HEREDITARY PROPERTIES, QUASI-RANDOM GRAPHS AND INDUCED SUBGRAPHS

Miklós Simonovits and Vera T. Sós
Mathematical Institute of the
Hungarian Academy of Sciences, Budapest

[1]

Abstract

This is a continuation of our work on quasi-random graph properties. The class of quasi-random graphs is defined by certain equivalent graph properties possessed by random graphs. One of the most important of these properties is that every fixed sample graph L_ν, has the same frequency in G_n and in each not too small induced subgraph F_h of G_n as in the p-random graph. (This holds for induced and not necessarily induced containment.) Earlier we proved for *not necessarily induced* subgraphs the converse assertion: if the frequency of just one fixed L_ν in large induced subgraphs $F_h \subseteq G_n$ is the same as for the random graphs, then (G_n) is quasi-random. Here we shall investigate the analogous problem for *induced* subgraphs L_ν. In such cases it may happen that (G_n) is not quasi-random but the union of two quasi-random graph sequences (with distinct attached probabilities.) So we are interested in the following question:

> For which graphs L is it true that if the number of induced copies of L in every induced $F_h \subseteq G_n$ is asymptotically the same as in a p-random graph,[2] then (G_n) is the union of (at most) two quasi-random graph sequences.

[1]Research supported by the Hungarian Research Council Grants, T026069, T038210, and OTKA T032236.

[2]i.e., up to an error term $o(n^{v(L)})$.

We shall prove that if, e.g., L is a regular graph, then this is the case. We shall reduce the problem to solving a system of polynomials. This gives a "simple" algorithm to decide the problem for every given L_ν.

1. NOTATION

We shall use notations that are mostly standard. For a (simple) graph G, $v(G)$ and $e(G)$ denote the *number* of vertices and edges, $V(G)$ and $E(G)$ denote the *set* of vertices and edges respectively. The (first) subscript *in case of graphs* will almost always denote the number of vertices. If $X \subseteq V(G)$, then $e(X)$ denotes the number of edges of the subgraph induced by X, and $G[X]$ denotes the subgraph of G induced by X. Given two disjoint subsets $X, Y \subseteq V(G)$, then $e(X, Y)$ denotes the number of edges joining X and Y.

Mostly we shall have a *sample graph* $L = L_\nu$ with ν vertices, ($V(L) = \{a_1, a_2, \dots, a_\nu\}$), and a graph G with some copies of L. The vertices of a copy $L \subseteq G$ will typically be denoted by $\{b_1, b_2, \dots, b_\nu\}$.

- A *not necessarily induced* (abbreviated to NNI) labelled copy is given by a function $\psi : V(L) \to V(G)$ mapping different a_i's into different $b_i := \psi(a_i)$'s, where we assume (only) that if $(a_i, a_j) \in E(L)$, then $(\psi(a_i), \psi(a_j)) \in E(G)$. Denote by $\mathbf{N}(L \subseteq G)$ the number of labelled *not necessarily induced* copies of L in G.
- A *labelled induced* copy of $L \subseteq G$ is given by a function $\psi : V(L) \to V(G)$ mapping different a_i's into different b_i's, where $(\psi(a_i), \psi(a_j)) \in E(G)$ iff $(a_i, a_j) \in E(L)$. Denote the number of labelled *induced* copies of $L \subseteq G$ by $\mathbf{N}^*(L \subseteq G)$. If we wish to *emphasize* that $L \subseteq G$ is an induced graph, we shall write $L \overset{*}{\subseteq} G$.

We shall use $u_n \sim v_n$ if $u_n/v_n \to 1$ as $n \to \infty$.

The complementary graph of H is denoted by $\overline{H}$.

2. INTRODUCTION

This paper is strongly connected to our previous papers [8, 9]: it is a continuation of [9]. Therefore we give here only a shortened introduction. For a longer one see [9].

One of the important questions of modern mathematics and computer science is, how random-like objects can be generated in nonrandom ways and when an individual event could be considered random, and in which sense.

A. Thomason, [12, 13], P. Frankl, V. Rödl and R. Wilson [6], and F. R. K. Chung, R. L. Graham, and R. M. Wilson [1], gave some characterization of random-like graph sequences.

Our starting point is a theorem of Chung, Graham and Wilson [6]. There some graph properties $\mathbf{P}$ are considered, all possessed by (binomially distributed) random graphs and at the same time equivalent to each other in some well-defined sense. *A graph sequence is called p-quasi-random if it satisfies one of these properties*, (and therefore all the others as well).

Here we need only two of the quasi-random properties. Let $p \in (0,1)$. Let $\nu = v(L)$.[3]

We consider the following property of a graph sequence (G_n):

$\mathbf{P}_1^*(\nu)$: for fixed $\nu \geq 4$, *for all* graphs L_ν

$$\mathbf{N}^*(L_\nu \subseteq G_n) = (1+o(1))n^\nu p^{e(L_\nu)}(1-p)^{\binom{\nu}{2}-e(L_\nu)} \quad \text{as} \quad n \to \infty. \tag{1}$$

Obviously, $\mathbf{P}_1^*(\nu)$ says that the graph G_n contains each graph L_ν of order ν with the same frequency as the p-random graph. Property $\mathbf{P}_1^*(\nu)$ refers to the induced copies. We define the analogous property for "not necessarily induced" (NNI) copies:

$\mathbf{P}_1(\nu)$: for fixed $\nu \geq 4$, *for all* graphs L_ν

$$\mathbf{N}(L_\nu \subseteq G_n) = (1+o(1))n^\nu p^{e(L_\nu)} \quad \text{as} \quad n \to \infty. \tag{2}$$

Trivially, the above two properties are equivalent for fixed ν and p.

According to the Chung-Graham-Wilson theorem, both $\mathbf{P}_1(\nu)$ and $\mathbf{P}_1^*(\nu)$ are quasi-random properties. This implies

Corollary 2.1. *If (1) – or (2) – holds for a given $\nu \geq 4$ for every graph L_ν (of ν vertices), then it holds for arbitrary other graphs L_μ, (for arbitrary $\mu \geq 3$), e.g.,*

$$\mathbf{N}^*(L_\mu \subseteq G_n) = (1+o(1))n^\mu p^{e(L_\mu)}(1-p)^{\binom{\mu}{2}-e(L_\mu)}. \tag{3}$$

In [8] we proved that the Szemerédi partition of graphs is crucial the theory of quasi-random graphs.

