Combinatorics, Probability and Computing (2020), **29**, pp. 616–632
doi:10.1017/S0963548320000048

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Estimating parameters associated with monotone properties

Carlos Hoppen[1], Yoshiharu Kohayakawa[2], Richard Lang[3], Hanno Lefmann[4] and Henrique Stagni[2,*]

[1]Instituto de Matemática, UFRGS, Avenida Bento Gonçalves, 9500, 91501-970 Porto Alegre, RS, Brazil, [2]Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão 1010, 05508-090 São Paulo, Brazil, [3]Combinatorics and Optimization, University of Waterloo, N2L 3G1 Waterloo ON, Canada and [4]Fakultät für Informatik, Technische Universität Chemnitz, Straße der Nationen 62, 09111 Chemnitz, Germany

*Corresponding author. Email: stagni@gmail.com

(Received 24 July 2017; revised 2 July 2019; first published online 24 March 2020)

Abstract

There has been substantial interest in estimating the value of a graph parameter, *i.e.* of a real-valued function defined on the set of finite graphs, by querying a randomly sampled substructure whose size is independent of the size of the input. Graph parameters that may be successfully estimated in this way are said to be *testable* or *estimable*, and the *sample complexity* $q_z = q_z(\varepsilon)$ of an estimable parameter z is the size of a random sample of a graph G required to ensure that the value of $z(G)$ may be estimated within an error of ε with probability at least 2/3. In this paper, for any fixed monotone graph property $\mathcal{P} = \mathrm{Forb}(\mathcal{F})$, we study the sample complexity of estimating a bounded graph parameter $z_{\mathcal{F}}$ that, for an input graph G, counts the number of *spanning* subgraphs of G that satisfy $\mathcal{P}$. To improve upon previous upper bounds on the sample complexity, we show that the vertex set of any graph that satisfies a monotone property $\mathcal{P}$ may be partitioned equitably into a constant number of classes in such a way that the cluster graph induced by the partition is not far from satisfying a natural weighted graph generalization of $\mathcal{P}$. Properties for which this holds are said to be *recoverable*, and the study of recoverable properties may be of independent interest.

2010 MSC Codes: Primary 68W20; Secondary 68W25, 68Q25, 68R10, 05C35, 05C85

1. Introduction and main results

In the past two decades, a lot of effort has been put into finding constant-time randomized algorithms (conditional on sampling) to gauge whether a combinatorial structure satisfies some property, or to estimate the value of some numerical function associated with this structure. In this paper we focus on the graph case and, as usual, we consider algorithms that have the ability to query whether any desired pair of vertices in the input graph is adjacent or not. Let $\mathcal{G}$ be the set of finite simple graphs and let $\mathcal{G}(V)$ be the set of such graphs with vertex set V. We shall consider subsets $\mathcal{P}$ of $\mathcal{G}$ that are closed under isomorphism, which we call *graph properties*. To avoid technicalities, we restrict ourselves to graph properties $\mathcal{P}$ such that $\mathcal{P} \cap \mathcal{G}(V) \neq \emptyset$ whenever $V \neq \emptyset$. For instance, this includes all non-trivial *monotone* and *hereditary* graph properties, which are graph properties that are inherited by subgraphs and by induced subgraphs, respectively. Here, we will focus on monotone properties. The prototypical example of a monotone property is $\mathrm{Forb}(F)$, the class of all graphs that do not contain a copy of a fixed graph F. More generally, if $\mathcal{P}$ is a monotone

© Cambridge University Press 2020

https://doi.org/10.1017/S0963548320000048 Published online by Cambridge University Press

property and $\mathcal{F}$ contains all minimal graphs that are not in $\mathcal{P}$, then the graphs that lie in $\mathcal{P}$ are precisely those that do not contain a copy of an element of $\mathcal{F}$. This class of graphs will be denoted by $\mathcal{P} = \mathrm{Forb}(\mathcal{F})$. The elements of $\mathrm{Forb}(\mathcal{F})$ are said to be *$\mathcal{F}$-free*.

A graph property $\mathcal{P}$ is said to be *testable* if, for every $\varepsilon > 0$, there exist a positive integer $q_{\mathcal{P}} = q_{\mathcal{P}}(\varepsilon)$, called the *query complexity*, and a randomized algorithm $\mathcal{T}_{\mathcal{P}}$, called a *tester*, which may perform at most $q_{\mathcal{P}}$ queries in the input graph, satisfying the following property. For an n-vertex input graph Γ, the algorithm $\mathcal{T}_{\mathcal{P}}$ distinguishes with probability at least $2/3$ between the cases in which Γ satisfies $\mathcal{P}$ and in which Γ is ε-far from satisfying $\mathcal{P}$, that is, in which no graph obtained from Γ by the addition or removal of at most $\varepsilon n^2/2$ edges satisfies $\mathcal{P}$. This may be stated in terms of graph distances: given two graphs Γ and Γ' on the same vertex set $V(\Gamma) = V(\Gamma')$, we may define the *normalized edit distance* between Γ and Γ' by

$$\mathrm{d}_1(\Gamma, \Gamma') = \frac{2}{|V|^2}|E(\Gamma) \triangle E(\Gamma')|,$$

where $E(\Gamma)\triangle E(\Gamma')$ denotes the symmetric difference of their edge sets. If $\mathcal{P}$ is a graph property, we let the distance between a graph Γ and $\mathcal{P}$ be

$$\mathrm{d}_1(\Gamma, \mathcal{P}) = \min\{\mathrm{d}_1(\Gamma, \Gamma') \colon V(\Gamma') = V(\Gamma) \text{ and } \Gamma' \in \mathcal{P}\}.$$

For instance, if $\Gamma = K_n$ and $\mathcal{P} = \mathrm{Forb}(K_3)$, Turán's theorem ensures that $\binom{n}{2} - \lfloor n^2/4 \rfloor$ edges need to be removed to produce a graph that is K_3-free. In particular, $\mathrm{d}_1(K_n, \mathrm{Forb}(K_3)) \to 1/2$. A graph property $\mathcal{P}$ is testable if there is a tester with bounded query complexity that distinguishes with probability at least $2/3$ between the cases $\mathrm{d}_1(\Gamma, \mathcal{P}) = 0$ and $\mathrm{d}_1(\Gamma, \mathcal{P}) > \varepsilon$.

The systematic study of property testing was initiated by Goldreich, Goldwasser and Ron [25], and there is a very rich literature on this topic. For instance, regarding testers, Goldreich and Trevisan [26] showed that it is sufficient to consider simpler canonical testers, namely those that randomly choose a subset X of vertices in Γ and then verify whether the induced subgraph $\Gamma[X]$ satisfies some related property $\mathcal{P}'$. For example, if the property being tested has edge density $1/2$, then the algorithm chooses a random subset X of appropriate size and checks whether the edge density of $\Gamma[X]$ is within, say, $\varepsilon/2$ of $1/2$. Regarding testable properties, Alon and Shapira [6] proved that every monotone graph property is testable, and, more generally, that the same holds for hereditary graph properties [5]. For more information about property testing, we refer the reader to [24] and the references therein.

In a similar vein, a function $z\colon \mathcal{G} \to \mathbb{R}$ from the set $\mathcal{G}$ of all finite graphs into the real numbers is called a *graph parameter* if it is invariant under relabelling of vertices. A graph parameter $z\colon \mathcal{G} \to \mathbb{R}$ is *estimable* if, for every $\varepsilon > 0$ and every large enough graph Γ with probability at least $2/3$, the value of $z(\Gamma)$ can be approximated up to an additive error of ε by an algorithm that only has access to a subgraph of Γ induced by a set of vertices of size $q_z = q_z(\varepsilon)$, chosen uniformly at random. The query complexity of such an algorithm is $\binom{q_z}{2}$ and the size q_z is called its *sample complexity*. Estimable parameters have been considered in [18] and are often called *testable parameters*. When the parameter is the distance to a given property, estimating it is closely related to the notion of *tolerant testing*, which was introduced by Parnas, Ron and Rubinfeld [33], and is a generalization of standard property testing. Let $0 \leqslant \varepsilon_1 < \varepsilon_2 \leqslant 1$. An $(\varepsilon_1, \varepsilon_2)$-tolerant tester for a graph property $\mathcal{P}$ is an algorithm that receives a graph Γ as input and distinguishes between the cases $\mathrm{d}_1(\Gamma, \mathcal{P}) \leqslant \varepsilon_1$ and $\mathrm{d}_1(\Gamma, \mathcal{P}) \geqslant \varepsilon_2$ with probability at least $2/3$ and constant query complexity. In a seminal piece of work, Fischer and Newman [18] proved that every testable graph property $\mathcal{P}$ has a $(d - \varepsilon, d)$-tolerant tester, for every positive d and ε. The distance from a graph to $\mathcal{P}$ can then be estimated by successively running such tolerant testers. This strategy has been improved by Alon, Shapira and Sudakov [7] in the case of monotone properties $\mathcal{P}$, but, as we shall discuss below in more detail, it leads to an algorithm whose query complexity depends on the constants given by the Szemerédi Regularity Lemma. Going back to general parameters, we mention that Borgs, Chayes, Lovász, Sós and Vesztergombi [11, Theorem 6.1] gave a complete characterization

https://doi.org/10.1017/S0963548320000048 Published online by Cambridge University Press

of the graph parameters that are estimable, which, in particular, also implies that the distance from monotone graph properties is estimable. Their work uses the concept of graph limits and does not give explicit bounds on the query complexity required for this estimation.

We obtain results for the bounded graph parameter defined, for a graph family $\mathcal{F}$, as the number of $\mathcal{F}$-free spanning subgraphs of the input graph Γ. Recall that $G' = (V', E')$ is a *spanning subgraph* of a graph $G = (V, E)$ if $V' = V$ and $E' \subseteq E$.

Formally, given a graph $\Gamma \in \mathcal{G}$ and a family $\mathcal{F}$ of graphs, we denote the set of all $\mathcal{F}$-free spanning subgraphs of Γ by

$$\mathrm{Forb}(\Gamma, \mathcal{F}) = \{G \text{ is a spanning subgraph of } \Gamma : G \in \mathrm{Forb}(\mathcal{F})\},$$

and we consider the graph parameter

$$z_{\mathcal{F}}(\Gamma) = \frac{2}{|V(\Gamma)|^2} \log_2 |\mathrm{Forb}(\Gamma, \mathcal{F})|. \tag{1.1}$$

For example, if $\mathcal{F} = \{K_3\}$ and $\Gamma = K_n$, computing $z_{\mathcal{F}}$ requires estimating the number of K_3-free subgraphs of K_n up to a multiplicative error of $2^{o(n^2)}$:

$$z_{\mathcal{F}}(K_n) = \frac{2}{n^2} \log_2 |\mathrm{Forb}(\Gamma, \mathcal{F})| = \frac{2}{n^2} \log_2 2^{\binom{n}{2}/2 + o(n^2)} \to \frac{1}{2}.$$

This was done by Erdős, Kleitman and Rothschild [16] for $\mathcal{F} = \{K_k\}$; see also Erdős, Frankl and Rödl [15] for F-free subgraphs. Counting problems of this type have been studied by several people. Consider for instance the work of Prömel and Steger [34, 35], the logarithmic density in Bollobás [10], and some more recent results about the number of n-vertex graphs avoiding copies of some fixed forbidden graphs [8, 9]. Algorithmic aspects have been investigated by Duke, Lefmann and Rödl [14] and, quite recently, by Fox, Lovász and Zhao [20].

