What is the furthest graph from a hereditary property?

Noga Alon * Uri Stav †

May 26, 2006

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 furthest graph on n vertices from $\mathcal{P}$ and what is the largest possible edit distance from $\mathcal{P}$? Denote this maximal distance by $ed(n, \mathcal{P})$. This question is motivated by algorithmic edge-modification problems, in which one wishes to find or approximate the value of $E_{\mathcal{P}}(G)$ given an input graph G.

A *monotone* graph property is closed under removal of edges and vertices. Trivially, for any monotone property, the largest edit distance is attained by a complete graph. We show that this is a simple instance of a much broader phenomenon. A *hereditary* graph property is closed under removal of vertices. We prove that for any hereditary graph property $\mathcal{P}$, a random graph with an edge density that depends on $\mathcal{P}$ essentially achieves the maximal distance from $\mathcal{P}$, that is: $ed(n, \mathcal{P}) = E_{\mathcal{P}}(G(n, p(\mathcal{P}))) + o(n^2)$ with high probability. The proofs combine several tools, including strengthened versions of the Szemerédi Regularity Lemma, properties of random graphs and probabilistic arguments.

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.

*Schools of Mathematics and Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel and IAS, Princeton, NJ 08540, USA. Email: nogaa@tau.ac.il. Research supported in part by the Israel Science Foundation, by a USA-Israeli BSF grant, by NSF grant CCR-0324906, by a Wolfensohn fund and by the State of New Jersey

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

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 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 extremal 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 in hand. Narrowing our discussion to hereditary properties is one of the mildest and yet natural restrictions. These properties play an important role in various areas of research in graph theory as well as in theoretical and applied computer science. Due to the simple nature of these properties they also arise in Chemistry, Biology, Social Science as well as in many other areas. Some of these properties are the well studied graph properties of being Perfect, Chordal, Interval, Comparability, Permutation and more. In fact, almost all interesting graph properties are hereditary. The recent results of [4] 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.

1.2 The new results

The main result of this paper, Theorem 1.2, is that for any hereditary graph 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 $\mathcal{P}$.

The proof of the main result follows a method used by Alon and Shapira in [3], which is based on a strengthened version of the Szemerédi Regularity Lemma (proved in [2]). Using this method, we prove the following theorem:

Theorem 1.1. *Let $\mathcal{P}$ be an arbitrary hereditary graph property, and $\varepsilon > 0$. Then there is $n_{1.1}(\mathcal{P}, \varepsilon)$ such that for any $n \geq n_{1.1}(\mathcal{P}, \varepsilon)$ there is $p = p_{1.1}(\mathcal{P}, \varepsilon, n)$ satisfying with high probability*

$$ed(n, \mathcal{P}) \leq E_{\mathcal{P}}(G(n, p)) + \varepsilon n^2 \tag{1}$$

Note that the value of p in (1) depends on n. Thus, some additional effort is needed in order to

deduce that there is a *single* value of p that suits *all* n.

Theorem 1.2. *Let $\mathcal{P}$ be an arbitrary hereditary graph property. Then there exists $p = p_{1.2}(\mathcal{P}) \in [0,1]$, such that with high probability*

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

To be completely formal, Theorem 1.2 should be read as follows: for a hereditary property $\mathcal{P}$, there is $p = p_{1.2}(\mathcal{P})$ such that for any $\varepsilon > 0$, a graph $G \sim G(n,p)$ satisfies $ed(n,\mathcal{P}) \leq E_{\mathcal{P}}(G) + \varepsilon n^2$ with probability that tends to 1 as n tends to infinity.

In the rest of the paper we write $p(\mathcal{P})$ for $p_{1.2}(\mathcal{P})$. Note that Theorem 1.2 implicitly asserts the existence of $p(\mathcal{P})$, but it supplies neither a general tool for determining its value nor a general way to compute the maximum possible edit distance. It seems to be a challenging task to express the extremal probability $p(\mathcal{P})$ as a "natural" function of $\mathcal{P}$. Yet this is possible for several large families of hereditary properties, see Section 5 for more details.

1.3 Related work

The study of extremal edge modification problems for *monotone* graph properties was initiated by Turán ([25]). In this case one only deletes edges, since a monotone property is closed under removal of edges. Thus, trivially, the furthest graph from such properties is a complete graph. The contribution of Theorem 1.2 is extending this fact to arbitrary *hereditary* properties, where in general the role of the complete graph is played by some random graph $G(n,p)$.

As for the edit distance itself, Turán's Theorem and its various extensions (most notably by Erdős and Stone [16], and by Erdős and Simonovits [14]) show that for any monotone graph property $\mathcal{M}$, defined by the (possibly infinite) set of its forbidden weak subgraphs $\mathcal{F}_{\mathcal{M}}$ [1]: $ed(n,\mathcal{M}) = (\frac{1}{r} - o(1))\binom{n}{2}$ where $r = min\{\chi(F) - 1 \mid F \in \mathcal{F}_{\mathcal{M}}\}$.

For a fixed graph H, denote by $\mathcal{P}^*_H$ the hereditary property that contains all graphs excluding an *induced* copy of H. Axenovich, Kézdi and Martin recently showed in [7] that $ed(n,\mathcal{P}^*_H)$ is bounded by a function of H as follows. They define a graph parameter $\chi_B(H)$ [2] and show that if $\chi_B(H) = k+1$ then $(\frac{1}{2k} - o(1))\binom{n}{2} < ed(n,\mathcal{P}^*_H) \leq \frac{1}{k}\binom{n}{2}$. The lower bound is obtained by the random graph $G(n,1/2)$. Therefore, whenever the lower bound is asymptotically tight, it follows that in our notation $p(\mathcal{P}^*_H) = \frac{1}{2}$. The gap left in the general bound is settled in [7] for some families of graphs. In particular, for self-complementary graphs (i.e. $H = \overline{H}$) it is shown there that $ed(n,\mathcal{P}^*_H) = (\frac{1}{2k} - o(1))\binom{n}{2}$, and hence $p(\mathcal{P}^*_H) = \frac{1}{2}$.

The problem we address and the techniques we use relate and extend work in different paths of research on hereditary graph properties. In particular:

[1] That is, a graph G belongs to $\mathcal{M}$ iff it excludes subgraphs isomorphic to any of the members of $\mathcal{F}_{\mathcal{M}}$.

[2] $\chi_B(H)$ is the least integer $k+1$ such that for any pair (r,s) satisfying $r+s = k+1$, the vertices of H can be partitioned into $r+s$ sets, r of which inducing an empty graph in H and s inducing a complete graph. This parameter was first defined by Prömel and Steger in [21], where it was called $\tau(H)$.

Growth of hereditary graph properties: The edit distance of a hereditary property is closely related to the so called *speed* of the property, which is measured by the function $|\mathcal{P}^n|$. Scheinerman and Zito ([23]) showed that $|\mathcal{P}^n|$ belongs to one of few possible classes of functions. Several other papers sharpen these results, concentrating on, e.g., sparse hereditary properties (Balogh, Bollobás and Weinreich [8], [9]), dense hereditary properties (Bollobás and Thomason [10], [11] and Alekseev [1]) and properties of the type $\mathcal{P}_H^*$ (Prömel and Steger [20] ,[21], [22]).