---

[3] Sometimes we shall use $\eta = e(L_\nu)$, in other cases we shall write $e(L_\nu)$.

Given a graph G, with two disjoint subsets of vertices, X and Y, the edge-density between X and Y is defined as

$$d(X,Y) = \frac{e(X,Y)}{|X||Y|}.$$

Definition 2.2 (ε-Regularity). Given a graph G and two disjoint vertex sets $X, Y \subseteq V(G)$, we shall call the pair (X,Y) ε–regular, if for every $X' \subset X$ and $Y' \subset Y$ satisfying $|X'| > \varepsilon|X|$ and $|Y'| > \varepsilon|Y|$, we have

$$|d(X',Y') - d(X,Y)| < \varepsilon.$$

Our main result in [8] was

Theorem A (Simonovits, T. Sós). *(G_n) is p–quasi–random iff*

> $\mathbf{P}_S(p)$: *For every $\varepsilon > 0$ and κ there exist two integers, $\Omega(\varepsilon,\kappa)$ and $n_0(\varepsilon,\kappa)$ such that, for $n > n_0$, $V(G_n)$ has a partition into k classes $U_1,\dots,U_k$, with $\kappa < k < \Omega(\varepsilon,\kappa)$, $||U_i| - n/k| < \varepsilon n/k$[4] such that for all but at most εk^2 pairs (i,j), $1 \le i < j \le k$,*
>
> $$(U_i,U_j) \text{ is } \varepsilon - \text{regular, and} \qquad |d(U_i,U_j) - p| < \varepsilon.$$

In our previous paper [9] we investigated those properties $\mathbf{P}$ which *do not imply* quasi-randomness of graph sequences (G_n) on their own, but do imply if they are assumed not only for the whole graph G_n but also for every sufficiently large *induced* subgraph $F_h \overset{*}{\subseteq} G_n$. We called such properties *Hereditarily Extended Properties.* To consider such extensions is motivated by the fact that sufficiently large induced subgraphs of random-like graphs must also be random-like: being a random graph is a "hereditary property".

Denote by $\beta_L(p)$ and $\gamma_L(p)$ the "densities" of *labelled induced* and *labelled not necessarily induced* copies of L in a p–random graph, respectively:

$$\beta_L(p) = p^{e(L_\nu)}(1-p)^{\binom{\nu}{2}-e(L_\nu)} \quad \text{and} \quad \gamma_L(p) = p^{e(L_\nu)}. \tag{4}$$

[4]Sometimes we use the error-term 1, here the error term $\varepsilon n/k$, "for historical reasons": they are obviously equivalent in this context.

In [9] we have considered graph sequences for which, for a fixed L_ν,

$$\text{for every induced subgraph} \quad F_h \overset{*}{\subseteq} G_n \qquad (5)$$
$$\mathbf{N}(L_\nu \subseteq F_h) = \gamma_L(p)h^\nu + o(n^\nu).$$

Of course, (5) holds for any sequence of p-random graphs, or, more generally, for any sequence of p-quasi-random graphs (G_n). The question is if (5) implies p-quasi-randomness.

Observe that in (5) we used $o(n^\nu)$ instead of $o(h^\nu)$, i.e. for small values of h we allow a relatively much larger error-term. As soon as $h = o(n)$, condition (5) is automatically fulfilled. One of our main results in [9] was

Theorem B. *Let L_ν be a fixed sample-graph, with $e(L_\nu) > 0$, and $p \in (0,1)$ be fixed. Let (G_n) be a sequence of graphs for which (5) holds. Then (G_n) is p–quasi–random.*

Consequently, (5) holds for every other graph L_μ.

Theorem B means that instead of assuming that for a $\nu \geq 4$, (1) holds *for every graph on ν vertices* it is enough to assume it just for one specific L_ν, but in the stronger, *hereditarily extended sense* of (5). Moreover, Theorem B holds even for $\nu = 3$.

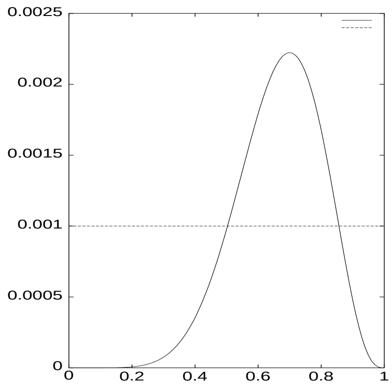
3. NEW RESULTS

The aim of this paper is to investigate phenomena analogous to the one described in Theorem B for the *induced case*, i.e., when (5) is replaced by

$$\text{for every induced subgraph} \quad F_h \overset{*}{\subseteq} G_n \qquad (6)$$
$$\mathbf{N}^*(L_\nu \subseteq F_h) = \beta_L(p)h^\nu + o(n^\nu).$$

We shall see that the situation for the induced case is much more involved, because, if $G_{n,p}$ is a p-random graph, then the expected number of $\mathbf{N}^*(L_\nu \subseteq G_{n,p})$ is not monotone for fixed n while p increases.[5]

[5]Here we use the $\overset{*}{\subseteq}$ in two places: $F_h \overset{*}{\subseteq} G_n$ and $L_\nu \overset{*}{\subseteq} F_h$. They are completely different: the question does not make sense if we replace $F_h \overset{*}{\subseteq} G_n$ by $F_h \subseteq G_n$.



Clearly, $\beta_L(p)$ (in (4)) is a function of p which is monotone increasing in $\left[0, e(L_\nu)/\binom{\nu}{2}\right]$, monotone decreasing in $\left[e(L_\nu)/\binom{\nu}{2}, 1\right]$ and vanishes in $p=0$ and in $p=1$. For every $p \in \left(0, e(L_\nu)/\binom{\nu}{2}\right)$ there is a *unique* probability $\overline{p} \in \left(e(L_\nu)/\binom{\nu}{2}, 1\right)$ yielding the same expected value. Therefore the hereditarily assumed number of induced copies does not determine the probability uniquely, unless $p = e(L_\nu)/\binom{\nu}{2}$.