As we shall see, estimating graph parameters $z_{\mathcal{F}}(\Gamma)$ is related to estimating distances of graphs from the corresponding graph property $\mathcal{P} = \mathrm{Forb}(\mathcal{F})$. Improving upon the approach of Fischer and Newman [18], Alon, Shapira and Sudakov [7, Theorem 1.2] proved that the distance to every monotone graph property $\mathcal{P}$ is estimable using a natural algorithm, which simply computes the distance from the induced sampled graph to $\mathcal{P}$. However, one disadvantage of this approach is that the accuracy of the estimate relies heavily on stronger versions of Szemerédi's Regularity Lemma [3, 37]. Therefore the query complexity is at least of the order $\mathtt{TOWER}(\mathtt{poly}(1/\varepsilon))$, by which we mean a tower of twos of height that is polynomial in $1/\varepsilon$. Moreover, it follows from a result of Gowers [27] that any approach based on Szemerédi's Regularity Lemma cannot lead to a bound that is better than $\mathtt{TOWER}(\mathtt{poly}(1/\varepsilon))$.

In this paper we introduce the concept of *recoverable* graph properties. Roughly speaking, given $\varepsilon > 0$, an 'ε-recovering' partition $\mathcal{V}$ of a graph $G \in \mathcal{P}$ is a partition from which one can recover the fact that G is (in particular) ε-close to $\mathcal{P}$ by inspecting the set of densities between the classes of $\mathcal{V}$ only. In other words, for such a partition $\mathcal{V}$, the 'cluster graph' $G/\mathcal{V}$ (see Section 2 for the definition) is a witness to the fact that G is ε-close to $\mathcal{P}$. Given a function $f : (0, 1] \to \mathbb{R}$, we say that a graph property $\mathcal{P}$ is 'f-recoverable' if every graph $G \in \mathcal{P}$ has one such ε-recovering partition into at most $f(\varepsilon)$ classes, for every $\varepsilon > 0$. The precise definitions of an ε-recovering partition and an f-recoverable graph property are given in Definitions 3.4 and 3.5, respectively.

Theorem 1.1. *Let* $\mathrm{Forb}(\mathcal{F})$ *be an* f*-recoverable graph property, for some function* $f : (0, 1] \to \mathbb{R}$. *Then, for every* $\varepsilon > 0$ *there exists* n_0 *such that, for any graph* Γ *with* $|V(\Gamma)| \geqslant n_0$, *the graph parameter* $z_{\mathcal{F}}$ *defined in* (1.1) *can be estimated within an additive error of* ε *with sample complexity* $\mathtt{poly}(f(\varepsilon/6)/\varepsilon)$.

https://doi.org/10.1017/S0963548320000048 Published online by Cambridge University Press

Given a monotone property Forb($\mathcal{F}$), the parameter $z_{\mathcal{F}}$ is actually closely related to the parameter $d_{\mathcal{F}}\colon \Gamma \mapsto \mathrm{d}_1(\Gamma, \mathrm{Forb}(\mathcal{F}))$. In fact we shall prove the following simple consequence of Theorem 1.1.

Corollary 1.2. *Let* Forb($\mathcal{F}$) *be an f-recoverable graph property, for some function $f\colon (0,1] \to \mathbb{R}$. Then, for every $\varepsilon > 0$ there exists n_0 such that, for any graph Γ with $|V(\Gamma)| \geqslant n_0$, the graph parameter $d_{\mathcal{F}}\colon \Gamma \mapsto d_1(\Gamma, \mathrm{Forb}(\mathcal{F}))$ can be estimated within an additive error of ε with sample complexity* $\mathtt{poly}(f(\varepsilon/6)/\varepsilon)$.

Although one could apply strong versions of regularity to show that every monotone property Forb($\mathcal{F}$) is f-recoverable, this would provide an upper bound of at least $\mathtt{TOWER}(\mathtt{poly}(1/\varepsilon))$ for the function f. We find a connection between the notion of recoverability and the graph Removal Lemma, which can lead to better bounds on the function $f(\varepsilon)$. The Removal Lemma for triangles was proved by Ruzsa and Szemerédi in [36]. The more general version was first stated explicitly in the literature by Alon, Duke, Lefmann, Rödl and Yuster [2] and by Füredi [22]. The following version, which holds for arbitrary families of graphs, was first proved in [6].

Lemma 1.3 (Removal Lemma). *For every $\varepsilon > 0$ and every (possibly infinite) family $\mathcal{F}$ of graphs, there exist $M = M(\varepsilon, \mathcal{F})$, $\delta = \delta(\varepsilon, \mathcal{F}) > 0$ and $n_0 = n_0(\varepsilon, \mathcal{F})$ such that the following holds. If a graph G on $n \geqslant n_0$ vertices satisfies $d_1(G, \mathrm{Forb}(\mathcal{F})) \geqslant \varepsilon$, then there exists $F \in \mathcal{F}$ with $|V(F)| \leqslant M$ such that G has at least $\delta n^{|V(F)|}$ copies of F.* □

Conlon and Fox [13] showed that Lemma 1.3 holds with $1/\delta$, $n_0 \leqslant \mathtt{TOWER}(\mathtt{poly}(1/\varepsilon))$. Although this remains the best known bound for the general case, there are many families $\mathcal{F}$ for which Lemma 1.3 holds with a significantly better dependence on ε. For families $\mathcal{F} = \{F\}$ where F is an arbitrary graph, Fox [19] (see also [32]) showed that Lemma 1.3 holds with both $1/\delta$ and n_0 bounded by $\mathtt{TOWER}(O(\log(1/\varepsilon)))$; as a consequence, this same bound holds for every *finite* family $\mathcal{F}$. Moreover, if F is bipartite, then $1/\delta$ and n_0 are polynomial in $1/\varepsilon$ and, although it is not possible to get polynomial bounds when F is not bipartite, the best known lower bound for $1/\delta$ is only quasi-polynomial in $1/\varepsilon$ (see Alon [1]). Lemma 1.3 also holds with $1/\delta$, M, $n_0 \leqslant \mathtt{poly}(1/\varepsilon)$ for certain *infinite* families $\mathcal{F}$. For instance, results from [25] provide such polynomial bounds when Forb($\mathcal{F}$) is the property of 'being k-colourable' (for every fixed positive integer k) or the property of 'having a bisection of size at most ρn^2' (for every fixed $\rho > 0$) or many other properties that can be expressed as 'partition problems'.

We show that every monotone graph property Forb($\mathcal{F}$) is f-recoverable for some function f that is only exponential in the bounds given by the Removal Lemma for the family $\mathcal{F}$. For the proof, we use a weighted version of the Removal Lemma (see Lemma 3.6).

Theorem 1.4. *For every family $\mathcal{F}$ of graphs, the property* Forb($\mathcal{F}$) *is f-recoverable for $f(\varepsilon) = n_0 2^{\mathtt{poly}(M/\delta)}$, where δ, M and n_0 are as in Lemma 3.6 with input $\mathcal{F}$ and ε.*

The case of finite $\mathcal{F}$ is an instance where the bounds given by Lemma 3.6 relate polynomially to the bounds given by Lemma 1.3. In particular, Theorem 1.4, together with the above-mentioned bounds for Lemma 1.3 obtained by Fox [19] for finite families $\mathcal{F}$, implies that Forb($\mathcal{F}$) is f-recoverable with $f(\varepsilon) = \mathtt{TOWER}(\mathtt{poly}(\log(1/\varepsilon)))$.

The remainder of the paper is structured as follows. In Section 2 we introduce notation and describe some tools that we use in our arguments. In Section 3 we introduce the concept of recoverable graph properties and prove Theorem 1.4. Theorem 1.1 is a consequence of Theorem 4.1, which is the main result in Section 4. In Section 5 we prove Theorem 3.2, which is the technical tool for establishing Theorem 4.1. Corollary 1.2 is proved in Section 6. We finish the paper with some concluding remarks in Section 7.

https://doi.org/10.1017/S0963548320000048 Published online by Cambridge University Press

2. Notation and tools

A *weighted graph* R over a (finite) set V of vertices is a symmetric function from $V \times V$ to $[0, 1]$. A weighted graph R may be viewed as a complete graph (with loops) in which a weight $R(i, j)$ is given to each edge $(i, j) \in V(R) \times V(R)$, where $V(R)$ denotes the vertex set of R. The set of all weighted graphs with vertex set V is denoted by $\mathcal{G}^*(V)$ and we define $\mathcal{G}^*$ as the union of all $\mathcal{G}^*(V)$ for V finite. In particular, a *graph* G is a weighted graph such that $G(i, i) = 0$, for every $i \in V(G)$, and either $G(i, j) = 1$ or $G(i, j) = 0$ for every $(i, j) \in V(G) \times V(G)$, $i \neq j$. For a weighted graph $R \in \mathcal{G}^*(V)$ and for sets $A, B \subset V$, we denote $e_R(A, B) = \sum_{(i,j)\in A\times B} R(i, j)$ and $e(R) = e_R(V, V)/2$. Given a graph $G = (V, E)$ and vertex sets $U, W \subseteq V(G)$, let $E_G(U, W) = \{(u, w) \in E\colon u \in U, w \in W\}$ and $e_G(U, W) = |E_G(U, W)|$.

An *equipartition* $\mathcal{V} = \{V_i\}_{i=1}^k$ of a weighted graph R is a partition of its vertex set $V(R)$ such that $|V_i| \leqslant |V_j| + 1$ for all $(i, j) \in [k] \times [k]$. We sometimes abuse terminology and say that $\mathcal{V}$ is a partition of R.

Let $\mathcal{V} = \{V_1, \ldots, V_k\}$ be an equipartition into k classes of a graph $G = (V, E)$. The *cluster graph* of G by $\mathcal{V}$ is a weighted graph $G/\mathcal{V} \in \mathcal{G}^*([k])$ such that $G/\mathcal{V}\,(i, j) = e_G(V_i, V_j)/(|V_i|\,|V_j|)$ for all $(i, j) \in [k] \times [k]$. For a fixed integer $K > 0$, the set of all equipartitions of a vertex set V into at most K classes will be denoted by $\Pi_K(V)$. We also define the set $G/\Pi_K = \{G/\mathcal{V} : \mathcal{V} \in \Pi_K(V(G))\}$ of all cluster graphs of G whose vertex set has size at most K.

