The largest subgraph without a forbidden induced subgraph

Jacob Fox[*]   Rajko Nenadov[†]   Huy Tuan Pham[‡*]

Abstract

We initiate the systematic study of the following Turán-type question. Suppose Γ is a graph with n vertices such that the edge density between any pair of subsets of vertices of size at least t is at most $1 - c$, for some t and $c > 0$. What is the largest number of edges in a subgraph $G \subseteq \Gamma$ which does not contain a fixed graph H as an induced subgraph or, more generally, which belongs to a hereditary property $\mathcal{P}$? This provides a common generalization of two recently studied cases, namely Γ being a (pseudo-)random graph and a graph without a large complete bipartite subgraph. We focus on the interesting case where H is a bipartite graph.

We determine the answer up to a constant factor with respect to n and t, for certain bipartite H and for Γ either a dense random graph or a Paley graph with a square number of vertices. In particular, our bounds match if H is a tree, or if one part of H has d vertices complete to the other part, all other vertices in that part have degree at most d, and the other part has sufficiently many vertices. As applications of the latter result, we answer a question of Alon, Krivelevich, and Samotij on the largest subgraph with a hereditary property which misses a bipartite graph, and determine up to a constant factor the largest number of edges in a string subgraph of Γ. The proofs are based on a variant of the dependent random choice and a novel approach for finding induced copies by inductively defining probability distributions supported on induced copies of smaller subgraphs.

1 Introduction

The Turán number $\mathrm{ex}(n, H)$ is the maximum number of edges in a graph on n vertices which does not contain H as a subgraph. The Erdős-Stone-Simonovits theorem [13, 14] asymptotically determines Turán numbers for fixed non-bipartite H. It states that $\mathrm{ex}(n, H) = (1 - \frac{1}{\chi(H)-1} + o(1))\binom{n}{2}$, where $\chi(H)$ is the chromatic number of H, and the $o(1)$ term goes to 0 for H fixed as $n \to \infty$. For H bipartite, the Erdős-Stone-Simonovits theorem gives $\mathrm{ex}(n, H) = o(n^2)$, and it is further known that there is $\varepsilon = \varepsilon(H) > 0$ such that $\mathrm{ex}(n, H) \leq n^{2-\varepsilon}$. However, it remains a challenging open problem to give good estimates of Turán numbers of bipartite graphs. A prominent case is complete bipartite graphs $K_{s,r}$ with $s \leq r$, for which the Kővári-Sós-Turán theorem [26] gives $\mathrm{ex}(n, K_{s,r}) = O_r(n^{2-1/s})$. This bound is known to be sharp up to the constant factor if $s = 2$ or $s = 3$, or r is sufficiently large as a function of s (see [4, 8, 25]).

A property of graphs is *hereditary* if it is closed under taking induced subgraphs. A hereditary property of graphs is nontrivial if it contains all empty graphs and misses some graph. Throughout we assume all hereditary properties of graphs are nontrivial. For a hereditary property $\mathcal{P}$ of graphs and a graph Γ, let $\mathrm{ex}(\Gamma, \mathcal{P})$ be the maximum number of edges of a subgraph $G \subseteq \Gamma$ that belongs to $\mathcal{P}$. Note that in the

[*]Department of Mathematics, Stanford University, Stanford, CA. Email: **jacobfox@stanford.edu**. Research supported by NSF Award DMS-2154129.

[†]School of Computer Science, University of Auckland, New Zealand. Email: **rajko.nenadov@auckland.ac.nz**. Research supported by the New Zealand Marsden Fund.

[‡]Department of Mathematics, Stanford University, Stanford, CA. Email: **huypham@stanford.edu**. Research supported by a Clay Research Fellowship and a Stanford Science Fellowship.

special case that Γ is the complete graph K_n and $\mathcal{P}$ is the property of not containing H as a subgraph, the definition coincides with that of $\mathrm{ex}(n, H)$.

Alon, Krivelevich, and Samotij [2] proved the following extension of the Erdős-Stone-Simonovits theorem: For $0 < p \leq 1$ fixed, if $k = k(\mathcal{P}) \geq 2$ is the minimum chromatic number of a graph that does not belong to $\mathcal{P}$, then the Erdős-Renyi random graph $\Gamma = G(n, p)$ satisfies $\mathrm{ex}(\Gamma, \mathcal{P}) = (1 - \frac{1}{k-1} + o(1))p\binom{n}{2}$ with high probability. While the proof they give does not extend from Erdős-Renyi graphs to quasirandom graphs, a different proof based on Szemeredi's regularity lemma mentioned in the concluding remarks of their paper does extend to quasirandom graphs, but gives weak quantitative bounds. They left as an open problem to improve the bound $\mathrm{ex}(\Gamma, \mathcal{P}) = o(n^2)$ in the case $k = 2$ to $\mathrm{ex}(\Gamma, \mathcal{P}) \leq n^{2-\varepsilon}$ for some $\varepsilon = \varepsilon(\mathcal{P}) > 0$. We verify that this is the case.

Our results hold for a much larger class of graphs Γ. We say that a graph Γ is (c, t)-*sparse* if for every pair of vertex subsets $A, B \subset V(G)$ (not necessarily disjoint) with $|A|, |B| \geq t$, we have $e(A, B) < (1-c)|A||B|$. Here, $e(A, B)$ counts the number of pairs $(a, b) \in A \times B$ such that $\{a, b\}$ is an edge in Γ. In particular, any edge which lies in $A \cap B$ is counted twice. Roughly, a graph is (c, t)-sparse if it doesn't have large, almost complete bipartite subgraphs. If a graph is (c, t)-sparse, then all of its subgraphs are as well.

Generalizing the question of Alon, Krivelevich, and Samotij [2], we propose the following Turán-type problem.

Problem 1.1. *Given a hereditary property $\mathcal{P}$ which misses a bipartite graph and a (c, t)-sparse graph Γ, determine $\mathrm{ex}(\Gamma, \mathcal{P})$.*

Note that, if $\Gamma = G(n, p)$ with $0 < p < 1$ fixed, with high probability Γ is (c, t)-sparse for any fixed $c < 1 - p$ and any $t \geq C' \log n$ where $C' = C'(p, c)$ is sufficiently large. Problem 1.1 thus entails the study of Turán-type problems in dense random (and pseudorandom) graphs. Problem 1.1 also entails the case Γ is $K_{t,t}$-free (taking $c = 1/t^2$), which has been the subject of recent progress [5, 7, 20, 24, 30, 31, 35]. We refer the reader to Subsection 1.4 for further details on the relation between our work and these results.

The following extension of the Kővári-Sós-Turán theorem is our first main result. We say a bipartite graph $H = (A \cup B, E)$ is d-*bounded* if there are $1 \leq \ell \leq d$ vertices in A which are complete to B, and all other vertices in A have degree at most d.

Theorem 1.2. *Let H be a d-bounded bipartite graph and $c > 0$. Then there are constants C and $\alpha > 0$ such that if Γ is a (c, t)-sparse graph with n vertices, for some t, then any subgraph $G \subseteq \Gamma$ with at least $Ct^{1/d}n^{2-1/d}$ edges contains $\alpha q^{e(H)} n^{v(H)}$ copies of H which are induced copies in Γ, where $q = 2e(G)/n^2$.*

The bound on the number of edges in Theorem 1.2 is best possible up to the constant factor C in various cases as discussed in the next subsection, thus giving a complete answer (up to the constant factor) to Problem 1.1 in these cases. Also, the bound on the number of induced copies of H in G is best possible up to the constant factor α if for instance G is a random graph with edge density q.

If $H = (A \cup B, E)$ is such that all vertices in A have degree at most d, then H is a subgraph of a d-bounded graph. Thus we immediately get the following corollary, albeit without a bound on the number of copies.

Corollary 1.3. *Let H be a bipartite graph with all vertices in one part having degree at most d. If Γ is a (c, t)-sparse graph with n vertices, for some $c > 0$ and t, then there exists $C = C(c, H) > 0$ such that any subgraph $G \subseteq \Gamma$ with at least $Ct^{1/d}n^{2-1/d}$ edges contains a copy of H which is induced in Γ.*

A result of Alon, Krivelevich, and Sudakov [3] and Furedi [18] states that if H satisfies the condition in Corollary 1.3, then $\mathrm{ex}(n, H) = O(n^{2-1/d})$. Therefore, Corollary 1.3 can be seen as an induced analogue. Moreover, the counting part of Theorem 1.2 provides an induced analogue of a result of Conlon, Fox, and Sudakov [10] on Sidorenko's conjecture. The proof of Theorem 1.2 uses a variant of the dependent

2

random choice motivated by the one from [10], however we encounter a number of challenges that do not arise in the non-induced setup. As in [10], we construct appropriate embeddings of the vertices in B of H by selecting vertices in the common neighborhood of an ℓ-tuple of vertices. We then consider, for each vertex a in A of H with degree at most d, a large set of potential options for embedding a. However, compared to the non-induced setting, in order to guarantee that we have the correct number of options viable for creating an induced copy of H, we need to make several crucial modifications of the dependent random choice argument. In particular, we cannot obtain the desired guarantee when selecting uniformly at random ℓ vertices and looking for the embedding of B in their common neighborhood. Instead, an important idea in the proof is to choose each of the ℓ vertices one at a time, with each selection preceded by a degree regularization procedure.

Note that Corollary 1.3 solves the problem of Alon, Krivelevich, and Samotij [2]. Indeed, if $\Gamma = G(n, p)$ with $0 < p < 1$ fixed, with high probability the number of edges of G is $(1 + o(1))p\binom{n}{2}$, and, as earlier remarked, Γ is (c, t)-sparse for any fixed $c < 1 - p$ and any $t \geq C' \log n$ where $C' = C'(p, c)$ is sufficiently large. If $k(\mathcal{P}) = 2$, then there is a bipartite graph H not in $\mathcal{P}$, and for d being, say, the maximum degree of H, we can apply Corollary 1.3 to deduce $\mathrm{ex}(\Gamma, \mathcal{P}) \leq n^{2-1/d}$. In particular, $\mathrm{ex}(\Gamma, \mathcal{P}) \leq n^{2-\varepsilon}$ for some $\varepsilon = \varepsilon(\mathcal{P}) > 0$.

The rest of the introduction is organized as follows. In Subsection 1.1, we discuss several cases in which the lower bound on the number of edges of G in Theorem 1.2 is best possible. In Subsection 1.2, we discuss our result giving tight bounds on the size of a largest subgraph of a (c, t)-sparse graph forbidding an induced tree. In comparison to the ordinary Turán problem for trees, the situation here becomes drastically more involved. In Subsection 1.3, we discuss an application of Theorem 1.2 that obtains tight bounds on the size of a largest string subgraph of a (c, t)-sparse graph. We then discuss the relationship with the special case of forbidding a large complete bipartite subgraph in Subsection 1.4, which has been the subject of much recent interest.

1.1 Optimality of Theorem 1.2

For a graph H, let $\mathcal{P}_H$ be the hereditary graph property of not containing H as an induced subgraph. For Γ a graph with $p\binom{n}{2}$ edges, by averaging, we have $\mathrm{ex}(\Gamma, \mathcal{P}_H) \geq p\,\mathrm{ex}(n, H)$. Analogous to the case where H is not bipartite, it is natural to suspect that this is a good bound, at least for $\Gamma = G(n, p)$. Surprisingly, we show that this is not the case. Our general lower bound on $\mathrm{ex}(\Gamma, \mathcal{P}_H)$ is related to the chromatic number of the complement of Γ. In what follows, we are mainly concerned with $H = K_{s,r}$ as $\mathrm{ex}(n, K_{s,r})$ is well understood for r sufficiently large in s.

