The maximum edit distance from hereditary graph properties

Noga Alon * Uri Stav †

September 21, 2007

Abstract

For a graph property $\mathcal{P}$, the *edit distance* of a graph G from $\mathcal{P}$, denoted $E_{\mathcal{P}}(G)$, is the minimum number of edge modifications (additions or deletions) one needs to apply to G in order to turn it into a graph satisfying $\mathcal{P}$. What is the largest possible edit distance of a graph on n vertices from $\mathcal{P}$? Denote this distance by $ed(n, \mathcal{P})$.

A graph property is *hereditary* if it is closed under removal of vertices. In [7], the authors show that for any hereditary property, a random graph $G(n, p(\mathcal{P}))$ essentially achieves the maximal distance from $\mathcal{P}$, proving: $ed(n, \mathcal{P}) = E_{\mathcal{P}}(G(n, p(\mathcal{P}))) + o(n^2)$ with high probability. The proof implicitly asserts the existence of such $p(\mathcal{P})$, but it does not supply a general tool for determining its value or the edit distance.

In this paper, we determine the values of $p(\mathcal{P})$ and $ed(n, \mathcal{P})$ for some subfamilies of hereditary properties including sparse hereditary properties, complement invariant properties, (r, s)-colorability and more. We provide methods for analyzing the maximum edit distance from the graph properties of being induced H-free for some graphs H, and use it to show that in some natural cases $G(n, 1/2)$ is *not* the furthest graph. Throughout the paper, the various tools let us deduce the asymptotic maximum edit distance from some well studied hereditary graph properties, such as being Perfect, Chordal, Interval, Permutation, Claw-Free, Cograph and more. We also determine the edit distance of $G(n, 1/2)$ from *any* hereditary property, and investigate the behavior of $E_{\mathcal{P}}(G(n, p))$ as a function of p.

The proofs combine several tools in Extremal Graph Theory, including strengthened versions of the Szemerédi Regularity Lemma, Ramsey Theory and properties of random graphs.

*Schools of Mathematics and Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. Email: nogaa@tau.ac.il. Research supported in part by the Israel Science Foundation, by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University and by a USA Israeli BSF grant.

†School of Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. Email: uristav@tau.ac.il.

1 Introduction

1.1 Definitions and motivation

A **graph property** is a set of graphs closed under isomorphism. A graph property is **hereditary** if it is closed under removal of vertices (and *not* necessarily under removal of edges). Equivalently, such properties are closed under taking induced subgraphs.

Given two graphs on n vertices, G_1 and G_2, the **edit distance** between G_1 and G_2 is the minimum number of edge additions and/or deletions that are needed in order to turn G_1 into a graph isomorphic to G_2. We denote this quantity by $\boldsymbol{\Delta(G_1, G_2)}$.

For a given graph property $\mathcal{P}$, let $\mathcal{P}^n$ denote the set of (labelled) graphs on n vertices which satisfy $\mathcal{P}$. We want to investigate how far a graph G is from satisfying $\mathcal{P}$, and thus define the edit distance of a graph G from $\mathcal{P}$ by $\boldsymbol{E_{\mathcal{P}}(G)} = \min\{\Delta(G, G') \mid G' \in \mathcal{P}^{|V(G)|}\}$. In words, $E_{\mathcal{P}}(G)$ is the minimum edit distance of G to a graph satisfying $\mathcal{P}$.

In this paper we address the following extremal question: Given a hereditary graph property $\mathcal{P}$, what is the graph on n vertices with the largest edit distance from $\mathcal{P}$? That is, the graph to which one has to apply the largest number of edge modifications in order to obtain a member of $\mathcal{P}$. Denote the maximal possible distance by $\boldsymbol{ed(n, \mathcal{P})}$.

Although this question seems natural on its own, it is mainly motivated by problems in theoretical computer science. In the edge-modification problem of the property $\mathcal{P}$, one wants to determine $E_{\mathcal{P}}(G)$ given an input graph G. Clearly, the computational complexity of such an optimization problem strongly depends on the graph property at hand. Narrowing our discussion to hereditary properties is one of the mildest and yet natural restrictions. These properties have attracted the attention of researchers in various areas of graph theory and in theoretical and applied computer science (cf., for example, [24], [27], [31] and their references). Some of them are the following well studied graph properties:

- **Perfect Graphs:** A graph G is perfect if for every *induced* subgraph G' of G, the chromatic number of G' equals the size of the largest clique in G'.
- **Chordal Graphs:** A graph is chordal if it contains no *induced* cycle of length at least 4.
- **Interval Graphs:** A graph G on n vertices is an interval graph if there are closed intervals on the real line $I_1, \ldots, I_n$ such that $(i, j) \in E(G)$ if and only if $I_i \cap I_j \neq \emptyset$.
- **Permutation Graphs:** A graph G on n vertices is a permutation graph if there is a permutation σ of $\{1, \ldots, n\}$ such that $(i, j) \in E(G)$ if and only if (i, j) is an inversion under σ.
- **Cographs:** This class is defined recursively as follows. A graph consisting of a single vertex is a cograph, the complement of every cograph is also a cograph and so is the disjoint union of

any two cographs. Equivalently, as shown in [18], this is exactly the class of induced P_4-free graphs, where P_i denotes the path on i vertices.

In fact, almost all interesting graph properties are hereditary. The recent results of [5] on the approximability of edge-modification problems for monotone graph properties indicate that the extremal aspects of edge-modification problems for hereditary properties should be helpful in obtaining tools for establishing the hardness of such problems. We note that another motivation for this question comes from the investigation of Hamming distance between matrices, as described in [9] and [10].

1.2 Related work

In [7], the authors showed that for any hereditary property $\mathcal{P}$, the maximal distance from $\mathcal{P}$ is essentially achieved by a random graph $G(n,p)$ with an edge density that depends on $\mathcal{P}$.

Theorem 1.1. ([7]) *Let $\mathcal{P}$ be an arbitrary hereditary graph property. Then there exists $p = p(\mathcal{P}) \in [0,1]$, such that almost surely (that is, with probability that tends to 1 as n tends to infinity),*

$$ed(n,\mathcal{P}) = E_{\mathcal{P}}(G(n,p)) + o(n^2) \ . \tag{1}$$

The proof of Theorem 1.1 implicitly asserts the existence of such p, but it does not supply a general tool for determining its value or the edit distance. It is thus a natural question to seek for the value of $p(\mathcal{P})$ for hereditary properties $\mathcal{P}$, and to find the edit distance of $G(n,p(\mathcal{P}))$ from $\mathcal{P}$.

For example, consider the family of monotone graph properties. A graph property $\mathcal{M}$ is **monotone** if it is closed under removal of vertices and edges. Clearly, every monotone graph property is also hereditary, and when modifying a graph G in the most economical way in order to obtain a graph in $\mathcal{M}$ one only deletes edges from G. Hence, trivially, for any monotone graph property $\mathcal{M}$, the furthest graph from satisfying $\mathcal{M}$ is the complete graph. In our notations, this means that $p(\mathcal{M}) = 1$. Moreover, Turán's Theorem [34] and its various extensions (most notably by Erdős and Stone [23], and by Erdős and Simonovits [21]) show that for any monotone graph property $\mathcal{M}$: $ed(n,\mathcal{M}) = (\frac{1}{r} - o(1))\binom{n}{2}$ where $r = min\{\chi(F) - 1 \mid F \notin \mathcal{M}\}$.

Any hereditary property can be defined by its (possibly infinite) set of minimal forbidden induced subgraphs. This is a family of graphs $\boldsymbol{\mathcal{F}^*_{\mathcal{P}}}$ such that a graph G belongs to $\mathcal{P}$ if and only if it does not contain an induced copy of a member of $\mathcal{F}^*_{\mathcal{P}}$. If $\mathcal{F}^*_{\mathcal{P}}$ consists of a single graph $\mathcal{F}^*_{\mathcal{P}} = \{H\}$, then $\mathcal{P}$ contains all graphs excluding an induced subgraph isomorphic to H. We call these graphs **induced H-free**, and denote such a property by $\boldsymbol{\mathcal{P}^*_H}$. Practicing the above definitions, we note that any result involving a graph H immediately implies an analogous result for its complement (denoted $\overline{H}$), since $E_{\mathcal{P}^*_H}(G) = E_{\mathcal{P}^*_{\overline{H}}}(\overline{G})$ and hence $ed(n,\mathcal{P}^*_H) = ed(n,\mathcal{P}^*_{\overline{H}})$.

Axenovich, Kézdy and Martin addressed the extremal question of finding $ed(n,\mathcal{P}^*_H)$. They recently showed in [9] that $ed(n,\mathcal{P}^*_H)$ is bounded by a function of H as follows. Define $\chi_B(H)$ to

be the least integer $k+1$ such that for all (r,s) satisfying $r+s=k+1$ the vertices of H can be partitioned into $r+s$ sets, r of them spanning empty graphs and s spanning complete graphs.[1] For such H, they show that

$$(\frac{1}{2k}-o(1))\binom{n}{2} < ed(n,\mathcal{P}_H^*) \leq \frac{1}{k}\binom{n}{2} . \tag{2}$$

The gap left in this general bound is settled in [9] for some families of graphs. In particular, for self-complementary graphs (i.e. $H=\overline{H}$) it is shown there that $ed(n,\mathcal{P}_H^*)=(\frac{1}{2k}-o(1))\binom{n}{2}$. We note that the lower bound of (2) is attained by the random graph $G(n,1/2)$. Therefore, whenever the lower bound of (2) is asymptotically tight, it follows that, in our notation, $p=\frac{1}{2}$ satisfies (1) for $\mathcal{P}_H^*$.

1.3 Organization

The rest of the paper is organized as follows. In Section 2 we detail all the new results proved in this paper. In Section 3 we review the definitions and state the Regularity Lemmas which will be used in the proofs, as well as some auxiliary definitions and tools. Sections 4, 5, 6, 7, 8 and 9 contain the proofs of all the results. Section 10 consists of some concluding remarks and open problems.

2 The new results

We determine values of $p(\mathcal{P})$ and establish bounds on the maximum edit distance for the following subfamilies of hereditary properties. In this section we only state the results, while the proofs appear in the following sections.

2.1 Sparse hereditary properties

We write $|\mathcal{P}^n|$ for the number of *labelled* graphs on n vertices that satisfy $\mathcal{P}$. Scheinerman and Zito ([32]) showed that $|\mathcal{P}^n|$ belongs to one of few possible classes of functions. Several other papers sharpen these results, focusing on sparse hereditary properties (Balogh, Bollobás and Weinreich [11], [12]), dense hereditary properties (Bollobás and Thomason [14], [16] and Alekseev [1]) and properties of the type $\mathcal{P}_H^*$ (Prömel and Steger [28] ,[29], [30]).

Let us focus on the edit distance of "sparse" hereditary properties, for which $|\mathcal{P}^n|=2^{o(n^2)}$. Consider $\mathcal{P}^n$ as a set of points in the space of all n-vertex graphs. Intuitively, it seems that if there are "few" graphs in $\mathcal{P}^n$, then there should be some graph on n vertices that is "far" from all the points representing members of $\mathcal{P}^n$. Confirming this intuition, we show the following result. Denote by $e(G)=|E(G)|$ the number of edges of the graph G.

[1] This parameter was first defined by Prömel and Steger in [29], where it was denoted by $\tau(H)$

Theorem 2.1. *Let $\mathcal{P}$ be a hereditary property such that $|\mathcal{P}^n| = 2^{o(n^2)}$. Then precisely one of the following holds:*

1. *Any $G \in \mathcal{P}$ satisfies $e(G) = o(n^2)$, $ed(n, \mathcal{P}) = (1 - o(1))\binom{n}{2}$ and $p(\mathcal{P}) = 1$.*
2. *Any $G \in \mathcal{P}$ satisfies $e(\overline{G}) = o(n^2)$, $ed(n, \mathcal{P}) = (1 - o(1))\binom{n}{2}$ and $p(\mathcal{P}) = 0$.*
3. *For every n there are graphs $G_1, G_2 \in \mathcal{P}^n$ such that $e(G_1) = \Omega(n^2)$ and $e(\overline{G_2}) = \Omega(n^2)$, $ed(n, \mathcal{P}) = (\frac{1}{2} - o(1))\binom{n}{2}$ and $p(\mathcal{P}) = \frac{1}{2}$.*

There are many natural sparse hereditary properties for which an asymptotic result is immediately attained by applying the above theorem. For example, the following well known result refers to the natural graph properties of being Interval, Permutation or Cograph.