The *distance* between two weighted graphs $R, R' \in \mathcal{G}^*(V)$ on the same vertex set V is given by

$$\mathrm{d}_1(R, R') = \frac{1}{|V|^2} \sum_{(i,j)\in V\times V} |R(i, j) - R'(i, j)|.$$

Note that the above definition extends the one given in the previous section. For a property $\mathcal{H} \subseteq \mathcal{G}^*$ of weighted graphs, *i.e.* for a subset of the set of weighted graphs which is closed under isomorphisms, we define

$$\mathrm{d}_1(R, \mathcal{H}) = \min_{\substack{R'\in\mathcal{H}:\\ V(R')=V(R)}} \mathrm{d}_1(R, R').$$

Unless said otherwise, we will assume that $\mathcal{H}$ contains weighted graphs with vertex sets of all possible sizes.

Next, to set up the version of regularity (or Regularity Lemma) that we use in this work, we consider a second well-known distance between weighted graphs. Let $R_1, R_2 \in \mathcal{G}^*(V)$ be weighted graphs on the same vertex set. The *cut-distance* between R_1 and R_2 is defined as

$$\mathrm{d}_\square(R_1, R_2) = \frac{1}{|V|^2} \max_{S,T\subseteq V} |e_{R_1}(S, T) - e_{R_2}(S, T)|.$$

Let $\Gamma \in \mathcal{G}(V)$ and $\mathcal{V} = \{V_i\}_{i=1}^k$ be a partition of V. We define the weighted graph $\Gamma_\mathcal{V} \in \mathcal{G}^*(V)$ as the weighted graph such that $\Gamma_\mathcal{V}(u, v) = \Gamma/\mathcal{V}\,(i, j)$ if $u \in V_i$ and $v \in V_j$. Graph regularity lemmas ensure that, for any large graph Γ, there exists an equipartition $\mathcal{V}$ into a constant number of classes such that $\Gamma_\mathcal{V}$ is a faithful approximation of Γ. Here we use the regularity introduced by Frieze and Kannan [21]. Henceforth, we write $b = a \pm x$ for $a - x \leqslant b \leqslant a + x$.

Definition 2.1. A partition $\mathcal{V} = \{V_i\}_{i=1}^k$ of a graph Γ is *γ-FK-regular* if $\mathrm{d}_\square(\Gamma, \Gamma_\mathcal{V}) \leqslant \gamma$, or, equivalently, if for all $S, T \subseteq V(\Gamma)$ we have

$$e_\Gamma(S, T) = \left(\sum_{(i,j)\in[k]\times[k]} |S \cap V_i|\,|T \cap V_j|\ \Gamma/\mathcal{V}\,(i, j) \right) \pm \gamma |V(\Gamma)|^2.$$

The concept of *FK*-regularity is also known as *weak regularity*.

https://doi.org/10.1017/S0963548320000048 Published online by Cambridge University Press

Lemma 2.2 (Frieze–Kannan Regularity Lemma). *For every $\gamma > 0$ and every $k_0 > 0$, there exists $K = k_0 \cdot 2^{\mathtt{poly}(1/\gamma)}$ such that every graph Γ on $n \geqslant K$ vertices admits a γ-FK-regular equipartition into k classes, where $k_0 \leqslant k \leqslant K$.* □

We remark that Conlon and Fox [12] found graphs where the number of classes in any γ-FK-regular equipartition is at least $2^{1/(2^{60}\gamma^2)}$ (for an earlier result, see Lovász and Szegedy [31]).

3. Recoverable parameters

The main objective of this section is to introduce the concept of ε-recoverability and to state our main results in terms of it.

3.1 Estimation over cluster graphs

For a weighted graph $R \in \mathcal{G}^*(V)$ and a subset $Q \subseteq V$ of vertices, let $R[Q]$ denote the induced weighted subgraph of R with vertex set Q. Let us now define estimable parameters in the context of weighted graphs.

Definition 3.1. We say that a function $z\colon \mathcal{G}^* \to \mathbb{R}$ (also called a *weighted graph parameter*) is *estimable* with *sample complexity* $q\colon (0,1) \to \mathbb{N}$ if, for every $\varepsilon > 0$ and every weighted graph $\Gamma^* \in \mathcal{G}^*(V)$ with $|V| \geqslant q(\varepsilon)$, we have $z(\Gamma^*) = z(\Gamma^*[Q]) \pm \varepsilon$ with probability at least $2/3$, where Q is chosen uniformly from all subsets of V of size $q(\varepsilon)$.

The following result states that graph parameters that can be expressed as the solution of some optimization problem over G/Π_K can be estimated with a query complexity that is only polynomial in K and in the reciprocal of the error parameter ε. Here G/Π_K is the set of all cluster graphs of G of order at most K. We have been writing $\mathtt{poly}(x)$ for a polynomial in x. In what follows, $\mathtt{poly}(x, y)$ denotes a polynomial in the variables x and y.

Theorem 3.2. *Let $z\colon \mathcal{G} \to \mathbb{R}$ be a graph parameter and suppose that there are a weighted graph parameter $z^*\colon \mathcal{G}^* \to \mathbb{R}$ and constants $K > 0$ and $c > 0$ such that*

(1) $z(\Gamma) = \max_{R \in \Gamma/\Pi_K} z^*(R)$ *for every $\Gamma \in \mathcal{G}$, and*
(2) $|z^*(R) - z^*(R')| \leqslant c \cdot d_1(R, R')$ *for all weighted graphs $R, R' \in \mathcal{G}^*$ on the same vertex set.*

Then z is estimable with sample complexity $\varepsilon \mapsto \mathtt{poly}(K, c/\varepsilon)$.

The proof of Theorem 3.2 is rather technical and is therefore deferred to Section 5. Moreover, in Section 4 we show how to express the parameter $z_{\mathcal{F}}$ introduced in (1.1), in terms of the solution of a suitable optimization problem over the set Γ/Π_K of cluster graphs of Γ of order at most K.

3.2 Recovering partitions

We are interested in the property of graphs that are free of *copies* of members of a (possibly infinite) family $\mathcal{F}$ of graphs. To relate this property to a property of cluster graphs, we introduce some definitions. Let $\varphi\colon V(F) \to V(R)$ be a mapping from the set of vertices of a graph $F \in \mathcal{G}$ to the set of vertices of a weighted graph $R \in \mathcal{G}^*$. The *homomorphism weight* $\hom_\varphi(F, R)$ of φ is defined as

$$\hom_\varphi(F, R) = \prod_{(i,j) \in E(F)} R(\varphi(i), \varphi(j)).$$

https://doi.org/10.1017/S0963548320000048 Published online by Cambridge University Press

The *homomorphism density* $t(F,R)$ of $F \in \mathcal{G}$ in $R \in \mathcal{G}^*$ is defined as the average homomorphism weight of a mapping in $\Phi := \{\varphi\colon V(F) \to V(R)\}$, that is,

$$t(F,R) = \frac{1}{|\Phi|} \sum_{\varphi \in \Phi} \hom_\varphi(F,R).$$

Note that, if F and R are graphs, then $t(F,R)$ is approximately the density of copies of F in R (and converges to this quantity when the order of R tends to infinity). Since weighted graphs will represent cluster graphs associated with a partition of the vertex set of the input graph, it will be convenient to work with the following property of weighted graphs:

$$\mathrm{Forb}^*_{\mathrm{hom}}(\mathcal{F}) = \{R \in \mathcal{G}^*\colon t(F,R) = 0 \text{ for every } F \in \mathcal{F}\}.$$

Let $R, S \in \mathcal{G}^*(V)$ be weighted graphs on the same vertex set V. We say that S is a *spanning subgraph* of R, which will be denoted by $S \leqslant R$, if $S(i,j) \leqslant R(i,j)$ for every $(i,j) \in V \times V$. When there is no ambiguity, we will just say that S is a *subgraph* of R. We also define $\mathrm{Forb}^*_{\mathrm{hom}}(R,\mathcal{F}) = \{S \in \mathrm{Forb}^*_{\mathrm{hom}}(\mathcal{F})\colon S \leqslant R\}$.

The following result shows that having a cluster graph in $\mathrm{Forb}^*_{\mathrm{hom}}(\mathcal{F})$ witnesses membership in $\mathrm{Forb}(\mathcal{F})$.

Proposition 3.3. *Let $\mathcal{F}$ be a family of graphs and let $\mathcal{V}$ be an equipartition of a graph G. If $G/\mathcal{V} \in \mathrm{Forb}^*_{\mathrm{hom}}(\mathcal{F})$, then $G \in \mathrm{Forb}(\mathcal{F})$.*

Proof. Let $\mathcal{V} = \{V_i\}_{i=1}^k$ be an equipartition of G and let $R = G/\mathcal{V}$. Fix an arbitrary element $F \in \mathcal{F}$ and an arbitrary injective mapping $\varphi\colon V(F) \hookrightarrow V(G)$. Define the function $\psi\colon V(F) \to V(R)$ by $\psi(v) = i$ if $\varphi(v) \in V_i$. Now, if $t(F,R) = 0$, there must be some edge $(u,w) \in E(F)$ such that $R(\psi(u),\psi(w)) = 0$, thus $G(\varphi(u),\varphi(v)) = 0$ and hence $\hom_\varphi(F,G) = 0$. Since φ and F were taken arbitrarily, we must have $G \in \mathrm{Forb}(\mathcal{F})$. □

It is easy to see that the converse of Proposition 3.3 does not hold in general. Indeed, there exist graph families $\mathcal{F}$ and graphs $G \in \mathrm{Forb}(\mathcal{F})$ such that $G/\mathcal{V}$ is actually very *far* from being in $\mathrm{Forb}^*_{\mathrm{hom}}(\mathcal{F})$ for some equipartition $\mathcal{V}$ of G. As an example, let G be the n-vertex bipartite Turán graph $T_2(n)$ for the triangle K_3 with partition $V(G) = A \cup B$, and consider $\mathcal{V} = \{V_i\}_{i=1}^t$ with $V_i = A_i \cup B_i$, $i = 1,\dots,t$, where $\{A_i\}_{i=1}^t$ and $\{B_i\}_{i=1}^t$ are equipartitions of A and B respectively. Then $G/\mathcal{V}$ has weight $1/2$ on every edge, so that the distance of $G/\mathcal{V}$ to the family $\mathrm{Forb}^*_{\mathrm{hom}}(\{K_3\})$ tends to $1/4$ for t large by Turán's theorem. More generally, if $\mathcal{V}$ is a random equipartition of a triangle-free graph $G \in \mathrm{Forb}(\{K_3\})$ with large edge density, then with high probability the cluster graph $G/\mathcal{V}$ is still approximately $1/4$-far from being in $\mathrm{Forb}^*_{\mathrm{hom}}(\{K_3\})$.