Definition 3.1. Given a graph L_ν, the probabilities p and $\overline{p}$ are called *conjugate* if $\beta_L(p) = \beta_L(\overline{p})$, i.e.,

$$p^{e(L_\nu)}(1-p)^{\binom{\nu}{2}-e(L_\nu)} = \overline{p}^{e(L_\nu)}(1-\overline{p})^{\binom{\nu}{2}-e(L_\nu)}, \tag{7}$$

and $p \neq \overline{p}$[6].

Obviously, a random graph sequence with edge-probability u satisfies (6) iff $u \in \{p, \overline{p}\}$.

Example 3.2. If $e(L_\nu) = e(\overline{L_\nu})$, then for every p the conjugate probability is $\overline{p} = 1-p$. This is the case, e.g., if L_ν is self-complementary.

Assume that $\mathcal{G} = (G_n)$ is obtained by merging two infinite graph sequences: $\mathcal{G}_1$ being a p-quasi-random and $\mathcal{G}_2$ being a $\overline{p}$-quasi-random one. Then $\mathcal{G}$ satisfies (6) but is not quasi-random (unless $\overline{p} = p$). We shall call this the *case of merged sequences with conjugate probabilities.*

Given a graph L_ν, and a $p \in (0,1)$, we call a graph sequence a *Strong Counterexample* sequence for (L_ν, p) if it satisfies (6) but it is not a quasi-random graph sequence, not even a *merged sequence with conjugate probabilities.*

We will show that there are two reasons for the existence of strong counterexamples:

- There may occur "strange" *algebraic coincidences.*
- There are some *degenerate counterexamples.*

[6]For the "peak" $p = e(L_\nu)/\binom{\nu}{2}$, $p = \overline{p}$: p is "selfconjugate".

Remark 3.3. If (G_n) is a strong counterexample sequence for (L_ν, p), then the same sequence is also a strong counterexample sequence for $\overline{p}$. Further, the complementary graphs, $(\overline{G}_n)$ form a sequence of strong counterexamples for $\overline{L}_\nu$ and $1-p$ (and $1-\overline{p}$).

To formulate our main results, we generalize the notion of random graphs as follows (see [8] for a more general notion of the r-class generalized random graph).

Definition 3.4 (2-class generalized random graph). Define the graph $G_n = G(V_1, V_2, u, v, s)$ as follows: $V(G_n) = V_1 \cup V_2$ (where $V_1 \cap V_2 = \emptyset$). We join independently the pairs in V_1 with probability u, in V_2 with probability v and the pairs (x, y) for $x \in V_1$ and $y \in V_2$ with probability s. We shall call this graph *trivial* if $u = v = s$ and *non-trivial* otherwise.

Remark 3.5. If $u \in (0,1)$ is fixed and $|V_1| > cn$, then almost surely

$$\mathbf{N}^*(L_\nu \subseteq G(V_1, V_2, u, v, s)) - \mathbb{E}(\mathbf{N}^*(L_\nu \subseteq G(V_1, V_2, u, v, s))) = o(n^\nu).$$

So we do not have to distinguish whether we speak of the expected value or of the almost sure value.

Remark 3.6. Assume that $G_n = G(V_1, V_2, u, v, s)$ for $cn < |V_i| < (1-c)n$. If (G_n) satisfies (6) then the two parts $G[V_i]$ form random graphs satisfying (6) and therefore

$$\{u, v\} \subseteq \{p, \overline{p}\}. \tag{8}$$

Our main result is

Theorem 3.7 (Two-class counterexample). *If there is a strong counterexample sequence (G_n) for a fixed sample graph L and for a probability $p \in (0,1)$, then there is also a strong counterexample sequence of form $G_n = G(V_1, V_2, u, v, s)$ $(s \neq u)$ with $|V_1| \sim n/2$, and satisfying (8): $\{u, v\} \subseteq \{p, \overline{p}\}$.*

Remark 3.8. In such cases, i.e., if for some $c^* > 0$, for (L_ν, p) there is a strong counterexample sequence of form $G_n = G(V_1, V_2, u, v, s)$ $(s \neq u)$ with $|V_1| \in (c^* n, (1 - c^* n))$, we shall simply say that $G(n, V_1, V_2, u, v, s)$ is a strong conterexample for (L_ν, p).

The following theorem shows that for (P_3, p) and $(\overline{P}_3, p)$, for some $p \in (0,1)$, there exist strong counterexamples.

Theorem 3.9. *Let $L_\nu = P_3$. Then*

(a) For every $p \geq \frac{1}{\sqrt{3}}$, ($p \neq \frac{2}{3}$) there exists an $s \in [0,1]$, namely,

$$s = s(p) := 3p\frac{1-p}{3p-1} \tag{9}$$

such that the sequence $G_n = G(V_1^{(n)}, V_2^{(n)}, p, p, s)$ is a strong counterexample for P_3, assumed that for some constant $c^* > 0$ we have $|V_1^{(n)}|, |V_2^{(n)}| \geq c^* n$.

(b) Let for P_3, and $p_c = \frac{1}{\sqrt{3}} \approx 0.577$ the conjugate probability be $\overline{p}_c$.[7] *For every $p \leq \overline{p}_c$, ($p \neq \frac{2}{3}$), taking*

$$\overline{s} := 3\overline{p}\frac{1-\overline{p}}{3\overline{p}-1} \in [0,1], \tag{10}$$

the sequence $G_n = G(V_1^{(n)}, V_2^{(n)}, \overline{p}, \overline{p}, \overline{s})$ is a strong counterexample for P_3 assumed that for some constant $c^* > 0$ we have $|V_1^{(n)}|, |V_2^{(n)}| \geq c^* n$.[8]

This means that for $p \in (p_c, \overline{p}_c)$ we have two different strong counterexample sequences.

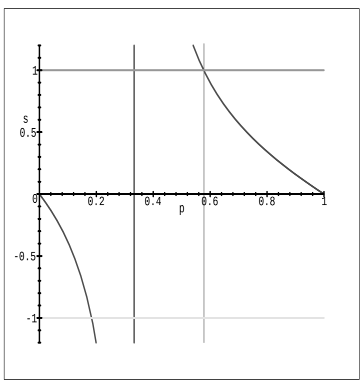


Fig 1. $s(p) := 3p\frac{1-p}{3p-1}$

To understand the situation, consider Figure 1 where one can see that $s = s(p)$ of Theorem 3.9, is negative in $(0, \frac{1}{3})$, then it becomes positive but larger than 1 and becomes a probability only for $p \geq \frac{1}{\sqrt{3}}$.