Lemma 2.2. *Let $\mathcal{P}$ be one of the three hereditary properties of Interval graphs, Permutation graphs or Cographs. Then $|\mathcal{P}^n| = 2^{\Theta(n \log n)}$.*

Corollary 2.3. *Let $\mathcal{P}$ be one of the three hereditary properties of Interval graphs, Permutation graphs or Cographs. Then $p(\mathcal{P}) = \frac{1}{2}$ and $ed(n, \mathcal{P}) = \frac{1}{2}\binom{n}{2} - O(n^{1.5}\sqrt{\log n}\,)$.*

This result can be extended to a much wider family of graph properties, defined by polynomials as follows. Let $\mathcal{Q} = \{Q_1, \ldots, Q_t\}$ be real polynomials in $2d$ variables, and let $b : \{-1, 0, 1\}^t \to \{0, 1\}$ be a binary mapping from all possible sign patterns of the polynomials. We say that an ordered pair of points $\underline{c}_i, \underline{c}_j \in \mathbb{R}^d$ *satisfies* $(\mathcal{Q}, b)$ if

$$b(sign(Q_1(\underline{c}_i, \underline{c}_j)), \ldots, sign(Q_t(\underline{c}_i, \underline{c}_j))) = 1 \ .$$

In words, $(\underline{c}_i, \underline{c}_j)$ satisfies $(\mathcal{Q}, b)$ if the sign pattern of the substitution of the coordinates of $\underline{c}_i$ and $\underline{c}_j$ in the polynomials has value 1 in b. We define the following class of graphs.

Definition 2.4. *For a pair $(\mathcal{Q}, b)$ as above, the labelled set of $n > 0$ points $\underline{c}_1, \ldots, \underline{c}_n \in \mathbb{R}^d$ defines a graph G on n vertices where (i, j) is an edge in G if and only if $(\underline{c}_i, \underline{c}_j)$ satisfies $(\mathcal{Q}, b)$, for $1 \leq i < j \leq n$. We denote by $\boldsymbol{\mathcal{P}_{\mathcal{Q},b}}$ the collection of all such graphs obtained by any n-tuple of points in $\mathbb{R}^d$, $n > 0$.*

For any $\mathcal{Q}$ and b, $\mathcal{P}_{\mathcal{Q},b}$ is a hereditary graph property. Note that if for *any* pair of points $\underline{c}_i, \underline{c}_j \in \mathbb{R}^d$ the value of $b(sign(Q_1(\underline{c}_i, \underline{c}_j)), \ldots, sign(Q_t(\underline{c}_i, \underline{c}_j)))$ is constant, then $\mathcal{P}_{\mathcal{Q},b}$ contains either only complete graphs or only empty graphs. In order to avoid this trivial case, if there is a pair of points $(\underline{c}_i, \underline{c}_j)$ that satisfies $(\mathcal{Q}, b)$ and a pair of points $(\underline{c}_i', \underline{c}_j')$ that does not satisfy $(\mathcal{Q}, b)$ then we say that $(\mathcal{Q}, b)$ is *non-trivial*.

Lemma 2.5. *Let $\mathcal{P}_{\mathcal{Q},b}$ be a graph property as defined above, where $(\mathcal{Q}, b)$ is* non-trivial. *Then $p(\mathcal{P}_{\mathcal{Q},b}) = \frac{1}{2}$ and $ed(n, \mathcal{P}_{\mathcal{Q},b}) = \frac{1}{2}\binom{n}{2} - O(n^{1.5}\sqrt{\log n}\,)$.*

We note that many natural families of intersection graphs of various geometric bodies may be represented by polynomials as above. See Example 4.5 in Section 4.

2.2 Complement invariant properties

A graph property $\mathcal{P}$ is complement invariant if it is closed under taking the complement of a graph, i.e. $G \in \mathcal{P}$ if and only if $\overline{G} \in \mathcal{P}$. In Section 5 we sketch the proof of a property of the expectation of $E_{\mathcal{P}}(G(n,p))$ as a function of p, based on the methods of [7]. It also yields the following result on complement invariant properties.

Theorem 2.6. *For any hereditary complement invariant graph property $\mathcal{P}$, $p = \frac{1}{2}$ satisfies condition (1) of Theorem 1.1.*

For example, it follows that $G(n, 1/2)$ is essentially the furthest graph from being perfect.

2.3 (r,s)-colorability

Due to the broad range of hereditary properties, when studying them one seeks a small family of hereditary properties that approximate every other one in some sense. The simplest such family is defined by (r,s)-colorability of graphs. These properties were introduced in several contexts, and seem to capture important algorithmic and extremal characteristics of hereditary properties. See e.g. Prömel and Steger ([29], [30]), Bollobás and Thomason ([16], [15]) and Alekseev [1]. We define them as follows.

Definition 2.7. *For any pair of integers (r,s), such that $r+s > 0$, we say that a graph $G = (V,E)$ is **(r,s)-colorable** if there is a partition of V into $r+s$ (possibly empty) subsets $I_1, \ldots, I_r, C_1, \ldots, C_s$ such that each I_k induces an independent set in G, and each C_k induces a clique in G.*

For example, $(r,0)$-colorable graphs are r-colorable graphs. We denote by $\boldsymbol{\mathcal{P}_{r,s}}$ the graph property comprising all the (r,s)-colorable graphs. Clearly, for any pair (r,s), the property $\mathcal{P}_{r,s}$ is hereditary.

Suppose that a graph H is **not** (r,s)-colorable. In this case, $\mathcal{P}_{r,s} \subseteq \mathcal{P}_H^*$ and hence $ed(n, \mathcal{P}_{r,s}) \geq ed(n, \mathcal{P}_H^*)$. An upper bound for $ed(n, \mathcal{P}_{r,s})$ therefore provides an upper bound for $ed(n, \mathcal{P}_H^*)$. In [16], Bollobás and Thomason showed that for any hereditary property $\mathcal{P}$, there is some $\mathcal{P}_{r,s} \subseteq \mathcal{P}$ such that the size of $\mathcal{P}$ is asymptotically close to the size of $\mathcal{P}_{r,s}$ in the logarithmic sense, that is $\log|\mathcal{P}^n| = (1-o(1))\log|\mathcal{P}_{r,s}^n|$. It is therefore not surprising that in some cases one can show for a graph G, that $E_{\mathcal{P}}(G)$ cannot be much smaller than $E_{\mathcal{P}_{r,s}}(G)$. This motivates the following extremal question: Given a pair of integers (r,s), $r+s>0$, what is the maximal edit distance of a graph G on n vertices from being (r,s)-colorable? We prove:

Theorem 2.8. *For any pair of integers (r,s), such that $r+s>0$,*

$$ed(n, \mathcal{P}_{r,s}) = \left(\frac{1}{(\sqrt{r}+\sqrt{s})^2} - o(1)\right)\binom{n}{2}.$$

The proof of Theorem 2.8 also provides the value of $p(\mathcal{P}_{r,s})$ for any (r,s). In Section 6 we describe another broader family of properties used for approximating hereditary properties, and a method for analyzing its edit distance.

2.4 The edit distance of $G(n,1/2)$ from hereditary properties

The importance of random graphs with respect to the edit distance problem, which is derived from Theorem 1.1, motivates another natural question: is there a robust method for determining the typical edit distance of $G(n,p)$ from an arbitrary hereditary property $\mathcal{P}$? As it turns out, this question is much easier when $p=\frac{1}{2}$.

Definition 2.9. *Let $\mathcal{P}$ be a hereditary property. Define the* ***binary chromatic number*** *of $\mathcal{P}$ as the least integer $k+1$ such that for any (r,s) satisfying $r+s=k+1$ there is a graph not in $\mathcal{P}$ that is (r,s)-colorable, and denote it by $\boldsymbol{\chi_B(\mathcal{P})}$. Equivalently,*

$$\boldsymbol{\chi_B(\mathcal{P})} = 1+\max\{r+s : \mathcal{P}_{r,s}\subseteq\mathcal{P}\}\ .$$

This definition extends the definition of the binary chromatic number for graphs from [9] and [29] since $\chi_B(\mathcal{P}_H^*)=\chi_B(H)$.

Theorem 2.10. *Let $\mathcal{P}$ be an arbitrary hereditary property, then with high probability*

$$E_{\mathcal{P}}(G(n,1/2)) = \Big(\frac{1}{2(\chi_B(\mathcal{P})-1)} \pm o(1)\Big)\binom{n}{2}\ .$$

Example 2.11. *Let $\mathcal{P}$ be the class of perfect graphs. In this case, e.g., $\mathcal{P}_{2,0}\subset\mathcal{P}$. On the other hand, C_5 is not perfect and whenever $r+s\geq 3$, C_5 is (r,s)-colorable and therefore $\mathcal{P}_{r,s}\nsubseteq\mathcal{P}$. Thus, $\chi_B(\mathcal{P})=3$ and since being perfect is complement invariant, by Theorem 2.6 and Theorem 2.10: $ed(n,\mathcal{P})=E_{\mathcal{P}}(G(n,\frac{1}{2}))=(\frac{1}{4}-o(1))\binom{n}{2}$.*

2.5 Induced H-freeness

Having proved general results on the edit distance from hereditary properties, we demonstrate that in some natural case a random graph with a density which differs from $\frac{1}{2}$ is the furthest from satisfying $\mathcal{P}$, even when $\mathcal{P}=\mathcal{P}_H^*$. The well studied (see e.g. [19]) family of (induced) claw-free graphs is a good example for that. A claw $K_{1,3}$ consists of a vertex connected to three other vertices (no two of which are connected). For short we omit the 'induced', and as usual call such graphs claw-free[2].

We first make the following observation on (r,s)-colorability of claw-free graphs. If a graph G is edgeless, i.e. $(1,0)$-colorable, then certainly it is claw free. On the other hand, if G is $(0,2)$-colorable, i.e. its vertices can be partitioned into two sets, each spanning a clique, then it is again claw-free. It is not difficult to verify that $(2,0)$, $(0,3)$, and $(1,1)$ colorable graphs may contain an induced copy of $K_{1,3}$, and therefore are not guaranteed to be claw-free. Hence, the binary chromatic number of $\mathcal{P}_{K_{1,3}}^*$ is 3. The general bounds of [9] show that $(\frac{1}{4}-o(1))\binom{n}{2}\leq ed(n,\mathcal{P}_{K_{1,3}}^*)\leq\frac{1}{2}\binom{n}{2}$. The lower bound is attained by $G(n,\frac{1}{2})$ (this also follows from Theorem 2.10). Nevertheless, this turns out to be far from optimal as we show that the extremal probability is in fact $\frac{1}{3}$.

[2] A graph without any *weak* copies of a claw has maximum degree at most two, and hence consists of a disjoint collection of cycles and paths.

Theorem 2.12. *Let $K_{1,3}$ denote a claw. Then $p(\mathcal{P}^*_{K_{1,3}}) = \frac{1}{3}$ and $ed(n, \mathcal{P}^*_{K_{1,3}}) = (\frac{1}{3} - o(1))\binom{n}{2}$.*

Similar results are also obtained for all the graphs H on at most four vertices.

A natural question which arises from the asymptotic results is whether it is possible to determine the exact value of $ed(n, \mathcal{P})$ for a given $\mathcal{P} = \mathcal{P}^*_H$. If H consists of 2 vertices, then it is either an edge or an independent set of size two. In these cases, $\mathcal{P}^*_H$ consists of either empty or complete graphs and thus the edit distance is $\binom{n}{2}$. Let us briefly review the graphs H on three vertices. The triangle K_3 corresponds to the classical result of Turán [34] (or its special case due to Mantel [26]). The other options are either a path P_3 consisting of three vertices or its complement. In [9] it is shown that $ed(n, \mathcal{P}^*_{P_3}) = \lfloor \frac{n}{2} \rfloor (\lceil \frac{n}{2} \rceil - 1)$, and hence when $|V(H)| \leq 3$ all values of $ed(n, \mathcal{P}^*_H)$ are known.

We thus discuss some graphs on four vertices. For the cycle of length four, denoted C_4, we show:

Theorem 2.13. *If n is even, then $ed(n, \mathcal{P}^*_{C_4}) = \binom{n/2}{2}$, and an extremal graph in this case is $K_{\frac{n}{2},\frac{n}{2}}$.*

The proof also yields:

Corollary 2.14. *Let $\mathcal{P}$ denote the class of all chordal graphs. Then $ed(2n, \mathcal{P}) = \binom{n}{2}$.*

We also prove an improved upper bound for the edit distance from being a cograph, which shows, together with Corollary 2.3:

Theorem 2.15. *For the property $\mathcal{P}^*_{P_4}$: $\frac{1}{2}\binom{n}{2} - O(n^{1.5}\sqrt{\log n}) < ed(n, \mathcal{P}^*_{P_4}) < \frac{1}{2}\binom{n}{2} - \Omega(n^{1.5})$.*

3 Regularity lemma background and preliminaries

In this section we discuss the basic notions of regularity, some of the basic applications of regular partitions and state the regularity lemmas that we use in the proofs of Theorems 2.10 and 2.12. See [25] for a comprehensive survey on the regularity lemma.