On the other hand, we will prove that there exist equipartitions for graphs in $\mathrm{Forb}(\mathcal{F})$ with respect to which an approximate version of the converse of Proposition 3.3 does hold, that is, we will prove that every graph in $\mathrm{Forb}(\mathcal{F})$ is not too far from having an equipartition into a constant number of classes that witnesses membership in $\mathrm{Forb}(\mathcal{F})$. We say that such an equipartition is *recovering* with respect to $\mathrm{Forb}(\mathcal{F})$. Let us make this more precise.

Definition 3.4. Let $\mathcal{P} = \mathrm{Forb}(\mathcal{F})$ be a monotone graph property. An equipartition $\mathcal{V}$ of a graph $G \in \mathcal{P}$ is *ε-recovering* for $\mathcal{P}$ if $\mathrm{d}_1(G/\mathcal{V}, \mathrm{Forb}^*_{\mathrm{hom}}(\mathcal{F})) \leqslant \varepsilon$.

It is well known that the regularity method implies that, for any fixed graph F, the property $\mathcal{P} = \mathrm{Forb}(\{F\})$ admits a recovering partition (see *e.g.* [30] or [38]). In fact this holds [6] for general monotone properties $\mathcal{P}$. However, regularity methods yield recovering partitions that have a number of parts that is at least of tower type. Here we focus on the size of such partitions for monotone properties $\mathcal{P}$.

https://doi.org/10.1017/S0963548320000048 Published online by Cambridge University Press

Definition 3.5. Let $\mathcal{P}$ be a graph property. For a fixed function $f\colon (0,1] \to \mathbb{R}$, we say that the class $\mathcal{P}$ is *f-recoverable* if, for every $\varepsilon > 0$, there exists $n_0 = n_0(\varepsilon)$ such that the following holds. For every graph $G \in \mathcal{P}$ on $n \geqslant n_0$ vertices, there is an equipartition $\mathcal{V}$ of G into at most $f(\varepsilon)$ classes which is ε-recovering for $\mathcal{P}$.

As a simple example, one can verify that the graph property $\mathcal{P} = \mathrm{Forb}(\mathcal{F})$ of being r-colourable is f-recoverable for $f(\varepsilon) = r/\varepsilon$. Here and in what follows, for simplicity, we ignore divisibility conditions and drop floor and ceiling signs. Let G be a graph in $\mathcal{P}$, with colour classes $C_1, \ldots, C_r$. Let $k = r/\varepsilon$. Start by fixing parts $V_1, \ldots, V_t$, for some $t \leqslant k$, of size n/k each, with each V_i contained in some C_j ($j = j(i)$), and leaving out fewer than n/k vertices from each C_j, $1 \leqslant j \leqslant r$. The sets V_i, $1 \leqslant i \leqslant t$, cover a subset C_j' of C_j and $X_j = C_j \setminus C_j'$ is left over. We then complete the partition by taking arbitrary parts $U_1, \ldots, U_{k-t}$ of size n/k each, forming a partition of $\bigcup_{1 \leqslant j \leqslant r} X_j$. The cluster graph $G/\mathcal{V}$ can be made r-partite by giving weight zero to every edge incident to vertices corresponding to $U_1, \ldots, U_{k-t}$. Therefore $G/\mathcal{V}$ is at distance at most $r/k = \varepsilon$ from being r-partite. Thus, $\mathrm{d}_1(G/\mathcal{V}, \mathrm{Forb}^*_{\mathrm{hom}}(\mathcal{F})) \leqslant \varepsilon$ as required.

We finish this section by noting that the definition of f-recoverable properties has some similarity to the notion of *regular-reducible* properties $\mathcal{P}$ defined by Alon, Fischer, Newman and Shapira [4]. When dealing with monotone properties $\mathcal{P} = \mathrm{Forb}(\mathcal{F})$, the main difference is that the notion of being regular-reducible requires that every graph $G \in \mathcal{P}$ should have a *regular* partition such that $G/\mathcal{V}$ is close to some property $\mathcal{H}^*$ of weighted graphs, while the definition of f-recoverable properties requires only that every G has a partition $\mathcal{V}$ (regular or not) such that $G/\mathcal{V}$ is close to $\mathrm{Forb}^*_{\mathrm{hom}}(\mathcal{F})$. Another difference is that $\mathcal{H}^*$ must be such that having a (regular) cluster graph in $\mathcal{H}^*$ witnesses only *closeness* to $\mathcal{P}$, while having a (regular or not) cluster graph in $\mathrm{Forb}^*_{\mathrm{hom}}(\mathcal{F})$ witnesses membership in $\mathcal{P}$.

3.3 Monotone graph properties are recoverable

Szemerédi's Regularity Lemma [37] can be used to show that every monotone (and actually every hereditary) graph property is f-recoverable, for $f(\varepsilon) = \mathtt{TOWER}(\mathtt{poly}(1/\varepsilon))$. In the remainder of this section we prove that monotone properties $\mathcal{P} = \mathrm{Forb}(\mathcal{F})$ are recoverable using a weaker version of regularity along with the Removal Lemma, which leads to an improvement on the growth of f for families $\mathcal{F}$ where the Removal Lemma is known to hold with better bounds than the Regularity Lemma.

We first derive a version of the Removal Lemma stated in the Introduction (Lemma 1.3) that applies to weighted graphs and homomorphic copies.

Lemma 3.6. *For every $\varepsilon > 0$ and every (possibly infinite) family $\mathcal{F}$ of graphs, there exist $\delta = \delta(\varepsilon, \mathcal{F})$, $M = M(\varepsilon, \mathcal{F})$ and $n_0 = n_0(\varepsilon, \mathcal{F})$ such that the following holds. If a weighted graph R with $|V(R)| > n_0$ satisfies $d_1(R, \mathrm{Forb}^*_{\mathrm{hom}}(\mathcal{F})) \geqslant \varepsilon$, then there is a graph $F \in \mathcal{F}$ with $|V(F)| \leqslant M$ such that $t(F, R) \geqslant \delta$.*

To prove Lemma 3.6, we use the following auxiliary result, which follows from work of Erdős and Simonovits [17]. For completeness we include its proof.

Proposition 3.7. *Let $\hat{F}$ and F be graphs in $\mathcal{G}$ such that there is a surjective homomorphism $\zeta\colon V(F) \to V(\hat{F})$. Then, for every graph H such that $t(\hat{F}, H) \geqslant \hat{\delta}$, we must have $t(F, H) \geqslant \hat{\delta}^{\ell}$, where $\ell = |V(F)|^{|V(\hat{F})|}$.*

Proof. We will consider the particular case in which F is obtained by blowing up a single vertex v of $\hat{F}$ into r distinct vertices $v_1, \ldots, v_r$ with the same adjacency as v, and hence we assume that $\zeta(v_j) = v$ for every $j = 1, \ldots, r$ and $\zeta(u) = u$ for every $u \notin \{v_1, \ldots, v_r\}$.

https://doi.org/10.1017/S0963548320000048 Published online by Cambridge University Press

Let $n = |V(H)|$, $\hat{a} = |V(\hat{F})|$, $a = |V(F)| = \hat{a} + r - 1$ and let $\hat{F}_- = \hat{F} - v$ be the graph on $\hat{a} - 1$ vertices obtained from $\hat{F}$ by deleting v. Let $N = t(\hat{F}_-, H)n^{\hat{a}-1}$ be the number of homomorphisms from $\hat{F}_-$ to H and let $\varphi_1, \dots, \varphi_N \in V(H)^{V(\hat{F}_-)}$ be an enumeration of such homomorphisms. Note that $N \leqslant n^{\hat{a}-1}$.

For every $i \in [N]$ and $u \in V(H)$, we consider the function φ_i^u that extends φ_i by mapping v to u. Define $Z_i = \{u \in V(H)\colon \hom_{\varphi_i^u}(\hat{F}, H) = 1\}$ and $z_i = |Z_i|$. We claim there are z_i^r ways of extending φ_i to a homomorphism from F to H. Indeed, every possible extension $\varphi_i'\colon V(F) \to V(H)$ of φ_i, such that $\varphi_i'(v_j) \in Z_i$, for every $j = 1, \dots, r$, satisfies $\hom_{\varphi_i'}(F, H) = 1$. Therefore we have $t(F, H)n^a \geqslant \sum_{i=1}^N z_i^r$. Since $g(x) = x^r$ is a convex function for $x \geqslant 0$ and $r \geqslant 1$, we get

$$t(F, H)n^a \geqslant N\left(\frac{1}{N}\sum_{i=1}^N z_i\right)^r = N^{1-r}\left(\sum_{i=1}^N z_i\right)^r.$$

Now we use the fact that $\sum_{i=1}^N z_i = t(\hat{F}, H)n^{\hat{a}} \geqslant \hat{\delta} n^{\hat{a}}$ and our previous bound $N \leqslant n^{\hat{a}-1}$ to obtain that

$$t(F, H)n^a \geqslant \hat{\delta}^r n^{\hat{a}+r-1} = \hat{\delta}^r n^a.$$

Therefore $t(F, H) \geqslant \hat{\delta}^r \geqslant \hat{\delta}^a$. The general case may be easily obtained by induction on the number of vertices of $\hat{F}$. □

Proof of Lemma 3.6. Let $\hat{\mathcal{F}}$ denote the set of all *homomorphic images* of members of $\mathcal{F}$, that is, the set of all graphs $\hat{F} \in \mathcal{G}$ such that there is a surjective homomorphism $F \to \hat{F}$, for some $F \in \mathcal{F}$. Let $\hat{M}$, $\hat{\delta}$ and $\hat{n}_0$ be as in Lemma 1.3 with input $\hat{\mathcal{F}}$ and $\varepsilon/2$. We take

$$M = \max_{\substack{\hat{F} \in \hat{\mathcal{F}}\colon \\ |V(\hat{F})| \leqslant \hat{M}}} \min_{\substack{F \in \mathcal{F}\colon \\ F \to \hat{F}}} |V(F)|,$$

$n_0 = \hat{n}_0$, $\delta = (\varepsilon/2)^{M^2}\hat{\delta}^{\ell}$, where $\ell = M^{\hat{M}}$.