Lemma 1.4. *Let $2 \leq s \leq r$ be integers with $r > 2$. For every graph Γ, we have*

$$\mathrm{ex}(\Gamma, \mathcal{P}_{K_{s,r}}) \geq \frac{1}{2}e(\Gamma)\mathrm{ex}(\chi(\overline{\Gamma}), K_{s,r})/\binom{\chi(\overline{\Gamma})}{2}.$$

We also provide a bound for a general graph H (see Corollary 2.4). These are proved in Section 2.

Closely related to both $\chi(\overline{\Gamma})$ and (c, t)-sparseness is $\omega(\Gamma)$, the size of a largest clique in Γ. We have $\chi(\overline{\Gamma}) \geq n/\omega(\Gamma)$ and if Γ is (c, t)-sparse for c which is not too small, then $t \geq \omega(\Gamma)$. The latter follows from $e(A, A) = (1 - 1/|A|)|A|^2$ where A is a largest clique in Γ. In the case of $\Gamma = G(n, p)$, for constant $0 < p < 1$, it is well-known that $\omega(\Gamma) = \Theta(\log n)$ and $\chi(\overline{\Gamma}) = (1 + o(1))n/\omega(\Gamma)$ with high probability (see [6]). Moreover, standard estimates on the edge distribution of $G(n, p)$ imply it is with high probability $(c, 2\omega(\Gamma))$-sparse, for some $c = c(p) > 0$. Combining Theorem 1.2 and Lemma 1.4 with known estimates on $\mathrm{ex}(n, K_{s,r})$, we thus obtain the following.

Theorem 1.5. *Let $2 \leq s \leq r$ be fixed with r sufficiently large in s. If $0 < p < 1$ is fixed and $\Gamma = G(n, p)$, then with high probability we have $\mathrm{ex}(\Gamma, \mathcal{P}_{K_{s,r}}) = \Theta\left(n^{2-1/s}(\log n)^{1/s}\right)$.*

We also obtain tight bounds in the case $H = K_{s,r}$, again for r sufficiently large in s, and Γ is a Paley graph P_q for q a perfect square. For a prime power $q \equiv 1 \pmod 4$, recall that the Paley graph P_q is a graph whose vertex set is the elements of a finite field with q elements, and two distinct vertices are adjacent if their difference is a perfect square. The clique number of P_q is known to be at most $\sqrt{q}$, with equality if q is a perfect square. Unlike in the case of random graphs, we do not know that $\chi(\overline{\Gamma}) = \Theta(n/\omega(G))$, but we can find a large subgraph $\Gamma' \subseteq \Gamma$ for which this holds and then apply Lemma 1.4 with Γ'. We defer the details to Section 2. Standard estimates on character sums (see Theorem 3.4.2 of [34]) implies that for each $c < 1/2$ there is $C = C(c)$ such that every Paley graph with n vertices is $(c, C\sqrt{n})$-sparse. Combined, this gives the following.

Theorem 1.6. *Let* $2 \leq s \leq r$ *be fixed with* r *sufficiently large in* s. *If* Γ *is a Paley graph with a square number* n *of vertices, then* $\mathrm{ex}(\Gamma, \mathcal{P}_{K_{s,r}}) = \Theta(n^{2-1/(2s)})$.

1.2 Induced trees in (c, t)-sparse graphs

For a fixed tree T, we consider the property $\mathcal{P}_T$ of not containing T as an induced subgraph. The following is our second main result.

Theorem 1.7. *For any tree* T *and* $c > 0$, *there exists* $C > 1$ *such that the following holds. Given a* (c, t)-*sparse graph* Γ *on* n *vertices, for some* t, *every subgraph* $G \subseteq \Gamma$ *with at least* Ctn *edges contains a copy of* T *which is induced in* Γ.

Assuming that T contains at least two edges, the disjoint union of cliques is an induced-T-free graph. A graph Γ that is a disjoint union of cliques of size $t/4$ is $(1/2, t)$-sparse, thus the bound on the number of edges required in Theorem 1.7 is best possible. For $\Gamma = G(n, p)$, by taking a subgraph which is a disjoint union of cliques, it follows that $\mathrm{ex}(\Gamma, \mathcal{P}_T) = \Omega(n \log_{1/p} n)$ with high probability. In particular, this shows $\mathrm{ex}(\Gamma, \mathcal{P}_T) = \Theta(n \log n)$ when $0 < p < 1$ is a constant. Similarly, when Γ is a Paley graph P_q for q a perfect square, we have $\mathrm{ex}(\Gamma, \mathcal{P}_T) = \Theta(n^{3/2})$.

A standard way of proving that the Turán number of a fixed tree T is $\Theta(n)$ proceeds by passing on to a subgraph with sufficiently large (constant) minimum degree which depends on T, and then constructing a copy of T greedily. However, we face significant challenges in guaranteeing that the tree is induced under the (c, t)-sparse condition particularly when the graph G has irregular degree distribution. In order to address this, our proof of Theorem 1.7 follows from a fairly involved procedure to glue together *distributions* $\lambda_{T'}$ supported on induced copies of subtrees T' of T. This is our main technical contribution, thus we give a brief overview of the strategy.

The crucial property satisfied by the distributions λ is that, for a tree T' obtained by removing a leaf from T, λ_T and $\lambda_{T'}$ satisfy appropriate consistency relations. The main step is then to inductively construct, for subtrees T_0 of T, the measure λ_{T_0} from $\lambda_{T'}$ for smaller subtrees T' of T_0. We identify two subtrees T_1, T_2 of T_0 such that each T_1, T_2 is obtained from T_0 by removing a (different) leaf. Let T' be the intersection of T_1 and T_2. We first sample an embedding $\rho_{T'}$ of T' from $\lambda_{T'}$. Granted that $\rho_{T'}$ lies in a good event, we consider two independent extensions of $\rho_{T'}$ to embeddings of T_1 and T_2, sampled according to λ_{T_1} and λ_{T_2} conditional on the image of T' being $\rho_{T'}$. If combining both extensions gives us a good embedding of T_0, then we successfully obtain a sample defining λ_{T_0}. Using consistency relations between $\rho_{T'}$ and ρ_{T_1}, and between $\rho_{T'}$ and ρ_{T_2}, and the (c, t)-sparse assumption, we can show that we successfully obtain a sample with a good probability. We then close the induction by verifying that the consistency relations hold between the defined distribution λ_{T_0}, and the distributions λ_{T_1} and λ_{T_2}.

1.3 Largest induced string subgraphs

Returning to the original set-up of Alon, Krivelevich, and Samotij, beyond studying the case of a single forbidden induced subgraph, estimating $\text{ex}(\Gamma, \mathcal{P})$ remains an interesting problem when not all bipartite graphs belong to $\mathcal{P}$.

A string graph is an intersection graph of arcwise connected sets in the plane. Let $\mathcal{S}$ denote the family of string graphs. Note that not every graph is a string graph. Indeed, subdividing each edge of a nonplanar graph at least once produces a graph which is not a string graph. In particular, subdividing each edge of the complete graph on five vertices yields a bipartite graph on fifteen vertices which is not a string graph. A nontrivial upper bound on $\text{ex}(\Gamma, \mathcal{S})$ for Γ a (c,t)-sparse graph follows from Theorem 1.2. However, we prove the following much better upper bound using an additional property of string graph, which roughly says that any string graph has a dense induced subgraph with about the same average degree.

Theorem 1.8. *For every $c > 0$ there exists C, such that if Γ is a (c,t)-sparse graph on n vertices then* $\text{ex}(\Gamma, \mathcal{S}) \leq Ctn$.

Similarly as in the case of trees, this bound is tight when Γ is a random graph or a Paley graph with a square number of vertices. As an application of Theorem 1.8, we discuss in Section 5 the determination of $\text{ex}(\Gamma, \mathcal{P})$ for $\mathcal{P}$ the property of being an incomparability graph, and Γ a random graph or a Paley graph with a square number of vertices.

1.4 Forbidding large complete bipartite subgraphs

A related induced Turán number was introduced and studied by Loh et al. [28]. For a positive integer n and graphs B and H, let $\text{ex}(n, B, H\text{-ind})$ be the maximum number of edges in a graph on n vertices which does not contain B as a subgraph and H as an induced subgraph. By averaging over subsets, if B is bipartite with parts of size $r, s \leq t$ and has m edges, then not containing B as a subgraph implies that the graph is (c,t)-sparse with $c = 1 - (m-1)/rs$. Furthermore, if $B = K_{t,t}$ is a complete balanced bipartite graph, then it is not hard to show that not containing B as a subgraph is equivalent to being (c,t)-sparse with $c = 1/t^2$. Several very recent papers [5, 7, 20, 24] appeared proving upper bounds on $\text{ex}(n, K_{t,t}, H\text{-ind})$ for various graphs H. The bounds from [24] establish that $\text{ex}(n, K_{t,t}, H\text{-ind}) = t^{O(|V(H)|)}n^{2-1/d}$ if H is bipartite and every vertex in one part of H has maximum degree at most d. The proof of Theorem 1.2 shows that we may take $C = c^{-O(|V(H)|^2)}$, and hence implies that $\text{ex}(n, K_{t,t}, H\text{-ind}) = t^{O(|V(H)|^2)}n^{2-1/d}$ if H is d-bounded. While it is plausible that the methods from [24] could also be extended to the case of d-bounded graphs, the main message here is that our bounds obtained for the more general setup of (c,t)-sparse graphs rather than $K_{t,t}$-free graphs.

Note that, for $0 < p < 1$ fixed, the Erdős-Renyi random graph $G(n,p)$ almost surely does not contain $K_{t,t}$ as a subgraph for a t which is logarithmic in n. Thus the main result of Hunter, Milojević, Sudakov, and Tomon [24] gives a weaker upper bound (off by a polylogarithmic in n factor) on $\text{ex}(\Gamma, \mathcal{P}_{K_{s,r}})$ with $\Gamma = G(n,p)$ for $0 < p < 1$ fixed, compared to Theorem 1.5. For Paley graphs with a square number n of vertices, which contains a clique of order $\sqrt{n}$, these results do not give a nontrivial upper bound on induced Turán numbers. In general, our setup is a natural strengthening where instead of forbidding large complete bipartite subgraphs we forbid large dense subgraphs. One attractive feature of this strengthening is that, in some cases, it allows us to prove bounds which are optimal in the size of such subgraphs. Consequently, this gives optimal bounds in the case the host graph is, say, a dense random graph. With this in mind, our focus was on optimizing the dependence on t. We did not try to optimize, or even calculate, the dependence on c.

Organization. In Section 2, we describe several constructions giving lower bounds on $\text{ex}(\Gamma, \mathcal{P}_H)$. In Section 3, we prove Theorem 1.2 based on a novel variation of dependent random choice. In Section 4, we

give the proof Theorem 1.7 on embedding induced trees in subgraphs of (c,t)-sparse graphs. In Section 5, we discuss applications of Theorem 1.2 to the extremal numbers of the properties of being string graphs, incomparability graphs and comparability graphs. In Section 6, we finish with some concluding remarks, highlighting open problems and discussing consequences of our results for induced Ramsey numbers.