Testing hereditary graph properties: Roughly speaking, a graph property $\mathcal{P}$ is *testable* if there is a probabilistic algorithm that samples a (small) portion of a (large) graph and decides whether the graph satisfies $\mathcal{P}$. The algorithm is expected to distinguish graphs that satisfy $\mathcal{P}$ from those that are far from it in the edit distance. Alon, Fischer, Krivelevich and Szegedy proved in [2] that for any fixed graph H, the property $\mathcal{P}_H^*$ is testable. In [3], Alon and Shapira extended this result, showing it applies to all hereditary graph properties. The technique used in [3] lays in the core of the proof of Theorem 1.1. Similar results on testing of hereditary properties were obtained by Lovász and Szegedy in [18] using convergent graph sequences (see also [13]). It should also be possible to apply this approach, initiated in [19], to give an alternative proof of Theorem 1.2.

Coloring random graphs with hereditary properties: In [12], Bollobás and Thomason estimated the coloring number of random graphs by graphs satisfying some hereditary property. As they noticed, finding the probability that a random graph $G(n,p)$ satisfies some hereditary property is much more difficult when $p \neq \frac{1}{2}$. Their approach is also based on the regularity lemma together with tools from the theory of Random Graphs, and is strongly related to some of our methods here.

1.4 Organization

The rest of the paper is organized as follows. In Section 2 we review the definitions and state the regularity lemmas which will be used in the rest of the paper. In Section 3 we prove Theorem 1.1 and in Section 4 we prove Theorem 1.2. Section 5 contains some concluding remarks and future work.

2 Regularity Lemma Background

In this section we discuss some of the basic applications of regular partitions and state the regularity lemmas that we use in the proof of Theorem 1.1. See [17] 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 connecting vertices in A and B, and $e(A,B)$ is the size of $E(A,B)$. The **edge density** of the pair (A,B) 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 2.1. *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.*

Lemma 2.2 below 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 (see, e.g., [3], [11], [17]).

Lemma 2.2. *For every real $0 < \eta < 1$ and integer $f \geq 1$ there exists $\gamma = \gamma_{2.2}(\eta, f)$ 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 nonempty 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.*

In fact, the statement of Lemma 2.2 could be strengthened to show that *many* induced copies of F exist in G. However, for our purposes a single induced copy suffices.

Remark 2.3. *Observe, that the function $\gamma_{2.2}(\eta, f)$ may and will be assumed to be monotone non-increasing in f. Also, for ease of future definitions we set $\gamma_{2.2}(\eta, 0) = 1$ for any $0 < \eta < 1$.*

Note, that in terms of regularity, Lemma 2.2 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 do not care if (i,j) is an edge of F.

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). The Regularity Lemma of Szemerédi can be formulated as follows.

Lemma 2.4 ([24]). *For every m and $\varepsilon > 0$ there exists a number $T = T_{2.4}(m, \varepsilon)$ with the following property: Any graph G on $n \geq T$ vertices, 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 $\varepsilon\binom{k}{2}$ of them, are ε-regular.*

The function $T_{2.4}(m, \varepsilon)$ may and is assumed to be monotone non-decreasing in m and monotone non-increasing in ε. Another lemma, which will be very useful in this paper is Lemma 2.5 below. Some versions of this lemma appear in various papers applying the regularity lemma (see e.g. the appendix of [3]).

Lemma 2.5. *For every ℓ and γ there exists $\delta = \delta_{2.5}(\ell, \gamma)$ such that for every graph G with $n \geq \delta^{-1}$ vertices there exist disjoint vertex sets $W_1, \ldots, W_\ell$ satisfying:*

1. $|W_i| \geq \delta n$.
2. **All** $\binom{\ell}{2}$ *pairs are γ-regular.*
3. *Either all pairs are with densities at least $\frac{1}{2}$, or all pairs are with densities less than $\frac{1}{2}$.*

Remark 2.6. *Observe, that the function $\delta_{2.5}(\ell, \gamma)$ may and will be assumed to be monotone non-increasing in ℓ and monotone non-decreasing in γ. For ease of future applications we will assume that for all ℓ and γ we have $\delta_{2.5}(\ell, \gamma) \leq 1/2$.*

Our main tool in the proofs, in addition to Lemmas 2.2 and 2.5 is Lemma 2.7 below, proved in [2]. This lemma can be considered as a strengthened variant of the standard regularity lemma. This variant has two advantages. The first advantage 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. The second advantage of this version is that one can define ε as a function of the size of the partition, rather than having to use a fixed ε as in Lemma 2.4. We denote such functions by $\mathcal{E}$ throughout the paper.

Lemma 2.7. ([2]) *For every integer m and every monotone non-increasing function $\mathcal{E} : \mathbb{N} \mapsto (0, 1)$ there is $S = S_{2.7}(m, \mathcal{E})$ which satisfies the following. For any graph G on $n \geq S$ vertices, there exists 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. $m \leq k \leq S$.
2. $U_i \subseteq V_i$ *for all* $i \geq 1$, *and* $|U_i| \geq n/S$.
3. *In the equipartition $\mathcal{B}$,* **all** *pairs are $\mathcal{E}(k)$-regular.*
4. *All but at most $\mathcal{E}(0)\binom{k}{2}$ of the pairs $1 \leq i < j \leq k$ are such that $|d(V_i, V_j) - d(U_i, U_j)| < \mathcal{E}(0)$.*

Remark 2.8. *For technical reasons (see the proof in [2]), Lemma 2.7 requires that for any $r > 0$ the function $\mathcal{E}(r)$ will satisfy $\mathcal{E}(r) \leq \min\{\mathcal{E}(0)/4, 1/4r^2\}$. However, we can always assume w.l.o.g. that $\mathcal{E}$ satisfies this condition because if it does not, then we can apply Lemma 2.7 with $\mathcal{E}'$ which is defined as $\mathcal{E}'(r) = \min\{\mathcal{E}(r), \mathcal{E}(0)/4, 1/4r^2\}$. We will thus disregard this technicality.*

The main power of Lemma 2.7 is that for *any* function $\mathcal{E}$ it allows us to find k sets of vertices $U_1, \ldots, U_k$, each of size $\Omega(n)$, such that all pairs (U_i, U_j) are $\mathcal{E}(k)$-regular. Note, that in Lemma 2.4 we first fix the regularity measure γ, and then get via the lemma k sets of vertices, where k can be very large in terms of γ.

3 Proof of Theorem 1.1

3.1 Preliminaries

We first need the following lemma stating that in a random graph, with high probability, the edge density inside and between any two sets of vertices is close to the density of the graph. The proof is a standard application of Chernoff's inequality.

