Sparse balanced partitions and the complexity of subgraph problems

Noga Alon [*]   Dániel Marx[†]

November 14, 2010

Abstract

We consider the problem of partitioning the vertices of a graph with maximum degree d into k classes V_1, ..., V_k of almost equal sizes in a way that minimizes the number of pairs (i, j) such that there is an edge between V_i and V_j. We show that there is always such a partition with $O(k^{2-2/d})$ adjacent pairs and this bound is tight. This problem is related to questions about the depth of certain graph embeddings, which have been used in the study of the complexity of subgraph and constraint satisfaction problems.

1 Introduction

If we randomly partition the vertices of a large graph G into a small number k of classes V_1, ..., V_k of roughly equal sizes, then we expect that every pair (V_i, V_j) of classes is adjacent (meaning that there is at least one edge with one endpoint in V_i and the other in V_j). This is true even if the graph G is sparse, for example, if it is a d-regular graph for any fixed positive d. The first question we investigate is whether it is always possible to partition the vertices in a balanced way into k classes such that the number of pairs of classes that are adjacent is significantly less than the total number $\binom{k}{2}$ of all pairs. Of course, no such partition is possible if G is a clique, thus this question makes sense only for sparser classes of graphs. We show that for graphs of maximum degree d, the answer is about $k^{2-2/d}$ in a fairly tight sense: every sufficiently large graph with maximum degree d has a partition where only $O(k^{2-2/d})$ pairs are used, while there are graphs for which $\Omega(k^{2-2/d})$ pairs are needed. The precise statement of the upper bound is the following:

Theorem 1.1. *For $d \geq 2$, there is a constant $c_d > 0$ such that for every $k > 0$, $d \geq 2$, and d-regular graph F with $n \geq n_0(k)$ vertices, the vertices of F can be partitioned into k classes V_1, ..., V_k, each of size at most $\lceil n/k \rceil$, such that there are at most $c_d \cdot k^{2-2/d}$ pairs (i, j) with $i \neq j$ for which V_i and V_j are adjacent.*

Theorem 1.1 is somewhat surprising: it says that every d-regular graph (even a random d-regular graph) is structured at a certain large scale. The actual statement we prove is stronger than Theorem 1.1: for every k and d, we give an explicit set $S_{k,d}$ of $O(k^{2-2/d})$ pairs such that every sufficiently large d-regular graph F has a partition with pattern S_k (meaning that $S_{k,d}$ describes which pairs of classes can be adjacent). The construction of $S_{k,d}$ is similar to the construction of universal graphs by Alon and Capalbo [1].

We prove the lower bound by showing that with high probability, a random d-regular graph needs $\Omega(k^{2-2/d})$ pairs. We say that an n-vertex random graph satisfies a property *asymptotically almost surely* (a.a.s for short), if the probability it satisfies it tends to 1 as n tends to infinity.

Theorem 1.2. *For every fixed $d > 2$ there is a constant $c_d > 0$ such that for every $k > k_0(d)$, and even $n > n_0(k)$, the following holds. Let F be a random d-regular graph on n vertices. Then a.a.s, for every partition $V_1, \ldots, V_k$ of the vertices of V into k classes with $|V_i| \leq 10n/k$ for every $1 \leq i \leq k$, there are at least $c_d \cdot k^{2-2/d}$ pairs (i, j) such that V_i and V_j are adjacent.*

We also prove a variant of Theorem 1.2 where we allow at most $n/k^{1-\epsilon}$ vertices in each class instead of $O(n/k)$, and we show that $\Omega(k^{2-2/d-3\epsilon})$ pairs are required, even after removing any set of ϵn edges of F.

Another way of looking at the partitioning problems treated in Theorems 1.1 and 1.2 is via homomorphisms. A *homomorphism* from a graph F to a graph H is a (not necessarily injective) mapping $\phi : V(F) \to V(H)$ such that if $uv \in E(H)$, then $\phi(u)\phi(v) \in E(F)$. The partitions in Theorems 1.1 and 1.2 can be interpreted as a homomorphism from F to some k-vertex graph H having a certain number of edges (and a loop at each vertex) and a balance requirement bounding the number of vertices in F that can be mapped to a vertex of H. It will be useful to keep this interpretation in mind, especially since we use techniques from [1] stated in terms of homomorphisms.

Embeddings of bounded depth. Our understanding of sparse partitions can be used to resolve problems arising in a different context. Recall that F is a *minor* of G if there is a mapping ϕ assigning disjoint connected subsets of G to each vertex of F such that for every edge uv of F there is an edge of G intersecting both $\phi(u)$ and $\phi(v)$. In [11], this notion was generalized in the following way: in an *embedding of depth d* we do not require the sets $\phi(u)$ to be disjoint, but we require that each vertex of G appears in the image of at most d vertices of F. For every edge uv of F, we require that $\phi(u)$ and $\phi(v)$ *touch,* that is, either they intersect or there is an edge between them. Clearly, F has an embedding of depth 1 into G if and only if F is a minor of G. The following result of [11] shows that, intuitively, larger treewidth[1] of G means that it has better embedding power:

Theorem 1.3 ([11])**.** *There is a function $m_0(G)$ and a universal constant c such that for every $k \geq 1$, if G is a graph with treewidth at least k and H is a graph with $|E(H)| =$*

---

[1]The exact definition of treewidth is not essential for the current paper, it is sufficient to know that treewidth is a graph measure and many algorithmic problems become easier on graphs of small treewidth, see e.g., [5, 4].

$m \geq m_0(G)$ and no isolated vertices, then H has an embedding into G with depth at most $\lceil cm \log k/k \rceil$.

Note that $|V(H)|/|V(G)|$ is a trivial lower bound on the depth of a embedding. Since there are graphs G where treewidth is linear in $|V(G)|$ and if H has no isolated vertices than $m = |E(H)| \geq |V(H)|/2$, the $\lceil cm \log k/k \rceil$ bound in the statement of Theorem 1.3 cannot be improved to $o(m/k)$. Thus Theorem 1.3 is tight, up to a $O(\log k)$ factor.

Theorem 1.3 was used in [11] as an essential tool to prove complexity results for constraint satisfaction and subgraphs problems (see later in the introduction). In the hope of making these results tighter, it was raised as an open question whether Theorem 1.3 remains true if the $\log k$ factor is removed from the bound on the depth. We answer this question in the negative: the bound $m \log k/k$ in Theorem 1.3 is tight (up to constant factors).

Theorem 1.4. *There is an infinite family $\mathcal{G}$ of graphs and a universal constant c, such that for every $G \in \mathcal{G}$ there exist arbitrarily large graphs H (without isolated vertices) such that if $m = |E(H)|$ and k is the treewidth of G, then every embedding of H into G has depth at least $cm \log k/k$.*

As the $\log k$ factor cannot be removed from Theorem 1.3 in general, we investigate families of graphs where this is possible and Theorem 1.3 holds in the strongest possible way. Let us say that a family $\mathcal{G}$ of graphs has the *tight embedding* property if Theorem 1.3 is true with a $\lceil cm/|V(G)| \rceil$ bound on the depth when G is restricted to the class $\mathcal{G}$. It can be shown that for such a class, the treewidth of every graph in the family has to be linear in its number of vertices. For example, line graphs of cliques form such a class: the line graph of the k-clique has $O(k^2)$ vertices, treewidth $\Theta(k^2)$, and it is shown in [11] that this class has the tight embedding property. Notice that the average degree of the line graph of the k-clique is $\Theta(k)$, i.e., square root of the number of vertices. Are there classes of graphs with the tight embedding property having significantly smaller average degree? We show that the average degree has to be polynomial in the number of vertices, but the exponent can be arbitrary small.