Notation. We follow standard graph theoretic notation. Given a graph G and a subset $S \subseteq V(G)$, we denote with $N_G^+(S)$ the union of neighborhoods of vertices in S, that is

$$N_G^+(S) = \bigcup_{v \in S} N_G(v).$$

Similarly, we let $N_G(S)$ denote the *common neighborhood* of vertices in S, that is

$$N_G(S) = \bigcap_{v \in S} N_G(v).$$

2 Lower bound on a largest subgraph in $\mathcal{P}_H$

Recall that any graph contains a bipartite subgraph with at least half of the edges. Therefore, Lemma 1.4 follows immediately from the following by taking the maximum bipartite subgraph F of a $K_{s,r}$-free graph on $\chi(\overline{\Gamma})$ vertices and $\mathrm{ex}(\chi(\overline{\Gamma}), K_{s,r})$ edges.

Lemma 2.1. *Let Γ be a graph such that $\chi(\overline{\Gamma}) = k$. Suppose F is a bipartite graph on k vertices which is $K_{s,r}$-free (as a subgraph), for some $2 \le s \le r$ such that $r \ne 2$. Then Γ has a subgraph G with at least $e(\Gamma)e(F)/\binom{k}{2}$ edges that is induced $K_{s,r}$-free.*

Proof. Since $\chi(\overline{\Gamma}) = k$, we can vertex partition Γ into k cliques $A_1, \ldots, A_k$. Consider a uniform random bijection f from $\{1, \ldots, k\}$ to $V(F)$. Let G be the subgraph of G consisting of all edges that are in the cliques A_i and any edge of Γ between A_i and A_j with $(f(i), f(j)) \in E(F)$.

We first show that G is necessarily induced $K_{s,r}$-free. Indeed, suppose there was an induced copy of $K_{s,r}$ in G. If each of the $s + r$ vertices are in distinct parts A_i, then the corresponding vertices of F would form a copy of $K_{s,r}$, thus a contradiction. No two vertices of an induced copy of $K_{s,r}$ from the same part in the bipartition can lie in the same A_i as A_i is a clique but they must be nonadjacent. So there must be a vertex from each part that maps to the same A_i, say v_1 and w_1 map to A_1. As $2 \le s \le r$ and $r > 2$, we can choose v_2, v_3 (distinct from v_1) from one part and w_2 (distinct from w_1) from the other part in the bipartition of H. As all three v_1, v_2, and v_3 have to lie in different A_j's, and w_2 cannot lie in A_1, we obtain three different A_j's which are pairwise connected by an edge. This implies F contains a triangle, contradicting that F is bipartite.

Each edge of Γ has probability at least $e(F)/\binom{k}{2}$ of surviving in G, and so the expected number of edges of G is at least $e(\Gamma)e(H)/\binom{k}{2}$. Hence there is a choice of such a G with at least this many edges. $\square$

We remark that the bipartite assumption on H in Lemma 2.1 is stronger than actually needed, but for an extremal problem, this assumption does not make a big difference, only changing the answer by a factor at most $1/2$.

As remarked earlier, for a graph Γ with n vertices we have $\chi(\overline{\Gamma}) \ge n/\omega(\Gamma)$. This bound is asymptotically sharp with high probability for $\Gamma = G(n,p)$ with $0 < p < 1$ fixed. In the case Γ is a vertex-transitive graph, by randomly picking out largest cliques we get that this lower bound on $\chi(\overline{\Gamma})$ is sharp up to a logarithmic

factor. In order to get an optimal bound in Theorem 1.6, it suffices to cover most of the vertices instead of all (or, more importantly, covered vertices need to span many edges).

Lemma 2.2. *Let Γ be a vertex transitive graph on n vertices. Then Γ has an induced subgraph Γ' with at least $e(\Gamma)/2$ edges satisfying $\chi(\overline{\Gamma'}) \leq (\ln 4)n/\omega(\Gamma)$.*

Proof. Since Γ is vertex transitive, it is d-regular for some d. Let $A_1, \ldots, A_k$ be cliques of size $\omega = \omega(\Gamma)$ in Γ sampled uniformly and independently. Each vertex belongs to the same number of cliques of size ω, thus the probability that a particular vertex does not belong to one A_i is exactly $1 - \omega/n$. By the independence of A_i's we have that it does not belong to any A_i with probability $(1 - \omega/n)^k$. Using the linearity of expectation, we conclude

$$\mathbb{E}\left[|V(\Gamma) \setminus (A_1 \cup \ldots \cup A_k)|\right] = n(1 - \omega/n)^k.$$

Therefore, for $k = (\ln 4)n/\omega$ there exist cliques $A_1, \ldots, A_k$ which cover all but at most $n/4$ vertices. Moreover, the subgraph Γ' induced by $A_1 \cup \ldots A_k$ can be partitioned into k cliques, thus $\chi(\overline{\Gamma'}) \leq k$. The vertices not in Γ' are incident to at most $dn/4$ edges, so $e(\Gamma') \geq e(\Gamma) - dn/4 = e(\Gamma)/2$. $\square$

For vertex transitive graphs Γ, we obtain a lower bound on $\mathrm{ex}(\Gamma, \mathcal{P}_{K_{s,r}})$ by applying Lemma 1.4 to the induced subgraph Γ' of Γ. Crucially, by passing on to Γ' we lose only a constant fraction of all the edges.

Proof of Theorem 1.6. Let $\Gamma = P_q$ for a square q. In this case it is known that $\omega(\Gamma) = \sqrt{q} = \sqrt{n}$. As a Paley graph is a Cayley graph, and these are known to be vertex transitive, by applying Lemma 2.2 we obtain a subgraph $\Gamma' \subseteq \Gamma$ with $e(\Gamma') \geq n(n-1)/8$ edges such that $\chi(\overline{\Gamma'}) \leq 2\sqrt{n}$. By applying Lemma 1.4 on Γ' instead of Γ, we obtain

$$\mathrm{ex}(\Gamma, \mathcal{P}_{K_{s,r}}) \geq \mathrm{ex}(\Gamma', \mathcal{P}_{K_{s,r}}) = \Omega(n^{2-1/(2s)}). \qquad \square$$

2.1 General graphs H

For a graph H and a vertex partition $\mathcal{P} : V(H) = V_1 \sqcup \ldots \sqcup V_k$, the quotient graph of H obtained from the partition $\mathcal{P}$ is the graph with vertex set $\{1, \ldots, k\}$ and where (i, j) with $i \neq j$ is an edge if some vertex in V_i is adjacent to a vertex in V_j. That is, the quotient graph is the graph obtained from H by contracting each of the sets V_i. If each V_i is a clique in H, then this is a *clique quotient of H*. For each graph H, let $\mathcal{F}_H$ denote the family of bipartite graphs that are clique quotients of H. Note that $H \in \mathcal{F}_H$ as H is a clique quotient of itself by taking the vertex partition into singletons. Note that if H is a bipartite graph, then the family $\mathcal{F}_H$ consists of exactly the bipartite graphs that can be obtained by contracting disjoint edges of H. The proof of Lemma 2.1 extends directly to show the following generalization.

Lemma 2.3. *Let Γ be a graph such that $\chi(\overline{\Gamma}) = k$. Let H be a graph and suppose F is a bipartite graph on k vertices which does not contain any graphs in $\mathcal{F}_H$ as a subgraph. Then Γ has a subgraph G with at least $e(\Gamma)e(F)/\binom{k}{2}$ edges that is induced H-free.*

Lemma 2.3 immediately implies the following generalization of Lemma 1.4.

Corollary 2.4. *For all graphs H and Γ, we have*

$$\mathrm{ex}(\Gamma, \mathcal{P}_H) \geq \frac{1}{2}e(\Gamma)\mathrm{ex}(\chi(\overline{\Gamma}), \mathcal{F}_H)/\binom{\chi(\overline{\Gamma})}{2}.$$

We refer the reader to a discussion in Section 6 where we make use of this corollary to give a lower bound on the number of edges when the forbidden subgraph is an even cycle C_{2k}.

3 Counting d-bounded graphs

In this section we prove Theorem 1.2. The following two lemmas will be useful to guarantee that the nonedges in the copy of H we find in a subgraph of the (c,t)-sparse graph Γ are nonedges of Γ.

Lemma 3.1. *Let Γ be a (c,t)-sparse graph with $c \leq 1/2$. Let $V, W \subseteq V(\Gamma)$ with $|W| \geq c^{-r}t$, for some positive integer r. If we pick a uniformly random r-tuple R of vertices in V, then*

$$\mathbb{P}_R\left[|W \setminus N_\Gamma^+(R)| < c^r|W|\right] \leq rt/|V|.$$

Proof. We pick vertices in a tuple R one at a time, picking v_i in step i. Let $W_0 = W$ and after picking v_i, let $W_i = W \setminus N_\Gamma^+(\{v_1, \ldots, v_i\})$. Assuming $|W_i| \geq c^{-i}|W| \geq t$, the number of vertices in V that do not have at least $c|W_i|$ nonneighbors in W_i is at most t. Hence, the probability that $|W_{i+1}| < c|W_i|$ conditioned on $|W_i| \geq c^{-i}|W|$ is at most $t/|V|$. Therefore, the probability that the number of common nonneighbors of R is less than $c^{-r}|W|$ is at most $rt/|V|$. $\qquad\square$

Lemma 3.2. *Let Γ be a (c,t)-sparse graph with $0 < c \leq 1/2$. Let $k \geq 2$ and $V_1, \ldots, V_k$ be vertex subsets of Γ, not necessarily disjoint, each of size at least $m \geq k^2 c^{-k^2} t$. A uniformly random k-tuple in $V_1 \times \cdots \times V_k$ forms an independent set of k distinct elements with probability at least c^{k^2}.*

Proof. Consider picking the vertices one at a time, picking $v_i \in V_i$ in step i. Having picked $v_1, \ldots, v_i$ for some $0 \leq i < k$, let $V_{j,i}$ for $i < j \leq k$ be the set of vertices in V_j not adjacent to $v_1, \ldots, v_i$. Note that $V_{j,0} = V_j$ for every j. For $0 \leq i < k$, call $v_{i+1} \in V_{i+1}$ *good* if the number of nonneighbors of v_{i+1} in $V_{j,i}$ is at least $c|V_{j,i}|$ for all $i + 1 < j \leq k$. Note that if $|V_{j,i}| \geq t$, then all but at most $t - 1$ vertices $v_{i+1} \in V_{i+1}$ have at least $c|V_{j,i}|$ nonneighbors in $V_{j,i}$. Hence, assuming $|V_{j,i}| \geq t$ for $i + 1 < j \leq k$, there are at most $(k - i)(t - 1)$ vertices in V_{i+1} which are not good. Note that if each $v_1, \ldots, v_i$ is good, then $|V_{j,i}| \geq c^i|V_j| \geq t$ for $i + 1 \leq j \leq k$. Hence the probability that not all $v_1, \ldots, v_{k-1}$ are good is at most

$$\sum_{i=1}^{k-1} \frac{(k-i)(t-1)}{m} = \binom{k}{2}(t-1)/m.$$

For $1 \leq i \leq k - 1$, if $v_1, \ldots, v_i$ are good, then $|V_{i+1,i}| \geq c^i|V_{i+1}|$. So conditioning on $v_1, \ldots, v_{k-1}$ being good, the probability that there are no edges between $v_1, \ldots, v_k$ is at least $\prod_{i=1}^k c^{i-1} = c^{\binom{k}{2}}$.