Lemma 3.1. *Assume $0 \leq p \leq 1$, and let $t : \mathbb{N} \to \mathbb{N}$ satisfy $t(n) = \omega(n^{1.5})$. Then with high probability, $G = G(n,p)$ satisfies*

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

Proof. Let A be some set of vertices in $G = G(n,p)$. By Chernoff's inequality (see e.g. pp. 266 in [5])

$$Pr\left[\mid e(A) - p\binom{|A|}{2} \mid > t(n)\right] < 2\exp\left\{\frac{-2(t(n))^2}{\binom{|A|}{2}}\right\} < 2\exp\left\{-\frac{4(t(n))^2}{n^2}\right\} = e^{-\omega(n)}$$

Similarly, for any disjoint sets A, B

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

The probability that this *does not* happen for *any* such set A nor for any pair of sets (A,B) is therefore at least $1 - 4^n e^{-\omega(n)}$, which tends to 1 as n grows.

■

We will also need the following simple fact about regular pairs, which shows that by taking two subsets of a regular pair, the subsets are also assured to be somewhat regular.

Claim 3.2. *If (A,B) is a γ-regular pair, and $A' \subseteq A$ and $B' \subseteq B$ satisfy $|A'| \geq \xi|A|$ and $|B'| \geq \xi|B|$ for some $\xi \geq \gamma$, then (A',B') is a $\max\{2\gamma, \gamma/\xi\}$-regular pair.*

Proof. As (A,B) is a γ-regular pair, for every pair of subsets of $A' \subseteq A$ with $|A'| \geq \xi|A| \geq \gamma|A|$ and $B' \subseteq B$ with $|B'| \geq \xi|B| \geq \gamma|B|$ we have $|d(A',B') - d(A,B)| \leq \gamma$. Note, that if A' and B' are as above, then for every pair of subsets $A'' \subseteq A'$ and $B'' \subseteq B'$ satisfying $|A''| \geq \frac{\gamma}{\xi}|A'|$ and $|B''| \geq \frac{\gamma}{\xi}|B'|$ also satisfy $|A''| \geq \gamma|A|$ and $|B''| \geq \gamma|B|$. Therefore, by the γ-regularity of (A,B) we have $|d(A'',B'') - d(A,B)| \leq \gamma$. We thus conclude that $|d(A'',B'') - d(A',B')| \leq 2\gamma$. Hence, (A',B') is $\max\{2\gamma, \gamma/\xi\}$-regular. ■

3.2 Definitions and an overview of the proof

The proof of Theorem 1.1 is somewhat technical and requires several definitions. We therefore outline in this subsection an overview of the proof, while stating the main definitions and skipping most of the details and calculations. The detailed proof, based on the following discussion, is given in subsection 3.3.

We first note that any hereditary graph property can be defined by the set of its forbidden induced subgraphs as follows:

Definition 3.3. *For a graph property $\mathcal{P}$, define the set of* ***forbidden induced subgraphs*** *for $\mathcal{P}$, denoted $\boldsymbol{\mathcal{F}_\mathcal{P}}$ to be the set of graphs which are minimal with respect to not satisfying property $\mathcal{P}$. In other words, a graph F belongs to $\mathcal{F}_\mathcal{P}$ if it does not satisfy $\mathcal{P}$, but any graph obtained from F by removing a vertex, satisfies $\mathcal{P}$.*

Clearly, a graph G belongs to the hereditary property $\mathcal{P}$ if and only if it does not contain an induced copy of any graph in $\mathcal{F} = \mathcal{F}_\mathcal{P}$. It will be more convenient to use this equivalent definition of $\mathcal{P}$ along the proof.

Let $\mathcal{P}$, ε and n be given, and define the following graphs:

1. G^* is the furthest graph on n vertices from $\mathcal{P}$, i.e. $ed(n,\mathcal{P}) = E_\mathcal{P}(G^*)$. Assume its edge density is $p = |E(G)|/\binom{n}{2}$. We shall prove the theorem for this p.
2. We pick $G = G(n,p)$, and assume it satisfies the assertions of Lemma 3.1 (with $t(n) = n^{1.6}$).
3. $G' \in \mathcal{P}$ is the closest graph to G in $\mathcal{P}$, hence $\Delta(G,G') = E_\mathcal{P}(G)$.

We thus need to show that the edit distance of G from G' is not much smaller than the edit distance of G^* from $\mathcal{P}$.

We apply the strengthened regularity lemma, Lemma 2.7, to G', obtaining a partition of its vertices into k clusters. On each cluster, we apply Lemma 2.5 [3]. As in many applications of the regularity lemma, we then obtain a graph G'', which is a "clean" version of G'. In G'', each of the k clusters spans a homogenous set (either a complete or an empty graph), and the bipartite graph between any pair of clusters is either empty, complete or "regular enough" with a "moderate" density (i.e., bounded away from 0 and 1). It is crucial that, as we shall prove, when cleaning G' we modify at most $\frac{\varepsilon}{2}n^2$ edges, i.e. $\Delta(G',G'') \leq \frac{\varepsilon}{2}n^2$.

We then use the following object to model the clean graph G''.

Definition 3.4. *A* ***colored regularity graph*** *K is a complete graph whose vertices are colored black or white, and whose edges are colored black, white or grey.*

[3] In fact, we apply it to each of the subsets U_i of the equipartition $\mathcal{B}$ returned by Lemma 2.7.

Note that neither the vertex nor the edge coloring is assumed to be legal in the standard sense. We denote the sets of black, white and grey edges of K by $EB(K)$,$EW(K)$ and $EG(K)$ respectively. Similarly, we write $VB(K)$ and $VW(K)$ for K's black and white vertices. The definition of colored regularity graphs should be considered with respect to G'', as we shall define a colored regularity graph K on k vertices as follows: any vertex in K represents a cluster in G'', and the coloring of K represents the edge density inside or between these clusters (black for dense, white for sparse). This seemingly rough model of G'' proves to be quite useful by the following.

Definition 3.5. *For a colored regularity graph K (where $k = |V(K)|$), **the graph property $\mathcal{P}_{K,n}$** consists of all graphs J on n vertices for which there is an equipartition $\mathcal{A} = \{A_i \mid 1 \leq i \leq k\}$ of the vertices of J satisfying the following conditions. For any $1 \leq i \leq k$, if $i \in VW(K)$ then A_i spans an empty graph in J, otherwise $i \in VB(K)$ and A_i spans a complete graph in J. For any $1 \leq i < j \leq k$:*

- *If $(i,j) \in EB(K)$ then (A_i, A_j) span a complete bipartite graph in J.*
- *If $(i,j) \in EW(K)$ then (A_i, A_j) span an empty bipartite graph in J.*
- *If $(i,j) \in EG(K)$ then there is no restriction on $E(A_i, A_j)$.*

*If all the above holds, we say that the equipartition $\mathcal{A}$ **witnesses** the membership of J in $\mathcal{P}_{K,n}$.*

$\mathcal{P}_{K,n}$ may be viewed as a set of graphs that share the same approximation by the colored regularity graph K. Clearly, G'' belongs to $\mathcal{P}_{K,n}$. Also note that $\mathcal{P}_{K,n}$ is *not* hereditary. We will justify this definition by showing that $\mathcal{P}_{K,n} \subset \mathcal{P}$. For this purpose, we now make the connection between forbidden induced subgraphs and colored regularity graphs:

Definition 3.6. *A **colored-homomorphism** from a (simple) graph F to a colored regularity graph K is a mapping $\varphi : V(F) \mapsto V(K)$, which satisfies the following:*

1. *If $(u,v) \in E(F)$ then either $\varphi(u) = \varphi(v) = t$ and t is colored black, or $\varphi(u) \neq \varphi(v)$ and $(\varphi(u), \varphi(v))$ is colored black or grey.*
2. *If $(u,v) \notin E(F)$ then either $\varphi(u) = \varphi(v) = t$ and t is colored white, or $\varphi(u) \neq \varphi(v)$ and $(\varphi(u), \varphi(v))$ is colored white or grey.*

Practicing the above definitions, we note the important fact that if some graph belongs to $\mathcal{P}_{K,n}$, and that graph contains an induced copy of F', then by mapping each vertex in the copy of F' to its cluster, we get that $F' \mapsto_c K$, that is, there is a colored homomorphism from F' to K. Furthermore, the above definition should be considered with Lemma 2.2 in mind. With the right choice of parameters, the existence of such a member in $\mathcal{P}_{K,n}$ with an induced subgraph F' that belongs to $\mathcal{F}$ will let us apply Lemma 2.2 to some of the clusters found by Lemma 2.5 in G', and conclude that in fact G' contains an induced copy of some graph in $\mathcal{F}$. This leads to a contradiction since $G' \in \mathcal{P}$, showing that indeed $\mathcal{P}_{K,n} \subset \mathcal{P}$.

The difficulty in the above argument lies in the "right choice of parameters". One of the limitations of Lemma 2.2 is that the number of vertices in the embedded graph should not be too large. In other words, the regularity measure required for Lemma 2.2 depends on the size f of F'. This seems to be impossible, since $\mathcal{F}$ might be infinite. For that purpose we make our last two definitions, which follow the main idea behind the use of Lemma 2.7 in [3].

Definition 3.7. *For any (possibly infinite) family of graphs $\mathcal{F}$, and any integer k, let $\boldsymbol{\mathcal{F}_k}$ be the following set of colored regularity graphs: A colored regularity graph K belongs to $\mathcal{F}_k$ if it has at most k vertices and there is at least one $F \in \mathcal{F}$ such that $F \mapsto_c K$.*

In the proof of Theorem 1.1, the set $\mathcal{F}_k$, defined above, will represent a subset of the colored regularity graphs of size at most k. Namely, those K for which there is at least one $F \in \mathcal{F}$ such that $F \mapsto_c K$. We now define the following function, which constitutes the dependence on $\mathcal{P}$ in the proof of Theorem 1.1.

Definition 3.8. *For any family of graphs $\mathcal{F}$ and integer k for which $\mathcal{F}_k \neq \emptyset$, let*

$$\boldsymbol{\Psi_{\mathcal{F}}(k)} := \max_{K \in \mathcal{F}_k} \min_{\{F \in \mathcal{F}: F \mapsto_c K\}} |V(F)|. \tag{3}$$

Define $\Psi_{\mathcal{F}}(k) = 0$ if $\mathcal{F}_k = \emptyset$. Therefore, $\Psi_{\mathcal{F}}(k)$ is monotone non-decreasing in k.

We will use Ψ when defining the function $\mathcal{E}$ for Lemma 2.7. It will assure that Lemma 2.5 will provide enough subsets in order to apply Lemma 2.2, since $\Psi_{\mathcal{F}}(k)$ upper bounds the size of F. This way, we will indeed obtain the above mentioned contradiction.

After showing that $\mathcal{P}_{K,n} \subset \mathcal{P}$, a random partition of the vertices of G^* will show that $E_{\mathcal{P}_{K,n}}(G^*) \leq E_{\mathcal{P}_{K,n}}(G) + \frac{\varepsilon}{2}n^2$. The left hand side is at least $E_{\mathcal{P}}(G)$, while the right hand side is at most $\Delta(G, G') + \Delta(G', G'') + \frac{\varepsilon}{2}n^2$, thus completing the proof.

3.3 The detailed proof of Theorem 1.1

Step 1: Applying the regularity lemmas to G'

We write $\mathcal{F}$ for $\mathcal{F}_{\mathcal{P}}$, the set of forbidden induced subgraphs for $\mathcal{P}$, and denote $\mathcal{F}_k$ and $\Psi_{\mathcal{F}}(k)$ as in Definitions 3.7 and 3.8 above.

Define the following functions of r:

$$\alpha(r) = \delta_{2.5}(\Psi_{\mathcal{F}}(r),\ \gamma_{2.2}(\varepsilon/10,\ \Psi_{\mathcal{F}}(r))), \tag{4}$$

$$\beta(r) = \alpha(r) \cdot \gamma_{2.2}(\varepsilon/10,\ \Psi_{\mathcal{F}}(r)), \tag{5}$$

and

$$\mathcal{E}(r) = \begin{cases} \varepsilon/10, & r = 0 \\ \min\{\beta(r),\ \varepsilon/10\}, & r \geq 1 \end{cases} \tag{6}$$

Set $m = 10/\varepsilon$ and $S(\varepsilon) = S_{2.7}(10/\varepsilon, \mathcal{E})$, hence $S(\varepsilon)$ is a function of ε and $\mathcal{P}$ only. We also set

$$n_{1.1}(\mathcal{P}, \varepsilon) = (2/\varepsilon)^{5/2} S(\varepsilon)^5 \tag{7}$$

and assume $n > n_{1.1}(\mathcal{P}, \varepsilon)$.

Consider G^* - the furthest graph on n vertices from $\mathcal{P}$. Assume its edge density is p. We shall prove the theorem for this p. Pick $G = G(n,p)$, and denote the closest graph to G in $\mathcal{P}$ by $G' \in \mathcal{P}$. Assume that G satisfies the conditions of Lemma 3.1 for $t(n) = n^{1.6}$, which indeed happens with high probability.

We apply Lemma 2.7 to G' with $\mathcal{E}$ and m as above. Lemma 2.7 provides a partition of $V(G') = V(G)$ into $\frac{10}{\varepsilon} \leq k \leq S(\varepsilon)$ clusters $V_1, \ldots, V_k$ (given by item (1) in Lemma 2.7). By item (2) of Lemma 2.7, for $1 \leq i \leq k$ we have sets $U_i \subseteq V_i$ each of size at least $n/S(\varepsilon)$. Also, by item (3) of Lemma 2.7, **every** pair (U_i, U_j) is $\beta(k)$-regular (recall that $\mathcal{E}(k) \leq \beta(k)$).