Theorem 1.5.   *1. If $\mathcal{G}$ has the tight embedding property, then there is a $\delta > 0$ such that every $G \in \mathcal{G}$ has average degree $\Omega(|V(G)|^\delta)$.*

   2. For every $\delta > 0$, there is a class $\mathcal{G}_\delta$ having the tight embedding property such that for every $G \in \mathcal{G}_\delta$, the average degree of G is $O(|V(G)|^\delta)$.

Complexity implications. The main goal of [11] was to understand the complexity of constraint satisfaction problems in terms of the treewidth of the primal graph. Rather than defining constraint satisfaction problems and going through the relevant background, we can discuss the problem in an essentially equivalent way in terms of (colored) subgraph problems. Given two graphs G and H, the Subgraph Isomorphism problem asks if G is a subgraph of H. In the colored (or more precisely, partitioned) version of the problem, the input contains a (not necessarily proper) coloring of the vertices of H, where the set of colors is the same

3

as the set of vertices of G, and we ask if G appears as a subgraph of H in such a way that every vertex v of G is mapped to a vertex with color v. In other words, the vertices of H are partitioned into $|V(G)|$ classes and we want to find a subgraph isomorphic to G such that the i-th vertex of G appears in the i-th class.

If G has k vertices and H has n vertices, then Colored Subgraph Isomorphism can be solved in time $n^{O(k)}$ by brute force. If G has small treewidth, then a more efficient solution is possible: for every fixed G, if G has treewidth at most w, then there is an $n^{O(w)}$ time algorithm for the problem [8, 2]. The main result of [11] shows that this is essentially best possible in the sense that there is no class of graphs where significant improvement is possible in the exponent. The result is proved under the complexity-theoretic assumption that there is no $2^{o(n)}$ time algorithm for n-variable 3SAT, which is also known as the Exponential Time Hypothesis (ETH), see [10].

Theorem 1.6 ([11]). *If there is a class $\mathcal{G}$ of graphs with unbounded treewidth and an arbitrary function f such that Colored Subgraph Isomorphism with the smaller graph G restricted to being in $\mathcal{G}$ can be solved in time $f(G)n^{o(w/\log w)}$ (where w is the treewidth of G), then ETH fails.*

It is conjectured in [11] that Theorem 1.6 holds even without the $\log w$ factor in the exponent.

Conjecture 1.7. *There is no class $\mathcal{G}$ of graphs with unbounded treewidth and no function f such that Colored Subgraph Isomorphism with the smaller graph G restricted to being in $\mathcal{G}$ can be solved in time $f(G)n^{o(w)}$ (where w is the treewidth of G).*

Conjecture 1.7 could be proved by showing that the $\log k$ factor in Theorem 1.3 is not needed. Unfortunately, by Theorem 1.4, this is not true. Therefore, the techniques presented in [11] are not sufficient to prove the conjecture. This does not invalidate the conjecture, but shows that if it is true, then substantially different techniques are needed for its proof.

As a special case of Conjecture 1.7, we would like to find classes of graphs where Colored Subgraph Isomorphism is "as hard as possible": classes for which there is no significantly better algorithm than trying all possibilities in $n^{O(|V(G)|)}$ time For example, it is true for the class of cliques: [6, 7] showed that there is no $f(k)n^{o(k)}$ time algorithm for the k-Clique problem, unless ETH fails. Moreover, as discussed in [11], if a class $\mathcal{G}$ has the tight embedding property, then Conjecture 1.7 holds for $\mathcal{G}$ (assuming ETH).

For the uncolored version of Subgraph Isomorphism, the hardness proof requires the additional condition that every graph is a core. Recall that a graph G is a *core* if every homomorphism from G to G is surjective, i.e., there is no homomorphism from G to a proper induced subgraph of G.

Theorem 1.8 ([11]). *Assume that ETH is true, let $\mathcal{G}$ be a class of graphs having the tight embedding property, and let f be an arbitrary function.*

(1) *There is no $f(G)n^{o(|V(G)|)}$ time algorithm for Colored Subgraph Isomorphism with the smaller graph G restricted to being in $\mathcal{G}$,*

(2) If every $G \in \mathcal{G}$ is a core, then the same is true for (uncolored) Subgraph Isomorphism.

Theorem 1.5(2) provide examples of relatively sparse classes that are "as hard as possible."

Theorem 1.9. *If there is a $\delta > 0$ and a function $f(G)$ such that (Colored) Subgraph Isomorphism can be solved in time $f(G)n^{o(|V(G)|)}$ when restricted to graphs G with average degree at most $|V(G)|^{\delta}$, then ETH fails.*

To prove Theorem 1.9 for (uncolored) Subgraph Isomorphism problem, we need some additional arguments: Theorem 1.8(2) applies only to graph classes that contain only cores, i.e., graphs that have no proper endomorphisms. By slightly modifying the construction of Theorem 1.5(2), we can ensure that the class $\mathcal{G}_{\delta}$ contains only cores, and the complexity result for (uncolored) Subgraph Isomorphism follows.

Theorem 1.9 leaves open the question whether there are really sparse (i.e., constant maximum degree) graph classes that are "as hard as possible" to find. As proved in Theorem 1.5(1), a graph class with constant average degree cannot have the tight embedding property, thus this approach cannot be used to construct sparse classes that are hard to find. Note that if a graph has a linear number of edges and treewidth is linear in the number of vertices, then it contains a large expander (cf. [9]). Thus it seems that the main question that lies at the heart of Conjecture 1.7 is whether it is possible to find a given k-vertex sparse expander in an n-vertex graph in time $n^{o(k)}$.

2 Upper Bound

In this section, we prove the existence of the partitions required by Theorem 1.1. The construction is similar to the sparse universal graph construction of [1]. Following [1], it will be convenient to consider the partitions as homomorphisms. We say that a homomorphism ϕ from F to H is *ϵ-balanced* if $(1 - \epsilon)|V(F)|/|V(H)| \leq |\phi^{-1}(v)| \leq (1 + \epsilon)|V(F)|/|V(H)|$ for every $v \in V(H)$. Our first result proves the existence of an ϵ-balanced homomorphism to a specific graph:

Theorem 2.1. *Let T be an arbitrary regular connected graph and let $\epsilon > 0$. Let H be the graph whose vertex set is $V(T)^d$ and two vertices are connected if and only if in at least two coordinates they are within distance 4 in T. Then every d-regular graph F with $n \geq n_0(T, d, \epsilon)$ vertices has an ϵ-balanced homomorphism f into H.*

Note that in particular every vertex of H is adjacent to itself, i.e., has a loop. Assuming that d and the degree of the regular graph T are fixed constants, every vertex of T has degree $O(|V(T)|^{d-2})$. As $|V(H)| = |V(T)|^d$, this means that H has $O(|V(H)|^{2-2/d})$ edges, which is precisely the right exponent for Theorem 1.1.