For $1 \leq i < j \leq k$, the probability that the same vertex is picked in V_i and V_j is at most $1/m$. Hence by the union bound, the probability that uniform random $v_1, \ldots, v_k$ are all distinct is at least $1 - \binom{k}{2}/m$. All together, the probability that the chosen vertices form an independent set of distinct vertices is at least

$$c^{\binom{k}{2}} - \binom{k}{2}t/m \geq c^{k^2}.$$

$\qquad\square$

Given a graph G and a subset $S \subseteq V(G)$, we denote with $d_{\text{avg}}(S)$ the average degree of vertices in S. We say that a bipartite graph G on vertex sets U and R is δ-reduced if $d(u) \in [\delta d_{\text{avg}}(U), \delta^{-1}d_{\text{avg}}(U)]$ for all $u \in U$. The following lemma provides a strong form of one-sided regularization.

Lemma 3.3. *Given $\kappa > 1$, let $0 < \delta < 1/4$ be so that $1 - 2\delta > \delta^{1-1/\kappa}$. Let G be a bipartite graph with vertex parts U and R. Then there exists $U' \subseteq U$ such that $G' = [U', R]$ is δ^2-reduced,*

$$d_{\text{avg}}(U') \geq \delta d_{\text{avg}}(U),$$

and for any real $x \geq \kappa$ we have

$$|U'|d_{\text{avg}}(U')^x \geq \delta^x|U|d_{\text{avg}}(U)^x.$$

Proof. Set $U_0 = U$, and consider the following process for $i \geq 0$. As long as

$$\sum_{\substack{u \in U_i \\ d(u) \leq \delta^{-1} D_i}} d(u) < 2\delta e(U_i, R), \tag{1}$$

where $D_i = d_{\mathrm{avg}}(U_i)$ is the average degree of vertices in U_i, set

$$U_{i+1} := \{u \in U_i : d(u) \geq \delta^{-1} D_i\}$$

and proceed to the next iteration. By Markov's inequality, we have $|U_{i+1}| \leq \delta |U_i|$. By (1), we have $|e(U_{i+1}, R)| \geq (1 - 2\delta)|e(U_i, R)|$.

Let i_0 denote the first index i for which (1) does not hold. Set

$$U' = \{u \in U_{i_0} : \delta D_{i_0} \leq d(u) \leq \delta^{-1} D_{i_0}\}.$$

Note that $G' = G[U', R]$ is δ^2-reduced. Furthermore, from the assumption that (1) fails for i_0 we have $e(U', R) \geq \delta e(U_{i_0}, R)$. We also have that $e(U_{i_0}, R) \geq (1 - 2\delta)^{i_0} e(G)$ and $|U'| \leq |U_{i_0}| \leq \delta^{i_0} |U|$. Then

$$d_{\mathrm{avg}}(U') = \frac{e(U', R)}{|U'|} \geq \frac{\delta e(U_{i_0}, R)}{|U_{i_0}|} \geq \delta \left(\frac{1 - 2\delta}{\delta}\right)^{i_0} d_{\mathrm{avg}}(U) \geq \delta d_{\mathrm{avg}}(U).$$

Similarly, for $x \geq \kappa$ we have

$$|U'| d_{\mathrm{avg}}(U')^x = \frac{e(U', R)^x}{|U'|^{x-1}} \geq \frac{\delta^x e(U_{i_0}, R)^x}{|U_{i_0}|^{x-1}} \geq \delta^x \left(\frac{1 - 2\delta}{\delta^{1-1/x}}\right)^{x i_0} |U| d_{\mathrm{avg}}(U)^x.$$

As $1 - 2\delta > \delta^{1-1/\kappa} \geq \delta^{1-1/x}$, we obtain the desired inequality. $\qquad\square$

Given a graph H, for the rest of the section we fix $\kappa = 1 + 1/v(H)$ and let δ be as in Lemma 3.3.

The following lemma is the heart of the proof of Theorem 1.2. Given a graph Γ with n vertices and a bipartite subgraph $G \subseteq \Gamma$ on vertex sets U and R with $|U| = m, |R| = r$, let $q = e(G)/(mr)$ be the edge density of G. We say that $S \subseteq R$ is (ε, T)-*rich*, for some $\varepsilon > 0$ and $T \subseteq U$, if

$$|N_G(S) \setminus N_\Gamma^+(T)| \geq \varepsilon m q^{|S|}.$$

For $1 \leq k \leq d$, we say that T is (β, ε, k)-*good* if the number of k-tuples of vertices in $N_G(T)$ that are not (ε, T)-rich is at most $\beta |N_G(T)|^k$. Note that we do not require elements in these tuples to be distinct.

Lemma 3.4. *For every $0 < c < 1/2$, $\beta > 0$, and positive integer d there are $C, \varepsilon > 0$ such that the following holds. Let Γ be a (c, t)-sparse graph with n vertices and $G \subseteq \Gamma$ be a δ-reduced bipartite subgraph of Γ with parts U, R with $|U| = m, |R| = r$, and at least $Crm^{1-1/d} t^{1/d}$ edges. Then, with probability at least $1/2$ over uniformly chosen $u \in U$, we have that u is (β, ε, k)-good for all $k \leq d$.*

Proof. Consider $k \in [d]$. Let $\mathcal{B}_k$ denote the set of all vertices u which are not (β, ε, k)-good, where $\varepsilon > 0$ will be chosen as a sufficiently small constant in terms of β, c and d.

We distinguish two types of k-tuples that are not (ε, u)-rich. Let $\mathcal{S}$ denote the family of k-tuples S in R such that $|N_G(S)| \geq c^{-d} \varepsilon m q^k$. Let $s_1(u)$ denote the number of k-tuples $S \subset N_G(u)$ for which S is not (ε, u)-rich and $S \in \mathcal{S}$. Let $s_2(u)$ denote the number of k-tuples $S \subset N_G(u)$ for which S is not (ε, u)-rich and $S \notin \mathcal{S}$.

We first bound $\mathbb{E}[s_1(u)]$. For each $S \in \mathcal{S}$, since Γ is (c, t)-sparse there are at most t vertices u for which S is not (ε, u)-rich. Hence,

$$\mathbb{P}[S \text{ is not } (\varepsilon, u)\text{-rich} \mid S \in N_G(u)] \leq \frac{t}{|N_G(S)|} \leq \frac{t}{c^{-d} \varepsilon m q^k}.$$

In particular, since G is δ-reduced,

$$\frac{c^{-d} \varepsilon m q^k}{t} \mathbb{E}[s_1(u)] \leq \mathbb{E}[|N_G(u)|^k] \in [(\delta D)^k, (\delta^{-1} D)^k],$$

9

where D is the average degree of vertices in U. Next, we bound $\mathbb{E}[s_2(u)]$. For $S \notin \mathcal{S}$ we have

$$\mathbb{P}[S \in N_G(u)] \leq c^{-d}\varepsilon q^k.$$

Hence,

$$\mathbb{E}[s_2(u)] \leq r^k c^{-d}\varepsilon q^k \leq c^{-d}\varepsilon(\delta^{-1}D)^k.$$

By Markov's inequality, with probability at least $1 - 1/(2d)$ over $u \in U$,

$$s_1(u) + s_2(u) \leq \frac{2dt}{c^{-d}\varepsilon mq^k}(\delta^{-1}D)^k + 2dc^{-d}\varepsilon(\delta^{-1}D)^k \leq 2d(\delta^{-1}D)^k \left(\frac{c^d t}{\varepsilon mq^k} + c^{-d}\varepsilon\right).$$

For ε sufficiently small and C sufficiently large, we then have $s_1(u) + s_2(u) \leq \beta|N_G(u)|^k$ with probability at least $1 - 1/(2d)$.

By the union bound, with probability at least $1/2$ we have that u is (β, ε, k)-good for all $k \in [1, d]$. $\qquad\square$

Let $H = (A \cup B, E)$ be a d-bounded bipartite graph with $A_0 \subseteq A$ denoting the set of vertices complete to B and $A_1 = A \setminus A_0$. Recall that, by the definition, every vertex in A_1 has degree at most d. Let $\ell = |A_0|$, $a = |A_1|$, $b = |B|$, and let e be the number of edges of H.

Claim 3.5. *Assume $\ell = 1$. Let G be a bipartite subgraph of Γ with vertex sets $|U| = m, |R| = r$ and at least $Crm^{1-1/d}t^{1/d}$ edges. Assume that G is δ-reduced, and $r \geq |V(H)|^2 c^{-|V(H)|^2}\delta t(m/t)^{1/d}$. Let $\beta > 0$ be sufficiently small in H, c and d. Suppose $u \in U$ is (β, ε, k)-good. Then, for some constant $\gamma = \gamma(\delta, c, \beta, \varepsilon, H) > 0$, there are at least*

$$\gamma m^a r^b q^{e(H)}$$

copies of H in G which are induced in Γ and contain u, where $q = e(G)/(mr)$ is the edge density of G.

Proof. Consider a vertex $u \in U$ which is (β, ε, k)-good for all $k \leq d$. We then consider embeddings of H where A_0 (of size 1) maps to u. We then embed B according to a uniformly random b-tuple $X \subseteq N_G(u)$. We first show that with decent probability, X satisfies the following properties:

1. all elements in X are distinct and independent in Γ,

2. every d-tuple Y in X is $(c^{b-d}\varepsilon, \{u\} \cup (X \setminus Y))$-rich.

The first property holds with probability at least c^{b^2} by Lemma 3.2. We show the second property holds with probability close to one by a union bound over Y. By the assumption that u is (β, ε, k)-good for $k \leq d$, one particular k-tuple Y in X fails to be (ε, u)-rich with probability at most β. Conditioning on Y being (ε, u)-rich, by Lemma 3.1 the probability of choosing the remaining vertices in X such that

$$|N_G(Y) \setminus N_\Gamma^+(\{u\} \cup (X \setminus Y))| < c^{b-k}\varepsilon mq^k$$

is at most $(b-k)t/(\varepsilon mq^k)$. Therefore, one particular Y fails to be $(\varepsilon, \{u\} \cup (X \setminus Y))$-rich with probability at most $\beta + bt/(\varepsilon mq^k)$. As $mq^d > Ct$, by taking C to be sufficiently large we can upper bound this probability by 2β. Finally, for $\beta > 0$ sufficiently small, by taking union bound over all d-tuples of X we have that X fails either property with probability at most

$$1 - c^{b^2} + \binom{b}{d}2\beta < 1 - c^{b^2}/2.$$

Next, we show that each X which satisfies the two properties gives rise to $c^{2ab}\varepsilon^a m^a q^{e(H)-b}$ copies of H in G which are induced in Γ. As there are $c^{b^2}|N_G(u)|^b/2 \geq (c^{b^2}/2)(\delta qr)^b$ such b-tuples X, this implies the claim.

Fix an ordering $v_1, \ldots, v_b$ of the vertices in B, and consider one such X. For every tuple of vertices Y in X, let $U(Y) = N_G(Y) \setminus N_\Gamma^+(\{u\} \cup (X \setminus Y))$ denote the 'unique' neighborhood of Y. Let f be an arbitrary

10

mapping of $v_1, \ldots, v_b$ to B and, for each $w \in A_1$, let $B_w = f(N_H(w))$ and $V_w = U(B_w)$. Now pick one vertex from each V_w for $w \in A_1$, uniformly at random. As $|V_w| \geq c^b \varepsilon m q^{|N_H(w)|}$, by Lemma 3.2 we have that all chosen vertices are distinct and independent in Γ with probability c^{a^2}. Assuming this is the case, we obtain a copy of H in G which is induced in Γ. Therefore, there are at least $c^{a^2}(c^b \varepsilon m)^a q^{e(H)-b}$ copies which contain u. $\qquad \square$

Theorem 1.2 follows immediately from Lemma 3.6 below, upon replacing G with a bipartite subgraph with at least $(1 - o(1))e(G)/2$ edges on parts U and R with $|U| = \lfloor n/2 \rfloor$ and $|R| = \lceil n/2 \rceil$.

