, where each coordinate is chosen to be 0 or 1 uniformly at random. We argue that if $c_2$ is large enough, then with positive probability, the required condition $(*)$ holds, up to scaling, and this will complete the proof.

Let $Y_{uv}$ be the number of coordinates in which $\beta(u)$ and $\beta(v)$ disagree. Clearly $\mathbb{E}[Y_{uv}] = \frac{1}{2}c_2 d$. Furthermore, by standard Chernoff bounds, there exists $c_3 = c_3(\nu) > 0$ such that

$$\Pr\left[Y_{uv} < \frac{1}{1+\nu}\,\mathbb{E}[Y_{uv}] \;\text{ or }\; Y_{uv} > (1+\nu)\,\mathbb{E}[Y_{uv}]\right] \le e^{-c_3 c_2 d}.$$

Let $\mathcal{E}_u$ be the event that there exists some $\varphi(v) \in B_2(\varphi(u),K)$, for which either $Y_{uv} < \frac{1}{1+\nu}\mathbb{E}[Y_{uv}]$ or $Y_{uv} > (1+\nu)\mathbb{E}[Y_{uv}]$. Since $|\{v \in V : \|\varphi(v) - \varphi(u)\|_2 \le K\}| \le (c_1 K)^d$, we can choose $c_2$ to be large enough so that (by a union bound) $\Pr[\mathcal{E}_u] \le e^{-c_3 c_2 d/2}$. Finally, note that each event $\mathcal{E}_u$ is mutually independent of all events $\mathcal{E}_v$ for which $\varphi(v) \notin B_2(\varphi(u), 2K)$. The number of such events is again at most $(c_1 K)^d$, and hence choosing $c_2 > 0$ to be large enough and applying the local lemma (Lemma 5.4), we see that, with positive probability, no event $\mathcal{E}_u$ occurs. In this case, the map $\beta$ satisfies $\sqrt{\frac{1}{1+\nu} c_2 d/2} \le \|\beta(u) - \beta(v)\|_2 \le \sqrt{(1+\nu) c_2 d/2}$ whenever $u \ne v$ and $\|\varphi(u) - \varphi(v)\|_2 \le K$. Scaling this map proves the existence of the required $\beta$ and completes the proof of the theorem. ∎

**Theorem 6.2.** *If $G = (V,E)$ is a $\Theta(k)$-degree expander with $1 \le k \le \log|V|$, then $\dim_2(G) = \Omega\big(\frac{\log|V|}{\log\log_k \log|V|}\big)$. For a $\Theta(\log|V|)$-degree expander, $\dim_2(G) = \Omega(\rho_G \log \rho_G)$.*

**Proof.** Similar to that of Theorem 3.2. ∎

**Acknowledgments.** We are indebted to Nati Linial for his part in initiating this research. We are particularly grateful to him for valuable discussions in an early stage of the work and for his suggestion to focus on trees.

## References


[1] N. Alon and J. H. Spencer: *The probabilistic method*, Wiley-Interscience [John Wiley & Sons], New York, second edition, 2000.

[2] P. Assouad: Plongements lipschitziens dans $\mathbf{R}^n$, *Bull. Soc. Math. France* **111(4)** (1983), 429–448.

[3] Y. Bartal: Probabilistic approximation of metric spaces and its algorithmic applications, in *37th Annual Symposium on Foundations of Computer Science*, pages 184–193, IEEE, 1996.

[4] J. A. Bondy and U. S. R. Murty: *Graph theory with applications*, American Elsevier Publishing Co., Inc., New York, 1976.

[5] A. Brandstädt, V. Chepoi and F. Dragan: Distance approximating trees for chordal and dually chordal graphs, *J. Algorithms* **30(1)** (1999), 166–184.

[6] V. Chepoi and F. Dragan: A note on distance approximating trees in graphs, *European J. Combin.* **21(6)** (2000), 761–766.

[7] F. R. K. Chung: Labelings of graphs, in *Selected topics in graph theory*, **3**, pages 151–168, Academic Press, San Diego, CA, 1988.

[8] P. Erdős, F. Harary and W. T. Tutte: On the dimension of a graph, *Mathematika* **12** (1965), 118–122.

[9] U. Feige: Approximating the bandwidth via volume respecting embeddings, *J. Comput. System Sci.* **60(3)** (2000), 510–539.

[10] J. Fakcharoenphol and K. Talwar: An improved decomposition theorem for graphs excluding a fixed minor, in *Proceedings of 6th Workshop on Approximation, Randomization, and Combinatorial Optimization*, Springer Lecture Notes in Computer Science **2764**, 36–46, 2003.

[11] A. Gupta, R. Krauthgamer and J. R. Lee: Bounded geometries, fractals, and low-distortion embeddings; in *Proceedings of the 44th Annual Symposium on Foundations of Computer Science*, 2003.

[12] J. Heinonen: *Lectures on analysis on metric spaces*, Universitext, Springer-Verlag, New York, 2001.

[13] P. Indyk: Algorithmic applications of low-distortion geometric embeddings, in *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 10–33, October 2001.

[14] P. Klein, S. A. Plotkin and S. Rao: Excluded minors, network decomposition, and multicommodity flow; in *25th Annual ACM Symposium on Theory of Computing*, pages 682–690, May 1993.

[15] N. Linial: Variation on a theme of Levin, in *Open Problems, Workshop on Discrete Metric Spaces and their Algorithmic Applications (J. Matoušek, ed.)*, Haifa, March 2002.

[16] N. Linial, E. London and Y. Rabinovich: The geometry of graphs and some of its algorithmic applications, *Combinatorica* **15(2)** (1995), 215–245.

[17] N. Linial and M. Saks: Low diameter graph decompositions, *Combinatorica* **13(4)** (1993), 441–454.

[18] L. Lovász and K. Vesztergombi: Geometric representations of graphs, in *Paul Erdős, Proc. Conf.*, Budapest, 1999.
[19] J. Matoušek: *Lectures on discrete geometry*, Vol. **212** of *Graduate Texts in Mathematics*, Springer-Verlag, New York, 2002.
[20] R. Motwani and P. Raghavan: *Randomized Algorithms*, Cambridge University Press, 1995.
[21] S. Rao: Small distortion and volume preserving embeddings for planar and Euclidean metrics, in *Proceedings of the 15th Annual Symposium on Computational Geometry*, pages 300–306, ACM, 1999.

Robert Krauthgamer

*IBM Almaden Research Center*
*650 Harry Road*
*San Jose, CA 95120*
*USA*
robi@almaden.ibm.com

James R. Lee

*Department of Computer Science and Engineering*
*Box 352350*
*University of Washington*
*Seattle, WA 98195-2350*
*USA*
jrl@cs.washington.edu