As examples, we get strong counterexample sequences for $p = \frac{4}{5}$, if $s = \frac{12}{35}$, or for $p = \frac{1}{\sqrt{3}}$, if $s = 1$. The sharpness of this theorem is expressed by Theorem 3.10 below: it asserts that essentially these are the only strong counterexample sequences for P_3:

Theorem 3.10 (Structure of P_3-counterexamples). *If for $\mathcal{G} = (G_n)$*

$$\mathbf{N}^*(P_3 \subseteq F_h) = p^2(1-p)h^3 + o(n^3). \tag{11}$$

holds for every $F_h \overset{*}{\subseteq} G_n$, then $\mathcal{G} = (G_n)$ can be split into four subsequences $\mathcal{G}_i$, where

[7] $\overline{p}_c = -\frac{1}{2\sqrt{3}} + \frac{1}{2} + \frac{\sqrt[4]{3}}{\sqrt{6}} = .7486098314.$

[8] Here $\overline{s}$ is not (necessarily) the conjugate probability of s!

(a) $\mathcal{G}_1$ *is p-quasi-random,*

(b) $\mathcal{G}_2$ *is* $\overline{p}$*-quasi-random,*

(c) For each $G_n \in \mathcal{G}_3$, $V(G_n)$ *can be partitioned into two parts:* $V(G_n) = V_1^n \cup V_2^n$ *so that both* $G_n[V_1^n]$ *and* $G_n[V_2^n]$ *are p-quasi-random,*[9] $d(V_1^n, V_2^n) = s + o(1)$, $s \neq p$, *and* V_1^n *and* V_2^n *are joined* $o(1)$*-regularly;*

(d) $\mathcal{G}_4$ *is like* $\mathcal{G}_3$*, but p and s are replaced by* $\overline{p}$ *and* $\overline{s}$*, respcetively.*

We think that P_3 and $\overline{P}_3$ are exceptional sample graphs:

Conjecture 3.11. *Let* L_ν *be fixed,* $\nu \geq 4$ *and* $p \in (0,1)$*. Let* (G_n) *be a graph sequence satisfying (6). Then* (G_n) *is the union of two sequences, one being p-quasi-random, the other* $\overline{p}$*-quasi-random (where one of these two sequences may be finite, or even empty).*

A possible weakening of Conjecture 3.11 could be that for given L_ν there are only finitely many values of p for which there exist strong counterexample sequences.

We can prove the conjecture only for some special cases.

Theorem 3.12 (Regular Graphs). *Given a regular sample graph* L_ν $(\nu \geq 4)$ *and a probability p, if for a graph sequence* (G_n) *(6) holds, then* (G_n) *is the union of a p-quasi-random and a* $\overline{p}$*-quasi-random graph sequences.*

Theorem 3.13. *Let* L_ν *be a graph,* $\nu = 4$ *or* $L_\nu = K(2,3)$*, and* $p \in (0,1)$*. If for a graph sequence* (G_n) *(6) holds, then* (G_n) *is the union of a p-quasi-random and a* $\overline{p}$*-quasi-random graph sequences.*

Theorem 3.13 will be proved in a continuation of this paper.

As we have mentioned, there is a singular, trivial case of counterexamples of which we would like to forget:

Construction 3.14 (Degenerate Counterexamples). If L_ν is connected, and $L_\nu \neq K_\nu$, and if G_n is the vertex-disjoint union of $\ell_n \geq 2$ complete graphs, then $\mathbf{N}^*(L_\nu \subseteq G_n) = 0$: (G_n) is a sequence of strong counterexamples for L_ν and $p = 0$.

[9] Here $G_n[V_1^n]$ is not a graph of n vertices!

To avoid this and similar counterexamples, we shall always assume that $p \in (0,1)$. We shall also always exclude $e(L_\nu) = 0$ and $e(\overline{L}_\nu) = 0$.

* * *

By Theorem 3.7, to prove that Conjecture 3.11 holds for some specific (L_ν, p), it is enough to prove that there are no two-class generalized random graph counterexamples. As we shall see, *this reduces to proving that some algebraic equations on (u, v, s) have only the trivial solutions $u = v = s$.* So, Theorem 3.7 can often be used to prove that Conjecture 3.11 holds for certain graphs.

If for some $p \in (0,1)$ there exists a counterexample sequence, we may restrict ourselves to the 2-class generalized random graph counterexample sequences $G_n(V_1, V_2, u, v, s)$ and these may be of three different types:

- "Counterexamples of first kind": $G_n(V_1, V_2, p, p, s)$.
- "Counterexamples of the second kind": $G_n(V_1, V_2, \overline{p}, \overline{p}, s)$.
- "Mixed case": $G_n(V_1, V_2, p, \overline{p}, s)$, $(p \neq \overline{p})$.

We shall see that for P_3 there are no "mixed" counterexamples. So Conjecture 3.11 would imply that there are no "mixed" counterexamples at all.

A trivial corollary of Theorem 3.7 is

Algorithm 3.15. *There is a finite algorithm such that if there is no strong counterexample for (L_ν, p), then the algorithm will "prove" this.*

Indeed, one can reduce the problem to deciding if a given system of polynomials has roots in a 3-dimensional cube or not. For more details, see Section 5.4. We do not claim that this algorithm is "efficient".

Remark 3.16. All the theorems of this paper are formulated for labelled graphs (induced or not necessarily induced), however, all our results easily extend to unlabelled graphs.