Lemma 3.6. *For every $0 < c < 1/2$, and positive integer d and $\ell \leq d$, there are $C_\ell, C'_\ell, \gamma_\ell > 0$ such that the following holds. Let Γ be a (c,t)-sparse graph with n vertices and $G \subseteq \Gamma$ be a bipartite subgraph with parts U, R with $|U| = m, |R| = r$ and edge density $q \geq C_\ell m^{-1/d} t^{1/d}$. Assume that $r \geq C'_\ell t(m/t)^{\ell/d}$. Let H be d-bounded with vertex parts A and B, ℓ vertices in A complete to B, $|A| = \ell + a$ and $|B| = b$. Then, there are at least $\gamma_\ell m^{a+\ell} r^b q^{e(H)}$ copies of H in G which are induced in Γ.*

Proof. We prove the statement by induction on ℓ. We let $\beta > 0$ be sufficiently small (depending on H, c and d) and let $\varepsilon > 0$ be given by Lemma 3.4. For $\ell \geq 1$, we choose the constants C_ℓ and C'_ℓ as follow.

$$C_\ell = 1/(2(1-\beta)\varepsilon\delta)^\ell, \quad C'_\ell = \delta^{-2\ell}, \quad \gamma_\ell = ((1-\beta)^{e(H)}\varepsilon^{e(H)}\delta^{2b}/2)^\ell.$$

Base case: $\ell = 1$.

Apply Lemma 3.3 to obtain a δ^2-reduced induced subgraph $G' = G[U', R]$ with $|U'| = m'$. In particular, for $x \geq 1 + 1/v(H)$ we have

$$m' \cdot d_{\mathrm{avg}}(U')^x \geq \delta^x m \cdot d_{\mathrm{avg}}(U)^x.$$

Dividing each side by r^x, we get

$$m'(q')^x \geq \delta^x m q^x, \tag{2}$$

where $q' = e(G')/(m'r)$ is the density of G'. By Lemma 3.4 and Claim 3.5, the number of copies of H in G' induced in Γ is at least

$$\gamma \frac{m'}{2}(m')^a r^b (q')^{e(H)} = \frac{\gamma}{2} r^b \left(m' \cdot (q')^{e(H)/(a+1)} \right)^{a+1} \geq \frac{\gamma \delta^{e(H)}}{2} r^b \left(m q^{e(H)/(a+1)} \right)^{a+1} = \gamma' m^{a+1} r^b q^{e(H)}.$$

The first inequality follows from (2) applied with $x = e(H)/(a+1) \geq 1 + 1/v(H)$.

Inductive step: Assume the conclusion of the lemma holds when H has $\ell - 1$ vertices complete to the other side.

Apply Lemma 3.3 to obtain a δ^2-reduced induced subgraph $G' = G[U', R]$ with $|U'| = m'$. By Lemma 3.4, at least $m'/2$ vertices $u \in U'$ are such that $N_{G'}(u)$ is (β, ε, k)-good for all $k \leq d$. In particular, for at least $(1 - \beta)|N_{G'}(u)|$ vertices $x \in N_{G'}(u)$, we have

$$|N_{G'}(x) \setminus N_\Gamma(u)| \geq \varepsilon m' q',$$

where $q' = e(G')/(m'r)$.

Let H' be the graph obtained from H by removing one of ℓ vertices in A complete to B. Let G'' be the induced subgraph of G between $N_{G'}(u)$ and $U' \setminus N_\Gamma(u)$. We then have

$$e(G'') = \sum_{x \in N_{G'}(u)} |N_{G'}(x) \setminus N_\Gamma(u)| \geq (1 - \beta)|N_{G'}(u)| \cdot \varepsilon m' q',$$

thus

$$q'' \geq (1 - \beta)\varepsilon q'. \tag{3}$$

11

Towards obtaining more precise estimate on the density q'' of G'', we have

$$\frac{e(G'')}{|N_{G'}(u)|(m')^{1-1/d}} \geq (1-\beta)\varepsilon\frac{m'q'}{r(m')^{1-1/d}}$$

$$\geq (1-\beta)\varepsilon\frac{d_{\mathrm{avg}}(U')(m')^{1/d}}{r}$$

$$= (1-\beta)\varepsilon\frac{d_{\mathrm{avg}}(U)m^{1/d}}{r}\left(\frac{d_{\mathrm{avg}}(U')^d m'}{d_{\mathrm{avg}}(U)^d m}\right)^{1/d}$$

$$\geq (1-\beta)\varepsilon\delta\frac{d_{\mathrm{avg}}(U)m^{1/d}}{r} = (1-\beta)\varepsilon\delta qm^{1/d} \geq C_\ell(1-\beta)\varepsilon\delta t^{1/d},$$

where the third inequality follows from the property of Lemma 3.3 applied with $x = d$. Hence, the edge density q'' of G'' satisfies $q'' \geq C_\ell(1-\beta)\varepsilon\delta(m')^{-1/d}t^{1/d}$. Our choice of C_ℓ guarantees that $C_\ell(1-\beta)\varepsilon\delta \geq C_{\ell-1}$. Moreover,

$$|N_{G'}(u)| \geq \delta d_{\mathrm{avg}}(U') \geq \delta^2 d_{\mathrm{avg}}(U) = \delta^2 rq \geq \delta^2 \cdot C'_\ell t(m/t)^{\ell/d} \cdot C_\ell(t/m)^{1/d},$$

which is at least $C'_{\ell-1}t(m/t)^{(\ell-1)/d}$ for our choice of C'_ℓ. Then, by the inductive hypothesis, we have that the number of copies of H' in G'' is at least

$$\gamma_{\ell-1}(m')^{a+\ell-1}|N_{G'}(u)|^b(q'')^{e(H)-b}$$

$$\overset{(3)}{\geq} \gamma_{\ell-1}(1-\beta)^{e(H)}\varepsilon^{e(H)} \cdot (m')^{a+\ell-1} \cdot |N_{G'}(u)|^b \cdot (q')^{e-b}$$

$$\geq \gamma_{\ell-1}(1-\beta)^{e(H)}\varepsilon^{e(H)} \cdot (m')^{a+\ell-1} \cdot \delta^{2b}d_{\mathrm{avg}}(U')^b \cdot (d_{\mathrm{avg}}(U')/r)^{e(H)-b}$$

$$\geq \gamma' \cdot m^{a+\ell-1} \cdot d_{\mathrm{avg}}(U)^{e(H)}r^{b-e(H)}$$

$$= \gamma' m^{a+\ell-1}r^b q^{e(H)},$$

where $\gamma' = \gamma_{\ell-1}(1-\beta)^{e(H)}\varepsilon^{e(H)}\delta^{2b}$. Summing over vertices $u \in U'$ such that $N_{G'}(u)$ is (β, ε, k)-good for all $k \leq d$, we obtain that the number of copies of H in G which are induced in Γ is at least

$$(m/2)\gamma' m^{a+\ell-1}r^b q^{e(H)} = \gamma_\ell m^{a+\ell}r^b q^{e(H)}.$$

This completes the inductive step. $\qquad\square$

4 Trees

The proof of Theorem 1.7 is based on a delicate inductive argument which 'glues' induced copies of subtrees of T. To this end, we start by introducing necessary notions. We only consider labelled copies of T, which we identify with an injective function $\phi\colon V(T) \to V(G)$. For a given copy ϕ of T and a leaf $v \in V(T)$, we denote by $\phi - v$ a copy of $T - v$ given by $\phi\restriction_{V(T)-v}$.

Unique neighbors. Given graphs G and Γ on the same vertex set, we say that a subset of vertices $S \subset V(G)$ is *c-unique* if for every $v \in S$ we have

$$|N_G(v) \setminus N_\Gamma^+(S - v)| \geq c^{|S|-1}|N_G(v)|.$$

We say that a copy ϕ of T is c-unique if the set $\phi(T)$ is c-unique.

Extending $T - v$ to T. Let T be a tree with at least two vertices, $v \in V(T)$ a leaf in T and $h_v \in V(T)$ the unique neighbor of v. Suppose λ_{T-v} and λ_T are probability measures over copies of $T - v$ and T, respectively, in some graph G. We say that λ_T (K, ε, κ)-*extends* λ_{T-v}, for $K, \varepsilon, \kappa > 0$, if the following two properties hold:

$$\lambda_T(\phi) \leq \frac{K\lambda_{T-v}(\phi - v)}{\deg_G(\phi(h_v))} \quad \text{for every } \phi \in \mathrm{supp}(\lambda_T) \tag{4}$$

12

and

$$\mathbb{P}_{\phi' \sim \lambda_{T-v}} \left[\lambda_T(\phi') \geq \varepsilon \lambda_{T-v}(\phi') \right] \geq 1 - \kappa, \tag{5}$$

where

$$\lambda_T(\phi') = \lambda_T(\{\phi \in \mathrm{supp}(\lambda_T) \colon \phi - v = \phi'\}).$$

Note that (4) implies

$$\lambda_T(\phi') \leq K \lambda_{T-v}(\phi') \quad \text{for every } \phi' \in \mathrm{supp}(\lambda_{T-v}). \tag{6}$$

The following two lemmas are indirectly related to these properties, thus we state them now.

Lemma 4.1. *Let Γ be a (c,t)-sparse graph. Suppose π_1 and π_2 are probability measures over $V(\Gamma)$ such that $\pi_i(v) < \gamma/t$ for all $v \in V(\Gamma)$ and $i \in \{1,2\}$, for some $0 < \gamma < 1$. Define the* good *set of vertices with respect to π_2 as*

$$D(\pi_2) = \{x \in V(\Gamma) \colon \pi_2(\{y \in V(\Gamma) \colon (x,y) \notin \Gamma\}) \geq c\}.$$

Then

$$\mathbb{P}_{x \sim \pi_1} \left[x \in D(\pi_2) \right] > 1 - \gamma.$$

Proof. Assume for the sake of contradiction that $\mathbb{P}_{x \sim \pi_1} \left[x \in D(\pi_2) \right] \leq 1 - \gamma$. Since $\pi_1(v) < \gamma/t$ for all v, there is a set $B_1 \subseteq \Gamma$ of size at least t for which $\pi_2(\{y \in V(\Gamma) \colon (x,y) \notin \Gamma\}) < c$, and hence $\pi_2(\{y \in V(\Gamma) \colon (x,y) \in \Gamma\}) > 1 - c$, for all $x \in B_1$. Thus, letting U_1 denote the uniform distribution on B_1,

$$\mathbb{E}_{x_1 \sim U_1, x_2 \sim \pi_2}[\mathbb{I}((x,y) \in \Gamma)] > 1 - c.$$