We now know the value of k, and apply Lemma 2.5 k times on the subgraphs induced in G' by each U_i, with $\ell = \Psi_{\mathcal{F}}(k)$ and $\gamma = \gamma_{2.2}(\varepsilon/10,\ \Psi_{\mathcal{F}}(k))$ in order to obtain the appropriate sets $W_{i,1}, \ldots, W_{i,\Psi_{\mathcal{F}}(k)} \subset U_i$, all of size at least $\alpha(k)|U_i|$. The following observation will be useful for the rest of the proof:

Claim 3.9. *All the pairs $(W_{i,i'}, W_{j,j'})$ are $\gamma_{2.2}(\varepsilon/10, \Psi_{\mathcal{F}}(k))$-regular. Also, if $i \neq j$ then we have $|d_{G'}(W_{i,i'}, W_{j,j'}) - d_{G'}(U_i, U_j)| \leq \varepsilon/10$.*

Proof: Consider first pairs that belong to the same set U_i. In this case, the fact that any pair $(W_{i,i'}, W_{i,j'})$ is $\gamma_{2.2}(\varepsilon/10, \Psi_{\mathcal{F}}(k))$-regular follows immediately from our choice of these sets, as we applied Lemma 2.5 on each set U_i with $\gamma = \gamma_{2.2}(\varepsilon/10,\ \Psi_{\mathcal{F}}(k))$. Consider now pairs that belong to different sets U_i, U_j. As was mentioned above, any pair (U_i, U_j) is $\beta(k)$-regular. Since each set $W_{i,j}$ satisfies $|W_{i,j}| \geq \alpha(k)|U_i|$, we get from Claim 3.2 and the definition of $\beta(k)$ that any pair $(W_{i,i'}, W_{j,j'})$ is at least $\max\{2\beta(k), \beta(k)/\alpha(k)\} \leq \gamma_{2.2}(\varepsilon/10, \Psi_{\mathcal{F}}(k))$-regular (here we use the fact that $\alpha(k) \leq 1/2$, which is guaranteed by Comment 2.6). Finally, as each of the sets $W_{i,j}$ satisfies $|W_{i,j}| \geq \alpha(k)|U_i| \geq \beta(k)|U_i| \geq \mathcal{E}(k)|U_i|$ we get from the fact that each pair (U_i, U_j) is $\mathcal{E}(k)$-regular that $|d_{G'}(W_{i,i'}, W_{j,j'}) - d_{G'}(U_i, U_j)| \leq \mathcal{E}(k) \leq \varepsilon/10$, thus completing the proof.

■

Step 2: Obtaining G''

We obtain from G' a new graph G'' (the "clean" version of G') by modifying the following edges, in the following order:

1. For $1 \leq i < j \leq k$ such that $|d_{G'}(V_i, V_j) - d_{G'}(U_i, U_j)| > \frac{\varepsilon}{10}$, for all $v \in V_i$ and $v' \in V_j$ the pair (v, v') becomes an edge if $d_{G'}(U_i, U_j) \geq \frac{1}{2}$, and becomes a non-edge if $d_{G'}(U_i, U_j) < \frac{1}{2}$. By item (4) of Lemma 2.7 there are no more than $\mathcal{E}(0)\binom{k}{2} = \frac{\varepsilon}{10}\binom{k}{2}$ such $1 \leq i < j \leq k$, hence we modify less than $\frac{\varepsilon}{10}\binom{k}{2}(\frac{n}{k})^2 < \frac{\varepsilon}{10}n^2$ edges.

2. For $1 \leq i < j \leq k$ such that $d_{G'}(U_i, U_j) < \frac{2}{10}\varepsilon$, all edges between V_i and V_j are removed. For all $1 \leq i < j \leq k$ such that $d_{G'}(U_i, U_j) > 1 - \frac{2}{10}\varepsilon$, all non-edges between V_i and V_j become edges. In this stage, if $d_{G'}(U_i, U_j) < \frac{2}{10}\varepsilon$, then by the modifications made in the first stage, we have $d_{G'}(V_i, V_j) < \frac{3}{10}\varepsilon$. Similarly, if $d_{G'}(U_i, U_j) > 1 - \frac{2}{10}\varepsilon$ then $d_{G'}(V_i, V_j) > 1 - \frac{3}{10}\varepsilon$. Thus, in this stage we make at most $\binom{k}{2}\frac{3}{10}\varepsilon\frac{n^2}{k^2} < \frac{3}{10}\varepsilon n^2$ changes.

3. If for a fixed i all densities of pairs from $W_{i,1}, \ldots, W_{i,\Psi_{\mathcal{F}}(k)}$ are less than $\frac{1}{2}$, all edges within V_i are removed. Otherwise, all the above densities are at least $\frac{1}{2}$ (by the choice of $W_{i,1}, \ldots, W_{i,\Psi_{\mathcal{F}}(k)}$ through Lemma 2.5), in which case all non-edges within V_i become edges. In this stage we apply at most $k\binom{n/k}{2} < \frac{n^2}{k}$ changes. By our choice of m for Lemma 2.7, we have $k > m = \frac{10}{\varepsilon}$ and hence we applied at most $\frac{\varepsilon}{10}n^2$ changes at this stage.

Summing the above, we conclude that altogether the number of edge modifications satisfies:

Claim 3.10. $\Delta(G', G'') < \frac{\varepsilon}{2}n^2$

We also explicitly note the following relations between the edge densities of sets in G' and G''. These are straight-forward results of the above construction, together with the properties of U_i and V_i guaranteed by Lemma 2.7 and Claim 3.9.

Claim 3.11. *The following hold in G' and G'':*

1. *For any $1 \leq i \leq k$, and any $1 \leq i' < j' \leq \Psi_{\mathcal{F}}(k)$, either $d_{G''}(W_{i,i'}, W_{i,j'}) = 1$ and $d_{G'}(W_{i,i'}, W_{i,j'}) \geq \frac{1}{2}$ or $d_{G''}(W_{i,i'}, W_{i,j'}) = 0$ and $d_{G'}(W_{i,i'}, W_{i,j'}) \leq \frac{1}{2}$.*

2. *For any $1 \leq i < j \leq k$, and $1 \leq i', j' \leq \Psi_{\mathcal{F}}(k)$, exactly one of the following holds:*

   (a) *$d_{G''}(V_i, V_j) = 1$ and $d_{G'}(W_{i,i'}, W_{j,j'}) \geq \frac{\varepsilon}{10}$.*

   (b) *$d_{G''}(V_i, V_j) = 0$ and $d_{G'}(W_{i,i'}, W_{j,j'}) \leq 1 - \frac{\varepsilon}{10}$.*

   (c) *$\frac{\varepsilon}{10} \leq d_{G''}(V_i, V_j) \leq 1 - \frac{\varepsilon}{10}$ and $\frac{\varepsilon}{10} \leq d_{G'}(W_{i,i'}, W_{j,j'}) \leq 1 - \frac{\varepsilon}{10}$.*

Step 3: The colored regularity graph K, and $\mathcal{P}_{K,n}$

We define a colored regularity graph K on the vertices $\{1, \ldots, k\}$ which models the structure of G'' as follows. We color $i \in V(K)$ white if V_i is edgeless in G''. Otherwise, V_i spans a complete graph in G'' and we color $i \in V(K)$ black. If $d_{G''}(V_i, V_j) = 0$ we color (i, j) white, if $d_{G''}(V_i, V_j) = 1$ we color (i, j) black, otherwise (i.e. $\frac{\varepsilon}{10} \leq d_{G''}(V_i, V_j) \leq 1 - \frac{\varepsilon}{10}$) we color (i, j) grey.