For a set of vertices $A \subseteq V$, we denote by $E(A)$ the set of edges of the graph induced by A in G. We also denote by $e(A)$ the size of $E(A)$. Similarly, for every two nonempty disjoint vertex sets A and B of a graph G, $E(A, B)$ stands for the set of edges of G between A and B, and $e(A, B)$ is the number of edges . The **edge density** of the pair is defined as $d(A, B) = e(A, B)/|A||B|$. When several graphs on the same set of vertices are involved, we write $d_G(A, B)$ to specify the graph to which we refer.

Definition 3.1. (γ-regular pair) *A pair (A, B) is* γ-regular*, if for any two subsets $A' \subseteq A$ and $B' \subseteq B$, satisfying $|A'| \geq \gamma|A|$ and $|B'| \geq \gamma|B|$, the inequality $|d(A', B') - d(A, B)| \leq \gamma$ holds.*

The following simple fact about regular pairs is very useful. It roughly states that in a regular pair there cannot be too many vertices with low degrees.

Fact 3.2. *Let (A, B) be a γ-regular pair with density η, and let $Y \subseteq B$ be of size at least $\gamma|B|$. Then all but at most $\gamma|A|$ of the vertices of A have at least $(\eta - \gamma)|Y|$ neighbors in Y.*

Proof: Assume that for some X, such that $|X| \geq \gamma|A|$, for all $v \in X$ the inequality does not hold. This means that there are less than $(\eta - \gamma)|X||Y|$ edges connecting vertices of X and Y. Hence, the pair (X, Y) contradicts the γ-regularity of the pair (A, B).

■

A very useful lemma that we use in this paper is Lemma 3.3 below. This is a version of the classical key lemma, which helps us find induced copies of some fixed graph F, whenever a family of vertex sets are pairwise regular "enough" and their densities correspond to the edge-set of F. Several versions of this lemma were previously proved in papers using the regularity lemma (e.g. [4], [16], [25]).

Lemma 3.3. *For every real $0 < \eta < 1$ and integer $f \geq 1$ there exists $\gamma = \gamma_{3.3}(\eta, f)$ with the following property. Suppose that F is a graph on f vertices $v_1, \ldots, v_f$, and that $U_1, \ldots, U_f$ is an f-tuple of disjoint vertex sets of a graph G such that for every $1 \leq i < j \leq f$ the pair (U_i, U_j) is γ-regular. Moreover, suppose that whenever $(v_i, v_j) \in E(F)$ we have $d(U_i, U_j) \geq \eta$, and whenever $(v_i, v_j) \notin E(F)$ we have $d(U_i, U_j) \leq 1 - \eta$. Then, some f-tuple $u_1 \in U_1, \ldots, u_f \in U_f$ spans an* **induced** *copy of F, where each u_i plays the role of v_i.*

Note, that in terms of regularity, Lemma 3.3 requires all the pairs (U_i, U_j) to be γ-regular. However, and this will be very important later in the paper, the requirements in terms of density are not very restrictive. In particular, if $\eta \leq d(U_i, U_j) \leq 1 - \eta$ then we don't care whether (i, j) is an edge of F or not.

A partition $\mathcal{A} = \{V_i \mid 1 \leq i \leq k\}$ of the vertex set of a graph is called an **equipartition** if $|V_i|$ and $|V_j|$ differ by no more than 1 for all $1 \leq i < j \leq k$ (so in particular each V_i has one of two possible sizes).

Our main tool in the proofs is Lemma 3.4 below. A stronger version of this lemma is proved in [3]. This lemma can be considered as a strengthened variant of the standard regularity lemma of Szemerédi [33]. The advantage of this version is obtaining a regular partition in which **all** pairs are regular, where - roughly speaking - we compromise on the densities of the edges sets and consider only an induced subgraph of our graph which represents well the whole graph.

Lemma 3.4. ([3]) *For every integer m and every $\gamma > 0$ there is $T = T_{3.4}(m, \gamma)$ which satisfies the following. Any graph G on $n \geq T$ vertices, has an equipartition $\mathcal{A} = \{V_i \mid 1 \leq i \leq k\}$ of $V(G)$ and an induced subgraph U of G, with an equipartition $\mathcal{B} = \{U_i \mid 1 \leq i \leq k\}$ of the vertices of U, that satisfy:*

1. *$m \leq k \leq T$.*
2. *$U_i \subseteq V_i$ for all $i \geq 1$, and $|U_i| \geq n/T$.*

3. *In the equipartition $\mathcal{B}$,* **all** *pairs are γ-regular.*

4. *All but at most $\gamma\binom{k}{2}$ of the pairs $1 \leq i < j \leq k$ are such that $|d(V_i, V_j) - d(U_i, U_j)| < \gamma$.*

The following is a version of the regularity lemma which applies to edge colored graphs. It asserts that there is an equipartition which is regular with respect to all the colors simultaneously. We denote by $d_c(X, Y)$ the density of the edges of color c between X and Y.

Lemma 3.5. ([25], Multi-Color Regularity Lemma) *For any $\gamma > 0$ and integers m, r, there exists an integer $T = T_{3.5}(m, r, \gamma)$ with the following property: Any graph G on $n \geq T$ vertices, with edges colored by r colors, has an equipartition $\mathcal{A} = \{V_i \mid 1 \leq i \leq k\}$ of $V(G)$ with $m \leq k \leq T$, for which all pairs (V_i, V_j) but at most $\gamma\binom{k}{2}$ of them, satisfy the following regularity condition: for every $X \subseteq V_i$ and $Y \subseteq V_j$ of size $|X|, |Y| \geq \gamma|V_i|$, and every $1 \leq c \leq r$, we have $|d_c(X, Y) - d_c(V_i, V_j)| < \gamma$.*

3.1 Regularity graphs

The following definitions suggest very useful tools for modelling regular partitions of graphs, with respect to induced subgraphs.

Definition 3.6. *Suppose G is a graph, with vertex subsets $\mathcal{A} = \{U_1, \ldots, U_k\}$ and let $\eta > 0$. The* ***cluster graph*** *for the partition $\mathcal{A}$ with respect to η is a complete, labelled, edge colored graph K with $V(K) = \{1, \ldots, k\}$. The color of the edge (i, j) is white if $d(U_i, U_j) < \eta$, black if $d(U_i, U_j) > 1 - \eta$ and otherwise (i, j) is colored grey.*

For a regularity graph K, we denote by $EB(K)$, $EW(K)$ and $EG(K)$ the sets of black, white and grey edges of K respectively. This model relates to induced subgraphs by the following definition.

Definition 3.7. *For an arbitrary simple graph F, we say that a cluster graph K contains a* ***colored copy of F*** *if there is an injective mapping $\varphi : V(F) \mapsto V(K)$, which satisfies the following for every $u, v \in V(F)$:*

1. *If $(u, v) \in E(F)$ then $(\varphi(u), \varphi(v))$ is colored black or grey.*

2. *If $(u, v) \notin E(F)$ then $(\varphi(u), \varphi(v))$ is colored white or grey.*

The above definitions should be considered with Lemma 3.3 in mind, which leads to the following corollary.

Corollary 3.8. *Let $\eta > 0$ and f be some integer. Suppose G is a graph with vertex subsets $\mathcal{A} = \{U_1, \ldots, U_k\}$ such that for any $1 \leq i < j \leq k$, U_i and U_j is a $\gamma = \gamma_{3.3}(\eta, f)$-regular pair. Let K be a cluster graph of $\mathcal{A}$ with respect to η. If K contains a colored copy of a graph F with at most f vertices, then G contains an induced subgraph isomorphic to F.*

3.2 Random graphs

The following lemma states that in a random graph, with high probability, the edge density between and within any two large enough sets of vertices is close to the density of the graph. This lemma will be useful in various places along this paper. The proof is a standard application of Chernoff's inequality.

Lemma 3.9. *Assume $0 \leq p \leq 1$, and $f : \mathbb{N} \to \mathbb{N}$ satisfies $f(n) = \omega(n^{1.5})$. Then for a sufficiently large n, with high probability, $G = G(n,p)$ satisfies*

1. *For any set $A \subseteq V(G)$: $\mid e(A) - p\binom{|A|}{2} \mid \; < f(n)$.*
2. *For any pair of disjoint sets $A, B \subseteq V(G)$: $\mid e(A,B) - p|A||B| \mid \; < f(n)$.*

Proof. Let A be a fixed set of vertices in $G = G(n,p)$. By Chernoff's inequality (see e.g. pp. 266 in [6]):

$$Pr\left[\mid e(A) - p\binom{|A|}{2} \mid \; > f(n)\right] < 2\exp\left\{\frac{-2(f(n))^2}{\binom{|A|}{2}}\right\} < 2\exp\left\{-\frac{(f(n))^2}{n^2}\right\} = e^{-\omega(n)} \; .$$

Similarly, for any disjoint sets A, B

$$Pr\left[\mid e(A,B) - p|A||B| \mid \; > f(n)\right] < 2\exp\left\{\frac{-(f(n))^2}{|A||B|}\right\} < 2\exp\left\{-\frac{(f(n))^2}{n^2}\right\} = e^{-\omega(n)} \; .$$

The probability that *all* such sets A and pairs of sets (A,B) satisfy the conditions of the lemma is therefore at least $1 - 4^n e^{-\omega(n)}$, which tends to 1 as n grows. ■

4 Sparse hereditary properties

The following lemma is useful when the property $\mathcal{P}$ is sparse:

Lemma 4.1. $ed(n,\mathcal{P}) \geq \frac{1}{2}\binom{n}{2} - n\sqrt{\log|\mathcal{P}^n|}$.

Proof. We denote by $\Delta_\ell(G_1,G_2)$ the edit distance between two *labelled* graphs on the same set of vertices. By Chernoff's inequality, for every fixed graph G_0:

$$p = \Pr\left[\Delta_\ell(G(n,1/2),G_0) < \frac{1}{2}\binom{n}{2} - n\sqrt{\log|\mathcal{P}^n|}\;\right] < e^{\frac{-(n\sqrt{\log|\mathcal{P}^n|})^2}{2\binom{n}{2}}} < e^{-\log|\mathcal{P}^n|} = \frac{1}{|\mathcal{P}^n|} \; ,$$

and hence, by the union bound for all the graphs in $\mathcal{P}^n$,

$$\Pr\left[E_{\mathcal{P}}(G = G(n,1/2)) \geq \frac{1}{2}\binom{n}{2} - n\sqrt{\log|\mathcal{P}^n|}\;\right] \geq 1 - |\mathcal{P}^n|p > 0$$

which shows that indeed such a graph exists and completes the proof. ■

Proof of Theorem 2.1:

We first make the following observations on (infinite) hereditary properties. Denote the empty graph on n vertices by E_n. By Ramsey Theorem, at least one of K_n and E_n belongs to $\mathcal{P}^n$ for infinitely many n. Since $\mathcal{P}$ is hereditary, it in fact belongs to $\mathcal{P}^n$ for *every* n. We may assume, without loss of generality, that $E_n \in \mathcal{P}^n$. Otherwise, we proceed with $\mathcal{P}' = \{\overline{G} \ : \ G \in \mathcal{P}\}$ since clearly $p(\mathcal{P}) = 1 - p(\mathcal{P}')$ and $ed(n, \mathcal{P}) = ed(n, \mathcal{P}')$.

Hence $E_n \in \mathcal{P}^n$ for every n. Further assume the following: for any $\varepsilon > 0$ there is n_ε such that

$$\forall n \geq n_\varepsilon, \ G \in \mathcal{P}^n : \quad e(G) \leq \varepsilon \binom{n}{2} .$$

In this case, since all the graphs in $\mathcal{P}$ have very few edges $ed(n, \mathcal{P}) \geq E_{\mathcal{P}}(K_n) = (1 - o(1))\binom{n}{2}$. Hence indeed such $\mathcal{P}$ satisfies $p(\mathcal{P}) = 1$ which proves case 1 in Theorem 2.1, and the analogous case 2 for $\mathcal{P}'$.