Recall the standard fact that, any distribution P on a finite set V with $\max_{v \in V} P(v) \leq \theta$ can be written as a convex combination of distributions, each of which is uniform over its support (which is a subset of V) of size $\lfloor 1/\theta \rfloor$ (see [33, Lemma 6.10]). Applying this fact to π_2 with $\theta = \gamma/t$, and by averaging, we obtain that there is a subset $B_2 \subseteq \Gamma$ of size $\lfloor t/\gamma \rfloor \geq t$ such that, for U_2 the uniform distribution on B_2,

$$\mathbb{E}_{x_1 \sim U_1, x_2 \sim U_2}[\mathbb{I}((x,y) \in \Gamma)] > 1 - c.$$

However, this contradicts the (c,t)-sparse assumption on Γ. $\qquad \square$

Lemma 4.2. *Let Γ be a (c,t)-sparse graph. Suppose π_1 and π_2 are probability measures over $V(\Gamma)$ such that $\pi_i(v) < \gamma/t$ for all $v \in V(\Gamma)$ and $i \in \{1,2\}$, for some $0 < \gamma < 1$. Moreover, let $\mathcal{S} = \{S_v \colon v \in \mathrm{supp}(\pi_2)\}$ be a family of subsets of $V(\Gamma)$, each of size $|S_v| \geq t$. Denote with $D(\pi_2, \mathcal{S})$ the* good *set of vertices with respect to π_2 and the family $\mathcal{S}$, defined as*

$$D(\pi_2, \mathcal{S}) = \{x \in V(\Gamma) \colon \pi_2(\{y \in V(\Gamma) \colon |S_y \setminus N_\Gamma(x)| < c|S_y|\}) < \sqrt{\gamma}\}.$$

Then

$$\mathbb{P}_{x \sim \pi_1} \left[x \in D(\pi_2, \mathcal{S}) \right] > 1 - \sqrt{\gamma}.$$

Proof. For $x, y \in V(\Gamma)$, write $x \triangleright y$ as a shorthand for $|S_y \setminus N_\Gamma(x)| < c|S_y|$. As Γ is (c,t)-sparse and $|S_y| \geq t$, there are at most t vertices x such that $x \triangleright y$, for any $y \in V(\Gamma)$. Therefore, $\mathbb{P}_{x,y}[x \triangleright y] < \gamma$ with $x \sim \pi_1$ and $y \sim \pi_2$ sampled independently. Let $B = V(G) \setminus D(\pi_2, \mathcal{S}) = \{x \in V(\Gamma) \colon \pi_2(\{y \in V(\Gamma) \colon x \triangleright y\}) \geq \sqrt{\gamma}\}$. By Markov's inequality,

$$\mathbb{P}_{x \sim \pi_1}[x \in B] \leq \frac{\mathbb{P}_{x,y}[x \triangleright y]}{\sqrt{\gamma}} < \sqrt{\gamma}.$$

Thus,

$$\mathbb{P}_{x \sim \pi_1} \left[x \in D(\pi_2, \mathcal{S}) \right] > 1 - \sqrt{\gamma}.$$

13

$\square$

Peeling a tree. Let T be a (labelled) tree, and let $\mathcal{S}(T)$ denote the family of all subtrees of T. A mapping $\nu\colon \mathcal{S}(T) \to 2^{V(T)}$ is T-*peeling* if the following holds: (i) $\nu(\{v\}) = \{v\}$ for $v \in V(T)$, and (ii) for $|T'| \geq 2$ we have $\nu(T') = \{u, w\}$ where u and w are leaves in T' such that $u \in \nu(T' - w)$ and $w \in \nu(T' - u)$.

Lemma 4.3. *For every tree T there exists a T-peeling mapping.*

Proof. We prove the lemma by induction on the size of T. For $|T| \leq 2$, the claim trivially holds. Consider some $|T| \geq 3$, and pick a leaf $v \in T$. Set $T_v = T - v$, and let ν_v be a T_v-peeling mapping. For a subtree $T' \subseteq T$ which does not contain v we set $\nu(T') = \nu_v(T')$. For a subtree $T' \subseteq T$ which contains v, take $w \in \nu_v(T' - v)$ which is not adjacent to v and set $\nu(T) = \{v, w\}$. $\square$

We are now ready to prove Theorem 1.7.

Proof of Theorem 1.7. By sequentially removing vertices of degree less than Ct in G, we eventually end up with a graph with $n' > 1$ vertices, at least Ctn' edges, and minimum degree at least Ct. Therefore, without loss of generality we may assume that G has minimum degree at least Ct.

Consider a T-peeling mapping ν. We show that for each tree $T' \subseteq T$ there exists a probability measure $\lambda_{T'}$ over c-unique copies of T' in G which are induced in Γ, such that $\lambda_{T'}$ $(K_\ell, \varepsilon_\ell, \kappa_\ell)$-extends $\lambda_{T'-v}$ for each $T' \subseteq T$ with at least two vertices and $v \in \nu(T')$, where $\ell = |V(T')|$. As the parameters depend only on the size of a tree, we simply write $\lambda_{T'}$ extends $\lambda_{T'-v}$. Here $K_i > 1$ and $0 < \varepsilon_i, \kappa_i < 1$, for $2 \leq i \leq |V(T)|$, are defined as follows:

$$K_2 = 2, \varepsilon_2 = 1/3, \kappa_{|V(T)|} = 1/2$$
$$\varepsilon_i = c/(2K_{i-1}), K_i = 2K_{i-1}/(\varepsilon_{i-1}^2 c) \quad \text{for } i \geq 3$$
$$\kappa_i = \kappa_{i+1}/(2K_i) \quad \text{for } 2 \leq i < |V(T)|,$$

and C is chosen to be sufficiently large. We prove the existence of such measures by induction on the size of T', handling the base cases T' has one or two vertices first.

T' **is a single vertex.** Let $T' = \{v\}$. We define λ_v as

$$\lambda_v(x) = \frac{\deg_G(x)}{2e(G)}.$$

T' **is an edge.** Let T' be an edge $\{u, w\}$, and note that necessarily $\nu(T') = \{u, w\}$ (recall that T is a labelled tree, so we distinguish between u and w). Consider the following process: Sample $x_u \sim \lambda_u$, and then sample $x_w \in N_G(x_u)$ uniformly at random from $N_G(x_u)$. If $|N_G(x_u) \setminus N_\Gamma(x_w)| \geq c|N_G(x_u)|$ and $|N_G(x_w) \setminus N_\Gamma(x_u)| \geq c|N_G(x_w)|$, output a copy $\phi = \{u \mapsto x_u, w \mapsto x_w\}$ of T' and otherwise reject. When we output, ϕ is a c-unique copy of T', and as T' contains a single edge it is trivially induced in Γ. Define $\lambda_{T'}$ to be the corresponding output distribution. We will make use of the fact that first sampling $x_w \sim \lambda_w$ and then $x_u \in N_G(x_w)$ yields the same probability distribution. Indeed, the distribution can alternatively be described as sampling a uniformly random edge $\{x_u, x_w\}$ in G, then accept if $|N_G(x_u) \setminus N_\Gamma(x_w)| \geq c|N_G(x_u)|$ and $|N_G(x_w) \setminus N_\Gamma(x_u)| \geq c|N_G(x_w)|$.

For each $x \in V(\Gamma)$, as Γ is (c, t)-sparse and $|N_G(x)| \geq Ct$, we have $|N_G(x) \setminus N_\Gamma(y)| \geq c|N_G(x)|$ for all but at most t vertices $y \in N_G(x)$. As we can sample T' starting with either x_u or x_w, the above claim (with $x = x_u$ or $x = x_w$) implies that ϕ is c-unique with probability at least $1 - 2t/(Ct) > 1 - \kappa_2/2$. In other words, the probability p_{succ} of successfully outputting a copy of T' is at least $1 - \kappa_2/2 > 1/2$. We now verify that $\lambda_{T'}$ $(K_2, \varepsilon_2, \kappa_2)$-extends $\lambda_{T'-u} = \lambda_w$ and $\lambda_{T'-w} = \lambda_u$.

14

Property (4) states that for every (ordered) edge $e = (x_u, x_w) \in \mathrm{supp}(\lambda_{T'})$ we have

$$\lambda_{T'}(e) = \frac{1}{p_{\mathrm{succ}}} \cdot \frac{\lambda_u(x_u)}{\deg_G(x_u)} \leq \frac{2\lambda_u(x_u)}{\deg_G(x_u)},$$

which follows from the lower bound on the success probability and the definition of the process. The analogous statement holds for w in place of u.

Property (5) states that

$$\mathbb{P}_{x_u \sim \lambda_u} \left[\lambda_{T'}(u \mapsto x_u) \geq \lambda_u(x_u)/3 \right] \geq 1 - \kappa_2, \tag{7}$$

where $\lambda_{T'}(u \mapsto x_u) = \lambda_{T'}(\{(x_u, x_w) \in \mathrm{supp}(\lambda_{T'})\})$. Due to symmetry, this implies the analogous statement for λ_w. Let $B \subseteq V(G)$ denote the set of vertices such that for each $x \in B$ there are less than $\deg_G(x)/2$ vertices $y \in N_G(x)$ for which $|N_G(y) \setminus N_\Gamma(x)| \geq c|N_G(y)|$. For each $x \notin B$, there are at least $\deg_G(x)/2 - t > \deg_G(x)/3$ vertices $y \in N_G(x)$ such that (x, y) is c-unique. Therefore, for $x_u \notin B$ we have

$$\lambda_{T'}(u \mapsto x_u) \geq \sum_{\substack{y \in N_G(x_u) \\ (x,y)\ c\text{-unique}}} \lambda_u(x_u)/\deg_G(x_u) \geq \lambda_u(x_u)/3.$$

To prove (7), it suffices to show $\lambda_u(B) \leq \kappa_2$. Suppose otherwise, then

$$1 - p_{\mathrm{succ}} = \mathbb{P}_{x_u, x_w} \left[(x_u, x_w) \text{ not } c\text{-unique} \right]$$
$$\geq \mathbb{P}_{x_u \sim \lambda_u}(x_u \in B) \mathbb{P}_{x_u, x_w} \left[(x_u, x_w) \text{ not } c\text{-unique} \mid x_u \in B \right] > \kappa_2/2,$$

where the joint probability (x_u, x_w) is as given by the process. This is a contradiction with the previously established lower bound on p_{succ}.

T' **has at least three vertices.** Suppose now T' has $\ell \geq 3$ vertices and the statement holds for every proper subtree $T'' \subset T'$. Let $\nu(T') = \{u, w\}$, and let h_u and h_w be unique neighbors of u and w in T'. Note that h_u and h_w could be the same vertex. Set $R = T' - \nu(T)$, $R_u = T' - w$ and $R_w = T' - u$. By the inductive assumption and the definition of a T-peeling mapping, λ_{R_u} and λ_{R_w} extend λ_R.

Consider the following process:

(I) Sample $\rho \sim \lambda_R$. If $\lambda_{R_u}(\rho) < \varepsilon_{\ell-1}\lambda_R(\rho)$ or $\lambda_{R_w}(\rho) < \varepsilon_{\ell-1}\lambda_R(\rho)$, reject.

(II) Sample $\rho_u \sim \lambda_{R_u} \mid \rho$, that is, ρ_u is sampled according to λ_{R_u} conditioned on agreeing with ρ on R. Similarly, independently sample $\rho_w \sim \lambda_{R_w} \mid \rho$. Let $x_u = \rho_u(u)$ and $x_w = \rho_w(w)$, and define ϕ to be a mapping given by $\rho \cup \{u \mapsto x_u, w \mapsto x_w\}$. Reject if any of the following events happen: (i) $x_u = x_w$, (ii) ϕ is not c-unique, or (iii) x_u and x_w are adjacent in Γ.