Let R be a weighted graph such that $|V(R)| > n_0$ and $\mathrm{d}_1(R, \mathrm{Forb}^*_{\mathrm{hom}}(\mathcal{F})) \geqslant \varepsilon$. We first define a graph $H \in \mathcal{G}(V(R))$ such that $H(i, j) = 1$ if and only if $R(i, j) \geqslant \varepsilon/2$. It follows from $\mathrm{d}_1(R, \mathrm{Forb}^*_{\mathrm{hom}}(\mathcal{F})) \geqslant \varepsilon$ that $\mathrm{d}_1(H, \mathrm{Forb}(\hat{\mathcal{F}})) \geqslant \varepsilon/2$. Indeed, let $H' \in \mathrm{Forb}(\hat{\mathcal{F}})$ be arbitrary. Define R' such that $R'(i, j) = R(i, j)$ if $H'(i, j) = 1$ and $R'(i, j) = 0$ otherwise. By construction, $R' \in \mathrm{Forb}^*_{\mathrm{hom}}(\mathcal{F})$, and we infer

$$\begin{aligned}
\varepsilon \leqslant \mathrm{d}_1(R, R') &= \frac{1}{|V(R)|^2}\sum_{\substack{i \in V(R),\\ j \in V(R)}} |R(i, j) - R'(i, j)| \\
&= \frac{1}{|V(R)|^2}\left(\sum_{\substack{(i,j)\colon \\ H(i,j)=1,\\ H'(i,j)=0}} |R(i, j) - R'(i, j)| + \sum_{\substack{(i,j)\colon \\ H(i,j)=0,\\ H'(i,j)=0}} |R(i, j) - R'(i, j)|\right) \\
&\leqslant \frac{1}{|V(H)|^2}\sum_{\substack{i \in V(H)\\ j \in V(H)}} |H(i, j) - H'(i, j)| + \frac{1}{|V(R)|^2}\sum_{\substack{i \in V(R)\\ j \in V(R)}} \frac{\varepsilon}{2} \\
&= \mathrm{d}_1(H, H') + \frac{\varepsilon}{2},
\end{aligned}$$

and thus $d_1(H, H') \geqslant \varepsilon/2$.

https://doi.org/10.1017/S0963548320000048 Published online by Cambridge University Press

By Lemma 1.3 there must be $\hat{F} \in \hat{\mathcal{F}}$, with $|V(\hat{F})| \leqslant \hat{M}$, such that $t(\hat{F}, H) \geqslant \hat{\delta}$. By definition of M, there must be $F \in \mathcal{F}$ such that $|V(F)| \leqslant M$ and there is a surjective homomorphism $F \to \hat{F}$. It follows from Proposition 3.7 that $t(F, H) \geqslant \hat{\delta}^{\ell}$. Since

$$\hom_{\varphi}(F, R) \geqslant (\varepsilon/2)^{|E(F)|} \hom_{\varphi}(F, H) \geqslant (\varepsilon/2)^{M^2} \hom_{\varphi}(F, H)$$

for each $\varphi\colon V(F) \to V(R)$, we must have

$$t(F, R) = \frac{\sum_{\varphi} \hom_{\varphi}(F, R)}{|V(R)|^{|V(F)|}} \geqslant \left(\frac{\varepsilon}{2}\right)^{M^2} \cdot \frac{\sum_{\varphi} \hom_{\varphi}(F, H)}{|V(H)|^{|V(F)|}} \geqslant \left(\frac{\varepsilon}{2}\right)^{M^2} \cdot \hat{\delta}^{\ell} = \delta. \qquad \square$$

We will use the next result, which states that a graph has homomorphism densities close to those of the cluster graphs with respect to FK-regular partitions.

Lemma 3.8 ([11, Theorem 2.7(a)]). *Let $\mathcal{V}$ be a γ-FK-regular equipartition of a graph $G \in \mathcal{G}$. Then, for any graph $F \in \mathcal{G}$, we have $t(F, G) = t(F, G_{\mathcal{V}}) \pm 4e(F)\gamma = t(F, G/\mathcal{V}) \pm 4e(F)\gamma$.* $\square$

We are now ready to prove Theorem 1.4, which establishes that every monotone graph property is f-recoverable.

Proof of Theorem 1.4. Let δ, M and n_0 be as in Lemma 3.6 with input $\mathcal{F}$ and ε and let $\gamma = \delta/(3M)^2$. By Lemma 2.2, it suffices to show that any γ-FK-regular equipartition $\mathcal{V} = \{V_i\}_{i=1}^{k}$ of a graph $G \in \mathrm{Forb}(\mathcal{F})$ into $k \geqslant n_0$ classes is ε-recovering.

Let $R = G/\mathcal{V}$ and suppose for contradiction that $\mathrm{d}_1(R, \mathrm{Forb}^*_{\mathrm{hom}}(\mathcal{F})) \geqslant \varepsilon$. Then, by Lemma 3.6, we have $t(F, R) \geqslant \delta$ for some graph $F \in \mathcal{F}$ such that $|F| \leqslant M$. By Lemma 3.8, we have $t(F, G) \geqslant \delta - 2\gamma M^2 > 0$, a contradiction to $G \in \mathrm{Forb}(\mathcal{F})$. $\square$

4. Estimation of $|\mathrm{Forb}(\Gamma, \mathcal{F})|$

The objective of this section is to prove Theorem 1.1. To do this, we shall approximate the parameter $z_{\mathcal{F}}$ by the solution of an optimization problem as in Theorem 3.2. Recall that

$$\mathrm{Forb}^*_{\mathrm{hom}}(R, \mathcal{F}) = \{S \leqslant R\colon t(F, S) = 0 \text{ for every } F \in \mathcal{F}\},$$

and set

$$\overline{\mathrm{ex}}^*(R, \mathcal{F}) = \frac{2}{|V(R)|^2} \max_{S \in \mathrm{Forb}^*_{\mathrm{hom}}(R, \mathcal{F})} e(S), \tag{4.1}$$

which measures the largest edge density of a subgraph of R not containing a copy of any $F \in \mathcal{F}$.

We shall derive Theorem 1.1 from the following auxiliary result.

Theorem 4.1. *Let $\mathcal{F}$ be a family of graphs such that $\mathrm{Forb}(\mathcal{F})$ is f-recoverable for some function $f\colon (0, 1] \to \mathbb{R}$. Then, for any $\varepsilon > 0$, there exist $K = f(\mathtt{poly}(\varepsilon))$ and $n_0 = \mathtt{poly}(K, 1/\varepsilon)$ such that, for any graph Γ of order $n \geqslant n_0$, we have*

$$\frac{2}{n^2} \log_2 |\mathrm{Forb}(\Gamma, \mathcal{F})| = \max_{R \in \Gamma/\Pi_K} \overline{\mathrm{ex}}^*(R, \mathcal{F}) \pm \varepsilon. \tag{4.2}$$

We define the following subsets of edges of a weighted graph R:

$$E_0(R) = \{(i, j) \in V(R) \times V(R)\colon R(i, j) = 0\},$$
$$E_1(R) = \{(i, j) \in V(R) \times V(R)\colon R(i, j) > 0\}.$$

https://doi.org/10.1017/S0963548320000048 Published online by Cambridge University Press

We will also make use of the *binary entropy function*, defined by

$$H(x) = -x\log_2 x - (1-x)\log_2(1-x) \quad \text{for } 0 < x < 1.$$

Note that $H(x) \leqslant -2x\log_2 x$ for $x \leqslant 1/8$. This function has the property that (see *e.g.* [29, Corollary 22.2]) the following inequality holds for $\varepsilon = k/n \leqslant 1/2$:

$$\sum_{i=0}^{k} \binom{n}{i} \leqslant 2^{H(\varepsilon)n}. \tag{4.3}$$

Proof of Theorem 4.1. Let $\mathcal{F}$ be a family of graphs such that $\mathrm{Forb}(\mathcal{F})$ is f-recoverable, and fix $\varepsilon > 0$. Without loss of generality, $\varepsilon < 1$. Let $\varepsilon' = (\varepsilon/18)^2$. Using that $\log_2(1/x^2) \leqslant 1/x$ for $0 < x \leqslant 1/4$, and $H(y) \leqslant -2y\log_2 y$ for $0 < y \leqslant 1/8$, we infer

$$H(\varepsilon') + \varepsilon' \leqslant -2(\varepsilon/18)^2\log_2(\varepsilon/18)^2 + (\varepsilon/18)^2 \leqslant 2(\varepsilon/18) + (\varepsilon/18)^2 \leqslant \varepsilon/6. \tag{4.4}$$

We set $K = f(\varepsilon'^2)$ and $n_0 \geqslant 2K^2/\varepsilon$ large enough so that $\log_2 n_0/n_0 < \varepsilon/6$.

Let Γ be an n-vertex graph with $n \geqslant n_0$. We first show that

$$\frac{2}{n^2}\log_2 |\mathrm{Forb}(\Gamma, \mathcal{F})| \geqslant \max_{R \in \Gamma/\Pi_K} \overline{\mathrm{ex}}^*(R, \mathcal{F}) - \varepsilon.$$

Let $R = \Gamma/\mathcal{V}$ be an arbitrary cluster graph in Γ/Π_K with equipartition $\mathcal{V} = \{V_i\}_{i=1}^k$ for some $k \leqslant K$. Choose a weighted graph $S \in \mathrm{Forb}^*_{\mathrm{hom}}(R, \mathcal{F})$ such that $e(S) = (k^2/2)\,\overline{\mathrm{ex}}^*(R, \mathcal{F})$. Further, let $G \leqslant \Gamma$ be the subgraph of Γ such that $G(r,s) = 0$ if there is a pair $(i,j) \in E_0(S)$ such that $r \in V_i$ and $s \in V_j$ and $G(r,s) = \Gamma(r,s)$ otherwise. Thus we obtain G by deleting all edges from Γ between V_i and V_j if $(i,j) \in E_0(S)$.

Since $e(S)$ maximizes $\overline{\mathrm{ex}}^*(R, \mathcal{F})$, it follows that $G/\mathcal{V} = S$, which implies, by Proposition 3.3, that $G \in \mathrm{Forb}(\Gamma, \mathcal{F})$. Since every subgraph of G also lies in $\mathrm{Forb}(\Gamma, \mathcal{F})$, we obtain

$$\begin{aligned}
\log_2 |\mathrm{Forb}(\Gamma, \mathcal{F})| &\geqslant |e(G)| \\
&= \frac{1}{2}\sum_{(i,j)\in[k]\times[k]} S(i,j)|V_i||V_j| \\
&\geqslant \frac{(n-k)^2}{k^2} e(S) \\
&\geqslant \overline{\mathrm{ex}}^*(R, \mathcal{F})\frac{n^2}{2} - kn \\
&\geqslant (\,\overline{\mathrm{ex}}^*(R, \mathcal{F}) - \varepsilon)\frac{n^2}{2}.
\end{aligned}$$

Note that we used the facts that $e(S) \leqslant k^2/2$ and $n \geqslant n_0 \geqslant 2K^2/\varepsilon \geqslant 2k/\varepsilon$, as well as $|V_i| \geqslant n/k - 1$ for all i.