The proof of Theorem 2.1 is similar to that of the main result of [1]. In particular, we need the following tool. Let $\sigma : V(F) \to \{1, 2, \ldots, |V(F)|\}$ be an ordering of the vertices of F. The *bandwidth* of σ is the maximum length of an edge in this ordering, that is,

$\max_{uv\in E(F)} |\sigma(u) - \sigma(v)|$. The bandwidth of a graph F is the smallest bandwidth taken over all orderings σ of $V(F)$.

Theorem 2.2 ([1]). *Let $d \geq 2$ be an integer and let F be an arbitrary graph of maximum degree at most d. Then there are d spanning subgraphs F_1, ..., F_d, each of bandwidth at most 4, such that every edge of F lies in exactly two graphs F_i.*

Proof (of Theorem 2.1). Let F_1, ..., F_d be a decomposition of F as in Theorem 2.2, and let σ_i be an ordering of F_i having bandwidth at most 4. Independently for $i = 1, \ldots, d$, let us choose a random walk $w_i : \mathbb{N} \to V(T)$ in the r-regular graph T: we fix an arbitrary start vertex for each walk, and in each step, the probability of staying at the same vertex or moving to a particular neighbor is $1/(r+1)$. It is well known that this random walk converges to a uniform distribution, i.e., the probability of every vertex is $1/q$, where $q = |V(T)|$. Therefore, we can fix a constant t_0 depending on $|T|$, ϵ, and d such that no matter where we start the random walk, every vertex has probability between $(1 - \frac{\epsilon}{2})^{1/d}/q$ and $(1 + \frac{\epsilon}{2})^{1/d}/q$ after any number $t \geq t_0$ of steps.

We define the homomorphism ϕ by setting $\phi(v) = (w_1(\sigma_1(v)), \ldots, w_d(\sigma_d(v)))$. To see that it is a homomorphism, consider an edge $uv \in E(F)$. By assumption, there are two indices i_1, i_2 such that uv appears in F_{i_1}, F_{i_2}. This means that $|\sigma_{i_1}(u) - \sigma_{i_1}(v)| \leq 4$ and hence the distance of $w_{i_1}(\sigma_{i_1}(u))$ and $w_{i_1}(\sigma_{i_1}(v))$ is at most 4 in T. Similarly, the distance of $w_{i_2}(\sigma_{i_2}(u))$ and $w_{i_2}(\sigma_{i_2}(v))$ is at most 4 in T. In other words, there are at least two coordinates where the distance of $\phi(u)$ and $\phi(v)$ is at most 4 in T, implying that $\phi(u)$ and $\phi(v)$ are adjacent in H.

Finally, we show that ϕ is ϵ-balanced with high probability: for every $\delta > 0$, the probability that ϕ is not ϵ-balanced is at most δ, if n is sufficiently large. For every $0 \leq i \leq d$ and $\mathbf{a} = (a_1, \ldots, a_i) \in V(T)^i$, let $V_{\mathbf{a}} = \{(a_1, \ldots, a_i, b_{i+1}, \ldots, b_d) \mid b_{i+1}, \ldots, b_d \in V(T)\}$. We claim that with probability at least $1 - \delta$, for every $0 \leq i \leq d$ and $\mathbf{a} \in V(T)^i$, we have

$$(1 - \epsilon)^{i/d} n/q^i \leq |\phi^{-1}(V_{\mathbf{a}})| \leq (1 + \epsilon)^{i/d} n/q^i.$$

We say that $V_{\mathbf{a}}$ is *bad* if it does not satisfy this requirement. For $i = d$, the claim shows that $(1 - \epsilon)n/|V(H)| \leq |\phi^{-1}(\mathbf{a})| \leq (1 + \epsilon)n/|V(H)|$ for every $\mathbf{a} \in V(H)$, i.e., ϕ is ϵ-balanced. If $\mathbf{0}$ is the empty vector with zero components, then $V_{\mathbf{0}} = V(T)^d$ and $|\phi^{-1}(V_{\mathbf{0}})| = n$, hence $V_{\mathbf{0}}$ is not bad.

For every $\mathbf{a} = (a_1, \ldots, a_i)$, let us define $\mathbf{a}' = (a_1, \ldots, a_{i-1})$ (for $i = 1$, vector $\mathbf{a}'$ has zero components). We show that if n is sufficiently large, then the conditional probability that $V_{\mathbf{a}}$ is bad assuming that $V_{\mathbf{a}'}$ is not bad is at most δ/q^d (for $i = 1$, this is just the probability that $V_{\mathbf{a}}$ is bad). If some $V_{\mathbf{a}}$ is bad, then there has to be an $\mathbf{a}$ such that $V_{\mathbf{a}}$ is bad and $V_{\mathbf{a}'}$ is not bad. The probability that $V_{\mathbf{a}}$ is bad and $V_{\mathbf{a}'}$ is not bad is at most the conditional probability that we bounded by δ/q^d. Therefore, by a union bound, this shows that the probability that at least one bad event happens is at most δ.

Observe that whether $V_{\mathbf{a}'}$ is bad depends only on the walks w_1, ..., w_{i-1}, while whether $V_{\mathbf{a}}$ is bad depends only on the walks w_1, ..., w_i. We show that fixing the walks w_1, ..., w_{i-1} such that $V_{\mathbf{a}'}$ is not bad, the probability that walk w_i makes $V_{\mathbf{a}}$ bad is at most δ/q^d.

Let us enumerate the vertices v of $\phi^{-1}(V_{\mathbf{a}'})$ by increasing value of $\sigma_i(v)$. For $1 \leq s \leq t_0$, let $X^s = x_1, x_2, \ldots$ be the subsequence of this enumeration containing every t_0-th vertex in this enumeration, starting with the s-th. As $x_j \in \phi^{-1}(V_{\mathbf{a}'})$ for every j, we know that $w_{i'}(\sigma_{i'}(x_j)) = a_{i'}$ for every $1 \leq i' < i$. Thus $x_j \in \phi^{-1}(V_{\mathbf{a}})$ if and only if $w_i(\sigma_i(x_j)) = a_i$ also holds. Since $\sigma_i(x_j) \geq \sigma_i(x_{j-1}) + t_0$, the definition of t_0 ensures that the conditional probability $P(w_i(\sigma_i(x_j)) = a \mid w_i(\sigma_i(x_{j-1})) = b)$ is between $(1 - \frac{\epsilon}{2})^{1/d}/q$ and $(1 + \frac{\epsilon}{2})^{1/d}/q$ for every $a, b \in V(T)$. Let Y be an arbitrary subsequence $y_1, y_2, \ldots, y_{|Y|}$ of X^s. The probability of the event that $w_i(\sigma_i(y)) = a_i$ for every $y \in Y$ can be bounded from above by the product of $|Y|$ such conditional probabilities. Therefore, the probability that $|X^s \cap \phi^{-1}(V_{\mathbf{a}})| \geq (1 + \epsilon)^{1/d}|X^s|/q$ holds is not larger than the probability that the binomial random variable $B(|X^s|, (1 + \frac{\epsilon}{2})^{1/d}/q)$ is larger than $(1 + \epsilon)^{1/d}|X^s|/q$. From standard bounds, we know that if n (and hence $|X^s|$) is sufficiently large, then this probability can be bounded by an arbitrary small constant. Thus we can assume that this probability is at most $\delta/(2t_0 q^d)$. Therefore, by the union bound, the upper bound on $|X^s \cap \phi^{-1}(V_{\mathbf{a}})|$ holds for every $1 \leq s \leq t_0$ with probability at least $1 - \delta/(2q^d)$, hence we have

$$|\phi^{-1}(V_{\mathbf{a}})| \leq (1 + \epsilon)^{1/d}|\phi^{-1}(V_{\mathbf{a}'})|/q \leq (1 + \epsilon)^{i/d}|V(F)|/q^i,$$

where the second inequality uses the assumption that $V_{\mathbf{a}'}$ is not bad. Similarly, we can show that the lower bound on $|\phi^{-1}(V_{\mathbf{a}})|$ holds with probability at least $1 - \delta/(2q^d)$, hence the conditional probability that $V_{\mathbf{a}}$ is bad assuming $V_{\mathbf{a}'}$ is not bad is at most δ/q^d. $\qquad \square$