Otherwise, there is some $\varepsilon > 0$ and an infinite sequence of graphs $G_1, G_2, \ldots \in \mathcal{P}$ such that $e(G_i) > \varepsilon\binom{|V(G_i)|}{2}$. Again, since $\mathcal{P}$ is hereditary, by a successive removal of vertices with the lowest degree from G_i, we get

$$\forall n \geq 2 \quad \exists G \in \mathcal{P}^n : \quad e(G) > \varepsilon \binom{n}{2} . \tag{3}$$

In order to complete the proof of Theorem 2.1, we show that in this case the edit distance is upper bounded by $\frac{1}{2}\binom{n}{2}$ (Lemma 4.3 below). We will use the following simple fact:

Fact 4.2. *Suppose G_1, G_2 are graphs on n vertices, with edge densities d_1, d_2 respectively. Then $\Delta(G_1, G_2) \leq (d_1(1 - d_2) + (1 - d_1)d_2)\binom{n}{2}$.*

Proof. Consider a random labelling of the vertices of G_1 and G_2, and modify the edges and non-edges of G_1 that differ from G_2. For any pair of vertices (u, v), the probability that the edge (u, v) has to be modified is $(d_1(1 - d_2) + (1 - d_1)d_2)$. Therefore, the expected number of modifications is $(d_1(1 - d_2) + (1 - d_1)d_2)\binom{n}{2}$. Hence, some labelling of G_1 witnesses that indeed $\Delta(G_1, G_2) \leq (d_1(1 - d_2) + (1 - d_1)d_2)\binom{n}{2}$. ■

Lemma 4.3. *Let $\mathcal{P}$ be a hereditary property such that the empty graph on n vertices belongs to $\mathcal{P}$ for every n, and there is some $\varepsilon > 0$ satisfying (3). Then $ed(n, \mathcal{P}) \leq \frac{1}{2}\binom{n}{2}$.*

Proof. For any n and $\varepsilon > 0$, there is some m such that any graph G' on m vertices with $e(G') > \varepsilon\binom{m}{2}$ contains a complete bipartite graph $K_{\lceil \frac{n}{2} \rceil, \lfloor \frac{n}{2} \rfloor}$as a weak subgraph (see, e.g., Chapter 6 in [13]). Let $G' \in \mathcal{P}$ be the graph on m vertices with edge density at least ε which is guaranteed by (3). Since $\mathcal{P}$ is hereditary, the subgraph of G' which is induced by the n vertices of that $K_{\lceil \frac{n}{2} \rceil, \lfloor \frac{n}{2} \rfloor}$ also belongs to $\mathcal{P}$. Call it G_1, where the edge density d_1 of G_1 is at least $\frac{1}{2}$. Hence, G_1 and E_n both belong to $\mathcal{P}^n$. For any graph G on n vertices with edge density d, it now follows from Fact 4.2 that

$$E_{\mathcal{P}}(G) \leq min\{\Delta(G, G_1), \Delta(G, E_n)\} \leq min\{d(1 - d_1) + (1 - d)d_1, d\} \binom{n}{2} .$$

However, since $d_1 \geq \frac{1}{2}$, the function $f(d) = d(1-d_1) + (1-d)d_1 = (1-2d_1)d + d_1$ is monotone non-increasing. Hence, if $d \geq \frac{1}{2}$ then $f(d) \leq f(\frac{1}{2}) = \frac{1}{2}$, which shows that

$$\min_{0 \leq d \leq 1} \{d(1-d_1) + (1-d)d_1, d\} \leq \frac{1}{2} ,$$

thus completing the proof.

■

Lemma 4.3, together with Lemma 4.1, completes the proof of Theorem 2.1.

■

Proof of Lemma 2.2:

We make a "wasteful" counting which lets us deduce the asymptotic, making no attempt to optimize the constants. We start with an upper bound, which we derive separately for each of the three properties:

Cographs: Let us observe the following set of strings. We write a permutation of the numbers $\{1, \ldots, n\}$ in some order, between every two numbers we write one of two possible signs (namely either U or C) and put n pairs of parenthesis arbitrarily. We interpret those expressions as a recursive construction of a labelled cograph on n vertices, where U stands for taking the disjoint union, and C represents taking the complement of the disjoint union. Thus every such string represents at most one cograph, and some of the strings are "illegal". On the other hand, every labelled cograph is represented by at least one of those strings. There are less than $n!2^{n-1}3^{2n}$ such strings, which accumulates to $2^{O(n \log n)}$.

Interval graphs: Consider the set of intervals which correspond to the graph vertices. The edge set of the graph is uniquely determined by a sorting of the multiset $\{1, 1, 2, 2, \ldots, n, n\}$ which represents a sorting of the endpoints of the (labelled) intervals. There are $\frac{(2n)!}{2^n} = 2^{O(n \log n)}$ ways to sort this multiset, which thus bounds the number of labelled interval graphs.

Permutation graphs: Clearly each permutation graph on n vertices is represented by a single permutation on n elements, and hence again there are $2^{O(n \log n)}$ such graphs.

We now establish a lower bound for the three properties. Note that every graph consisting of a disjoint union of complete graphs is a cograph, an interval graph and a permutation graph. The number of such labelled graphs is exactly the n'th Bell number B_n, which satisfies $\frac{\ln(B_n)}{n \ln n} = 1 - o(1)$. Hence, we also have that $|\mathcal{P}^n| \geq 2^{\Omega(n \log n)}$ for each of those classes.

■

Corollary 2.3 immediately follows from Lemma 4.1 and Lemma 2.2.

Proof of Lemma 2.5:

To prove the upper bound, we first show that for any $n > 0$ there are graphs $G_1, G_2 \in \mathcal{P}^n_{\mathcal{Q},b}$ with edge densities $d(G_1) > \frac{1}{2}$ and $d(G_2) < \frac{1}{2}$. Since $(\mathcal{Q}, b)$ is non-trivial, there is a pair of points $\underline{c}_1, \underline{c}_2$ that satisfies $(\mathcal{Q}, b)$. We consider a graph $G_1 \in \mathcal{P}_{\mathcal{Q},b}$ which is defined by the n-tuple of points in which $\lceil \frac{n}{2} \rceil$ appearances of $\underline{c}_1$ are followed by $\lfloor \frac{n}{2} \rfloor$ appearances of $\underline{c}_2$. Thus, G_1 has at least $\lceil \frac{n}{2} \rceil \lfloor \frac{n}{2} \rfloor$ edges, and $d(G_1) > \frac{1}{2}$. On the other hand, by choosing a pair of points that does not satisfy $(\mathcal{Q}, b)$ we may construct a graph G_2 similarly. By Fact 4.2, for any graph G on n vertices, if $d(G) \geq \frac{1}{2}$ then $\Delta(G, G_1) \leq \frac{1}{2}\binom{n}{2}$, and if $d(G) \leq \frac{1}{2}$ then $\Delta(G, G_2) \leq \frac{1}{2}\binom{n}{2}$. Hence, any graph G on n vertices has $\min(\Delta(G, G_1), \Delta(G, G_2)) \leq \frac{1}{2}\binom{n}{2}$ and therefore $ed(n, \mathcal{P}_{\mathcal{Q},b}) \leq \frac{1}{2}\binom{n}{2}$.

To prove the lower bound, we use the fact that the number of labelled graphs in $\mathcal{P}_{\mathcal{Q},b}$ is at most the number of possible sign patterns for the polynomials $\mathcal{Q}$. Formally, a sign pattern of a set of m polynomials $Q_1, \ldots, Q_m$ in ℓ variables, is an m-tuple $\underline{s} \in \{-1, 0, 1\}^m$ such that for some $\underline{c} \in \mathbb{R}^\ell$: $s_i = sign(Q_i(\underline{c}))$ for every $1 \leq i \leq m$. The sign patterns of the polynomials is the set of all such vectors $\underline{s}$. The following claim was proved by Warren [35] for vectors $\underline{s} \in \{-1, 1\}^m$. In [2] it is observed that this can be extended to vectors $\underline{s} \in \{-1, 0, 1\}^m$ as well.

Claim 4.4. (Warren [35], see also [2]) *Let $Q_1, \ldots, Q_m$ be m real polynomials in ℓ real variables, and suppose the total degree of the Q_is is $D = \sum_{i=1}^m deg(Q_i)$. If $m \geq \ell$ then the number of all possible sign patterns for the Q_is is at most $(8eD/\ell)^\ell$.*

Thus, since in our case $m = \binom{n}{2}$, $\ell = dn$ and $D = O(n^2) < cn^2$ (for some constant $c = c(\mathcal{Q})$), we get that the number of possible sign patterns is at most

$$\left(\frac{8ecn^2}{dn}\right)^{dn} = 2^{O(n \log n)} .$$

and therefore $|\mathcal{P}^n_{\mathcal{Q},b}| \leq 2^{O(n \log n)}$. The lemma now follows from Lemma 4.1. ■

Example 4.5. *Many hereditary families of intersection graphs may be defined by polynomials. For instance, define the class of intersection graphs of balls in $\mathbb{R}^d$ as follows. We represent a ball in $\mathbb{R}^d$ by the $(d+1)$-tuple consisting of the coordinates of its center followed by its radius. Define a single polynomial in $2(d+1)$ variables*

$$Q_1(x_1, \ldots, x_d, r_x, y_1, \ldots, y_d, r_y) = \sum_{i=1}^{d} (x_i - y_i)^2 - (r_x + r_y)^2 ,$$

and b gives value 1 if the polynomial has negative *value, and 0 otherwise. Thus, indeed, two $(d+1)$-tuples satisfy $(\mathcal{Q}, b)$ if and only if the balls they represent intersect.*

Note that this observation (for $d = 1$), together with Lemma 2.5, gives an alternative proof for the case of Interval graphs in Lemma 2.2 and Corollary 2.3.

5 Complement invariant properties

As in [7], we define for any graph property $\mathcal{P}$, $n > 0$ and $p \in [0,1]$,

$$\boldsymbol{e_{n,p}(\mathcal{P})} = \frac{\mathbb{E}[E_{\mathcal{P}}(G(n,p))]}{\binom{n}{2}} .$$

In words, this is the expected fraction of the edges that need to be modified in $G(n,p)$ in order to obtain a graph in $\mathcal{P}$. When the context is clear, we write $e_{n,p}$ for $e_{n,p}(\mathcal{P})$. It is shown in [7] that for any hereditary graph property, the sequence $\{e_{n,p}\}_{n=1}^{\infty}$ is monotone and thus has a limit denoted e_p.

The proof of Theorem 1.1 in [7] shows that for any hereditary property, the edit distance of a random graph $G = G(n,p)$ from $\mathcal{P}$ is obtained roughly by modifying G into a graph conforming to some colored cluster graph F . That is, a graph that has a regular partition whose cluster graph is F. Thus, for a given p, $e_p = \lim_{n\to\infty} e_{n,p}$ is the minimum of the distance from an infinite set of colored cluster graphs[3]. Yet, up to $o(1)$, the expected normalized (i.e. divided by $\binom{n}{2}$) edit distance of $G(n,p)$ from conforming to each such cluster graph F is a linear function in p. This function only depends on the number of white and black edges and vertices in F. Hence, in fact, e_p is the minimum of an infinite set of linear functions. Any such function is concave.

Thus, omitting the details which are identical to those of the proof of Theorem 1.1, we conclude the following by the above discussion.

Theorem 5.1. *Let $\mathcal{P}$ be an arbitrary hereditary property, and let e_p be as defined above. Then e_p is a concave function of p in the segment $[0,1]$.*

Corollary 5.2. *For any hereditary property, condition 1 of Theorem 1.1 is satisfied by every p along a single subsegment of $[0,1]$ (which may be degenerated to a single point).*

Example 5.3. *Consider the property of being either an empty or a complete graph. $G = G(n,p)$ is turned into a graph satisfying this property by either removing all the edges from G, or adding all the non-edges to G. Thus, e_p is the minimum of the linear functions p and $1-p$, and the maximum of e_p is attained when $p = \frac{1}{2}$.*

Example 5.4. *Let $\mathcal{P}$ be the hereditary property of being a complete bipartite graph. In this case, one could turn a graph $G(n,p)$ into a complete bipartite graph by splitting its vertices into two sets of sizes λn and $(1-\lambda)n$ where $0 \le \lambda \le \frac{1}{2}$, and remove all the edges inside each of those sets while adding every non-edge between them. The expected normalized number of changes is $p(\lambda^2 + (1-\lambda)^2) + 2(1-p)\lambda(1-\lambda) = p(4\lambda^2 - 4\lambda + 1) + 2(\lambda - \lambda^2)$, and taking the minimum over λ for every p we get that*

[3]We use slightly different regularity graphs in [7]. There, on top of the edges, the vertices are also colored white or black. This represents either empty or complete graphs that are spanned by the partition clusters.

$$e_p = \begin{cases} p & p \le \frac{1}{2} \\ \frac{1}{2} & p \ge \frac{1}{2} \end{cases} .$$

Hence the maximum of e_p is attained along the segment $[\frac{1}{2}, 1]$.