4. THE COPY-POLYNOMIALS

We shall introduce some polynomials counting the induced copies of L_ν in $F_h \overset{*}{\subseteq} G_n = G(V_1, V_2, u, v, s)$. The simplest way to define them is as follows:

Definition 4.1 (Copy-polynomials). Let $L = L_\nu$ be a fixed "sample graph" and $k = 0, \dots, \nu$. For a fixed k we partition the vertices of L into two classes

A and B with $|A| = k$, $|B| = \nu - k$. Let $\eta = e(L_\nu)$. Then

$$\mathbb{P}^k_{u,v}(s) := \binom{\nu}{k} u^\eta (1-u)^{\binom{\nu}{2}-\eta} - \tag{12}$$
$$\sum_{\substack{A \subseteq V(L_\nu) \\ |A|=k}} u^{e(A)}(1-u)^{\binom{k}{2}-e(A)} v^{e(B)}(1-v)^{\binom{\nu-k}{2}-e(B)} s^{e(A,B)}(1-s)^{k(\nu-k)-e(A,B)}$$

Here the terms of the $\sum$ are the probabilities that if we choose k (labelled) points in V_1 and $\nu - k$ points in V_2, then we get an induced (labelled) L_ν. The first term counts these L_ν if $u = v = s$. The meaning of these polynomials is expressed in

Lemma 4.2 (Copy-polynomials). *Fix an L_ν and a $p \in (0,1)$. Assume that $|V_1|, |V_2| > c^* n$ for some fixed $c^* > 0$. Then a graph sequence (G_n) consisting of 2-class generalized random graphs $G(V_1, V_2, u, v, s)$ satisfies (6) almost surely iff (8) holds: $u, v \in \{p, \overline{p}\}$; further, s is a common zero of the corresponding system of polynomials of (12).*[10],[11]

Generally we shall be interested in the solutions of the system of polynomial equations

$$\mathbb{P}^k_{u,v}(s) = 0, \quad k = 0, \ldots, \nu, \quad \text{where} \quad u \in \{p, \overline{p}\}. \tag{13}$$

We may forget $k = \nu$, since $\mathbb{P}^k_{u,v}(s) = 0$ for all s. It is worth considering the cases $k = 0$ and $k = \nu - 1$ separately. For $k = 0$ we get back (7):

$$u^\eta (1-u)^{\binom{\nu}{2}-\eta} = v^\eta (1-v)^{\binom{\nu}{2}-\eta},$$

where $\eta = e(L_\nu)$, i.e., $v \in \{p, \overline{p}\}$ as well. What is much more important, $\mathbb{P}^{\nu-1}_{u,v}(s)$ does not contain v, since $|B| = 1$. Actually, if $V(L_\nu) = \{a_1, \ldots, a_\nu\}$ and d_i denotes the degree of a_i in L_ν, then (taking $B := \{a_i\}$) we get

Lemma 4.3. *$\mathbb{P}^{\nu-1}_{u,v}(s) = 0$ is equivalent to*

$$1 = \frac{1}{\nu} \sum_i \left(\frac{s}{u}\right)^{d_i} \left(\frac{1-s}{1-u}\right)^{\nu-1-d_i}. \tag{14}$$

Therefore, for a given s we can choose u only in finitely many ways.

---

[10] For a given p we have three choices for $\{u, v\}$ and they are considered as parameters: we have to solve systems of equations consisting of polynomials of one unknown s.

[11] The expression "almost surely" could mean here two different assertions: that for each fixed n "almost surely" or that generating such a generalized random graph for each n, we have the assertion for the obtained sequence of graphs, almost surely. However, here both assertions hold.

Proof of Lemma 4.3. If the degree sequence of L_ν is $(d_1, d_2, \dots, d_\nu)$, then, by $e(B) = 0$, so (12) reduces to

$$\begin{aligned}\mathbb{P}^{\nu-1}_{u,v}(s) &= \nu u^{\frac{1}{2}\sum d_i}(1-u)^{\binom{\nu}{2}-\frac{1}{2}\sum d_i} \\ &- \sum_{i=1}^{\nu} u^{\frac{1}{2}\sum d_i - d_i}(1-u)^{\binom{\nu-1}{2}-\frac{1}{2}\sum d_i + d_i} s^{d_i}(1-s)^{\nu-d_i-1}\end{aligned}$$

So $\mathbb{P}^{\nu-1}_{u,v}(s) = 0$ is equivalent to

$$0 = \nu(1-u)^{\nu-1} - \sum_{i=1}^{\nu} u^{-d_i}(1-u)^{d_i} s^{d_i}(1-s)^{\nu-d_i-1} \tag{15}$$

This proves Lemma 4.3. ∎

The following trivial lemma will be useful:

Lemma 4.4. *If $D = 2e(L_\nu)/\nu$ (i.e., D is the average degree of L_ν), then the conjugacy relation (7) is described by*

$$p^D(1-p)^{\nu-D-1} = \overline{p}^D(1-\overline{p})^{\nu-D-1}. \tag{16}$$

Proof. For $a = p^D(1-p)^{\nu-D-1} > 0$ and $b = \overline{p}^D(1-\overline{p})^{\nu-D-1} > 0$, $\mathbb{P}^0_{u,v}(s) = 0$ is equivalent to

$$a^{\frac{\nu-1}{2}} = b^{\frac{\nu-1}{2}},$$

what is equivalent to $a = b$. ∎

Symmetries of the Copy-polynomials

The vertices of L_ν are labelled. Each $L_\nu \overset{*}{\subseteq} G(V_1, V_2, u, v, s)$ defines a partition of $V(L_\nu)$. The partitions correspond to the 2^ν 0-1 sequences, and $\binom{\nu}{k}$ of them contribute to $\mathbb{P}^k_{u,v}(s)$. Replacing k by $\nu - k$ is equivalent to switching to the complementary set of A. Hence the system of Copy-polynomials is symmetric in the following sense:

$$\mathbb{P}^k_{u,v}(s) = \mathbb{P}^{\nu-k}_{v,u}(s). \tag{17}$$

So, in the *symmetric case*, when $u = v \in \{p, \overline{p}\}$, we have $\lfloor \frac{\nu}{2} \rfloor$ equations.

Motivation of Conjecture 3.11

One motivation for the conjecture is as follows.

If we count the number of equations for the induced case, then mostly we find that the system of polynomials is over-determined. Indeed, generally we have a fixed p which determines $\overline{p}$ and therefore we have to solve the three systems of equations.

Obviously, $u = v = s$ is a solution of (13). We wish to motivate the conjecture that there are no other solutions.

If p is fixed, u and v may have only two values, and then the unknown (variable) s must satisfy the system (13) of $\nu - 2$ Copy-polynomial equations, or in the symmetric case, $\lfloor \frac{\nu}{2} \rfloor$ equations. So for $\nu \geq 4$ we have at least 2 equations for s: more equations than unknowns. And this gets "worse" as ν increases. (On the other hand, as $\nu \to \infty$, the possibilities for L_ν grow exponentially. This could work against the conjecture.)