To obtain the result stated in Theorem 1.1, we need to improve Theorem 2.1 in two ways. First, Theorem 2.1 partitions the set of vertices into q^d classes for some integer q, while in Theorem 1.1 we allow an arbitrary number of classes. More importantly, we need to ensure that the partition is not only ϵ-balanced, but every class contains at most $\lceil n/k \rceil$ vertices. This problem can be solved by a technique of [1]: we define a bounded-degree expander on the classes and allow the vertices to move between neighboring classes to achieve a perfectly balanced partition.

Theorem 2.3. *For every $d > 2$ and $k > 0$, there is an integer $n_0(d, k)$ and a set $S_{d,k}$ of $O(k^{2-2/d})$ pairs (i, j) $(i, j \in [k])$ such that the following holds. If F is a graph on $n > n_0(d, k)$ vertices and maximum degree d, then the vertices of F can be partitioned into k sets $V_1, \ldots, V_k$, each of size at most $\lceil n/k \rceil$, such that if V_i and V_j are adjacent, then $(i, j) \in S_{d,k}$.*

Proof. Let $q = \lceil (20k)^{1/d} \rceil$. Because of the big-$O$ notation in the statement of the theorem, we can assume that k is sufficiently large and hence $q \geq 3$. Let T be the cycle on q vertices and let H be defined as in Theorem 2.1. Note that $q^d \geq 20k$ and $q^d < ((20k)^{1/d} + 1)^d \leq 21k$ if k is sufficiently large. Therefore, $20 \leq q^d/k \leq 21$ and the vertices of H have a partition $U_1, \ldots, U_k$ such that $20 \leq |U_i| \leq 21$ for every $1 \leq i \leq k$.

Let M be a bounded-degree expander on $[k]$ with the property that for every subset X of at most half the vertices of M, the set X has at least $|X|/9$ neighbors outside X. For every $X \subseteq [k]$, denote by $N_M[X]$ the closed neighborhood of X, i.e., the set of all vertices that are

in X or adjacent to a vertex of X. The set $S_{d,k}$ is constructed in the following way: the pair (i, j) $(i \neq j)$ is in $S_{d,k}$ if and only if there is a pair (i', j') such that

- $i' \in N_M[\{i\}]$,

- $j' \in N_M[\{j\}]$,

- $U_{i'}$ and $U_{j'}$ are adjacent in H (i.e., there is an edge between a vertex of $U_{i'}$ and a vertex of $U_{j'}$).

To bound the size of $S_{d,k}$, recall first that for a fixed d, each vertex of H has degree $O(q^{d-2})$. As the set $U_{i'}$ contains at most 21 vertices of H, there can be at most $21 \cdot O(q^{d-2}) = O(k^{1-2/d})$ values j' such that $U_{i'}$ and $U_{j'}$ are adjacent. Therefore, if the degree of M is bounded by a constant c, then each $1 \leq i \leq k$ can participate in at most $c \cdot O(k^{1-2/d}) \cdot c = O(k^{1-2/d})$ pairs of $S_{d,k}$. Thus the total number of pairs in $S_{d,k}$ is $O(k^{2-2/d})$, as required.

To show that the required partition $V_1, \ldots, V_k$ of $V(F)$ exists, set $\epsilon = 0.01$ and let us use Theorem 2.1 to obtain an ϵ-balanced homomorphism $\phi : V(F) \to V(H)$. This homomorphism ϕ defines a partition $V_1', \ldots, V_k'$ by setting $V_i' = \{v \in V(F) \mid \phi(v) \in U_i\}$. Note that

$$|V_i'| \leq 21 \cdot 1.01 n/q^d \leq 21/20 \cdot 1.01 (n/k) \leq 1.1(n/k)$$

and

$$|V_i'| \geq 20 \cdot 0.99 n/q^d \geq 20/21 \cdot 0.99 (n/k) \geq 0.9(n/k).$$

We make the partition more balanced by allowing each vertex to move to a class that is adjacent in M. Let us build a bipartite graph B where the first class is the set of vertices in F and the second class contains $\lceil n/k \rceil$ vertices representing each class V_i (i.e., the second class contains $k\lceil n/k \rceil$ vertices). The edges of B are defined as follows: $v \in V(F)$ and a vertex representing class V_i are adjacent if $v \in V_{i'}'$ for some i' such that i and i' are adjacent in M. We show that this bipartite graph has a matching covering $V(F)$. If this is true, then we obtain the partition $V_1, \ldots, V_k$ by putting vertex v to the class represented by its mate. It is clear that each class V_i contains at most $\lceil n/k \rceil$ vertices and a vertex of V_i and vertex of V_j can be adjacent only if $(i, j) \in S_{d,k}$.

We use Hall's Theorem to show that the bipartite graph B has a matching covering $V(F)$. For $S \subseteq V(F)$, let $N_B(S)$ be the neighbors of S in B. Note that the vertices in V_i' have the same neighborhood in B, thus it is sufficient to check the Hall condition for every subset of $S \subseteq V(F)$ that is the union of some classes V_i'. Let $S = \bigcup_{i \in X} V_i'$ be such a set for some $X \subseteq [k]$. If $|X| \leq k/2$, then $|N_M[X]| \geq \frac{10}{9}|X|$, hence

$$|N_B(S)| \geq \frac{10}{9}|X|\lceil n/k \rceil > 1.1(n/k)|X| \geq |S|.$$

On the other hand, if $|X| > k/2$, then let $Y = [k] \setminus N_M[X]$; clearly $|Y| < k/2$. Therefore, $N_M[Y] \geq \frac{10}{9}|Y|$ and

$$\left| \bigcup_{i \in N_M[Y]} V_i' \right| \geq \frac{10}{9}|Y| 0.9(n/k) = |Y|(n/k).$$

8

If $i \in N_M[Y]$, then V_i' is not in S, thus we can bound the size of S by

$$|S| \le n - |Y|(n/k) = (k - |Y|)(n/k) = |N_M[X]|(n/k) \le |N_M[X]|\lceil n/k \rceil = |N_B(S)|.$$

Thus the Hall condition holds in this case as well. □

3   Lower bound

For the proof of Theorem 1.2 stated in the introduction, we need a lower bound on the number of labeled d-regular graphs on n vertices. The asymptotic number of such graph is known [3], but a lower bound of the form $(n/\alpha_d)^{nd/2}$ (for some constant $\alpha_d > 0$ depending only on d) will be sufficient for our purposes. We sketch how such a bound can be obtained by considering only the bipartite d-regular graphs having two fixed bipartite classes of size exactly $n/2$. Each such bipartite graph can be obtained as the union of d matchings between the two bipartite classes; the number of possibilities for selecting d matchings is $((n/2)!)^d \ge (n/(2e))^{nd/2}$. However, this formula might overcount the number of bipartite graphs for two reasons: the matchings might not be disjoint (hence the union is not 3-regular) and the same bipartite graph might be obtained multiple times. It is known that the probability of d permutations being disjoint is at least some constant $c_d' > 0$, and each d-regular bipartite graph can be obtained at most $d^{nd/2}$ times (as each of the $nd/2$ edges can belong to one of the d matchings). Therefore, by setting α_d appropriately large, it is true that there are at least $(n/\alpha_d)^{nd/2}$ different d-regular bipartite graphs.