Proof of Theorem 2.6:

Clearly, if p is an extremal density for $\mathcal{P}$, then so is $1-p$, since for any graph G: $E_{\mathcal{P}}(G) = E_{\mathcal{P}}(\overline{G})$. Thus, by Corollary 5.2 the maximum is also attained at $p = \frac{1}{2}$. ■

6 (r,s)-colorability and $\mathcal{P}_{r,s}$

We prove the lower and the upper bounds of Theorem 2.8 separately.

Lemma 6.1. $ed(n, \mathcal{P}_{r,s}) \le \frac{1}{(\sqrt{r}+\sqrt{s})^2}\binom{n}{2}$.

Proof. We prove that for any graph $G = (V, E)$ on n vertices, $E_{\mathcal{P}_{r,s}}(G) \le \frac{1}{(\sqrt{r}+\sqrt{s})^2}\binom{n}{2}$.

Let d denote the density of G, that is $d = e(G)/\binom{n}{2}$. Consider a random partition of the vertices of G into $r+s$ subsets $I_1, I_2, \ldots, I_r, C_1, C_2, \ldots, C_s$ as follows. Each vertex $v \in V$ chooses its set independently, with the following distribution:

$$\begin{aligned} Pr[v \in I_k] &= \frac{1-d}{r(1-d)+sd} \quad k = 1, \ldots, r \ , \\ Pr[v \in C_j] &= \frac{d}{r(1-d)+sd} \quad j = 1, \ldots, s \ . \end{aligned} \tag{4}$$

In what follows we calculate the expected number of changes one should apply to G in order to make it (r,s)-colorable by turning every I_k into an independent set, and every C_j into a clique. For any pair of vertices (u, v) in G, the edge between them needs to be modified if either

- for some $1 \le k \le r$, u and v belong to I_k and $(u, v) \in E(G)$, or
- for some $1 \le j \le s$, u and v belong to C_j and $(u, v) \notin E(G)$

Hence the expected number of modifications is

$$\begin{aligned} \mathbb{E}[\#changes] &= \left(r\left(\frac{(1-d)}{r(1-d)+sd}\right)^2 d + s\left(\frac{d}{r(1-d)+sd}\right)^2 (1-d) \right)\binom{n}{2} \\ &= \frac{r(1-d)^2 d + sd^2(1-d)}{(r(1-d)+sd)^2}\binom{n}{2} \\ &= \frac{d(1-d)}{r(1-d)+sd}\binom{n}{2} . \end{aligned}$$

Let $f(d) = \frac{d(1-d)}{r(1-d)+sd}\binom{n}{2}$. By differentiating f, we obtain the extremum value of d:

$$d_{r,s} = \frac{\sqrt{r}}{\sqrt{r}+\sqrt{s}} \ .$$

This is the "worst" density for the parameters (r,s): a random partition of a graph with density $d_{r,s}$ is expected to require the largest number of edge modifications (for the specific distribution of $|I_k|, |C_j|$ we chose). For a graph G with density $d_{r,s}$ the expected number of edge modifications is

$$\begin{aligned}
\mathbb{E}[\#changes] &\leq f(d_{r,s}) \\
&= \frac{d_{r,s}(1-d_{r,s})}{r(1-d_{r,s})+sd_{r,s}}\binom{n}{2} \\
&= \left(\frac{\frac{\sqrt{r}\sqrt{s}}{(\sqrt{r}+\sqrt{s})^2}}{r\frac{\sqrt{s}}{\sqrt{r}+\sqrt{s}}+s\frac{\sqrt{r}}{\sqrt{r}+\sqrt{s}}}\right)\binom{n}{2} \\
&= \frac{1}{(\sqrt{r}+\sqrt{s})^2}\binom{n}{2} \ .
\end{aligned}$$

Hence, we showed that every graph on n vertices can be (r,s)-colored such that the number of edges and non-edges violating the coloring is at most $\frac{1}{(\sqrt{r}+\sqrt{s})^2}\binom{n}{2}$, which indeed proves that $ed(n,\mathcal{P}_{r,s}) \leq \frac{1}{(\sqrt{r}+\sqrt{s})^2}\binom{n}{2}$.

■

Lemma 6.2. $ed(n,\mathcal{P}_{r,s}) \geq (\frac{1}{(\sqrt{r}+\sqrt{s})^2} - o(1))\binom{n}{2}$.

Proof. Define $d_{r,s}$ as in the proof of Lemma 6.1. We show that with high probability $G = G(n,d_{r,s})$ has edit distance $(\frac{1}{(\sqrt{r}+\sqrt{s})^2} - o(1))\binom{n}{2}$ from $\mathcal{P}_{r,s}$. By applying Lemma 3.9 to G, with $f(n) = n^{1.6}$, with high probability, any vertex set $A \subseteq V(G)$ satisfies $\mid e(A) - p\binom{|A|}{2} \mid \ < n^{1.6}$.

Let $\hat{G} \in \mathcal{P}_{r,s}$ be the closest graph to G. Consider the (r,s)-coloring of $\hat{G}$ witnessing its membership in $\mathcal{P}_{r,s}$. Denote the sizes of the vertex classes in that coloring by $i_1,\ldots,i_r,c_1,\ldots,c_s$, where

$$\sum_{k=1}^{r} i_k + \sum_{k=1}^{s} c_k = n \tag{5}$$

It now follows that

$$E_{\mathcal{P}_{r,s}}(G) = \Delta(G,\hat{G}) = d_{r,s}\sum_{k=1}^{r}\binom{i_k}{2} + (1-d_{r,s})\sum_{k=1}^{s}\binom{c_k}{2} + O(n^{1.6}) \tag{6}$$

By convexity, all the quantities i_k are equal to each other, and so are all the quantities c_k. Therefore, the right hand side of (6), subject to (5), becomes a quadratic polynomial in one variable $x = i_k$. Optimizing it, we conclude that the expression (6) is minimized under the constraint (5) when the sizes of the color classes are distributed as in (4). Hence it also follows that with high probability G is at least $(\frac{1}{(\sqrt{r}+\sqrt{s})^2} - O(\frac{1}{n^{0.4}}))\binom{n}{2}$ far from $\mathcal{P}_{r,s}$.

■

Remark 6.3. *Note that the proof of Lemma 6.2 also shows that* $p_{1.1}(\mathcal{P}_{r,s}) = d_{r,s}$.

Remark 6.4. *A result similar to Lemma 6.1 is implicitly proved in [9], where it is used for proving the upper bound of (2) on the edit distance from* $\mathcal{P}_H^*$ *for appropriate graphs* H.

Remark 6.5. *As pointed out by Bollobás and Thomason ([15], [17]), for some parameters of graph properties the approximation achieved by* (r,s)*-colorability is not strong enough. In particular, for some hereditary properties, the probability* $\Pr[G(n,p) \in \mathcal{P}]$ *cannot be approximated by* $\Pr[G(n,p) \in \mathcal{P}_{r,s}]$ *for any* $\mathcal{P}_{r,s}$ *(when* $p \neq \frac{1}{2}$*). In [17] they describe a larger family of properties called* ***basic properties****, each of them defined by a colored graph as follows.*

Let T *be a complete colored labelled graph where* $V(T) = [t]$*. Each vertex is colored either white or black, and each edge is colored white, grey or black. The property* $\mathcal{P}_T$ *consists of the graphs* G *such that* $V(G)$ *can be partitioned into* t *sets,* $V_1, \ldots, V_t$*, where each such set corresponds to a vertex of* T*. This partition witnesses membership of* G *in* $\mathcal{P}_T$ *if*

- *For every* $1 \leq i \leq t$*, if the color of vertex* i *in* T *is black (white), then* V_i *spans a complete (empty) graph in* G*.*
- *For any* $1 \leq i < j \leq t$*, if the color of the edge* (i,j) *is black (white) then* (V_i, V_j) *span a complete (empty) bipartite graph in* G*. If the color of* (i,j) *is grey, then there is no restriction on* $E(V_i, V_j)$*.*

It is possible to show that for such properties, and any value of $0 \leq p \leq 1$*, the edit distance* $E_{\mathcal{P}_T}(G(n,p))$ *can be derived from* T*. In this case, the sizes of the vertex sets in the partition of* $V(G)$ *are chosen as to minimize a quadratic form, which depends on the (symmetric) colors matrix of* T*.*

7 The edit distance of $G(n, 1/2)$

In this section we prove Theorem 2.10. We will use the classical theorem of Erdős and Stone [23]. Denote by $K_t(\ell)$ the complete t-partite graph with ℓ vertices in each part.

Theorem 7.1. ([23]) *Let* $t \geq 2$*,* $\ell > 0$ *and* $\varepsilon > 0$*. There is* $n_{7.1}(t, \ell, \varepsilon)$*, such that any graph* G *on* $n > n_{7.1}(t, \ell, \varepsilon)$ *vertices that contains at least* $(1 - \frac{1}{t-1} + \varepsilon)\binom{n}{2}$ *edges, contains a (not necessarily induced) copy of* $K_t(\ell)$ *as a subgraph.*

Proof of Theorem 2.10:

Let $\mathcal{P}$ be a hereditary property, and $t = \chi_B(\mathcal{P})$ as previously defined. For a graph $G = G(n, 1/2)$, we assume that the assertions of Lemma 3.9 hold for $f(n) = n^{1.6}$.

An upper bound is obtained as follows: By the definition of $\chi_B(\mathcal{P})$, there are some r and s such that $r + s = t - 1$ and $\mathcal{P}_{r,s} \subseteq \mathcal{P}$. We split the vertices of G into $t-1$ equal sized sets arbitrarily, and

turn r of those sets into independent sets and the remaining s into cliques. We thus obtain a graph in $\mathcal{P}$ by changing $\frac{1}{2(t-1)}\binom{n}{2} + O(n^{1.6})$ edges in G. Hence, w.h.p. $E_{\mathcal{P}}(G) \leq (\frac{1}{2(\chi_B(\mathcal{P})-1)} + o(1))\binom{n}{2}$.

For the lower bound, we first make the following definition:

$$h = \max_{r+s=t} \min\{|V(H)| \ : \ H \in (\mathcal{P}_{r,s} \setminus \mathcal{P})\} \ .$$

Thus, for any pair (r, s) such that $r + s \geq t$, there is some graph H on at most h vertices that witnesses the fact that $\mathcal{P}_{r,s} \nsubseteq \mathcal{P}$.

We first sketch the proof of the lower bound, and then complete the missing details. Let ε be an arbitrarily small positive constant. We consider a regular partition of $\hat{G}$, the closest graph to G in $\mathcal{P}$. We obtain an induced subgraph U of $\hat{G}$ and an equipartition $\mathcal{B}$ of U's vertices into k clusters through Lemma 3.4. Let F be a cluster graph for $\mathcal{B}$. We will prove that F contains at least $(\frac{1}{t-1} - \frac{\varepsilon}{4})\binom{k}{2}$ white or black edges. Each such edge in F was achieved by applying roughly $\frac{1}{2}(\frac{n}{k})^2$ edge changes to G, which derives the lower bound, as $\mathcal{B}$ is similar to the equipartition of the whole graph.

Thus, we assume towards a contradiction that F contains at least $(1 - \frac{1}{t-1} + \frac{\varepsilon}{4})\binom{k}{2}$ grey edges. We aim to show that in this case F contains a colored copy of a forbidden graph H which will then lead to the contradiction. Focusing only on the grey edges in F, Theorem 7.1 implies that F contains a copy of a complete t-partite graph K, with 4^h vertices in each part, in which all the edges between vertices in different parts are grey. For the moment, consider the grey edges inside K's parts as if they were white. By applying the symmetric Ramsey Theorem, each one of the t parts of K contains a clique on h vertices, consisting of either white or black edges in F. This implies that for some r and s such that $r + s = t$, F contains an induced t-partite subgraph, where each part consists of h vertices, every edge between vertices in different parts is grey, and r of the parts induce a white clique while the other s induce a black clique. Therefore, by the definition of h and t, there is a colored copy of a graph $H \in (\mathcal{P}_{r,s} \setminus \mathcal{P})$ in F, which yields an induced copy of H in $\hat{G}$ by Corollary 3.8. This contradicts the assumption that $\hat{G} \in \mathcal{P}$. Note that for obtaining a colored copy of H, it is indeed possible to consider the grey edges inside the parts as if they were white.

We now complete the missing details, which enable the above discussion. Given ε, we set $\eta = \frac{\varepsilon}{4}$. In order to make the above possible, we need the cluster graph F to be of size at least $n_{7.1}(t, 4^h, \frac{\varepsilon}{4})$. We thus set $m = n_{7.1}(t, 4^h, \frac{\varepsilon}{4})$ and $\gamma = \min\{\frac{\varepsilon}{4}, \gamma_{3.3}(\eta, h)\}$. We assume $n > T_{3.4}(m, \gamma)$ and apply Lemma 3.4 to $\hat{G}$ with m and γ. Thus we obtain an equipartition $\mathcal{A} = \{V_i \mid 1 \leq i \leq k\}$ of $V(\hat{G})$ and an induced subgraph U of $\hat{G}$, with an equipartition $\mathcal{B} = \{U_i \mid 1 \leq i \leq k\}$. We construct a cluster graph F for the equipartition $\mathcal{B}$ with respect to η, and make the following observations.