If for some fixed sample graph L_ν and $p \in (0,1)$ Conjecture 3.11 does not hold, then Theorem 3.7 guarantees that there is a generalized random graph counterexample $(G(V_1, V_2, u, v, s))$ where $|V_1| = |V_2| > cn$, and we know that there are only 3 possibilities for $\{u, v\}$, but earlier we did not know the value of s. By Lemma 4.2, we know that $d(V_1, V_2) = s$ is one of the roots of the "corresponding" copy-polynomials.

4.1. Proof of Lemma 4.2

Take a $G := G(V_1, V_2, u, v, s)$. Let $X \subseteq V_1$, $Y \subseteq V_2$, $|X| = x$, $|Y| = y$.

We think of L_ν as a graph on $a_1, \dots, a_\nu$ and for any of the 2^ν possible 0-1 sequences we have a partition of $V(L_\nu)$ into A and B.

Let us count the expected value S_k of $L_\nu \overset{*}{\subseteq} G[X, Y, u, v, s) \subseteq G$, having k vertices in X, $\nu - k$ vertices in Y.

Put the corresponding k vertices $b_i = \psi(a_i)$ of A into V_1, the others into V_2. The vertices $a_i \in A$ can be put into $X \subseteq V_1$ in $\sim x^k$ ways. [12] The vertices $a_i \in B = V(L_\nu) - A$ can be chosen in $\sim y^{\nu-k}$ ways. Consider the sum in (the second line of)

$$S_k = x^k y^{\nu-k} \times$$

[12] The error comes from that we use $x(x-1)\dots(x-k+1) \sim x^k$.

$$\left(\sum_{\substack{A\subseteq V(L_\nu)\\|A|=k}} u^{e(A)}(1-u)^{\binom{k}{2}-e(A)}v^{e(B)}(1-v)^{\binom{\nu-k}{2}-e(B)}s^{e(A,B)}(1-s)^{k(\nu-k)-e(A,B)}\right)$$
$$+O(n^{\nu-1}).$$

In each term, the first two factors correspond to the probability that for the vertices $a_i \in A$ the images, $b_i = \psi(a_i) \in X \subseteq V_1$ are joined according to the pattern described by $L_\nu[A] \subseteq L_\nu$, the next two factors reflect the probability that $L_\nu[B] \subseteq L_\nu$ is mapped into $G_n[Y]$ appropriately; the last two factors express the probability that $\psi(A) \subseteq X$ is joined correctly to $\psi(B) \subseteq Y$. Here the sum is just the one in definition (12) of $\mathbb{P}^k_{u,v}(s)$.

Condition (6) holds iff for all the possible choices of x, y $(x + y \le n)$,

$$\sum_k S_k = u^\eta(1-u)^{\binom{\nu}{2}-\eta}(x+y)^\nu + o(n^\nu) = \sum_k \binom{\nu}{k} u^\eta(1-u)^{\binom{\nu}{2}-\eta}x^k y^{\nu-k} + o(n^\nu).$$

If $c^* > 0$ is fixed and $x, y > c^* n$, then the $o(n^\nu)$ term is negligible: the two sides are equal iff

$$S_k = \left(\sum_{|A|=k} \dots\right) x^k y^{\nu-k} = \binom{\nu}{k} u^\eta (1-u)^{\binom{\nu}{2}-\eta} x^k y^{\nu-k} \quad \text{for} \quad k = 0, \dots, \nu.$$

This gives just the Copy-polynomial system. ∎

Observe that the above argument also showed that if $\mathbb{E}(.)$ denotes the expected value, then (for $|X|, |Y| > cn$)

$$\mathbb{E}(\mathbf{N}^*(L_\nu \subseteq G(X, Y, u, v, s)) \sim \sum_{k=0}^{\nu} \mathbb{P}^k_{u,v}(s)|X|^k|Y|^{\nu-k}. \tag{18}$$

4.2. Breaking the symmetry

Assume that we wish to prove that for a given L_ν there are no strong counterexamples if $p \in (0, 1)$. Since the condition (6) does not distinguish p from $\overline{p}$, and if there were strong counterexamples, then there would be strong counterexamples of form $(G(V_1, V_2, u, v, s))$ where either u or v would be p or $\overline{p}$, therefore we can assume in our proof that $u = p$: It is enough to check graphs of form $G(V_1, V_2, p, q, s)$, where $q = p$ or $q = \overline{p}$. (This assumption may include that we replaced the original p by $\overline{p}$.)

5. PROOF OF THEOREM 3.7

5.1. Regularity Lemma, Szemerédi Partitions

An important tool in the proof of our theorem is the Szemerédi's Regularity Lemmma which will make possible for us to apply Theorem A to prove that some graph sequences are quasi-random. We have defined the edge-density $d(X,Y)$ and the "regular pairs" in the introduction.

Regularity Lemma (Szemerédi, [10]). *For every $\varepsilon > 0$ and integer κ there exist an $n_0(\varepsilon,\kappa)$ and an $\Omega(\varepsilon,\kappa)$ such that for $n > n_0$ for every graph G_n, the vertex set $V(G_n)$ can be partitioned into k subsets $U_1,\dots,U_k$ with $\kappa < k < \Omega(\varepsilon,\kappa)$ so that $||U_i| - n/k| < 1$ and all but at most εk^2 pairs (U_i,U_j) are ε–regular.*

Such partitions will be called Szemerédi Partitions, κ will be called the *lower bound* on the number of classes, ε the *precision* and (a minimum) $\Omega(\varepsilon,\kappa)$ the *upper bound function.*

Remark 5.1. In most applications of the Regularity Lemma one has to allow that $k = k_n$ depends on n. Here, when for some fixed $p > 0$ the densities are roughly equal: $d(U_i,U_j) = p + o(1)$, one can also choose a *fixed* $k = k_1(\varepsilon,\kappa)$.

Below we formulate a theorem, which says that even if a graph sequence (G_n) is a strong counterexample to Conjecture 3.11 for (L_ν,p), yet G_n must have a relatively simple structure: the Szemerédi Partitions of the graphs in such sequences use only densities which are roots of the Copy-polynomials.