Now let us prove that

$$\frac{2}{n^2}\log_2 |\mathrm{Forb}(\Gamma, \mathcal{F})| \leqslant \max_{R \in \Gamma/\Pi_K} \overline{\mathrm{ex}}^*(R, \mathcal{F}) + \varepsilon.$$

We first define $\mathcal{U} = \bigcup_{G \in \mathrm{Forb}(\Gamma,\mathcal{F})} G/\Pi_K$ to be the set of all possible cluster graphs of order at most K of graphs in $\mathrm{Forb}(\Gamma, \mathcal{F})$. Since $\mathrm{Forb}(\mathcal{F})$ is f-recoverable, we can define a function

$$\begin{aligned}
\eta\colon\ \mathrm{Forb}(\Gamma, \mathcal{F}) &\to \Pi_K \times \mathcal{U}, \\
G &\mapsto (\mathcal{V}, T),
\end{aligned}$$

https://doi.org/10.1017/S0963548320000048 Published online by Cambridge University Press

where $\mathcal{V}$ is an (ε'^2)-recovering partition of G into $k \leqslant K$ classes and $T = G/\mathcal{V}$. Clearly

$$|\operatorname{Forb}(\Gamma, \mathcal{F})| \leqslant |\Pi_K \times \mathcal{U}| \cdot \max_{(\mathcal{V},T)} |\eta^{-1}(\mathcal{V}, T)|. \tag{4.5}$$

Since each mapping from $V(\Gamma) \to [K]$ gives a partition of $V(\Gamma)$ into at most K classes, we have $|\Pi_K| \leqslant K^n \leqslant n^n$. Moreover, given an arbitrary graph $G \in \mathcal{G}(V)$ and any partition $\mathcal{V}$ of V, an edge $G/\mathcal{V}$ (i, j) may assume n^2 different values. Hence, we have $|\mathcal{U}| \leqslant n^{2K^2} \leqslant n^n$.

Finally we make the following claim, whose proof is deferred for a moment:

$$\log_2\Big(\max_{(\mathcal{V},T)} |\eta^{-1}(\mathcal{V}, T)|\Big) \leqslant \Big(\max_{R \in \Gamma/\Pi_K} \overline{\mathrm{ex}}^*(R, \mathcal{F}) + \frac{\varepsilon}{3}\Big)\frac{n^2}{2}. \tag{4.6}$$

Taking the logarithm of (4.5) and using (4.6), we get

$$\begin{aligned}\log_2 |\operatorname{Forb}(\Gamma, \mathcal{F})| &\leqslant \log_2 n^n + \log_2 n^n + \Big(\max_{R \in \Gamma/\Pi_K} \overline{\mathrm{ex}}^*(R, \mathcal{F}) + \frac{\varepsilon}{3}\Big)\frac{n^2}{2} \\ &\leqslant \Big(\max_{R \in \Gamma/\Pi_K} \overline{\mathrm{ex}}^*(R, \mathcal{F}) + \varepsilon\Big)\frac{n^2}{2} \quad (\text{as } \log_2 n/n \leqslant \varepsilon/6),\end{aligned}$$

as desired.

It remains to prove (4.6). To this end, fix $(\mathcal{V}, T)$ in the image of η and let $R = \Gamma/\mathcal{V}$. Choose $S' \in \operatorname{Forb}^*(R, \mathcal{F})$ such that $\mathrm{d}_1(T, S') \leqslant \varepsilon'^2$. This is possible because $\mathcal{V}$ is an (ε'^2)-recovering partition. Set $E_1 = E_1(S')$ and partition $E_0(S')$ into $E_0^+ := \{(i, j) \in E_0(S') \colon T(i, j) > \varepsilon'\}$ and $E_0^- = E_0(S') \setminus E_0^+$. Since there are

$$b(i, j) := \binom{|V_i|\,|V_j|R(i, j)}{|V_i|\,|V_j|T(i, j)}$$

ways to choose $|V_i|\,|V_j|T(i, j)$ edges out of the $|V_i|\,|V_j|R(i, j)$ edges between V_i and V_j in Γ, we obtain

$$|\eta^{-1}(\mathcal{V}, T)| \leqslant \prod_{1 \leqslant i < j \leqslant k} b(i, j) \leqslant \prod_{(i,j) \in E_1} \sqrt{b(i, j)} \prod_{(i,j) \in E_0^+} \sqrt{b(i, j)} \prod_{(i,j) \in E_0^-} \sqrt{b(i, j)}. \tag{4.7}$$

We now estimate the products in (4.7).

We can bound each of the factors $b(i, j)$ of E_1 by $2^{R(i,j)|V_i|\,|V_j|}$. Since $\mathrm{d}_1(T, S') \leqslant \varepsilon'^2$, we have $|E_0^+| \leqslant \varepsilon' k^2$, as otherwise

$$\mathrm{d}_1(T, S') \geqslant \sum_{(i,j) \in E_0^+} |T(i, j) - S'(i, j)| > |E_0^+|\varepsilon' \geqslant \varepsilon'^2 k^2,$$

which is a contradiction. Clearly, we have $|E_0^-| \leqslant k^2$. This allows us to upper-bound each of the factors of E_0^+ trivially by $2^{|V_i|\,|V_j|}$, and each of the factors of E_0^- by $2^{H(\varepsilon')|V_i|\,|V_j|}$ using (4.3).

Now let $S \in \operatorname{Forb}^*(R, \mathcal{F})$ be such that

$$S(i, j) = \begin{cases} 0 & \text{if } (i, j) \in E_0(S'), \\ R(i, j) & \text{otherwise.} \end{cases}$$

https://doi.org/10.1017/S0963548320000048 Published online by Cambridge University Press

Taking the logarithm of (4.7) and using $|V_i|\,|V_j| \leqslant (n+k)^2/k^2$, we get

$$\begin{aligned}\log_2 |\eta^{-1}(\mathcal{V},T)| &\leqslant \frac{1}{2}\sum_{(i,j)\in E_1} R(i,j)|V_i|\,|V_j| + \frac{1}{2}\sum_{(i,j)\in E_0^-} H(\varepsilon')|V_i|\,|V_j| + \frac{1}{2}\sum_{(i,j)\in E_0^+} |V_i|\,|V_j| \\ &\leqslant \left(\sum_{(i,j)\in E_1}\frac{R(i,j)}{k^2} + \sum_{(i,j)\in E_0^-}\frac{H(\varepsilon')}{k^2} + \sum_{(i,j)\in E_0^+}\frac{1}{k^2}\right)\frac{(n+k)^2}{2} \\ &\leqslant \left(\frac{1}{k^2}\sum_{(i,j)\in E_1} S(i,j) + H(\varepsilon') + \varepsilon'\right)\frac{(n+k)^2}{2} \\ &\leqslant (\,\overline{\mathrm{ex}}^*(R,\mathcal{F}) + H(\varepsilon') + \varepsilon')\frac{(n+k)^2}{2}.\end{aligned}$$

Now, by using the fact that $S \in \mathrm{Forb}^*(R,\mathcal{F})$ and that $H(\varepsilon')+\varepsilon' \leqslant \varepsilon/6$ (see (4.4)), we infer

$$\log_2 |\eta^{-1}(\mathcal{V},T)| \leqslant \left(\overline{\mathrm{ex}}^*(R,\mathcal{F}) + \frac{\varepsilon}{6}\right)\frac{(n+k)^2}{2} \leqslant \left(\overline{\mathrm{ex}}^*(R,\mathcal{F}) + \frac{\varepsilon}{3}\right)\frac{n^2}{2},$$

which implies (4.6). □

Proof of Theorem 1.1. Let $\mathcal{F}$ be a family of graphs such that $\mathrm{Forb}(\mathcal{F})$ is f-recoverable. Set $K = f(\mathtt{poly}(\varepsilon))$ and $n_0 = \mathtt{poly}(K, 1/\varepsilon)$ given by Theorem 4.1 applied to $\varepsilon/3$. Theorem 4.1 ensures that, whenever Γ is a graph on $n \geqslant n_0$ vertices, we have

$$\left|\frac{2}{n^2}\log_2 |\,\mathrm{Forb}(\Gamma,\mathcal{F})| - \max_{R\in \Gamma/\Pi_K} \overline{\mathrm{ex}}^*(R,\mathcal{F})\right| \leqslant \frac{\varepsilon}{3}. \tag{4.8}$$

Let $\hat{z}\colon \mathcal{G}\to\mathbb{R}$ be the graph parameter defined by $\hat{z}(\Gamma) = \max_{R\in\Gamma/\Pi_K} z^*(R)$, where $z^*(R) = \overline{\mathrm{ex}}^*(R,\mathcal{F})$. We claim that, given R and R' in $\mathcal{G}^*(V)$, we have $|z^*(R)-z^*(R')| \leqslant \mathrm{d}_1(R,R')$. Indeed, assume without loss of generality that $z^*(R) \geqslant z^*(R')$, and fix a subgraph $S \leqslant R$ such that $S \in \mathrm{Forb}^*_{\mathrm{hom}}(R,\mathcal{F})$ and $z^*(R) = 2e(S)/|V|^2$. If $S \in \mathrm{Forb}^*_{\mathrm{hom}}(R',\mathcal{F})$ we are done, so assume that this is not the case. Let S' be a subgraph of S and R' maximizing $e(S')$, *i.e.* $S'(i,j) = \min\{S(i,j), R'(i,j)\}$. Clearly

$$e(S') \geqslant e(S) - \frac{1}{2}\sum_{(i,j)\in V\times V} |R(i,j)-R'(i,j)| = e(S) - \frac{|V|^2}{2}\,\mathrm{d}_1(R,R'),$$

so that $0 \leqslant z^*(R) - z^*(R') \leqslant 2(e(S)-e(S'))/|V|^2 \leqslant \mathrm{d}_1(R,R')$.