Consider the property $\mathcal{P}_{K,n}$ as in Definition 3.5, where $G'' \in \mathcal{P}_{K,n}$.

Claim 3.12. $\mathcal{P}_{K,n} \subset \mathcal{P}$

Proof. Let J be a graph in $\mathcal{P}_{K,n}$. Assume, towards a contradiction, that J contains an induced copy of a forbidden graph $F' \in \mathcal{F}$. We shall show that in this case G' must also contain a graph from $\mathcal{F}$, which contradicts our assumption that $G' \in \mathcal{P}$.

Fix some equipartition witnessing $J \in \mathcal{P}_{K,n}$. Consider mapping each vertex v in the induced copy of F' in J to $i \in \{1, 2, \ldots, k\}$ such that v belongs to the i'th cluster in the equipartition. By the definition of $\mathcal{P}_{K,n}$, this mapping shows that $F' \mapsto_c K$ and hence $K \in \mathcal{F}_k$. The definition of $\Psi_{\mathcal{F}}(k)$ guarantees that there is some $F \in \mathcal{F}$ such that $f = |V(F)| \leq \Psi_{\mathcal{F}}(k)$ and $F \mapsto_c K$. Denote the vertex set of F by $\{1, 2, \ldots, f\}$, and let $\varphi : V(F) \mapsto V(K)$ be the colored homomorphism from F to K. We will now consider the sets $W_{\varphi(1),1}, \ldots, W_{\varphi(f),f}$ and show that by applying Lemma 2.2 on those sets we obtain an induced copy of F in G'. By our construction of K, and the definition of colored homomorphism, we conclude from Claim 3.11 that

- If $(i,j) \in E(F)$ then $d_{G'}(W_{\varphi(i),i}, W_{\varphi(j),j}) \geq \frac{\varepsilon}{10}$.
- If $(i,j) \notin E(F)$ then $d_{G'}(W_{\varphi(i),i}, W_{\varphi(j),j}) \leq 1 - \frac{\varepsilon}{10}$.

Moreover, by Claim 3.9 any pair of these sets is at least $\gamma_{2.2}(\frac{\varepsilon}{10}, \Psi_{\mathcal{F}}(k))$-regular in G'. Thus, by Lemma 2.2 indeed G' contains an induced copy of F, which completes the proof of the claim.

■

Step 4: The edit distance of G^* from $\mathcal{P}_{K,n}$

Claim 3.13. *With high probability, $E_{\mathcal{P}_{K,n}}(G^*) \leq E_{\mathcal{P}_{K,n}}(G) + \frac{\varepsilon}{2}n^2$.*

Proof. Consider a uniformly random equipartition of the vertices of G^* into k sets. The expected number of edge modifications one needs to apply to G^* so that this partition would witness membership in $\mathcal{P}_{K,n}$ is

$$D = (1-p)\binom{\frac{n}{k}}{2}|VB(K)| + p\binom{\frac{n}{k}}{2}|VW(K)| + (1-p)\Big(\frac{n}{k}\Big)^2|EB(K)| + p\Big(\frac{n}{k}\Big)^2|EW(K)|$$

It is therefore witnessed by some equipartition that $E_{\mathcal{P}_{K,n}}(G^*) \leq D$.

Yet, by Lemma 3.1, for *any* equipartition of G into k sets, the number of edge modifications needed for it to witness membership in $\mathcal{P}_{K,n}$ is at least $D - t(n)k^2$. By our assumption on n in (7), we have $t(n)k^2 \leq n^{1.6}n^{2/5}\frac{\varepsilon}{2} = \frac{\varepsilon}{2}n^2$, and indeed w.h.p. $D \leq E_{\mathcal{P}_{K,n}}(G) + \frac{\varepsilon}{2}n^2$.

■

Hence, with high probability:

$$\begin{aligned} ed(n,\mathcal{P}) &= E_{\mathcal{P}}(G^*) \leq E_{\mathcal{P}_{K,n}}(G^*) \leq E_{\mathcal{P}_{K,n}}(G) + \frac{\varepsilon}{2}n^2 \\ &\leq \Delta(G,G') + \Delta(G',G'') + \frac{\varepsilon}{2}n^2 \\ &\leq E_{\mathcal{P}}(G) + \varepsilon n^2 \end{aligned}$$

■

Remark 3.14. *Note that along the proof of Theorem 1.1 the only dependence on G is in the proof of Claim 3.13. Therefore, any pseudo-random graph with edge density p that satisfies the conditions of Lemma 3.1 for some $t(n) = o(n^2)$, can play the role of $G(n,p)$.*

4 Proof of Theorem 1.2

It would be more convenient to express the intermediate results along the proof of Theorem 1.2 in terms of the expected edit distance of the random graph from $\mathcal{P}$. We will later show that the edit distance of the random graph is concentrated near its expected value, hence this relaxation is possible. We thus 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 $e_{n,p}(\mathcal{P})$.

Let us first rephrase Theorem 1.1 as follows. Clearly, we may assume that the function $n_{1.1}(\mathcal{P},\varepsilon)$ is monotone non-increasing in ε. Moreover, since the value of $E_{\mathcal{P}}(G)$ is always bounded between 0 and $\binom{n}{2}$, Theorem 1.1 also implies the following for the expected value of $E_{\mathcal{P}}(G(n,p))$.

Corollary 4.1. *Let $\mathcal{P}$ be an arbitrary graph property, and $\varepsilon > 0$. Then any $n \geq n_{1.1}(\mathcal{P},\varepsilon)$ and $p = p_{1.1}(\mathcal{P},\varepsilon/2,n)$ as above, also satisfy*

$$ed(n,\mathcal{P}) \leq \mathbb{E}\Big[E_{\mathcal{P}}\Big(G(n,p)\Big)\Big] + \varepsilon n^2 = e_{n,p}(\mathcal{P})\binom{n}{2} + \varepsilon n^2$$

The main difficulty arising when one tries to base a proof of Theorem 1.2 on Theorem 1.1 is finding a single value of p which suits any ε and any (large enough) n. We overcome this difficulty by showing that the expected edit distance of random graphs from a hereditary property is continuous in the following senses:

1. The exact size of the graph n has a limited influence on the edit distance, i.e. if n_1 and n_2 are close, then $e_{n_1,p}$ and $e_{n_2,p}$ are close.
2. A small change in p causes a small change in the edit distance, i.e. if p_1 and p_2 are close, then e_{n,p_1} and e_{n,p_2} are close.

The first ingredient of the continuity results from the following lemma.