Definition 5.2. Given a graph Q, we shall say that it satisfies the *density condition* $\mathcal{D}(p,\varepsilon)$ if for every $X \subseteq V(Q)$, $|X| > \varepsilon v(Q)$ we have

$$(p-\varepsilon)\binom{|X|}{2} \le e(Q[X]) \le (p+\varepsilon)\binom{|X|}{2}. \tag{19}$$

Theorem 5.3. *Let L_ν and $p \in (0,1)$ be fixed. Let the graph sequence (G_n) satisfy (6). For every $\omega_n \to \infty$ (as $n \to \infty$) there exists an $\varepsilon_n \to 0$ (as $n \to \infty$) such that (G_n) has a Regular Partition $U_1,\dots,U_{k_n}$ for some $k_n \le \omega_n$, where*

- *the graphs $G_n[U_i]$ satisfy the density condition (19) with p, or $\overline{p}$ in* *the classes, and*

- *(all) the pairs* (U_a, U_b) $(1 \le a < b \le k_n)$ *are* ε_n*-regular, and for each density* $d(U_a, U_b)$ *there is a root* $s_j^{a,b}$ *of the "corresponding" Copy-polynomials, such that*

$$\left| d(U_a, U_b) - s_j^{a,b} \right| < \varepsilon_n.$$

We shall not prove Theorem 5.3 here, primarily because it is slightly technical (and if, by chance, it turns out that Conjecture 3.11 holds, then this theorem and some possible sharpenings of it will become uninteresting).

5.2. Preparation to prove Theorem 3.7

The standard counting technique, connected to the Szemerédi Lemma, shows that

Lemma 5.4 (Approximate Counting). *There is a function* $f_\nu(\varepsilon) \to 0$ *(as* $\varepsilon \to 0$*) with the following property. Let* $U_1, \dots, U_k$ *be an* ε*-regular partition of* G_n*,* $k > \frac{1}{\varepsilon}$*, and, for some index set* $I \subseteq [1, k]$*,* $Z_M \subseteq G_n$ *be spanned by* $\bigcup_{i \in I} U_i$*. Assume that all the pairs* (U_i, U_j) *for* $i, j \in I$ $(i \ne j)$ *are* ε*-regular. If we replace the edges in* Z_M *between* U_i *and* U_j *independently by random edges of probability* $d(U_i, U_j)$*, and we change arbitrarily the edges with endvertices in the same classes* U_i*, then almost surely, in the resulting* W_M*,*

$$|\mathbf{N}^*(L_\nu \subseteq W_M) - \mathbf{N}^*(L_\nu \subseteq Z_M)| < \left(\frac{\binom{\nu}{2}}{|I|} + f_\nu(\varepsilon) \right) M^\nu. \tag{20}$$

Observe, that (20) is trivial if $|I| < \binom{\nu}{2}$. We may and shall assume that

$$f_\nu(\varepsilon) > \nu^2 \varepsilon. \tag{21}$$

Here $f_\nu(\varepsilon)$ corresponds to the "errors" coming from the application of the Regularity Lemma and the approximation by random graphs, $\binom{\nu}{2} M^\nu / |I|$ estimates the number of copies of L_ν's having at least two points in the same U_i, $(i = 1, \dots, |I|)$: if $a_\xi, a_\zeta \in V(L_\nu)$ are mapped into $b_\xi, b_\zeta \in U_\ell$, (i.e., into the same class), then each b_i $(i = 1, \dots, \nu)$ can be selected in at most M ways, and having selected b_ξ fixes U_ℓ, so b_ζ can be selected in at most $|U_\ell| \le M/|I|$ ways.

5.3. Proof of Theorem 3.7

The threshold R_t

Let t be an arbitrary integer, $\varepsilon_t = \frac{1}{t}$. We fix a Ramsey number r_t, as the minimum integer such that every t-edge-coloring of a K_{r_t} contains a monochromatic K_t.

Next fix a number $R := R_t$, such that, for any graph H_k, if $k > R_t$ and $e(H_k) > k^2/t^2$, then $H_k \supseteq K_2(r_t, r_t)$. This can be done, using the Kővári-T. Sós-Turán theorem [7] according to which,

$$\text{if} \quad K(a,b) \nsubseteq G_n \quad \text{then} \quad e(G_n) < \frac{1}{2}\sqrt[a]{b-1}\, n^{2-1/a} + \frac{a-1}{2}n. \tag{22}$$

Therefore, we can fix an R_t for which,

$$\text{if} \quad k > R_t, \quad \text{then} \quad \frac{3}{4}k^{2-1/r_t} + \frac{r_t - 1}{2}k < \frac{k^2}{t^2}. \tag{23}$$

This R_t is what we wished to define.

Types of Regular Partitions

Assume that for a fixed $\tau > 0$ and for $\varepsilon = \varepsilon_t \ll \tau$ for each G_n of a sequence (G_n) we have an ε-regular partition, (for n sufficiently large). Our partition of $V(G_n)$ can have (at most) εk^2 non-ε-regular pairs (U_i, U_j). For each such pair we delete all the edges joining U_i to U_j. That may change $\mathbf{N}^*(L_\nu \subseteq G_n)$ by at most εn^ν and we may forget about this slight difference. After this all the pairs become ε-regular.

An ε-Regular Partition $(U_1, \dots, U_k)$ of a graph G_n will be classified as follows:

(a) At least τk^2 ε-regular pairs (U_i, U_j) satisfy

$$|d(U_i, U_j) - p| > \tau \quad \text{and} \quad |d(U_i, U_j) - \overline{p}| > \tau. \tag{24}$$

(b) Assume that (a) does not hold, but

$$\begin{cases} |d(U_i, U_j) - p| < \tau, & \text{for at least } \tau k^2 \text{ pairs } (U_i, U_j) \text{ and} \\ |d(U_i, U_j) - \overline{p}| < \tau, & \text{for some other } \tau k^2 \text{ pairs } (U_i, U_j). \end{cases} \tag{25}$$

(c) Neither (a), nor (b) hold.

Classification of graph sequences

Below certain subgraphs and subsets obtained from some graphs $G_n \in \mathcal{G}$ may depend on n but we suppress indicating this dependence in our notation. Also, ignoring the simpler case $p = \overline{p}$, assume that $p \neq \overline{p}$.