This allows us to apply Theorem 3.2 to conclude that $\hat{z}$ is estimable with sample complexity $q(\varepsilon) = \mathtt{poly}(K, 1/\varepsilon)$. Let Q be chosen uniformly from all subsets of V of size

$$q' = \max\{q(\varepsilon/3), n_0\}$$

and set $\overline{\Gamma} = \Gamma[Q]$. It follows that, with probability at least $2/3$, we have $|\hat{z}(\overline{\Gamma}) - \hat{z}(\Gamma)| \leqslant \varepsilon/3$. By (4.8) we have $|2n^{-2}\log_2|\,\mathrm{Forb}(\Gamma,\mathcal{F})| - \hat{z}(\Gamma)| \leqslant \varepsilon/3$. On the other hand, we can also apply (4.8) to $\overline{\Gamma}$ to obtain $|\hat{z}(\overline{\Gamma}) - 2q'^{-2}\log_2|\,\mathrm{Forb}(\overline{\Gamma},\mathcal{F})|\,| \leqslant \varepsilon/3$. By adding the last three inequalities, we get that

$$\left|\frac{2}{n^2}\log_2|\,\mathrm{Forb}(\Gamma,\mathcal{F})| - \frac{2}{q'^2}\log_2|\,\mathrm{Forb}(\overline{\Gamma},\mathcal{F})|\right| \leqslant \varepsilon,$$

as required. □

https://doi.org/10.1017/S0963548320000048 Published online by Cambridge University Press

5. Proof of Theorem 3.2

Here we will prove Theorem 3.2. Its proof is based on the following lemma, which asserts that the set of cluster graphs of a graph Γ is very 'similar' to the set of cluster graphs of 'large enough' samples of Γ.

Lemma 5.1. *Given $K > 0$, $\varepsilon > 0$ there exists $q = \mathtt{poly}(K, 1/\varepsilon)$ such that the following holds. Consider a graph Γ whose vertex set V has cardinality $n \geqslant q$ and a random subgraph $\overline{\Gamma} = \Gamma[\overline{V}]$, where $\overline{V}$ is chosen uniformly from all subsets of V of size q. Then, with probability at least $2/3$, we have*

(1) *for each $\mathcal{V} \in \Pi_K(V)$, there is a $\overline{\mathcal{V}} \in \Pi_K(\overline{V})$ with $d_1(\,\Gamma/\mathcal{V}\,,\,\overline{\Gamma}/\overline{\mathcal{V}}\,) \leqslant \varepsilon$, and*
(2) *for each $\overline{\mathcal{V}} \in \Pi_K(\overline{V})$, there is a $\mathcal{V} \in \Pi_K(V)$ with $d_1(\,\Gamma/\mathcal{V}\,,\,\overline{\Gamma}/\overline{\mathcal{V}}\,) \leqslant \varepsilon$.*

Lemma 5.1 is proved in [A. Shapira and H. Stagni, in preparation]. We now deduce Theorem 3.2 from Lemma 5.1.

Proof of Theorem 3.2. Fix $\varepsilon > 0$ and an input graph $\Gamma \in \mathcal{G}(V)$. Let q be as in Lemma 5.1 with input K and ε/c. Choose Q uniformly from all subsets of V of size q and set $\overline{\Gamma} = \Gamma[Q]$. We will show that $z(\Gamma) = z(\overline{\Gamma}) \pm \varepsilon$ with probability at least $2/3$.

Let $\mathcal{V} \in \Pi_K(V)$ be an equipartition of Γ such that $z(\Gamma) = z^*(\,\Gamma/\mathcal{V}\,)$. By Lemma 5.1(1), with probability at least $2/3$, there is a partition $\overline{\mathcal{V}}$ of $\overline{\Gamma}$ such that $\mathrm{d}_1(\,\Gamma/\mathcal{V}\,,\,\overline{\Gamma}/\overline{\mathcal{V}}\,) < \varepsilon/c$. By the second condition on z^* in the statement of Theorem 3.2, we have $|z^*(\,\overline{\Gamma}/\overline{\mathcal{V}}\,) - z^*(\,\Gamma/\mathcal{V}\,)| \leqslant \varepsilon$, and therefore $z(\overline{\Gamma}) \leqslant z^*(\,\overline{\Gamma}/\overline{\mathcal{V}}\,) \leqslant z^*(\,\Gamma/\mathcal{V}\,) + \varepsilon = z(\Gamma) + \varepsilon$.

A symmetric argument, relying on Lemma 5.1(2), shows that $z(\Gamma) \leqslant z(\overline{\Gamma}) + \varepsilon$. □

6. Proof of Corollary 1.2

For a graph Γ and a family of graphs $\mathcal{F}$, let

$$\mathrm{ex}\,(\Gamma, \mathcal{F}) = \max\{e(G)\colon\, G \in \mathrm{Forb}(\Gamma, \mathcal{F})\}. \tag{6.1}$$

Moreover, let

$$\overline{\mathrm{ex}}\,(\Gamma, \mathcal{F}) = \frac{2}{n^2}\,\mathrm{ex}\,(\Gamma, \mathcal{F}) \tag{6.2}$$

be a normalized version of $\mathrm{ex}\,(\Gamma, \mathcal{F})$. Since $\mathrm{Forb}(\Gamma, \mathcal{F})$ is a monotone property, it is easy to see that the graph G attaining the maximum in (6.1) must also minimize the distance $\mathrm{d}_1(\Gamma, G)$ over every $G \in \mathrm{Forb}(\Gamma, \mathcal{F})$. In other words, we must have

$$\mathrm{d}_1(\Gamma, \mathcal{F}) = \frac{2}{n^2}\,e(\Gamma) - \overline{\mathrm{ex}}\,(\Gamma, \mathcal{F}). \tag{6.3}$$

Moreover, if G is a graph attending the maximum of (6.2), then *every* subgraph of G is also in $\mathrm{Forb}(\Gamma, \mathcal{F})$. Hence, it is clear that

$$\frac{2}{n^2}\log_2 |\,\mathrm{Forb}(\Gamma, \mathcal{F})| =: z_{\mathcal{F}}(\Gamma) \geqslant \overline{\mathrm{ex}}\,(\Gamma, \mathcal{F}). \tag{6.4}$$

By using the regularity method, Erdős, Frankl and Rödl [15] showed that, for $\Gamma = K_n$ and $\mathcal{F} = \{F\}$, the trivial lower bound given in (6.4) is actually asymptotically tight, *i.e.* that $z_{\mathcal{F}}(K_n) \leqslant \overline{\mathrm{ex}}\,(K_n, \mathcal{F}) + o(1)$. One may extend these results to arbitrary graphs Γ and arbitrary families $\mathcal{F}$, by using a stronger version of the Regularity Lemma, given in [3].

For completeness and to get a bound for n that does not depend upon the Regularity Lemma, we show the following result, which is a consequence of Theorem 4.1.

https://doi.org/10.1017/S0963548320000048 Published online by Cambridge University Press

Corollary 6.1. *Let $\mathcal{F}$ be a family of graphs such that* $\mathrm{Forb}(\mathcal{F})$ *is f-recoverable for some function $f\colon (0,1] \to \mathbb{R}$. Then, for any $\varepsilon > 0$, there exists $n_0 = \mathtt{poly}(f(\mathtt{poly}(\varepsilon)))$ such that any graph Γ of order $n \geqslant n_0$ satisfies*

$$\overline{\mathrm{ex}}(\Gamma, \mathcal{F}) \leqslant z_{\mathcal{F}}(\Gamma) \leqslant \overline{\mathrm{ex}}(\Gamma, \mathcal{F}) + \varepsilon.$$

Proof. By the above discussion, it suffices to prove the upper bound $z_{\mathcal{F}}(\Gamma) \leqslant \overline{\mathrm{ex}}(\Gamma, \mathcal{F}) + \varepsilon$. Let K be as in Theorem 4.1. Let $\mathcal{V} = \{V_i\}_{i=1}^k \in \Pi_K$ be an equipartition that maximizes $\overline{\mathrm{ex}}^*(\Gamma/\mathcal{V}, \mathcal{F})$ over all equipartitions of Γ of size at most K. Let $S \in \mathrm{Forb}^*_{\mathrm{hom}}(\Gamma/\mathcal{V}, \mathcal{F})$ be such that $e(S) = (k^2/2)\,\overline{\mathrm{ex}}^*(\Gamma/\mathcal{V}, \mathcal{F})$. By the maximality of S, we can assume without loss of generality that, for every $1 \leqslant i \leqslant j \leqslant k$, either $S(i,j) = \Gamma/\mathcal{V}(i,j)$ or $S(i,j) = 0$. Furthermore, $S(i,i) = 0$ for all $1 \leqslant i \leqslant k$. Consider the subgraph G of Γ that we get after removing every edge between V_i and V_j, for each $1 \leqslant i \leqslant j \leqslant k$ satisfying $S(i,j) = 0$. By construction we have $\Gamma/\mathcal{V} = S$, which implies that $G \in \mathrm{Forb}(\Gamma, \mathcal{F})$ by Proposition 3.3. Thus,

$$\begin{aligned}
\overline{\mathrm{ex}}(\Gamma, \mathcal{F}) &\geqslant \frac{2}{n^2} e(G) \\
&= \frac{2}{n^2} \sum_{1 \leqslant i < j \leqslant k} S(i,j) \frac{n^2}{k^2} \\
&= \frac{2}{k^2} e(S) \\
&= \overline{\mathrm{ex}}^*(\Gamma/\mathcal{V}, \mathcal{F}) \\
&= \max_{R \in \Gamma/\Pi_K} \overline{\mathrm{ex}}^*(R, \mathcal{F}) \\
&\geqslant z_{\mathcal{F}}(\Gamma) - \varepsilon,
\end{aligned}$$

where the last inequality follows from Theorem 4.1. □

Proof of Corollary 1.2. From Theorem 1.1 and Corollary 6.1, it follows that one can estimate $\overline{\mathrm{ex}}(\Gamma, \mathcal{F})$ with an additive error of $\varepsilon/2$ and with sample complexity $\mathtt{poly}(f(\mathtt{poly}(\varepsilon)))$ by simply estimating $z_{\mathcal{F}}(\Gamma)$ with an additive factor of $\varepsilon/4$. Furthermore, $\Gamma \mapsto (2/n^2)e(\Gamma)$ can be estimated with an additive error of $\varepsilon/2$ with sample complexity polynomial in $1/\varepsilon$. Hence, it follows from (6.3) that $\mathrm{d}_1(\Gamma, \mathrm{Forb}(\mathcal{F}))$ is estimable with an additive error of ε with sample complexity $\mathtt{poly}(f(\mathtt{poly}(\varepsilon)))$. □

7. Concluding remarks

In this paper we have introduced the concept of f-recoverability of a graph property $\mathcal{P}$. Using this concept, and the fact that any monotone property $\mathcal{P} = \mathrm{Forb}(\mathcal{F})$ is recoverable for a function f whose size is given by the Graph Removal Lemma, we found a probabilistic algorithm to estimate the number of $\mathcal{F}$-free subgraphs of a large graph G whose sample complexity does not depend on regularity.