Proof (of Theorem 1.2). Let us fix a sufficiently small positive c_d. Let us call a d-regular graph $F = (V, E)$ on n vertices *bad* if there is a partition $V_1, \ldots, V_k$ such that each V_i is of size at most $10n/k$, and if S is defined as the set that contains the pair $\{i, j\}$ $(i \ne j)$ if and only if V_i and V_j are adjacent, then $|S| \le c_d k^{2 - \frac{2}{d}}$. We estimate the number of bad graphs as follows. The number of allowed partitions can be bounded by the number k^n of all partitions, and the number of possibilities for the set S can be generously bounded by 2^{k^2}. For a fixed partition and set S, we bound the number of bad graphs by considering all possibilities for the edges. Each edge is either fully contained in some V_i, or the endpoints are in V_i and V_j for some $\{i, j\} \in S$. Since each V_i has size at most $10n/k$, there are at most $(|S| + k) \cdot 100n^2/k^2 \le 100c_d n^2/k^{\frac{2}{d}} + 100n^2/k \le 200c_d n^2/k^{\frac{2}{d}}$ such edges (where the last inequality holds if k is sufficiently large compared to c_d). Thus we can bound the number of bad graphs by

$$k^n \cdot 2^{k^2} \cdot \binom{200c_d n^2/k^{\frac{2}{d}}}{dn/2} \le k^n \cdot 2^{k^2} \cdot \left(\frac{400c_d en}{dk^{\frac{2}{d}}} \right)^{dn/2} = n^{dn/2} \cdot 2^{k^2} \cdot \left(\frac{400c_d e}{d} \right)^{dn/2} \ll \left(\frac{n}{2\alpha_d} \right)^{dn/2}.$$

We used $\binom{a}{b} \le (ae/b)^b$ in the first inequality and in the last inequality we assume that c_d is sufficiently small and n is sufficiently large. As the number of d-regular graphs is at least $(n/\alpha_d)^{nd/2}$, this shows that the probability of a random d-regular graph being bad goes to zero. □

The following version of Theorem 1.2 is stronger in the sense that we allow larger color classes and a set of at most ϵn exceptional edges that do not respect the pairs S, but it gives a slightly weaker bound of $k^{2-\frac{2}{d}-3\epsilon}$ on the size of S. To give a different perspective, we state it in terms of colors, where there is a bound on the color classes and on the pairs of colors that can appear on the edges.

Theorem 3.1. *For every fixed integer $d > 2$, real $\epsilon < 1/4$, integer $k > k_0(\epsilon, d)$ and for every even $n > n_0(k)$ the following holds. Let F be a random d-regular graph on n vertices. Then a.a.s., for every (not necessarily proper) coloring of the vertices of F by k colors, so that each color appears at most $n/(k^{1-\epsilon})$ times, and for any choice of a set S of at most $k^{2-\frac{2}{d}-3\epsilon}$ pairs of colors, there are at least ϵn edges of F whose endpoints have colors x, y with $x \neq y$ and $\{x, y\} \notin S$.*

Proof (of Theorem 3.1). Let us call a d-regular graph $F = (V, E)$ on n vertices *bad* if there is a coloring, a set S, and a subset E' of at most ϵn edges such that each color appears on at most $n/(k^{1-\epsilon})$ vertices, $|S| \leq k^{2-\frac{2}{d}-3\epsilon}$, and for every edge in $E \setminus E'$, the two endpoints either have the same color or colored by a pair from S. We estimate the number of bad graphs as follows. The number of allowed colorings can be bounded by the number k^n of total colorings, and the number of possibilities for the set S can be generously bounded by 2^{k^2}. For a fixed coloring and set S, we bound the number of bad graphs by considering all possibilities for the set E' and for the set $E \setminus E'$. An edge of E' can be any of the $\binom{n}{2} < n^2$ possible edges, while the colors of the endpoints of each edge of $E \setminus E'$ have to be in S or have to be the same. Since each color appears on at most $n/k^{1-\epsilon}$ vertices, there are at most $k^{2-\frac{2}{d}-3\epsilon} \cdot (n/k^{1-\epsilon})^2 = n^2/k^{\frac{2}{d}+\epsilon}$ such edges. Thus we can bound the number of bad graphs by

$$
k^n \cdot 2^{k^2} \cdot \binom{n^2}{\epsilon n} \cdot \binom{n^2/k^{\frac{2}{d}+\epsilon}}{dn/2 - \epsilon n} \leq k^n \cdot 2^{k^2} \cdot \left(\frac{en}{\epsilon}\right)^{\epsilon n} \cdot \left(\frac{en}{(d/2-\epsilon)k^{\frac{2}{d}+\epsilon}}\right)^{(d/2-\epsilon)n}
$$
$$
\leq n^{dn/2} \cdot 2^{k^2} \cdot g(d)^n \cdot k^{n(1-(\frac{2}{d}+\epsilon)(\frac{d}{2}-\epsilon))} = n^{dn/2} \cdot 2^{k^2} \cdot g(d)^n \cdot k^{-((\frac{d}{2}-\frac{2}{d}-\epsilon)\epsilon n)}
$$
$$
\leq n^{dn/2} \cdot 2^{k^2} \cdot g(d)^n \cdot k^{-((\frac{d}{2}-\frac{2}{d}-\frac{1}{4})\epsilon n)} \ll (n/\alpha_d)^{nd/2}.
$$

for some function $g(d)$ depending only on d. In the first inequality, we used $\binom{a}{b} \leq (ae/b)^b$; in the second inequality, we used the fact that $(1/x)^x$ can be bounded by a constant. For the last inequality, let us observe that for every $d \geq 3$, $\delta := \frac{d}{2} - \frac{2}{d} - \frac{1}{4}$ is positive. Thus if k is sufficiently large compared to d and $1/\epsilon$, and n is sufficiently large compared to k and ϵ, then $k^{\delta \epsilon n}$ dominates 2^{k^2}, $g(d)^n$, and $\alpha_d^{nd/2}$. As the number of d-regular graphs is at least $(n/\alpha_d)^{nd/2}$, this shows that the probability of a random d-regular graph being bad goes to zero. $\square$

4 Bounded depth embeddings

We can use the lower bound of Section 3 to obtain lower bounds on the depths of certain embeddings. Our first result shows that Theorem 1.3 of [11] is tight.