1. By Lemma 3.4 all but at most $\frac{\varepsilon}{4}\binom{k}{2}$ pairs $1 \leq i < j \leq k$ satisfy $|d_{\hat{G}}(V_i, V_j) - d_{\hat{G}}(U_i, U_j)| < \frac{\varepsilon}{4}$.
2. If (i, j) is either a black or a white edge in F, then either $d_{\hat{G}}(U_i, U_j) > 1-\eta$ or $d_{\hat{G}}(U_i, U_j) < \eta$ respectively.

3. By Lemma 3.9, for a sufficiently large n, $|d_G(V_i, V_j) - \frac{1}{2}| < \frac{\varepsilon}{4}$.

For any pair $1 \leq i < j \leq k$ satisfying condition (1), if (i, j) is not a grey edge of F, then there were at least $(\frac{1}{2} - \frac{3\varepsilon}{4})\frac{n^2}{k^2}$ modifications in $E(V_i, V_j)$. Hence, if there are at least $(\frac{1}{t-1} - \frac{\varepsilon}{4})\binom{k}{2}$ black or white edges in F, then at least $(\frac{1}{t-1} - \frac{\varepsilon}{2})\binom{k}{2}$ of them satisfy (1) above. For a sufficiently large n, and since $t \geq 2$:

$$E_{\mathcal{P}}(G) = \Delta(G, \hat{G}) \geq \left(\frac{1}{t-1} - \frac{\varepsilon}{2}\right)\binom{k}{2}\left(\frac{1}{2} - \frac{3\varepsilon}{4}\right)\frac{n^2}{k^2} > \left(\frac{1}{2(t-1)} - \varepsilon\right)\binom{n}{2}.$$

Otherwise, there are at least $(1 - \frac{1}{t-1} + \frac{\varepsilon}{4})\binom{k}{2}$ grey edges in F, which yields a contradiction by the above discussion.

■

Remark 7.2. *The proof resembles proofs of [29] and [16], which also apply some versions of Erdős-Stone Theorem and Ramsey Theorem. It is used there to show some results on the speed of hereditary properties. The treatment here is somewhat simpler, because of the application of the stronger Lemma 3.4 instead of using the standard regularity lemma.*

8 Exact asymptotic for $\mathcal{P}^*_H$ for small graphs

8.1 The edit distance from being claw free

We begin the proof of Theorem 2.12 by showing an upper bound for any graph.

Lemma 8.1. *For any graph G on n vertices, $E_{\mathcal{P}^*_{K_{1,3}}}(G) \leq \frac{1}{3}\binom{n}{2}$.*

Proof. Recall that if a graph is either $(1, 0)$-colorable or $(0, 2)$-colorable then it is claw free. Let G be a graph on n vertices with edge density $d = |E(G)|/\binom{n}{2}$. If $d \leq \frac{1}{3}$, we remove all the edges of G and thus turn it into a claw free graph, changing at most $\frac{1}{3}\binom{n}{2}$ edges this way. Otherwise, we randomly split its vertices into two equal size sets. We add all the missing edges inside each set, and hence again turn it into a claw free $(0, 2)$-colorable graph. For each $1 \leq i < j \leq n$, the indicator of adding the edge (i, j) to the graph has value 1 with probability less than $\frac{1}{2}(1 - d) \leq \frac{1}{3}$. Hence, the expected number of edges added is at most $\frac{1}{3}\binom{n}{2}$, which completes the proof.

■

We now turn to the more challenging part of the proof of Theorem 2.12, in which we prove that $G(n, \frac{1}{3})$ is far from being claw free:

Lemma 8.2. *With high probability, $E_{\mathcal{P}^*_{K_{1,3}}}(G = G(n, \frac{1}{3})) \geq (\frac{1}{3} - o(1))\binom{n}{2}$.*

Proof. We first describe an overview of the proof. Let $G = G(n, \frac{1}{3})$, and $\hat{G} \in \mathcal{P}^*_{K_{1,3}}$ be the closest claw free graph to G. By applying the strengthened regularity lemma, Lemma 3.4, to $\hat{G}$ we obtain an induced subgraph and an equipartition $\{U_1, \ldots, U_k\}$. This partition defines a cluster graph F on k vertices. Our choice of parameters will let us deduce by Corollary 3.8 that F does not contain a colored copy of $K_{1,3}$. We then make some observations on the structure of F. The main results follow some assertions on cycles of length four which consist of grey edges in F. These basic observations are followed by an application of the multi-colored regularity lemma to the cluster graph F, to obtain even stronger results on F. These results, which are the core of the proof of the theorem, are then translated to several constraints on the number of edges of each color in F. Roughly, we will show that $2|EB(F)| + |EW(F)| \approx \binom{k}{2}$. In words, this last equations actually shows that for each grey edge in F, one must "pay" the price of a black edge. Counting the edge modifications that result from each black and white edge in F will show that w.h.p. the number of changes applied to G in order to obtain $\hat{G}$ is at least $(\frac{1}{3} - o(1))\binom{n}{2}$.

We now turn to the detailed proof. We shall prove that for any $\varepsilon > 0$, for a large enough n (depending on ε), with high probability, $\Delta(G, \hat{G}) \geq (\frac{1}{3} - \varepsilon)\binom{n}{2}$. Define $\eta = \frac{\varepsilon}{5}$, and $\gamma = min\{\frac{\varepsilon}{5}, \gamma_{3.3}(\eta, 4)\}$. Also let $m = max\{T_{3.5}(\frac{25}{\varepsilon}, 3, \frac{\varepsilon^2}{50}) \cdot \frac{10}{\varepsilon}, \frac{20}{\varepsilon}\}$. We assume $n > T_{3.4}(m, \gamma)$, and apply Lemma 3.4 to the graph $\hat{G}$ with the parameters γ and m as defined above. We thus obtain a regular partition $\mathcal{A} = \{V_i \mid 1 \leq i \leq k\}$ of $V(\hat{G})$ and an induced subgraph U of $\hat{G}$, with an equipartition $\mathcal{B} = \{U_i \mid 1 \leq i \leq k\}$. We construct a cluster graph F for $\mathcal{B}$ with respect to η. Note that since $\hat{G}$ does not contain any induced copy of $K_{1,3}$, then by Corollary 3.8, F does not contain a colored copy of $K_{1,3}$.

In the first stage of the proof we make some observations on the cycles of length four (C_4) consisting of grey edges in F. In what follows, it will be convenient to denote the vertices of $K_{1,3}$ by $V(K_{1,3}) = \{h_1, h_2, h_3, h_4\}$ where h_1 is connected to all the others, i.e. $E(K_{1,3}) = \{(h_1, h_2), (h_1, h_3), (h_1, h_4)\}$. For a C_4 which consists of the edges $\{(w, x), (x, y), (y, z), (z, w)\}$, we refer to the pairs (w, y) and (x, z) as the *middle edges* of that C_4.

Proposition 8.3. *Assume $\{x, y, z, w\}$ form a cycle of length four in F, such that all the edges $(x, y), (y, z), (z, w), (w, x)$ are grey. Then either the middle edges $(x, z), (y, w)$ are both white or they are both black.*

Proof. Assume, towards a contradiction and w.l.o.g., that (x, z) is white and (y, w) is black. Then, by setting $\varphi(h_1) = y, \varphi(h_2) = x, \varphi(h_3) = z, \varphi(h_4) = w$, we obtain a colored copy of $K_{1,3}$ in the graph spanned by $\{x, y, z, w\}$ in F. If at least one of the middle edges is grey, then no matter what the color of the other middle edge is, a proper contradicting colored copy can be found, since a grey edge can play the role of both black and white edges. We are thus left with the possibilities that either both middle edges are black or they are both white.

■

Proposition 8.4. *Assume $\{w, x, y, z\}$ form a cycle of length four in F, such that all the edges*

$(w,x),(x,y),(y,z),(z,w)$ *are grey, and both middle edges* $(x,z),(w,y)$ *are white. Then for any other vertex* $t \in V(F) \setminus \{w,x,y,z\}$*, none of its edges to* $\{w,x,y,z\}$ *is grey.*

Proof. Assume, towards a contradiction, that (t,x) is grey in F. If both edges (t,y) and (t,w) are either white or grey, then $\varphi(h_1) = x, \varphi(h_2) = t, \varphi(h_3) = y, \varphi(h_4) = w$ defines a colored copy of $K_{1,3}$ in F. Hence, at least one of these two edges must be black. W.l.o.g., assume (t,w) is black. We now consider two possibilities for the color of the edge (t,z). If this edge is either white or grey, we set $\varphi(h_1) = w, \varphi(h_2) = t, \varphi(h_3) = x, \varphi(h_4) = z$. If (t,z) is black, we set $\varphi(h_1) = t, \varphi(h_2) = x, \varphi(h_3) = z, \varphi(h_4) = w$. In both cases we obtain a colored copy of $K_{1,3}$ in F, which completes the proof.

■

Claim 8.5. $2|EB(F)| + |EW(F)| > (\frac{1}{2} - \frac{\varepsilon}{5})k^2$.

Proof. Let F_1 be an induced subgraph of F, that consists of all the grey C_4's in F, for which both middle edges are white. We denote the graph spanned by the remaining vertices in F by F_2. Also denote by $k_1 = |V(F_1)|$ the number of vertices in F_1, and by $k_2 = |V(F_2)|$, hence $k = k_1 + k_2$. By Proposition 8.3, in any grey C_4 in F_2 both middle edges are black. Moreover, by Proposition 8.4, the grey edges of F_1 form a vertex disjoint collection of C_4s, and none of the edges connecting vertices from F_1 and F_2 is grey. Therefore, in F there are exactly $|V(F_1)|$ grey edges that touch some vertex in $V(F_1)$.

We first take care of the case where F_2 is not large enough for applying the regularity lemma, that is $k_2 < T_{3.5}(\frac{25}{\varepsilon}, 3, \frac{\varepsilon^2}{50})$. In this case, by our choice of m for Lemma 3.4, and since $k \geq m \geq T_{3.5}(\frac{25}{\varepsilon}, 3, \frac{\varepsilon^2}{50}) \cdot \frac{10}{\varepsilon}$, we get that $k_2 < \frac{\varepsilon}{10}k$ and hence $|EG(F)| < k_1 + kk_2 < \frac{\varepsilon}{5}k^2$ which implies the claim.

Otherwise, we focus on the graph F_2. Intuitively, in F_2, any grey C_4 forces the existence of some black edges. Nevertheless, before we can formulate the tradeoff between the number of edges of each color, we need to make one stronger observation on the structure of F_2. This will be achieved by applying the regularity lemma on F_2. Since F_2 is a colored graph, it is necessary to use the multi-color version of the regularity lemma in order to prove the following proposition.

Proposition 8.6. *By recoloring at most* $\frac{\varepsilon}{20}k_2^2$ *of the grey and white edges of* F_2 *in black, it is possible to obtain a graph* F_2' *in which the following condition is satisfied: if* (x,y) *and* (x,z) *are grey edges in* F_2'*, then* (y,z) *is black in* F_2'*.*

Proof. We apply Lemma 3.5 to the graph F_2, with $\gamma' = \frac{\epsilon^2}{50}$, $r = 3$, $m' = \frac{25}{\varepsilon}$, and obtain an equipartition $\mathcal{C} = \{X_i \mid 1 \leq i \leq l\}$ of $V(F_2)$, with $m' \leq l \leq T_{3.5}(m', 3, \gamma')$ clusters. Let $\eta' = \frac{\varepsilon}{50}$. We now recolor some edges of F_2 as follows. Suppose $x \in X_i$ and $x' \in X_j$:

1. If $i = j$, we recolor (x,x') black.

2. If (X_i, X_j) is not γ'-regular with respect to some color, we recolor (x, x') black.

3. If $d_{grey}(X_i, X_j) < \eta'$, and (x, x') is grey, we recolor (x, x') black.

4. If $d_{white}(X_i, X_j) < \eta'$, and (x, x') is white, we recolor (x, x') black.

Denote the graph we obtain after recoloring all these edges by F_2'. Indeed we have recolored at most

$$l\binom{k_2/l}{2} + \gamma\binom{l}{2}\left(\frac{k_2}{l}\right)^2 + 2\eta'\binom{l}{2}\left(\frac{k_2}{l}\right)^2 \leq \frac{k_2^2}{2l} + \gamma'\frac{k_2^2}{2} + \eta' k_2^2 \leq \frac{\varepsilon}{20}k_2^2$$

edges in F_2.