(III) Output the copy ϕ of T'.

Note that if we did not reject in step (I) then $\lambda_{R_v}(\rho) > 0$ implies $\lambda_{R_v} \mid \rho$ is defined, for $v \in \{u, w\}$. Events (i)–(iii) ensure that ϕ is a c-unique copy of T' in G which is induced in Γ, thus we define $\lambda_{T'}$ to be the corresponding output distribution. It remains to show that the probability p_{succ} of outputting a copy of T' in step (3) is non-zero and $\lambda_{T'}$ extends $\lambda_{T'-u} = \lambda_{R_w}$ and $\lambda_{T'-w} = \lambda_{R_u}$.

For brevity, throughout the proof we denote with $\pi_{v,\rho} = \lambda_{R_v} \mid \rho$ the probability measure corresponding to x_v with respect to ρ, for $v \in \{u, w\}$.

Claim 4.4. *Suppose $\lambda_{R_u}(\rho) \geq \varepsilon_{\ell-1}\lambda_R(\rho)$. Then for every $x \in N_G(\rho(h_u))$ we have*

$$\pi_{u,\rho}(x) \leq \frac{L}{\deg_G(\rho(h_u))} \leq \frac{L}{Ct}.$$

where $L = K_{\ell-1}\varepsilon_{\ell-1}^{-1}$. The analogous statement holds for replacing u in each instance by w.

Proof. Consider some $x \in N_G(\rho(h_u))$ (otherwise $\pi_{u,\rho}(x) = 0$) and let $\rho_x = \rho \cup \{u \mapsto x\}$. Then, as λ_{R_u} extends λ_R, we have

$$\pi_{u,\rho}(x) = \frac{\lambda_{R_u}(\rho_x)}{\lambda_{R_u}(\rho)} \overset{(4)}{\leq} \frac{K_{\ell-1}\lambda_R(\rho)}{\deg_G(\rho(h_u))} \cdot \frac{1}{\lambda_{R_u}(\rho)},$$

thus the claim follows from $\lambda_{R_u}(\rho) \geq \varepsilon_{\ell-1}\lambda_R(\rho)$ and the minimum degree assumption. $\square$

Success probability p_{succ}. By (5), the probability of rejecting in step (I) is at most $2\kappa_{\ell-1}$. Let us condition on ρ not being rejected, that is, $\lambda_{R_u}(\rho) \geq \varepsilon_{\ell-1}\lambda_R(\rho)$ and $\lambda_{R_w}(\rho) \geq \varepsilon_{\ell-1}\lambda_R(\rho)$. By Claim 4.4 we have $x_u = x_w$ with probability at most L/C. Next, as ρ_u is a c-unique copy of R_u and Γ is (c,t)-sparse, for all but at most $(\ell-1)t$ choices of $x_w \in N_G(\rho(h_w))$ we have

$$|N_G(\rho_u(v)) \setminus \left(N_\Gamma^+(\rho_u(R_u - v)) \cup N_\Gamma(x_w)\right)| \geq c^{\ell-1}|N_G(\rho_u(v))| \quad \text{for every } v \in R_u.$$

We implicitly assume here $C > c^{-\ell}$, so that the right hand side of the previous inequality is at least t. Analogous statement holds for $x_u \in N_G(\rho(h_u))$. By Claim 4.4 and the fact that x_u and x_w are sampled independently, we conclude that ϕ is not c-unique with probability at most $2(\ell-1)L/C$. Finally, by Lemma 4.1 applied with $\pi_{u,\rho}$ (as π_1) and $\pi_{w,\rho}$ (as π_2), x_u and x_w are not adjacent in Γ with probability at least $(1-L/C)c$. All together, for sufficiently large C we have $p_{\text{succ}} > c/2$.

Next, we verify that $\lambda_{T'}$ $(K_\ell, \varepsilon_\ell, \kappa_\ell)$-extends λ_{R_u} and λ_{R_w}. The two cases are identical, thus we only do it for λ_{R_u}.

Property (4). Consider some $\phi \in \text{supp}(\lambda_{T'})$, and let $x_u = \phi(u)$ and $x_w = \phi(w)$. Then $\rho = \phi - \{u, w\}$ is not rejected in step (I), as otherwise ϕ would not be produced by the procedure. We have

$$\lambda_{T'}(\phi) = \frac{1}{p_{\text{succ}}} \cdot \lambda_R(\rho)\pi_{u,\rho}(x_u)\pi_{w,\rho}(x_w)$$
$$\leq \frac{2}{c} \cdot \lambda_R(\rho)\frac{\lambda_{R_u}(\phi - w)}{\lambda_{R_u}(\rho)}\pi_{w,\rho}(x_w) \leq \frac{2}{c\varepsilon_{\ell-1}}\lambda_{R_u}(\phi - w)\pi_{w,\rho}(x_w),$$

where the last inequality follows from $\lambda_{R_u}(\rho) \geq \varepsilon_{\ell-1}\lambda_R(\rho)$. Claim 4.4 applied on $\pi_{w,\rho}(x_w)$ gives (4).

Property (5). Consider a copy $\rho_u \in \text{supp}(\lambda_{R_u})$ of R_u. Set $x_u = \rho_u(u)$ and let $\rho = \phi_u - u$ be a corresponding copy of R. For $x \in \text{supp}(\pi_{w,\rho})$ define $S_x = N_G(x) \setminus N_\Gamma^+(\rho(R))$, and denote the family of all such subsets by $\mathcal{S}$. By the fact that $\rho \cup \{w \mapsto x\}$ is c-unique for every $x \in \text{supp}(\pi_{w,\rho})$, we have $|S_x| \geq t$. Suppose

$$\lambda_{R_u}(\rho) \geq \varepsilon_{\ell-1}\lambda_R(\rho) \quad \text{and} \quad x_u \in D(\pi_{w,\rho}) \cap D(\pi_{w,\rho}, \mathcal{S}), \tag{8}$$

where $D(\pi_{w,\rho})$ is as defined in Lemma 4.1, and $D(\pi_{w,\rho}, \mathcal{S})$ as defined in Lemma 4.2. Then, along similar lines as when estimating p_{succ}, we have

$$\mathbb{P}_{x_w \sim \pi_{w,\rho}}[\rho_u \cup \{w \mapsto x_w\} \in \text{supp}(\lambda_{T'})] \geq c - \frac{\ell t L}{Ct} > c/2, \tag{9}$$

for C sufficiently large. Briefly, c is the bound on the probability that x_w and x_u are not adjacent in Γ (coming from $x_u \in D(\pi_{w,\rho})$), and the remaining term takes into account the probability that x_u and x_w are distinct and $\phi_u \cup \{w \mapsto x_w\}$ is c-unique (recall that ρ_u is c-unique by the definition). Therefore,

$$\lambda_{T'}(\rho_u) \geq \lambda_R(\rho)\pi_{u,\rho}(x_u)\pi_{w,\rho}(\{x_w \colon \rho_u \cup \{w \mapsto x_w\} \in \text{supp}(\lambda_{T'})\})$$
$$\overset{(9)}{\geq} \lambda_R(\rho)\frac{\lambda_{R_u}(\rho_u)}{\lambda_{R_u}(\rho)} \cdot c/2 \overset{(6)}{\geq} \frac{c}{2K_{\ell-1}}\lambda_{R_u}(\rho_u).$$

In the last inequality we used the assumption that λ_{R_u} extends λ_R. To summarise, if ρ_u satisfies (8) then

$\lambda_{T'}(\rho_u) \geq \varepsilon_\ell \lambda_{R_u}(\rho_u)$ for $\varepsilon_\ell = c/(2K_{\ell-1})$. We estimate the probability that ρ_u does not satisfy (8):

$$\mathbb{P}_{\rho_u \sim \lambda_{R_u}} [\lambda_{R_u}(\rho) < \varepsilon_{\ell-1}\lambda_R(\rho) \vee x_u \notin D(\pi_{w,\rho}) \cap D(\pi_{w,\rho,}, \mathcal{S})] =$$
$$\mathbb{P}_{\rho_u \sim \lambda_{R_u}} [\lambda_{R_u}(\rho) < \varepsilon_{\ell-1}\lambda_R(\rho)] + \mathbb{P}_{\rho_u \sim \lambda_{R_u}} [x \notin D(\pi_{w,\rho}) \cap D(\pi_{w,\rho,}, \mathcal{S}) \mid \lambda_{R_u}(\rho) \geq \varepsilon_{\ell-1}\lambda_R(\rho)].$$

By (5) and (6), the first term is at most $\kappa_{\ell-1}K_{\ell-1}$. The second term is at most $L/C + \sqrt{L/C}$ by Lemma 4.1 and Lemma 4.2. For C sufficiently large and $\kappa_{\ell-1}$ sufficiently small with respect to $K_{\ell-1}$, we get a desired upper bound on $\rho \sim \lambda_{R_u}$ failing (5). This finishes the proof. $\square$

5 Applications of Theorem 1.2 for certain hereditary families

In this section, we prove Theorem 1.8 which states that $\mathrm{ex}(\Gamma, \mathcal{S}) = O_c(tn)$ for Γ a (c, t)-sparse graph and $\mathcal{S}$ the family of string graphs. The following lemma from [15] was proved using an iterative application of Lee's separator lemma for string graphs.

Lemma 5.1. *(Lemma 3, page 224 of [15]) For each $\varepsilon > 0$ there is $C = C(\varepsilon)$ such that every string graph which has average degree d has an induced subgraph on at most Cd vertices which has average degree at least $(1 - \varepsilon)d$.*

Lemma 5.1 shows that string graphs have dense induced subgraphs with average degree close to that of the original graph. Together with Corollary 1.3, this implies Theorem 1.8.

Proof of Theorem 1.8. Let Γ be a (c, t)-sparse graph on n vertices and $G \subseteq \Gamma$ be a subgraph which is a string graph with the maximum number of edges. Let d be the average degree of G. We may suppose suppose $d \geq At$ for some large constant A dependent on c as otherwise we are done. Applying Lemma 5.1 with $\varepsilon = 1/2$, there is an induced subgraph G' of G with average degree at least $d/2$ and with at most Cd vertices. As G' is a subgraph of Γ, G' is (c, t)-sparse. As G' is an induced subgraph of G, G' is a string graph. From Corollary 1.3 applied to G' as Γ (noting that G' is a string graph and hence does not contain any induced copy of the bipartite graph B which is the 1-subdivision of K_5), we get that G' has small edge density, contradicting that its edge density is at least $(d/2)/(Cd) = 1/(2C)$. $\square$

As an application of Theorem 1.8, we determine the extremal number within a constant factor when $\mathcal{P}$ is the property of being an incomparability graph of a partially ordered set, in which case for $\Gamma = G(n, p)$ with $0 < p < 1$ fixed, with high probability we have $\mathrm{ex}(\Gamma, \mathcal{P}) = \Theta(n \log n)$. Indeed, by [21, 29, 32], any incomparability graph is a string graph. Thus, $\mathrm{ex}(\Gamma, \mathcal{P}) = O(n \log n)$ by Theorem 1.8. On the other hand, we saw that Γ contains a subgraph G with $\Theta(n \log n)$ edges, which is a disjoint union of cliques of size $\Theta(\log n)$. Such a graph G is an incomparability graph of a partially ordered set. The same argument shows that if $\Gamma = P_q$ is the Paley graph for a square q, then $\mathrm{ex}(\Gamma, \mathcal{P}) = \Theta(q^{3/2})$.