Theorem 4.1. *Let H be a 3-regular graph with k vertices. Then, for all even $n > n_0(k)$, there exists a 3-regular graph F on n vertices so that any embedding of F into H is of depth at least $\Omega(\frac{n \log k}{k})$.*

Proof. Let $d = 3$ and $\epsilon = 1/100$, and let F be a random cubic graph on n vertices satisfying the requirements of Theorem 3.1. Suppose for contradiction that there is an embedding ϕ of F into H having depth less than $\frac{\epsilon^2 n \log k}{3k}$. Let V' be the set of all vertices of F that are mapped to sets of size at least $\epsilon \log k$; clearly, $|V'| < \epsilon n/3$. Let E' contain all edges of F that touch V', we have $|E'| < \epsilon n$.

For each vertex v of F, choose an arbitrary vertex $f(v)$ of $\phi(v)$ and consider it as a coloring of F with k colors (corresponding to the vertices of H) having the property that no color is used more than $\frac{\epsilon^2 n \log k}{3k} < n/k^{1-\epsilon}$ times (we can assume that k is sufficiently large since otherwise the theorem automatically holds due to the Ω notation). Let S be the set of all pairs of colors x, y (i.e., pairs of vertices of H) such that the distance of x and y in H is at most $2\epsilon \log k = 0.02 \log k$. Since H is 3-regular, $|S| \leq O(k \cdot 2^{2\epsilon \log k}) \leq O(k^{1.02}) < k^{4/3-3\epsilon}$. Therefore, by Theorem 3.1, there must be at least ϵn edges of F whose endpoints are colored by a pair of (two different) colors such that this pair does not appear in S. As $|E'| < \epsilon n$, there is such an edge $uv \in E \setminus E'$, that is, $u, v \notin V'$. Therefore, both $\phi(u)$ and $\phi(v)$ have size at most $\epsilon \log k$ and $\{f(u), f(v)\} \notin S$ implies that the distance of $f(u)$ and $f(v)$ is more than $2\epsilon \log k$. This means that $\phi(u)$ and $\phi(v)$ cannot touch, contradicting the definition of embedding. $\qquad\square$

To obtain Theorem 1.4, it is sufficient to take $\mathcal{G}$ to be a class of 3-regular expanders. It is well known that the treewidth of an expander is linear in the number of vertices (cf. [9]), and Theorem 1.4 follows.

Theorem 4.1 shows that a very sparse (3-regular) class of graphs cannot have the tight embedding property. How dense should a class be to have this property? The lower bound on the depth in Theorem 4.1 is a logarithmic factor larger than the trivial lower bound $\Omega(n/k)$ and it is matched by the embedding result of Theorem 1.3. Therefore, it might be a reasonable educated guess to expect that an extra logarithmic factor appears here as well and an average degree of $\log |V(G)|$ is sufficient for the tight embedding property. However, our second negative result shows that the number of edges has to be polynomially larger than linear, i.e., the average degree has to be $|V(G)|^\delta$ for some $\delta > 0$.

Theorem 4.2. *For every $\delta > 0$ and $k > k_0(\delta)$ the following holds. Let H be a graph with k vertices and at most $k^{1+\delta}$ edges. Then, for all even $n > n_0(k)$, there exists a 3-regular graph F on n vertices so that any embedding of F into H is of depth at least $\Omega(\frac{n}{k\delta})$.*

Proof. Let $d = 3$ and $\epsilon = 1/100$, and let F be a random cubic graph on n vertices satisfying the requirements of Theorem 3.1. Assume that k is sufficiently large to ensure that $k^\delta > 1/\delta$ holds. Let D be the set of those vertices of H that have degree at least k^δ/δ, we have $|D| \leq 2\delta k$.

Suppose for contradiction that there is an embedding ϕ of F into H having depth less than $\frac{\epsilon^2 n}{6k\delta}$. Let V' be the set of all vertices of F that are mapped to sets of size at least ϵ/δ;

11

clearly, $|V'| < \epsilon n/6$. Let V'' contain those vertices of $V \setminus V'$ whose images intersect D, we have $|V''| \leq \frac{\epsilon^2 n}{6k\delta}|D| \leq \epsilon^2 n/3 < \epsilon n/6$. Let E' contain all edges of F that touch $V' \cup V''$; clearly, we have $|E'| < \epsilon n$.

For each vertex v of F, choose an arbitrary vertex $f(v)$ of $\phi(v)$ and consider it as a coloring of F with k colors having the property that no color is used more than $\frac{\epsilon^2 n}{6k\delta} < n/k^{1-\epsilon}$ times (assuming that k is sufficiently large compared to $1/\delta$). Let S be the set of all pairs of colors x, y (i.e., pairs of vertices of H) such that the distance of x and y in $H \setminus D$ is at most $2\epsilon/\delta = 0.02/\delta$. Since every vertex of $H \setminus D$ has degree at most $k^\delta/\delta < k^{2\delta}$ (using $k^\delta > 1/\delta$), $|S| \leq O(k \cdot k^{2\delta \cdot 0.02/\delta}) = O(k^{1.04}) < k^{4/3-3\epsilon}$. Therefore, by Theorem 3.1, there must be at least ϵn edges of F whose endpoints are colored by a pair of (two different) colors not appearing in S. As $|E'| < \epsilon n$, there is such an edge $uv \in E \setminus E'$, that is, $u, v \notin V' \cup V''$. Therefore, both $\phi(u)$ and $\phi(v)$ have size at most ϵ/δ and they are disjoint from D. Furthermore, $\{f(u), f(v)\} \notin S$ implies that the distance between $f(u)$ and $f(v)$ is more than $2\epsilon/\delta$ in $H \setminus D$. This means that $\phi(u)$ and $\phi(v)$ cannot touch, contradicting the definition of embedding. $\square$

Theorem 4.2 shows that if for every $\delta > 0$, the class $\mathcal{G}$ contains infinitely many graphs G with average degree at most $|V(G)|^\delta$, then there is no constant c such that it is true that every graph F has an embedding into every $G \in \mathcal{G}$ with depth $c|E(F)|/|V(G)|$, or in other words, $\mathcal{G}$ does not have the tight embedding property. Thus if $\mathcal{G}$ has the tight embedding property, then there is a $\delta > 0$ such that there are only finitely many graphs $G \in \mathcal{G}$ with average degree at most $|V(G)|^\delta$. Therefore, we can say that every graph $G \in \mathcal{G}$ has average degree $\Omega(|V(G)|^\delta)$ (by choosing the constant hidden in the Ω notation appropriately), proving Theorem 1.5(1).

To prove Theorem 1.5(2), we construct a family of graphs having the tight embedding property. This family is based on a product construction similar to the one appearing in the proof of Theorem 2.1. This class in some sense generalizes line graphs of cliques, and we prove the tight embedding property similarly to the way it is proved for line graphs of cliques in [11].

Let $G[k, d]$ be the graph whose vertex set is $[k]^d$ and two vertices $(a_1, \ldots, a_d), (b_1, \ldots, b_d) \in [k]^d$ are adjacent if there is exactly one value $1 \leq i \leq d$ such that $a_i \neq b_i$. Note that $G[k, d]$ has k^d vertices and is $d(k-1)$-regular.