Assume there is a grey C_4 in F_2', with one of its middle edges either white or grey. By Proposition 8.3 and since black edges in F_2 remain black in F_2', this implies that there is some grey C_4 in F_2, in which both middle edges are white. However this contradicts our construction of F_2. Therefore, to complete the proof of the proposition, we will show that if (x_1, x_2) and (x_1, x_3) are grey edges in F_2', and the edge (x_2, x_3) is either white or grey, then there must also exist a grey C_4 in F_2', with one of its middle edges either white or grey. This is done by a standard usage of regularity, as follows.

Since there are no grey edges inside any X_i in F_2', it must be that $\{x_1, x_2, x_3\}$ come from three different clusters, say $x_1 \in X_1, x_2 \in X_2, x_3 \in X_3$. Moreover, after the recoloring, $d_{grey}(X_1, X_2) \geq \eta'$ and $d_{grey}(X_1, X_3) \geq \eta'$.

Assume (x_2, x_3) is grey, and hence $d_{grey}(X_2, X_3) \geq \eta'$. We now consider only the grey edges in F_2'. By Fact 3.2, there are at least $(1 - 2\gamma')|X_2| > 0$ vertices in X_2 which are connected (by grey edges) to at least $(\eta' - \gamma')$ vertices **both** in X_1 and in X_3. Pick such a vertex $a_2 \in X_2$. Denote by $A_1 \subseteq X_1$ the set of grey neighbors of a_2 in X_1, and by $A_3 \subseteq X_3$ the set of grey neighbors of a_2 in X_3. Since $|A_1|, |A_3| \geq (\eta' - \gamma')|X_i| > \gamma'|X_i|$, by the definition of regularity, $d_{grey}(A_1, A_3) \geq (\eta' - \gamma')$. Therefore there must be some vertex $a_3 \in A_3$ with at least two neighbors in A_1. Call those vertices a_1 and a_1'. It now follows that $\{a_1, a_2, a_1', a_3\}$ spans a grey C_4 in F_2', and (a_2, a_3) - one of its middle edges - is grey.

The case where (x_2, x_3) is white is settled similarly. This time, we only consider the grey edges in $E(X_1, X_2)$, and $E(X_1, X_3)$ and the white edges in $E(X_2, X_3)$, and obtain a grey C_4 with a white middle edge. This completes the proof of Proposition 8.6.

■

Having finished all the necessary preparations, we are now ready to count the edges of each color in F. We first calculate it for F_2'. Let M be a maximal matching in the grey edges of F_2', consisting of t edges ($t \leq \frac{k_2}{2}$). Denote the set of endpoints of the edges in M by V_M, and by $V_{\overline{M}}$ the rest of the vertices of F_2'. We make the following observations on the edges in F_2':

1. By the maximality of M, if $x, y \in V_{\overline{M}}$, then (x, y) is either white or black.

2. Assume $(x,y) \in M$, and $z \in V_{\overline{M}}$. Then by Proposition 8.6, either both edges $(x,z),(y,z)$ are not grey, or at least one of them is black.

3. Assume $(x,y),(u,v) \in M$. Note that if, e.g., (x,u) is grey, then by Proposition 8.6 both (x,v) and (u,y) must be black. Hence there are two options for the remaining four edges connecting $\{x,y,u,v\}$: either none of them is grey, or at least two of them are black.

Except for the edges of M, each edge of F_2' is relevant to exactly one of the above observations. It therefore follows that

$$\begin{aligned} 2|EB(F_2')| + |EW(F_2')| &\geq \binom{|V_{\overline{M}}|}{2} + 2t|V_{\overline{M}}| + 4\binom{t}{2} \\ &= \binom{k_2 - 2t}{2} + 2t(k_2 - 2t) + 4\binom{t}{2} \\ &= \frac{1}{2}(k_2^2 + 4t^2 - 4tk_2 - k_2 + 2t) + 2tk_2 - 4t^2 + 2t^2 - 2t \\ &= \frac{1}{2}(k_2^2 - k_2) - t \\ &\geq \binom{k_2}{2} - \frac{1}{2}k_2 \ . \end{aligned}$$

Since F_2' is obtained by recoloring at most $\frac{\varepsilon}{20}k_2^2$ edges in F_2 black, when going back to F_2, we have

$$2|EB(F_2)| + |EW(F_2)| \geq 2|EB(F_2')| + |EW(F_2')| - \frac{\varepsilon}{10}k_2^2 \geq \binom{k_2}{2} - \frac{1}{2}k_2 - \frac{\varepsilon}{10}k_2^2 \ .$$

Recall that there are at most k_1 grey edges touching vertices of F_1. Hence, besides the edges within F_2, there are at least $\binom{k_1}{2} - k_1 + k_1k_2$ black and white edges in F. This gives the following bound for F (recall that $k \geq m \geq \frac{20}{\varepsilon}$):

$$\begin{aligned} 2|EB(F)| + |EW(F)| &\geq \binom{k_2}{2} - \frac{1}{2}k_2 - \frac{\varepsilon}{10}k_2^2 + \binom{k_1}{2} - k_1 + k_1k_2 \\ &= \binom{k}{2} - \frac{1}{2}k_2 - k_1 - \frac{\varepsilon}{10}k_2^2 \\ &\geq \Big(\frac{1}{2} - \frac{\varepsilon}{5}\Big)k^2 \ . \end{aligned}$$

This completes the proof of Claim 8.5. ■

We have thus finished the discussion on the cluster graph F, and are ready to show the lower bound for the distance between G and $\hat{G}$. We use the following observations.

1. By Lemma 3.4, all but at most $\gamma\binom{k}{2} \leq \frac{\varepsilon}{5}\binom{k}{2}$ pairs $1 \leq i < j \leq k$ satisfy $|d_{\hat{G}}(V_i,V_j) - d_{\hat{G}}(U_i,U_j)| < \frac{\varepsilon}{5}$.

2. If (i,j) is either a black or a white edge in F, then either $d_{\hat{G}}(U_i,U_j) > 1-\eta$ or $d_{\hat{G}}(U_i,U_j) < \eta$ respectively.

3. By Lemma 3.9, with $p = \frac{1}{3}$ and $f(n) = n^{1.6}$, with high probability, $|d_G(V_i,V_j) - \frac{1}{3}| < \frac{\varepsilon}{5}$

Hence, for all but at most $\frac{\varepsilon}{5}\binom{k}{2}$ black edges (i,j) in F, there were at least $((1-\eta)-\frac{1}{3}-\frac{\varepsilon}{5}-\frac{\varepsilon}{5})\frac{n^2}{k^2} = (\frac{2}{3}-\frac{3\varepsilon}{5})\frac{n^2}{k^2}$ modifications in $E(V_i,V_j)$. Similarly, for all but at most $\frac{\varepsilon}{5}\binom{k}{2}$ white edges in F, there were at least $(\frac{1}{3}-\frac{3\varepsilon}{5})\frac{n^2}{k^2}$ modifications in $E(V_i,V_j)$. Combining this with Claim 8.5, we get

$$\begin{aligned}
\Delta(G,\hat{G}) &\geq \left(|EB(F)| - \frac{\varepsilon}{5}\binom{k}{2}\right)\left(\frac{2}{3}-\frac{3\varepsilon}{5}\right)\frac{n^2}{k^2} + \left(|EW(F)| - \frac{\varepsilon}{5}\binom{k}{2}\right)\left(\frac{1}{3}-\frac{3\varepsilon}{5}\right)\frac{n^2}{k^2} \\
&\geq \frac{n^2}{k^2}\left[|EB(F)|\left(\frac{2}{3}-\frac{3\varepsilon}{5}\right) + \left(\left(\frac{1}{2}-\frac{\varepsilon}{5}\right)k^2 - 2|EB(F)|\right)\left(\frac{1}{3}-\frac{3\varepsilon}{5}\right) - \frac{\varepsilon}{5}\binom{k}{2}\right] \\
&\geq \frac{n^2}{k^2}\left(\frac{3\varepsilon}{5}|EB(F)| + \left(\frac{1}{2}-\frac{\varepsilon}{5}\right)\left(\frac{1}{3}-\frac{3\varepsilon}{5}\right)k^2 - \frac{\varepsilon k^2}{10}\right) \\
&> \frac{n^2}{k^2}\left(\left(\frac{1}{6}-\frac{\varepsilon}{2}\right)k^2\right) \geq \left(\frac{1}{3}-\varepsilon\right)\binom{n}{2}.
\end{aligned}$$

■

8.2 Other graphs on four vertices

By our earlier observations, we already have tight asymptotic results on $ed(n,\mathcal{P}_H^*)$ for $H = K_4, P_4$, claw and their complements. The case $H = C_4$ is discussed in Section 9. In order to analyze $ed(n,\mathcal{P}_H^*)$ for all graphs H on at most four vertices, we are left with two additional graphs: the first is obtained by removing one edge from a clique of size four, denote this graph by $K_4 - e$, and the other is a triangle with a fourth vertex which is connected to exactly one vertex of the triangle, denoted $K_3 + e$. Assuming the reader is familiar with the proof of Theorem 2.12, we only sketch the proofs of the following theorems, and emphasize the differences.

Theorem 8.7. *Let $H = K_4 - e$. Then $p(\mathcal{P}_H^*) = \frac{2}{3}$ and $ed(n,\mathcal{P}_H^*) = (\frac{1}{3} - o(1))\binom{n}{2}$.*

Proof. **(sketch)** Any graph that is either $(2,0)$-colorable or $(0,1)$-colorable is in $\mathcal{P}_H^*$. Therefore, by either turning a graph G into a clique (if the density of G is at least $2/3$) or otherwise turning it into a bipartite graph we get that for any graph G on n vertices, $E_{\mathcal{P}_H^*}(G) \leq \frac{1}{3}\binom{n}{2}$.

On the other hand, we show that with high probability $E_{\mathcal{P}_H^*}(G(n,\frac{2}{3})) \geq (\frac{1}{3}-o(1))\binom{n}{2}$. Following the method of Lemma 8.2, we obtain a regular partition of $\hat{G} \in \mathcal{P}_H^*$ and construct an appropriate edge-colored cluster graph F on k vertices. In this case, for any C_4 consisting of grey edges in F, the middle edges must both be white in order to avoid a colored copy of H. Note that this observation is stronger than the one we made in Lemma 8.2, and allows us to skip the splitting of F into two separate graphs. Applying the multi-color regularity lemma to F, we get that

after recoloring $o(k^2)$ edges in F, whenever (x,y) and (x,z) are grey then (y,z) must be white (similar to Proposition 8.6). This is translated to a constraint on the number of edges which gives $|EB(F)| + 2|EW(F)| \geq (1-o(1))\binom{k}{2}$. Random graph calculations then complete the proof of the theorem.

■

Theorem 8.8. *Let $H = K_3 + e$. Then $p(\mathcal{P}^*_H) = \frac{2}{3}$ and $ed(n, \mathcal{P}^*_H) = (\frac{1}{3} - o(1))\binom{n}{2}$.*

Proof. **(sketch)** Following the proof of Theorem 8.7 one should note that the restrictions on (r,s)-colorability and on the illegal colored subgraphs of F holds also for H. Hence the same proof applies also for this case.

■

9 Improved asymptotic results

9.1 The case $H = C_4$

In this section we prove Theorem 2.13. We first prove the lower bound as follows.

Lemma 9.1. *$E_{\mathcal{P}^*_{C_4}}(K_{n,n}) \geq \binom{n}{2}$, and in particular $ed(2n, \mathcal{P}^*_{C_4}) \geq \binom{n}{2}$.*

Proof. Starting with $K = K_{n,n}$ on the classes of vertices A and B, let H be an induced C_4-free graph obtained from K by a minimum number of changes (additions and deletions of edges). Let G be the graph consisting of all edges of H that belong to K as well, and let k be the size of a maximum matching M in G. Suppose the matching M consists of the edges $a_1b_1, a_2b_2, \ldots, a_kb_k$, $(a_i \in A, b_i \in B)$, and let the other vertices of A and B be $a_{k+1}, a_{k+2}, \ldots a_n$ and $b_{k+1}, b_{k+2}, \ldots b_n$ (the case $k = n$ is also possible, of course). Since M is a maximum matching, there are no edges a_pb_q with both p and q bigger than k. In addition, for each $p > k$ and $i \leq k$, either the edge a_ib_p or the edge b_ia_p is missing in G (since otherwise there is an augmenting path $b_pa_ib_ia_p$, contradicting the maximality of M). This means that to obtain H from K we have deleted at least $(n-k)^2 + k(n-k)$ edges that are incident with at least one vertex not saturated by M.