In contrast, if $\mathcal{P}$ is the family of comparability graphs (complements of incomparability graphs) or perfect graphs, we have with high probability, for $\Gamma = G(n, p)$, $\mathrm{ex}(\Gamma, \mathcal{P})$ is asymptotically $pn^2/4$ by the theorem of Alon, Krivelevich and Samotij [2], as these families contain all bipartite graphs, but there are graphs of chromatic number 3 like the five-cycle which are not in the family.

6 Concluding remarks

Improved bound for Paley graphs. It is conjectured that the clique number of Paley graphs with a prime number q of vertices is much smaller than the upper bound $\sqrt{q}$. This bound has only been improved recently by a constant factor [9, 23]. An improvement on the bound in Theorem 1.6 when q is not a square

would imply the clique number of the Paley graph is $o(\sqrt{q})$. Furthermore, the Paley Graph Conjecture implies that such graphs should be (c, q^ε)-sparse for any fixed $c < 1/2$ and $\varepsilon > 0$ and q sufficiently large (see [34] for more). Together with our results, this would imply for n prime and Γ a Paley graph on n vertices that $\mathrm{ex}(\Gamma, \mathcal{P}_{K_{s,r}}) = n^{2-1/s+o(1)}$ for fixed r and s with r sufficiently large in s and $n \to \infty$.

Even cycles. Corollary 2.4 shows that for general graphs H and Γ, we can lower bound $\mathrm{ex}(\Gamma, \mathcal{P}_H)$ via the extremal numbers of the family of clique quotients of H. The corollary is interesting when H is bipartite. For some H, we understand $\mathcal{F}_H$ quite well. For example, if $H = C_{2\ell}$ is the even cycle on 2ℓ vertices, then $\mathcal{F}_H$ consists of all cycles of length at least ℓ and at most 2ℓ. It is known [1] that if a graph on n vertices has girth larger than 2ℓ (so the shortest cycle has more than 2ℓ vertices), then it has at most $\frac{1}{2}n^{1+1/\ell} + \frac{1}{2}n$ edges. It is conjectured that this upper bound is tight up to a constant factor, and this conjecture holds for $\ell \in \{2, 3, 5\}$, see [19] for more on this problem. We thus conjecture the following:

Conjecture 6.1. *For every $c > 0$ and integer ℓ there exists $C > 1$, such that if Γ is a (c, t)-sparse graph then $\mathrm{ex}(\Gamma, \mathcal{P}_{C_{2\ell}}) \leq Ct^{1-1/\ell}n^{1+1/\ell}$.*

For many bipartite H, like complete bipartite graphs $K_{s,r}$ with $2 \leq s \leq r$ and $r \geq 3$, every clique quotient of H apart from H is non-bipartite, and hence, for such graphs H, the construction used to prove Corollary 2.4 gives a natural bound of

$$\mathrm{ex}(\Gamma, \mathcal{P}_H) \geq \frac{1}{2}e(\Gamma)\mathrm{ex}(\chi(\overline{\Gamma}), H) / \binom{\chi(\overline{\Gamma})}{2}.$$

It seems challenging to prove a matching upper bound in general.

Induced subdivisions. Given a graph H, let us denote by $\mathcal{D}_H$ the property of containing a *subdivision* of H as an induced subgraph. In the context of induced Turán-type problem, this family of graphs has received considerable attention recently [7, 11, 20, 27]. In particular, a result of Hunter, Milojević, Sudakov, and Tomon [24] states that if G has $Ct^{O(|V(H)|)}n$ edges and no $K_{t,t}$ as a subgraph (not necessarily induced), then G contains an induced subdivision of H. Recasting the problem in our setup, it is natural to believe that the following is true.

Conjecture 6.2. *For every $c > 0$ and a graph H, there exists $C > 1$ such that if Γ is a (c, t)-sparse graph then $\mathrm{ex}(\Gamma, \mathcal{D}_H) \leq Ctn$.*

Multicolor induced Ramsey numbers. For a graph H and positive integer r, the induced Ramsey number $r_{\mathrm{ind}}(H, r)$ is the minimum N for which there is a graph G on N vertices such that in every r-edge-coloring of G there is a monochromatic induced copy of H. The existence of induced Ramsey numbers was proved in the early 1970s, and it has remained a challenging open problem to estimate these numbers. As shown in [16], like for Ramsey numbers, if H is non-bipartite, then

$$2^r \leq r_{\mathrm{ind}}(H, r) \leq r^{C(H)r}$$

for some constant $C = C(H)$.

One of the consequences of our results implies that for bipartite H, the multicolor induced Ramsey number has polynomial growth in the number of colors, similar to Ramsey numbers. Further, we can determine the induced Ramsey number for certain fixed bipartite graphs H up to a lower order factor.

Theorem 6.3. *Let d be a positive integer, and fix a d-bounded bipartite graph H which contains a $K_{d,s}$ with s sufficiently large. Then*

$$r_{ind}(H, r) = \tilde{\Theta}(r^d).$$

The upper bound follows from our results by considering a stronger density-type variant of induced Ramsey numbers studied by Dudek, Frankl, and Rödl [12]. Say that a graph G has the *induced (ε, H) density property* if for every subgraph G' of G with $e(G') \geq \varepsilon e(G)$, there is an induced copy of H in G with all edges in G'. If $\varepsilon = 1/r$ and G has the induced (ε, H) density property, then there is a monochromatic induced copy of H in the densest color in any r-edge coloring of G, thus showing that $r_{\mathrm{ind}}(H, r) \leq |G|$. Taking $G = G(n, 1/2)$ shows that Theorem 1.5 gives the desired upper bound in Theorem 6.3.

The lower bound in Theorem 6.3 follows from the lower bound on the ordinary Ramsey number for H. Using the symmetric hypergraph theorem (see Chapter 4.4 of [22]), we can color the edges of the complete graph on n vertices with at most $\frac{n^2}{\mathrm{ex}(n, H)} \log n$ colors such that each color class is H-free. This gives the desired lower bound on the Ramsey number (and hence the induced Ramsey number).

References

[1] N. Alon, S. Hoory, and N. Linial, The Moore bound for irregular graphs, *Graphs Combin.* **18** (2002), 53–57.

[2] N. Alon, M. Krivelevich, and W. Samotij, Largest subgraph from a hereditary property in a random graph, *Discrete Math.* **346** (2023), Paper No. 113480.

[3] N. Alon, M. Krivelevich, and B. Sudakov, Turán numbers of bipartite graphs and related Ramsey-type questions, *Combin. Probab. Comput.* **12** (2003), 477–494.

[4] N. Alon, L. Rónyai, and T. Szabó, Norm-graphs: variations and applications, *J. Combin. Theory Ser. B* **76** (1999), 280–290.

[5] M. Axenovich and J. Zimmermann, Induced Turán problem in bipartite graphs, arXiv:2401.11296.

[6] B. Bollobás, The chromatic number of random graphs, *Combinatorica* **8** (1988), 49–-55.

[7] R. Bourneuf, M. Bucić, L. Cook and J. Davies, On polynomial degree-boundedness, arXiv:2311.03341.

[8] B. Bukh, Extremal graphs without exponentially small bicliques, arXiv:2107.04167.

[9] D. Di Benedetto, J. Solymosi, E. P. White, On the directions determined by a Cartesian product in an affine Galois plane, *Combinatorica* **41** (2021), 755–763.

[10] D. Conlon, J. Fox, and B. Sudakov, An approximate version of Sidorenko's conjecture. *Geom. Funct. Anal.* **20** (2010), 1354–-1366.

[11] X. Du, A. Girão, Z. Hunter, R. McCarty, and A. Scott, Induced C4-free subgraphs with large average degree, arxiv:2307.08361.

[12] A. Dudek, P. Frankl, and V. Rödl, Some recent results on Ramsey-type numbers, *Discrete Appl. Math.* **161** (2013), 1197–1202.

[13] P. Erdős and A. M. Stone, On the structure of linear graphs, *Bull. Amer. Math. Soc.* **52** (1946), 1087–1091.

[14] P. Erdős and M. Simonovits, A limit theorem in graph theory, *Studia Sci. Math. Hungar.* **1** (1966), 51–57.

[15] J. Fox, J. Pach, and A. Suk, Quasiplanar graphs, string graphs, and the Erdős-Gallai problem, Graph drawing and network visualization, 219–231. Lecture Notes in Comput. Sci., 13764 Springer, Cham, 2023.

[16] J. Fox and B. Sudakov, Density theorems for bipartite graphs and related Ramsey-type results, *Combinatorica* **29** (2009), 153–196.

[17] J. Fox and B. Sudakov, Dependent random choice, *Random Structures Algorithms* **38** (2011), 68–99.

[18] Z. Füredi, On a Turán type problem of Erdős, *Combinatorica* **11** (1991), 75–79.

[19] Z. Füredi and M. Simonovits, The history of degenerate (bipartite) extremal graph problems, in Erdős centennial, Bolyai Soc. Math. Stud., vol. 25 (János Bolyai Math. Soc., Budapest, 2013), pp. 169–264.

[20] A. Girão and Z. Hunter, Induced subdivisions in $K_{s,s}$-free graphs with polynomial average degree, arXiv:2310.18452.

[21] M. C. Golumbic, D. Rotem, and J. Urrutia, Comparability graphs and intersection graphs, Discrete Math. 43 (1983), 37–46.

[22] R. L. Graham, B. L. Rothschild, and J. H. Spencer. *Ramsey theory*, volume 20. John Wiley & Sons, 1991.

[23] B. Hanson, and G. Petridis, Refined estimates concerning sumsets contained in the roots of unity, *Proc. Lond. Math. Soc. (3)* **122** (2021), 353–358.

[24] Z. Hunter, A. Milojević, B. Sudakov and I. Tomon, Kővari-Sós-Turán theorem for hereditary families, arXiv:2401.10853

[25] J. Kollár, L. Rónyai, and T. Szabó, Norm-graphs and bipartite Turán numbers, *Combinatorica* **16** (1996), 399–406.

[26] T. Kővári, V. T. Sós, and P. Turán. On a problem of K. Zarankiewicz, *Colloq. Math.* **3** (1954), 50–57.

[27] D. Kühn and D. Osthus, Induced subdivisions in $K_{s,s}$-free graphs of large average degree, *Combinatorica* **24** (2004), 287–304.

[28] P.-S. Loh, M. Tait, C. Timmons, and R. M. Zhou, Induced Turán numbers, *Combin. Probab. Comput.* **27** (2018), 274–288.

[29] L. Lovász, Perfect graphs, in: Selected Topics in Graph Theory, vol. 2, Academic Press, London, 1983, 55–87.

[30] A. Milojević, B. Sudakov, and I. Tomon, Incidence bounds via extremal graph theory, arXiv:2401.06670

[31] A. Milojević, B. Sudakov, and I. Tomon, Point-variety incidences, unit distances and Zarankiewicz's problem for algebraic graphs, arXiv:2403.08756

[32] J. Pach and G. Tóth, Comment on Fox News, *Geombinatorics* **15** (2006), 150–154.

[33] S. P. Vadhan, Pseudorandomness, *Foundations and Trends in Theoretical Computer Science* **7** (2012), No. 1–3, 1–336.

[34] C. H. Yip, On the clique number of Paley graphs and generalized Paley graphs, MSc thesis, University of British Columbia (2021).

[35] J. Zimmermann, Induced Turán problems, BSc thesis, Karlsruhe Institute of Technology (2024).