Lemma 4.2. *For any pair of integers $m < n$, a hereditary property $\mathcal{P}$ and $p \in [0,1]$:*

$$e_{m,p}(\mathcal{P}) \leq e_{n,p}(\mathcal{P})$$

Proof. We pick a random graph G_m in $G(m,p)$ by first obtaining G_n from $G(n,p)$, and then choosing a random subset of m of its vertices with uniform distribution. We can now edit G_m as follows: consider changing G_n into a graph satisfying $\mathcal{P}$, and apply to G_m the modifications which fall into the subgraph that G_m induces. Since $\mathcal{P}$ is hereditary, the modified m-vertex graph is also a member of $\mathcal{P}$. The expected number of modifications applied to edges of G_m this way is

$$\frac{\binom{m}{2}}{\binom{n}{2}}\,\mathbb{E}\Big[E_{\mathcal{P}}(G_n)\Big] = \binom{m}{2} e_{n,p}(\mathcal{P})$$

Hence the expected fraction of modifications needed in order to turn G_m into a graph in $\mathcal{P}$ is *at most* $e_{n,p}(\mathcal{P})$.

■

Thus, for an arbitrary hereditary property, $e_{1,p}$, $e_{2,p}$,... form a bounded monotone non-decreasing sequence.

Corollary 4.3. *For any hereditary property $\mathcal{P}$ and $p \in [0,1]$, the limit $\lim_{n\to\infty} e_{n,p}(\mathcal{P})$ exists.*

We denote this limit by $\boldsymbol{e_p(\mathcal{P})} := \lim_{n\to\infty} e_{n,p}(\mathcal{P})$.

We now address the second component of the continuity.

Lemma 4.4. *Let $\mathcal{P}$ be an arbitrary hereditary property, and $p_1, p_2 \in [0,1]$, then for any integer n:*

$$|e_{n,p_1}(\mathcal{P}) - e_{n,p_2}(\mathcal{P})| \leq |p_1 - p_2|$$

Proof. W.l.o.g., assume $e_{n,p_1} < e_{n,p_2}$. For a graph $G = G(n,p_2)$, we apply the following two steps in order to modify its edges turning it into a graph satisfying $\mathcal{P}$:

1. If $p_1 \leq p_2$, we decrease the edge density of G by randomly, independently, removing every edge with probability $\frac{p_2-p_1}{p_2}$. If $p_1 > p_2$, we increase the edge density of G by adding any non-edge of $G(n,p_2)$ with probability $\frac{p_1-p_2}{1-p_2}$. In both cases we obtain the graph distribution of $G(n,p_1)$.

2. We now turn it into a graph in $\mathcal{P}$ in the most economical way.

The expected total number of edge modifications is $(|p_2-p_1|+e_{n,p_1})\binom{n}{2}$. Thus $e_{n,p_2} \leq |p_2-p_1|+e_{n,p_1}$.

■

Taking limits, this implies the following.

Corollary 4.5. *For any hereditary property $\mathcal{P}$ and $p_1, p_2 \in [0,1]$*

$$|e_{p_1}(\mathcal{P}) - e_{p_2}(\mathcal{P})| \leq |p_1 - p_2|$$

For the proof of Theorem 1.2, we integrate the two ingredients of the continuity of $e_{n,p}$ as follows.

Lemma 4.6. *Let $\mathcal{P}$ be an arbitrary hereditary property, and $\eta > 0$. Then there is $n_{4.6}(\mathcal{P},\eta)$ such that for any $n > n_{4.6}(\mathcal{P},\eta)$ and any $p \in [0,1]$: $|e_p(\mathcal{P}) - e_{n,p}(\mathcal{P})| < \eta$.*

Proof. Let $M = \lfloor 4/\eta \rfloor$, and define $q_1, \ldots, q_M \in [0,1]$ by $q_i = i(\frac{\eta}{4})$. By Corollary 4.3, there are integers $\ell_1, \ldots, \ell_M$ such that for every $1 \leq i \leq M$: any $\ell > \ell_i$ satisfies $|e_{\ell,q_i} - e_{q_i}| < \frac{\eta}{2}$. Define $n_{4.6}(\mathcal{P},\eta) := \max\{\ell_i \mid 1 \leq i \leq M\}$.

Suppose $n > n_{4.6}(\mathcal{P}, \eta)$, and $p \in [0,1]$. Then there is some $1 \leq i \leq M$ such that $|p - q_i| \leq \frac{\eta}{4}$. Thus, by Lemma 4.4 and Corollary 4.5:

$$|e_p - e_{n,p}| \leq |e_p - e_{q_i}| + |e_{q_i} - e_{n,q_i}| + |e_{n,q_i} - e_{n,p}| < \frac{\eta}{4} + \frac{\eta}{2} + \frac{\eta}{4} \leq \eta$$

■

We will now formulate the last tool for the proof of Theorem 1.2, which is the concentration of the edit distance. Note that for any graph property $\mathcal{P}$, the function $G \mapsto E_{\mathcal{P}}(G)$ satisfies the edge Lipschitz condition, i.e. whenever G and G' differ in at most one edge, then $|E_{\mathcal{P}}(G) - E_{\mathcal{P}}(G')| \leq 1$. The following well known result enables the use of martingales for such graph theoretic functions:

Theorem 4.7. ***(Theorem 7.2.3 in [5])*** *When a graph parameter satisfies the edge Lipshitz condition, the corresponding edge exposure martingale, (as defined in [5], Chapter 7) satisfies* $|X_{i+1} - X_i| \leq 1$.

It therefore follows from Azuma's inequality, applied to the edge exposure martingale of the random graph $G(n,p)$, that the following holds.

Lemma 4.8. *Let $\mathcal{P}$ be a hereditary property and $p \in [0,1]$. Then for any n, and $\lambda \geq 0$:*

$$\Pr\left[\mid \mathbb{E}[E_{\mathcal{P}}(G(n,p))] - E_{\mathcal{P}}(G(n,p)) \mid > \lambda\sqrt{\tbinom{n}{2}} \right] \leq 2e^{-\frac{\lambda^2}{2}}$$

Proof of Theorem 1.2 :

For $i = 1, 2, \ldots$ define $\varepsilon_i := \inf\{2\delta \mid n_{1.1}(\mathcal{P}, \delta) \leq i\}$. Since $n_{1.1}(\mathcal{P}, \delta)$ is monotone non-increasing in δ, it follows that $\{\varepsilon_i\}$ is also monotone non-increasing and converges to 0. For any integer i, we apply Theorem 1.1 with $\mathcal{P}$, $\frac{\varepsilon_i}{2}$, $i \geq n_{1.1}(\mathcal{P}, \frac{\varepsilon_i}{2})$ and obtain $p_i = p_{1.1}(\mathcal{P}, \frac{\varepsilon_i}{2}, i)$. The bounded sequence $p_1, p_2, \ldots$ has a convergent subsequence $p_{i_1}, p_{i_2}, \ldots$ with a limit $p := \lim_{k\to\infty} p_{i_k}$. We shall prove that this p satisfies the condition of Theorem 1.2.