Theorem 4.3. *For integers $k, d > 0$ and every graph F with $m > m_0(k, d)$ edges and no isolated vertices, there is an embedding of depth $O(dm/k^d)$ from F into $G[k, d]$.*

Proof. First we argue that it is sufficient to prove the theorem for graphs F having maximum degree at most 3. Otherwise, let us construct F' by replacing every vertex v of F having degree $d(v)$ with a path $v_1, \ldots, v_{d(v)}$ of $d(v)$ vertices and let every edge incident to v use a different copy of v on the path. Clearly, F' has maximum degree at most 3 and has at most $3m$ edges. If there is an embedding ϕ' from F' into H, then it can be turned into an embedding ϕ from F into H by setting $\phi(v) = \bigcup_{i=1}^d \phi'(v_i)$. It is clear that the depth of ϕ is not larger than the depth of ϕ'. Thus in the following, we assume that F has maximum degree at most 3.

Let us partition the vertices of F into k^d classes $V_{\mathbf{a}}$ ($\mathbf{a} \in [k]^d$), each of size at most $\lceil |V|/k^d \rceil$, in an arbitrary way. Let us orient the edges of F arbitrarily and let $E_{\mathbf{a},\mathbf{b}}$ be the set of edges going from $V_{\mathbf{a}}$ to $V_{\mathbf{b}}$. For every $\mathbf{a}, \mathbf{b} \in [k]^d$, let us partition $E_{\mathbf{a},\mathbf{b}}$ into classes $E_{\mathbf{a},\mathbf{b}}^{\mathbf{c}}$ ($\mathbf{c} \in [k]^d$), each of size at most $\lceil |E_{\mathbf{a},\mathbf{b}}|/k^d \rceil$, in an arbitrary way. Let $E_{\mathbf{a},*}^{\mathbf{c}} = \bigcup_{\mathbf{b} \in [k]^d} E_{\mathbf{a},\mathbf{b}}^{\mathbf{c}}$ and $E_{*,\mathbf{b}}^{\mathbf{c}} = \bigcup_{\mathbf{a} \in [k]^d} E_{\mathbf{a},\mathbf{b}}^{\mathbf{c}}$. Note that

$$|E_{\mathbf{a},*}^{\mathbf{c}}| = \sum_{\mathbf{b} \in [k]^d} |E_{\mathbf{a},\mathbf{b}}^{\mathbf{c}}| \le \sum_{\mathbf{b} \in [k]^d} \lceil |E_{\mathbf{a},\mathbf{b}}|/k^d \rceil \le \sum_{\mathbf{b} \in [k]^d} |E_{\mathbf{a},\mathbf{b}}|/k^d + k^d \le 3|V_{\mathbf{a}}|/k^d + k^d$$

$$\le 3\lceil |V|/k^d \rceil/k^d + k^d \le 4|V|/k^{2d},$$

where the third inequality uses the fact every vertex in $V_{\mathbf{a}}$ has degree at most 3 and the last inequality uses that $|V|$ is sufficiently large. A similar bound holds for $|E_{*,\mathbf{b}}^{\mathbf{c}}|$.

For $\mathbf{a} = (a_1, \ldots, a_d)$ and $\mathbf{b} = (b_1, \ldots, b_d)$, we denote by $W_{\mathbf{a},\mathbf{b}}$ the walk whose i-th vertex ($0 \le i \le d$) is $(b_1, \ldots, b_i, a_{i+1}, \ldots, a_d)$. Note that if $a_i = b_i$, then the $(i-1)$-st and the i-th vertices are the same. Clearly, $W_{\mathbf{a},\mathbf{b}}$ is connected and contains $\mathbf{a}$ and $\mathbf{b}$.

We define the embedding ϕ in the following way. First, if $v \in V_{\mathbf{a}}$, then let $\phi(v)$ contain vertex $\mathbf{a}$. If an edge of $E_{\mathbf{a},\mathbf{b}}^{\mathbf{c}}$ leaves v, then we add $W_{\mathbf{a},\mathbf{c}}$ to $\phi(v)$; if an edge of $E_{\mathbf{b},\mathbf{a}}^{\mathbf{c}}$ enters v, then we add $W_{\mathbf{c},\mathbf{a}}$ to $\phi(v)$. Observe that this gives a correct embedding: $\phi(v)$ is connected and if an edge of $E_{\mathbf{a},\mathbf{b}}^{\mathbf{c}}$ connects x and y, then $\phi(x)$ contains $W_{\mathbf{a},\mathbf{c}}$ and $\phi(y)$ contains $W_{\mathbf{c},\mathbf{b}}$, hence $\phi(x)$ and $\phi(y)$ intersect in vertex $\mathbf{c}$.

To bound the depth of the embedding ϕ, let us estimate the number of vertices of F whose images contain a particular vertex $\mathbf{g} = (g_1, \ldots, g_d)$. Vertex $\mathbf{g}$ is in $\phi(v)$ if a walk $W_{\mathbf{x},\mathbf{y}}$ containing $\mathbf{g}$ was added to $\phi(v)$. For every $0 \le i \le d$, there are exactly k^d pairs $(\mathbf{x}, \mathbf{y})$ such that the i-th vertex of $W_{\mathbf{x},\mathbf{y}}$ is $\mathbf{g}$: namely, the pairs $(\mathbf{x}, \mathbf{y})$ with $\mathbf{x} = (x_1, \ldots, x_i, g_{i+1}, \ldots, g_d)$, $\mathbf{y} = (g_1, \ldots, g_i, y_{i+1}, \ldots, y_d)$ for arbitrary $x_1, \ldots, x_i, y_{i+1}, \ldots, y_d$. Therefore, there are at most $(d+1)k^d$ pairs $(\mathbf{x}, \mathbf{y})$ such that $W_{\mathbf{x},\mathbf{y}}$ contains $\mathbf{g}$. The path $W_{\mathbf{x},\mathbf{y}}$ is added to $\phi(v)$ only if an edge of $E_{\mathbf{x},*}^{\mathbf{y}}$ or an edge of $E_{*,\mathbf{y}}^{\mathbf{x}}$ is incident to v. Therefore, the path $W_{\mathbf{x},\mathbf{y}}$ is used at most $|E_{\mathbf{x},*}^{\mathbf{y}}| + |E_{*,\mathbf{y}}^{\mathbf{x}}|$ times. This means that the depth of vertex $\mathbf{g}$ is at most

$$(d+1)k^d(|E_{\mathbf{x},*}^{\mathbf{y}}| + |E_{*,\mathbf{y}}^{\mathbf{x}}|) \le 2(d+1)k^d \cdot 4|V|/k^{2d} = O(dm/k^d),$$

if m is sufficiently large. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Consider the graph class $\mathcal{G}[d]$ containing $G[k,d]$ for every $k \ge 1$. By Theorem 4.3, the class $\mathcal{G}[d]$ has the tight embedding property. A graph $G[k,d] \in \mathcal{G}[d]$ has k^d vertices and its average degree is $d(k-1) = O(k)$. Thus $\mathcal{G}_\delta = \mathcal{G}[\lceil 1/\delta \rceil]$ satisfies the requirements of Theorem 1.5(2).

By Theorem 1.8(1), if a graph class $\mathcal{G}$ has the tight embedding property, then there is no $f(G)n^{o(|V(G)|)}$ time algorithm for the special case of Colored Subgraph Isomorphism with the smaller graph G restricted to $\mathcal{G}$. Therefore, Theorem 1.9 follows for Colored Subgraph Isomorphism.