For each $1 \leq i < j \leq k$, let C_{ij} denote the induced copy of C_4 (in K) on the vertices a_i, b_i, a_j, b_j. We know that the edges a_ib_i and a_jb_j of this cycle are also in G, hence among the four other pairs $a_ib_j, a_jb_i, a_ia_j, b_ib_j$ at least one non-edge of K is an edge of H or one edge of K is a non-edge of G. As all these $\binom{k}{2}$ fourtuples of pairs are pairwise disjoint (no pair belongs to two of them), this accounts to another $\binom{k}{2}$ modifications between K and H. Moreover, each edge omission here is an omission of an edge with both ends saturated by M, hence it has not been counted before.

Altogether we conclude that the edit distance between K and H is at least

$$(n-k)^2 + k(n-k) + \binom{k}{2} \geq \binom{n-k}{2} + k(n-k) + \binom{k}{2} = \binom{n}{2}.$$

■

At this point we note that C_4 is not $(1,1)$-colorable and therefore every $(1,1)$-colorable graph does not contain an induced C_4, namely $\mathcal{P}_{1,1} \subseteq \mathcal{P}^*_{C_4}$. The graphs in $\mathcal{P}_{1,1}$ are called **split graphs**.

Lemma 9.2. $ed(2n, \mathcal{P}^*_{C_4}) \leq \binom{n}{2}$.

Proof. Consider some partition of the vertex set of G into two n-vertex sets V_1, V_2. W.l.o.g, $e(V_1) \geq e(V_2)$. Adding all the missing edges in V_1, and removing all the edges in V_2, turns G into a split graph. We have thus changed at most $\left(\binom{n}{2} - e(V_1)\right) + e(V_2) \leq \binom{n}{2}$ edges. As any split graph is also induced C_4-free, we actually obtained a graph in $\mathcal{P}^*_{C_4}$, proving the lemma.

■

This completes the proof of Theorem 2.13.

Proof of Corollary 2.14:

Clearly, every chordal graph is also induced C_4-free. Hence, $ed(2n, \mathcal{P}^*_{C_4}) \leq ed(2n, \mathcal{P})$. Moreover, every split graph is chordal, and thus Lemma 9.2 also applies for $\mathcal{P}$, which completes the proof of the corollary. ■

9.2 The case $H = P_4$

Corollary 2.3 gives the lower bound for Theorem 2.15. We thus complete the proof of Theorem 2.15 by proving the upper bound. We use the recursive definition of cographs and the following result of Erdős, Goldberg, Pach and Spencer [20]:

Theorem 9.3. (Erdős et al, [20]) *There exists an $\eta > 0$ such that in any graph on n vertices and $n < e \leq \frac{1}{2}\binom{n}{2}$ edges, one can find two disjoint subsets of vertices S and T such that $|S| = |T| = \frac{n}{4}$ and*

$$|e(S) - e(T)| > \eta\sqrt{en}$$

Lemma 9.4. *For a sufficiently large n, any graph on n vertices is at most $\frac{1}{2}\binom{n}{2} - \frac{1}{5}\eta n^{1.5}$ far from being a cograph, where η is the constant from Theorem 9.3.*

Proof. Let G be a graph on n vertices. Without loss of generality, assume $e(G) \leq \frac{1}{2}\binom{n}{2}$. Otherwise, we may consider its complement since being a cograph is a complement invariant property. If $e(G) \leq \frac{1}{2}\binom{n}{2} - \eta n^{1.5}$, we remove all the edges from G thus trivially turning it into a cograph. Hence, $\frac{1}{2}\binom{n}{2} - \eta n^{1.5} < e(G) \leq \frac{1}{2}\binom{n}{2}$. By Theorem 9.3, we find disjoint vertex sets S and T satisfying (for a sufficiently large n) $|e(S) - e(T)| > \eta\sqrt{en} > \frac{1}{5}\eta n^{1.5}$. We modify G to obtain a graph $\hat{G}$ as follows:

1. If $e(S) \leq \frac{1}{2}\binom{|S|}{2}$, remove all the edges inside S. Otherwise, add all the missing edges inside S.

2. If $e(T) \leq \frac{1}{2}\binom{|T|}{2}$, remove all the edges inside T. Otherwise, add all the missing edges inside T.

3. Do the same for the remainder of G, which consists of edges connecting S and T and edges with at least one endpoint in $V(G) \setminus (S \cup T)$.

$\hat{G}$ is a cograph since, up to taking a complement, it is a disjoint union of S, T and $V \setminus (S \cup T)$ where each of these three sets induces either a clique or an independent set in $\hat{G}$. Moreover, we change at most $\frac{1}{2}(\binom{n}{2} - \binom{|S|}{2} - \binom{|T|}{2})$ edges in step 3, and at most $\frac{1}{2}(\binom{|S|}{2} + \binom{|T|}{2}) - \frac{1}{5}\eta n^{1.5}$ edges in steps 1 and 2. Thus indeed we make at most $\frac{1}{2}\binom{n}{2} - \frac{1}{5}\eta n^{1.5}$ modifications.

■

10 Concluding remarks and future work

- It seems that for many of the natural hereditary properties $\mathcal{P}$, the tools and methods we describe in this paper allow one to find $ed(n, \mathcal{P})$ and $p(\mathcal{P})$. Yet this may still require a substantial ad hoc effort, as in the proofs of Section 8. It would be interesting to find a robust method for such analysis which applies to all hereditary properties. A milder task would be to establish such a method for all the properties $\mathcal{P}^*_H$. Recent results of Balogh and Martin, and of Marchant and Thomason, based on earlier work of Richer, provide the value of $p(\mathcal{P}^*_{K_{3,3}})$, but the general problem remains wide open.

- Other Turán type problems on hereditary properties also arise naturally, extending well known analogous results for monotone properties. In particular:

  - Which are the graphs in $\mathcal{P}$ that are the closest to $G(n, p(\mathcal{P}))$? Theorem 2.10 shows that this question is much easier when $p(\mathcal{P}) = \frac{1}{2}$.
  - What is the exact furthest graph from $\mathcal{P}$?
  - Consider a monotone property which contains all graphs excluding a (weak) copy of a fixed graph H. Some extremal features were proved for the case of H having a color critical edge (e.g. [8], [22]). What are the analogs of these special graphs when forbidding an *induced* copy H?

  Related questions for the properties $\mathcal{P}^*_H$, were addressed by Prömel and Steger in [29]. Their results might hint on possible answers.

- In [5], Alon, Shapira and Sudakov describe, for every monotone property $\mathcal{M}$ and $\varepsilon > 0$, a polynomial time algorithm for approximating the edit distance of a given input graph on n vertices from $\mathcal{M}$. The algorithm obtains an additive approximation within εn^2 of the correct edit distance. A slightly different version of their algorithm provides an approximation algorithm for edge-modification problems in the broader setting of hereditary properties.

The authors of [5] also characterize the properties for which the above mentioned algorithm achieves essentially the best possible approximation, that is, the monotone properties $\mathcal{M}$ for which it is NP-hard to approximate $E_{\mathcal{M}}(G)$ to within an additive error of $n^{2-\varepsilon}$, for any $\varepsilon > 0$. In a future work, we (partially) extend these results to hereditary properties, relying in part on the ideas of the present paper. The proofs are based on a refinement of Theorem 2.10 which determines the structure of the closest graph in $\mathcal{P}$ to $G(n, 1/2)$ for some hereditary properties.

References

[1] V. E. Alekseev, On the entropy values of hereditary properties, Discrete Math. Appl. 3 (1993), 191-199.

[2] N. Alon, Tools from higher algebra, in "Handbook of Combinatorics" (R.L. Graham, M. Grötschel and L. Lovàsz eds.), North Holland (1995), Chapter 32, pp. 1749-1783.

[3] N. Alon, E. Fischer, M. Krivelevich and M. Szegedy, Efficient testing of large graphs, Proc. of the 40 IEEE FOCS (1999), 656–666. Also: Combinatorica 20 (2000), 451-476.

[4] N. Alon and A. Shapira, A characterization of the (natural) graph properties testable with one-sided error, Proc. of the 46 IEEE FOCS (2005), 429-438.

[5] N. Alon, A. Shapira and B. Sudakov, Additive approximation for edge-deletion problems, Proc. of the 46 IEEE FOCS (2005), 419-428. Also: Annals of Math., to appear.

[6] N. Alon and J. H. Spencer,**The Probabilistic Method**, Second Edition, Wiley, New York (2000).

[7] N. Alon and U. Stav, What is the furthest graph from a hereditary property?, Random Structures and Algorithms, to appear.

[8] B. Andrásfai, P. Erdős and V. Sós, On the connection between chromatic number, maximal clique and minimal degree of a graph, Discrete Math. 8 (1974), 205-218.

[9] M. Axenovich, A. Kézdy and R. Martin, The editing distance in graphs, J. of Graph Theory, submitted.

[10] M. Axenovich and R. Martin, Avoiding patterns in matrices via a small number of changes, SIAM J. Discrete Math. 20(1) (2006), 49-54.

[11] J. Balogh, B. Bollobás and D. Weinreich, The speed of hereditary properties of graphs, J. of Combinatorial Theory, Ser. B 79 (2000), 131-156.

[12] J. Balogh, B. Bollobás and D. Weinreich, The penultimate rate of growth for graph properties, European J. of Combinatorics 22(3)(2001), 277-289.

[13] B. Bollobás, **Extremal Graph Theory**, Academic Press (1978).

[14] B. Bollobás and A. Thomason, Projections of bodies and hereditary properties of hypergraphs, J. London Math. Soc. 27(1995), 417-424.

[15] B. Bollobás and A. Thomason, Generalized chromatic numbers of random graphs, Random Structures and Algorithms 6 (1995), 353-356.

[16] B. Bollobás and A. Thomason, Hereditary and monotone properties of graphs, in "The Mathematics of Paul Erdős II" (R.L. Graham and J. Nešetřil, eds.) Algorithms and Combinatorics 14 (1997), 70-78.

[17] B. Bollobás and A. Thomason, The structure of hereditary properties and colourings of random graphs, Combinatorica 20 (2000), 173-202.

[18] D.G. Corneil, H. Lerchs, L. Stewart Burlingham, Complement reducible graphs, Discrete Applied Mathematics. 3 (1981), 163-174.

[19] M. Chudnovsky, P. D. Seymour, The structure of claw-free graphs, Surveys in Combinatorics 2005, London Mathematical Society Lecture Note Series 327, 153-171.

[20] P. Erdős, M. Goldberg, J. Pach and J. Spencer, Cutting a graph into two dissimilar halves, J. of Graph Theory 12(1988), 121-131.

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

[22] P. Erdős and M. Simonovits, On a valence problem in extremal graph theory, Discrete Math. 5 (1973), 323-334.

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

[24] M.C. Golumbic, **Algorithmic Graph Theory and Perfect Graphs**, Academic Press, New York (1980).

[25] J. Komlós and M. Simonovits, Szemerédi's Regularity Lemma and its applications in graph theory, in "ombinatorics, Paul Erdös is Eighty", Vol II (D. Miklós, V. T. Sós, T. Szönyi eds.), János Bolyai Math. Soc., Budapest (1996), 295-352.

[26] W. Mantel, Problem 28, Wiskundige Opgaven 10 (1907), 60-61.

[27] T. A. McKee and F.R. McMorris, **Topics in Intersection Graph Theory**, SIAM, Philadelphia, PA, 1999.

[28] H.J. Prömel and A. Steger, Excluding induced subgraphs: quadrilaterals, Random Strutures and Algorithms 2 (1991), 55-71.

[29] H.J. Prömel and A. Steger, Excluding induced subgraphs II: extremal graphs, Discrete Applied Mathematics 44 (1993), 283-294.

[30] H.J. Prömel and A. Steger, Excluding induced subgraphs III: a general asymptotic, Random Strutures and Algorithms 3 (1992), 19-31.

[31] J. L. Ramírez-Alfonsín (Editor), B. A. Reed (Editor), **Perfect Graphs**, Wiley, 2001.

[32] E. R. Scheinerman and J. Zito, On the size of hereditary classes of graphs, J. of Combinatorial Theory, Ser. B 61 (1994), 16-39.

[33] E. Szemerédi, Regular partitions of graphs, In: *Proc. Colloque Inter. CNRS* (J. C. Bermond, J. C. Fournier, M. Las Vergnas and D. Sotteau, eds.) (1978), 399-401.

[34] P. Turán, On an extremal problem in graph theory (in Hungarian), Mat. Fiz. Lapok 48 (1941), 436-452.

[35] H.E. Warren, Lower Bounds for approximation by nonlinear manifolds, Trans. Amer. Math. Soc. 133 (1968), 167-178.