Given an arbitrarily small $\varepsilon > 0$, consider a large enough i_k such that: (i) $|p_{i_k} - p| \leq \frac{\varepsilon}{5}$, (ii) $\varepsilon_{i_k} \leq \frac{\varepsilon}{10}$, (iii) $i_k \geq n_{4.6}(\mathcal{P}, \frac{\varepsilon}{5})$. For $n \geq i_k$ it now follows from the application of Lemma 1.1 and Corollary 4.1 that:

$$ed(n, \mathcal{P}) \leq \binom{n}{2} e_{n,p_n}(\mathcal{P}) + \varepsilon_n n^2 \leq \binom{n}{2} e_{n,p_n}(\mathcal{P}) + \frac{\varepsilon}{5} n^2$$

By Lemma 4.6 $|e_{p_n} - e_{i_k,p_n}| < \frac{\varepsilon}{5}$, and by Lemma 4.2 $e_{i_k,p_n} \leq e_{n,p_n} \leq e_{p_n}$. Therefore, $|e_{i_k,p_n} - e_{n,p_n}| < \frac{\varepsilon}{5}$ and

$$ed(n, \mathcal{P}) \leq \binom{n}{2} e_{i_k,p_n}(\mathcal{P}) + \frac{2\varepsilon}{5} n^2$$

Yet Corollary 4.1 also implies $\binom{i_k}{2} e_{i_k,p_n}(\mathcal{P}) \leq ed(i_k, \mathcal{P}) \leq \binom{i_k}{2} e_{i_k,p_{i_k}}(\mathcal{P}) + \varepsilon_{i_k} i_k^2$ and hence $|e_{i_k,p_n} - e_{i_k,p_{i_k}}| \leq \frac{\varepsilon}{5}$ and

$$ed(n, \mathcal{P}) \leq \binom{n}{2} e_{i_k,p_{i_k}}(\mathcal{P}) + \varepsilon_{i_k} n^2 + \frac{2\varepsilon}{5} n^2 \leq \binom{n}{2} e_{i_k,p_{i_k}}(\mathcal{P}) + \frac{3\varepsilon}{5} n^2$$

Now, by Lemma 4.4 for p and p_{i_k} we get

$$ed(n,\mathcal{P}) \leq \binom{n}{2} e_{i_k,p}(\mathcal{P}) + |p_{i_k} - p| \binom{n}{2} + \frac{3\varepsilon}{5} n^2 \leq \binom{n}{2} e_{i_k,p}(\mathcal{P}) + \frac{4\varepsilon}{5} n^2 \leq \binom{n}{2} e_{n,p}(\mathcal{P}) + \frac{4\varepsilon}{5} n^2$$

Moreover, by Lemma 4.8, the probability that $E_{\mathcal{P}}(G(n,p))$ is $\frac{\varepsilon}{5}n^2$-far from its expected value is at most $2e^{-\frac{\varepsilon^2 n^2}{25}}$ which tends to 0 as n grows. Hence indeed, with probability tending to 1, $ed(n,\mathcal{P}) \leq E_{\mathcal{P}}(G(n,p)) + \varepsilon n^2$.

■

5 Concluding remarks and future work

- The natural question arising from Theorem 1.2 is whether one can determine the value of the extremal probability $p(\mathcal{P})$ for some (or all !) hereditary properties. In a future work, we determine $p(\mathcal{P})$ for several interesting families of hereditary properties:

  Sparse hereditary properties: If there are at most $2^{o(n^2)}$ graphs on n vertices satisfying $\mathcal{P}$, then the edit distance $ed(n,\mathcal{P})$ is either $(1-o(1))\binom{n}{2}$ or $(\frac{1}{2}-o(1))\binom{n}{2}$, the extremal probability $p(\mathcal{P})$ is either $0, \frac{1}{2}$ or 1, and there is a simple criterion to decide which of these is the correct value.

  Self complementary properties: We say that a property $\mathcal{P}$ is self-complementary if for any graph G, $G \in \mathcal{P}$ if and only if $\overline{G} \in \mathcal{P}$. For example, perfect graphs form such a property. We show that $p(\mathcal{P})$ for these properties equals $\frac{1}{2}$, extending the result of [7] for $\mathcal{P}^*_H$ where $H = \overline{H}$.

  (r,s)-colorability: For a pair of integers (r,s), the property $\mathcal{P}_{r,s}$ consists of all graphs whose vertices can be partitioned into $r+s$ sets, r of them spanning empty graphs and s spanning complete graphs. We define explicit functions $d(r,s)$ and $c(r,s)$ and prove that $p(\mathcal{P}_{r,s}) = d(r,s)$ and $ed(n,\mathcal{P}_{r,s}) = (c(r,s)-o(1))\binom{n}{2}$.

  Induced H-freeness: We prove that for the claw $K_{1,3}$, $p(\mathcal{P}^*_{K_{1,3}}) = \frac{1}{3}$ and $ed(n,\mathcal{P}^*_{K_{1,3}}) = (\frac{1}{3}-o(1))\binom{n}{2}$, thus showing that for some H, $p(\mathcal{P}^*_H) \neq \frac{1}{2}$. We achieve similar asymptotic results for other small graphs, and sharpen the estimation of $ed(n,\mathcal{P}^*_{C_4})$ and $ed(n,\mathcal{P}^*_{P_4})$.

- 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 $G(n,p(\mathcal{P}))$? (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. [6], [15]). What is the analogue 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 [21]. Their results might hint on possible answers.

- In [4], 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 [4] also characterize the properties for which the above mentioned algorithm achieves essentially the best possible approximation, that is, the monotone properties $\mathcal{P}$ for which it is NP-hard to approximate $E_{\mathcal{P}}(G)$ 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.

Acknowledgement: Part of this research was carried out during a visit of the second author at the IAS, Princeton, and he would like to thank his hosts at the IAS for their hospitality.

References

[1] V. E. Alekseev, Range of values of entropy of hereditary classes of graphs, Discrete Math. Appl., 3 (1993), 191-199.

[2] N. Alon, E. Fischer, M. Krivelevich and M. Szegedy, Efficient testing of large graphs, Proc. of 40^{th} FOCS, New York, NY, IEEE (1999), 656–666. Also: Combinatorica 20 (2000), 451-476.

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

[4] N. Alon, A. Shapira and B. Sudakov, Additive approximation for edge-deletion problems, Proc. of the 46 IEEE FOCS, IEEE (2005), 419-428.

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

[6] 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.

[7] M. Axenovich, A. Kézdy and R. Martin, The editing distance in graphs, to appear.

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

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

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

[11] 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 Springer-Verlag (1997), 70-78.

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

[13] C. Borgs, J. Chayes, L. Lovász, V.T. Sós, B. Szegedy and K. Vesztergombi, Graph limits and parameter testing, Proc. of STOC 2006, to appear.

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

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

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

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

[18] L. Lovász and B. Szegedy, Graph limits and testing hereditary graph properties, preprint.
http://research.microsoft.com/users/lovasz/heredit-test.pdf

[19] L. Lovász and B. Szegedy, Limits of dense graph sequences, preprint.
http://research.microsoft.com/users/lovasz/limits.pdf

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

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

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

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

[24] 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.

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