We also proved that the number of $\mathcal{F}$-free subgraphs of a large graph G is closely related to the edit distance $\mathrm{d}_1(G, \mathrm{Forb}(\mathcal{F}))$ from G to $\mathrm{Forb}(\mathcal{F})$. In particular, our results give upper bounds on the sample complexity required to estimate $\mathrm{d}_1(G, \mathrm{Forb}(\mathcal{F}))$. Our bounds depend on the notion of f-recoverability of $\mathrm{Forb}(\mathcal{F})$, and are at most doubly exponential in the bounds given by the Removal Lemma for $\mathcal{F}$. Such bounds improve upon previous ones in [7] and [18], for the specific case of monotone properties. The approaches in [7] and [18] depend on the Regularity Lemma, yielding tower-type bounds for the sample complexity, even for properties admitting an efficient Removal Lemma (see [23]).

Going back to the notion of f-recoverability, we note that, being a new concept, little is known about it, and we believe that it would be interesting to investigate this notion in more

https://doi.org/10.1017/S0963548320000048 Published online by Cambridge University Press

detail. For instance, in our proof that any monotone property Forb($\mathcal{F}$) is f-recoverable, we found ε-recovering partitions $\mathcal{V}$ that were γ-FK-regular (in fact we showed that *any* such partition is ε-recovering), where $\gamma(\varepsilon)$ is chosen in such a way that the Removal Lemma applies. On the other hand, our discussion after Definition 3.5 implies that the property of being r-colourable is ε-recoverable with sample complexity r/ε, and thus we may find an ε-recovering partition whose size is less than the size required to ensure the existence of an FK-regular partition. It is natural to ask for properties that can be recovered by small partitions; more precisely, one could ask for a characterization of properties that are $f(\varepsilon)$-recoverable for $f(\varepsilon)$ polynomial in $1/\varepsilon$.

Here we have restricted ourselves to monotone graph properties. We should mention that the parameter

$$\hat{z}_{\mathcal{P}}(\Gamma) = \frac{2}{|V(\Gamma)|^2} \log_2 |\{G \leqslant \Gamma : G \in \mathcal{P}\}|$$

might not even be estimable for arbitrary (non-monotone) properties $\mathcal{P}$. For instance, if $\mathcal{P}$ is the hereditary property of graphs having no independent sets of size three, then the complete graph K_n and the graph $K_n - E(K_3)$, which is obtained from K_n by removing the edges of a triangle, have quite a different number of spanning subgraphs satisfying $\mathcal{P}$, namely $2^{n^2/4}$ and 0, respectively, although their edit distance is negligible. It follows from [11, Theorem 6.1] that $\hat{z}_{\mathcal{P}}$ is not estimable.

Nevertheless, the definition of f-recoverable can be extended to cope with general hereditary properties, and along with Theorem 3.2 it provides a way of estimating other interesting hereditary properties. In particular, this framework is used in a follow-up paper [C. Hoppen, Y. Kohayakawa, R. Lang, H. Lefmann, and H. Stagni, submitted for publication] to estimate the edit distance to any fixed hereditary property with a sample complexity similar to the one obtained here.

Acknowledgements

This study was financed in part by CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil, Finance Code 001. CNPq is the National Council for Scientific and Technological Development of Brazil. FAPESP is the São Paulo Research Foundation. Some results in this paper appeared in a preliminary form in the proceedings of APPROX/RANDOM 2016 [28]. C. Hoppen acknowledges the support of CNPq (308054/2018-0). C. Hoppen and H. Lefmann acknowledge the support of CAPES and DAAD via PROBRAL (CAPES Proc. 88881.143993/2017-01 and DAAD 57391132). C. Hoppen, Y. Kohayakawa and H. Stagni acknowledge the support of FAPESP (2013/03447-6). Y. Kohayakawa was partially supported by CNPq (311412/2018-1 and 423833/2018-9). H. Stagni was supported by FAPESP (2015/15986-4, 2017/02263-0) and CNPq (141970/2015-4). R. Lang acknowledges support by the Millennium Nucleus Information and Coordination in Networks ICM/FIC RC130003.

The authors are grateful to the referees for their useful comments.

References

[1] Alon, N. (2002) Testing subgraphs in large graphs. *Random Struct. Alg.* **21** 359–370.
[2] Alon, N., Duke, R. A., Lefmann, H., Rödl, V. and Yuster, R. (1994) The algorithmic aspects of the regularity lemma. *J. Algorithms* **16** 80–109.
[3] Alon, N., Fischer, E., Krivelevich, M. and Szegedy, M. (2000) Efficient testing of large graphs. *Combinatorica* **20** 451–476.
[4] Alon, N., Fischer, E., Newman, I. and Shapira, A. (2009) A combinatorial characterization of the testable graph properties: It's all about regularity. *SIAM J. Comput.* **39** 143–167.
[5] Alon, N. and Shapira, A. (2008) A characterization of the (natural) graph properties testable with one-sided error. *SIAM J. Comput.* **37** 1703–1727.
[6] Alon, N. and Shapira, A. (2008) Every monotone graph property is testable. *SIAM J. Comput.* **38** 505–522.
[7] Alon, N., Shapira, A. and Sudakov, B. (2009) Additive approximation for edge-deletion problems. *Ann. of Math.* **170** 371–411.

https://doi.org/10.1017/S0963548320000048 Published online by Cambridge University Press

[8] Balogh, J., Bollobás, B. and Simonovits, M. (2004) The number of graphs without forbidden subgraphs. *J. Combin. Theory Ser. B* **91** 1–24.
[9] Balogh, J. and Samotij, W. (2011) The number of $K_{s,t}$-free graphs. *J. Lond. Math. Soc. (2)* **83** 368–388.
[10] Bollobás, B. (1998) Hereditary properties of graphs: Asymptotic enumeration, global structure, and colouring. *Documenta Math.* Extra Vol. **ICM III** 333–342.
[11] Borgs, C., Chayes, J. T., Lovász, L., Sós, V. T. and Vesztergombi, K. (2008) Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Adv. Math.* **219** 1801–1851.
[12] Conlon, D. and Fox, J. (2012) Bounds for graph regularity and removal lemmas. *Geom. Funct. Anal.* **22** 1191–1256.
[13] Conlon, D. and Fox, J. (2013) Graph removal lemmas. In *Surveys in Combinatorics 2013*, Vol. 409 of London Mathematical Society Lecture Note Series, Cambridge University Press, pp. 1–49.
[14] Duke, R. A., Lefmann, H. and Rödl, V. (1995) A fast approximation algorithm for computing the frequencies of subgraphs in a given graph. *SIAM J. Comput.* **24** 598–620.
[15] Erdős, P., Frankl, P. and Rödl, V. (1986) The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent. *Graphs Combin.* **2** 113–121.
[16] Erdős, P., Kleitman, D. J. and Rothschild, B. L. (1976) Asymptotic enumeration of K_n-free graphs. In *Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973), Vol. II*, Accademia Nazionale dei Lincei, Rome, pp. 19–27.
[17] Erdős, P. and Simonovits, M. (1983) Supersaturated graphs and hypergraphs. *Combinatorica* **3** 181–192.
[18] Fischer, E. and Newman, I. (2007) Testing versus estimation of graph properties. *SIAM J. Comput.* **37** 482–501.
[19] Fox, J. (2011) A new proof of the graph removal lemma. *Ann. of Math. (2)* **174** 561–579.
[20] Fox, J., Lovász, L. M. and Zhao, Y. (2017) On regularity lemmas and their algorithmic applications. *Combin. Probab. Comput.* **26** 481–505.
[21] Frieze, A. and Kannan, R. (1999) Quick approximation to matrices and applications. *Combinatorica* **19** 175–220.
[22] Füredi, Z. (1995) Extremal hypergraphs and combinatorial geometry. In *Proceedings of the International Congress of Mathematicians 1994*, Birkhäuser, pp. 1343–1352.
[23] Gishboliner, L. and Shapira, A. (2017) Removal lemmas with polynomial bounds. In *STOC'17: Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, ACM, pp. 510–522.
[24] Goldreich, O. (2010) *Property Testing: Current Research and Surveys*, Vol. 6390 of Lecture Notes in Computer Science, Springer.
[25] Goldreich, O., Goldwasser, S. and Ron, D. (1998) Property testing and its connection to learning and approximation. *J. Assoc. Comput. Mach.* **45** 653–750.
[26] Goldreich, O. and Trevisan, L. (2003) Three theorems regarding testing graph properties. *Random Struct. Alg.* **23** 23–57.
[27] Gowers, W. T. (1997) Lower bounds of tower type for Szemerédi's uniformity lemma. *Geom. Funct. Anal.* **7** 322–337.
[28] Hoppen, C., Kohayakawa, Y., Lang, R., Lefmann, H. and Stagni, H. (2016) Estimating parameters associated with monotone properties. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2016)* (K. Jansen, C. Mathieu, J. D. P. Rolim and C. Umans, eds), Vol. 60 of Leibniz International Proceedings in Informatics (LIPIcs), Schloss Dagstuhl–Leibniz-Zentrum für Informatik, pp. 35:1–35:13.
[29] Jukna, S. (2001) *Extremal Combinatorics: With Applications in Computer Science*, second edition (2011), Texts in Theoretical Computer Science, Springer.
[30] Komlós, J. and Simonovits, M. (1996) Szemerédi's regularity lemma and its applications in graph theory. In *Combinatorics: Paul Erdős is Eighty (Keszthely 1993)*, Vol. 2 of Bolyai Society Mathematical Studies, János Bolyai Mathematical Society, pp. 295–352.
[31] Lovász, L. and Szegedy, B. (2007) Szemerédi's lemma for the analyst. *Geom. Funct. Anal.* **17** 252–270.
[32] Moshkovitz, G. and Shapira, A. (2019) A sparse regular approximation lemma. *Trans. Amer. Math. Soc.* **371** 6779–6814.
[33] Parnas, M., Ron, D. and Rubinfeld, R. (2006) Tolerant property testing and distance approximation. *J. Comput. System Sci.* **72** 1012–1042.
[34] Prömel, H. J. and Steger, A. (1992) The asymptotic number of graphs not containing a fixed color-critical subgraph. *Combinatorica* **12** 463–473.
[35] Prömel, H. J. and Steger, A. (1996) Counting H-free graphs. *Discrete Math.* **154** 311–315.
[36] Ruzsa, I. Z. and Szemerédi, E. (1978) Triple systems with no six points carrying three triangles. In *Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976)*, Vol. 18 of Colloquia Mathematica Societatis János Bolyai, North-Holland, pp. 939–945.
[37] Szemerédi, E. (1976) Regular partitions of graphs. *Colloq. Internat. CNRS* **260** 399–401.
[38] Tao, T. (2006) A variant of the hypergraph removal lemma. *J. Combin. Theory Ser. A* **113** 1257–1280.

Cite this article: Hoppen C, Kohayakawa Y, Lang R, Lefmann H and Stagni H (2020). Estimating parameters associated with monotone properties. *Combinatorics, Probability and Computing* **29**, 616–632. https://doi.org/10.1017/S0963548320000048

https://doi.org/10.1017/S0963548320000048 Published online by Cambridge University Press