In order to prove Theorem 1.9 for (uncolored) Subgraph Isomorphism, we have to use Theorem 1.8(2). Therefore, we need classes $\mathcal{G}_\delta$ that contain only cores (recall that a graph is

13

a core if it has no endomorphism to any of its proper induced subgraphs). Notice that $G[k,d]$ is not a core: it is k-colorable (let the color of a vertex be the sum of the coordinates modulo k) and contains a k-clique. However, by attaching a "rigid" graph to $G[k,d]$, we can make it a core and this modification can be done in such a way that the size of the graph does not increase by too much, thus the class retains the tight embedding property. Therefore, the following theorem proves Theorem 1.9 for (uncolored) Subgraph Isomorphism.

Theorem 4.4. *For every $d \geq 2$, there is a class $\mathcal{G}[d]$ of graphs having the tight embedding property such that every $G \in \mathcal{G}[d]$ is a core and has maximum degree $O(|V(G)|^{1/d})$.*

Proof. For every $k \geq 1$, the class $\mathcal{G}[d]$ contains a graph $G'[k,d]$, which is a supergraph of $G[k,d]$. Let D be a triangle-free graph with chromatic number 4 such that any proper induced subgraph of D is 3-colorable; for example, the Grötzsch graph is such a graph. Let $v_1, \ldots, v_n$ be the vertices of $G[k,d]$. The graph $G'[k,d]$ is obtained by extending $G[k,d]$ with the following vertices and edges:

- a clique K_1 of size $k+4$,

- a clique K_2 of size $k+1$,

- a copy of D, with every vertex adjacent with every vertex of K_2,

- a path $u_1, \ldots, u_{6n+1}$ where u_1 is a vertex of K_1 and u_{6n+1} is a vertex of K_2,

- for every $1 \leq i \leq 6n$, a vertex w_i that is adjacent with u_i and u_{i+1},

- for every $1 \leq i \leq n$, a vertex z_i that is adjacent with v_i and u_{6i}.

The maximum degree of $G'[k,d]$ is $\max\{d(k-1)+1, k+|D|+1\}$ and the number of vertices increases only by a constant factor. Therefore, the only thing we have to show is that $G'[k,d]$ is a core. Consider a homomorphism ϕ from $G'[k,d]$ to itself. K_1 is the only clique of size $k+4$ in the graph: the maximum clique size of $G[k,d]$ is k, and clique K_2 can be extended by at most 2 vertices of D to a larger clique. Therefore, ϕ is a permutation on K_1. Similarly, ϕ must map K_2 to a clique of size $k+1$, which is either a subset of K_1 or a subset of $K_2 \cup D$. However, since $K_2 \cup D$ is not $(k+4)$-colorable (as D is not 3-colorable), the closed neighborhood of every vertex in K_2 is not $(k+4)$-colorable. The closed neighborhood of every vertex in K_1 is $(k+4)$-colorable, thus ϕ cannot map a vertex of K_2 to a vertex of K_1. Therefore, ϕ must map every vertex of K_2 to $K_2 \cup D$. As every vertex of D is adjacent to every vertex of K_2, this also means that ϕ maps every vertex of D to $K_2 \cup D$. Since every proper subset of $K_2 \cup D$ is $(k+4)$-colorable, ϕ is a permutation of $K_2 \cup D$.

If a vertex is in a triangle, then ϕ must map this vertex to a vertex that is also in a triangle. Therefore, ϕ must map the path $u_1, \ldots, u_{6n+1}$ into a walk on $6n+1$ vertices from $\phi(u_1) \in K_1$ to $\phi(u_2) \in K_2 \cup D$ such that each vertex is in a triangle. This means that the walk cannot use the vertices z_i, hence the only possibility is that $\phi(u_i) = u_i$ for every $1 \leq i \leq 6n+1$. It also follows that $\phi(w_i) = w_i$ for every $1 \leq i \leq 6n$.

14

We show that $\phi(v_i) = v_i$ and $\phi(z_i) = z_i$ for every $1 \leq i \leq n$. Let v_j be a neighbor of v_i. There is a path u_{6i}, z_i, v_i, v_j, z_j, u_{6j} of length 5 between u_{6i} and u_{6j}. Homomorphism ϕ must map this path to a walk. Removing z_i or z_j makes the distance of u_{6i} and u_{6j} at least 6, thus the walk has to use both z_i and z_j. Now the only possibility is that the walk is the same as the path. This shows that ϕ is a permutation on $G'[k,d]$. $\qquad\square$

5   Conclusions

As an important ingredient in the hardness results of [11], an appropriate notion of embedding was defined and it was proved that embeddings with certain properties exist. The more efficient embedding we are able to find, the tighter the hardness results are. Thus obtaining tighter complexity results was the motivation for the purely combinatorial question of whether the logarithmic factor in Theorem 1.3 can be removed. It turned out that understanding a different kind of combinatorial question (sparse balanced partitions) allows us to resolve this question. We proved both positive and negative results on the existence of sparse balanced partitions. The positive results use techniques and ideas related to yet another combinatorial problem: the construction of sparse universal graphs.

The combinatorial results of the paper do not answer Conjecture 1.7, the main complexity question left open in [11]. However, the negative result in Theorem 1.4 shows the limitations of the techniques of [11] and implies that simple combinatorial embeddings are not sufficient to prove Conjecture 1.7. Therefore, substantially different methods would be required to prove the conjecture in the positive. It seems that the critical question that has to be understood first is the exact complexity of finding sparse expanders: Is there an $n^{o(k)}$ time algorithm that decides if a given k-vertex bounded-degree expander appears as subgraph?

References

[1] N. Alon and M. Capalbo. Sparse universal graphs for bounded-degree graphs. *Random Structures Algorithms*, 31(2):123–133, 2007.

[2] N. Alon, R. Yuster, and U. Zwick. Color-coding. *J. Assoc. Comput. Mach.*, 42(4):844–856, 1995.

[3] E. A. Bender and E. R. Canfield. The asymptotic number of labeled graphs with given degree sequences. *J. Combinatorial Theory Ser. A*, 24(3):296–307, 1978.

[4] H. L. Bodlaender. A tourist guide through treewidth. *Acta Cybernet.*, 11(1-2):1–21, 1993.

[5] H. L. Bodlaender. Treewidth: algorithmic techniques and results. In *Mathematical Foundations of Computer Science 1997 (Bratislava)*, pages 19–36. Springer, Berlin, 1997.

[6] J. Chen, B. Chor, M. Fellows, X. Huang, D. Juedes, I. Kanj, and G. Xia. Tight lower bounds for certain parameterized NP-hard problems. In *Proceedings of 19th Annual IEEE Conference on Computational Complexity*, pages 150–160, 2004.

[7] J. Chen, X. Huang, I. A. Kanj, and G. Xia. Linear FPT reductions and computational lower bounds. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 212–221, New York, 2004. ACM.

[8] E. C. Freuder. Complexity of k-tree structured constraint satisfaction problems. In *Proc. of AAAI-90*, pages 4–9, Boston, MA, 1990.

[9] M. Grohe and D. Marx. On tree width, bramble size, and expansion. *Journal of Combinatorial Theory Ser. B*, 99(1):218–228, 2009.

[10] R. Impagliazzo, R. Paturi, and F. Zane. Which problems have strongly exponential complexity? *J. Comput. System Sci.*, 63(4):512–530, 2001.

[11] D. Marx. Can you beat treewidth? *Theory of Computing*, 6(1):85–112, 2010.