Let $L = L_\nu$ and $p \in (0,1)$ be fixed. We shall create a 2-dimensional, infinite matrix of graphs, $\mathbb{B}$. For every pair of integers, (T,t) it will have a "box", $\mathbb{B}_{T,t}$ which may be empty or may contain a graph $G_n \in \mathcal{G}$, with an ε_t-regular partition. (Remember that $\varepsilon_t = \frac{1}{t}$, and think of T as $T \ll t$.)

For a given t define R_t as described earlier. Next find $\Omega = \Omega_t$ as the "upper bound function" in the Regularity Lemma, corresponding to ε_t and $\kappa := R_t$. Finally, denote by $\mathcal{G}[t]$ the subsequence of graphs $G_n \in \mathcal{G}$ for which

$$\begin{aligned} &\text{for every} \quad F_h \overset{*}{\subseteq} G_n \quad \text{with} \quad h > n/\Omega_t, \\ &|\mathbf{N}^*(L_\nu \subseteq F_h) - \beta_L(p)h^\nu| < \varepsilon_t h^\nu. \end{aligned} \tag{26}$$

(Note that by $h > n/\Omega_t$, and $|\mathbf{N}^*(L_\nu \subseteq F_h) - \beta_L(p)h^\nu| = o(n^\nu)$, (26) holds fro every sufficiently large n.) For each $\tau = \frac{1}{T}$ and each $\varepsilon = \frac{1}{t}$ we check if there exist graphs $G_n \in \mathcal{G}[t]$ having ε_t-Regular Partitions for $\kappa = R_t$, satisfying either (a) or (b).

If there exist no such graphs, or if $\tau < \varepsilon_t$, we define $\mathbb{B}_{T,t} = \emptyset$. If there exist such graphs G_n, we put one of them into $\mathbb{B}_{T,t}$, and also fix a corresponding ε_t-regular partition of it, $\{U_1, \dots, U_k\}$.

We distinguish two cases:

(i) There exists a T for which infinitely many $\mathbb{B}_{T,t}$ are non-empty. (In other words, there exists an $\tau > 0$ for which we have infinitely many t with corresponding graphs G_{n_t,ε_t}.)

Mostly we shall neglect indicating the dependence on t.

(ii) For any $\tau > 0$, we have only finitely many non-empty boxes $\mathbb{B}_{T,t}$: if $\varepsilon_t < \varepsilon(\tau)$ then for any G_n with sufficiently large n every ε_t-regular partition $\{U_1, \dots, U_k\}$ has at most τk^2 pairs (U_i, U_j) satisfying (24) or (25).

In Case (i) we shall provide the 2-class counterexamples; in Case (ii) we shall prove that every $\mathcal{G} = (G_n)$ satisfying (6) is the union of a p-quasi-random and a $\overline{p}$-quasi-random graph sequences.

We start with the simpler case.

Settling Case (ii)

Let $\mathcal{G} = (G_n)$ be a graph sequence satisfying (6). Decompose $\mathcal{G}$ into $\mathcal{G}_1 \cup \mathcal{G}_2$ by putting into $\mathcal{G}_1$ those G_n whose density $2e(G_n)/n^2$ is nearer to p

then to $\overline{p}$. We show that $\mathbf{P}_S(p)$ holds for G_n. Therefore, by Theorem A, $\mathcal{G}_1$ is p-quasi-random. Similarly, $\mathcal{G}_2$ is $\overline{p}$-quasi-random.

Fix an $\tau < \frac{1}{10}(p-\overline{p})$. We may restrict ourselves to the case $\tau = \frac{1}{T}$. Let $\tau > 0$ be sufficiently small. Since the T^{th} row of our matrix has only finitely many nonempty "boxes", therefore for any sufficiently small $\varepsilon > 0$ the corresponding "box" is empty: every ε-Regular Partition of G_n is of Type (c) in our classification. We may assume that n is sufficiently large, therefore, by (6), we also have (26). So the only reason why we have not put this G_n into this "box" is that it does not satisfy (24), nor (25). Hence for all but at most $2\tau k^2$ ε-regular pairs (U_i, U_j) we have $|d(U_i, U_j) - p| \le \tau$ and consequently, $\mathbf{P}_S(p)$ holds. (In principe these densities could also be near $\overline{p}$ but then the edge-density of G_n would be nearer to $\overline{p}$ than to p.) This shows that $\mathcal{G}$ is the union of two quasi-random graph sequences.

Settling case (i)

Now we know that there is a T for which the T^{th} row of our matrix contains infinitely many nonempty boxes. Fix this T and the corresponding $\tau = \frac{1}{T} > 0$. Choose infinitely many graphs G_n from this row: they form a sequence $\mathcal{G} = (G_n)$.[13] We recall that each graph is fixed also with a Regular Partition. We shall distinguish two subcases:

(α) For this T^{th} row of the matrix, there exist a sequence of integers $t \to \infty$ and the corresponding graphs G_{n_t, ε_t} whose ε_t-regular partition is of Type (a) (see (24)). (This is the most important case.)

(β) In the T^{th} row of the matrix, there exist infinitely many G_n whose ε-regular partition is of Type (b) (see (25)).

In these two cases we shall find the promised 2-class counterexample sequences of type $G(V_1, V_2, u, v, s)$.

Details of Case (α) [14]

(α_1) Now we assumed that there exists an $\tau > 0$ and infinitely many G_n with the ε-regular partition

$$V(G_n) = U_1 \cup \ldots \cup U_k$$

having at least τk^2 ε-regular pairs (U_i, U_j) satisfying (24).

[13] This $\mathcal{G}$ forms a strong counterexample, but we shall not need this directly.

[14] In this case R could be replaced by r: a single application of Ramsey Theorem to define R_t would suffice.

Restrict ourselves to a fixed G_n and the corresponding Regular Partition $(U_1, \dots, U_k)$. Consider the graph H_k the vertices of which are the classes U_i and the edges of which are the regular pairs satisfying (24).[15] We shall color these edges (=pairs) with their *rounded densities*: (U_i, U_j) gets color $\chi(U_i, U_j) := \frac{1}{t}\lceil t \cdot d(U_i, U_j) \rceil$. This way we get a $\le t$-colored H_k. (The edges of G_n corresponding to irregular pairs were deleted and therefore the irregular pairs get color 0. In principle H_k could have $t+1$ colors, but many of the colors, corresponding to densities near to p or $\overline{p}$, are excluded, by (24).)

Let H_k^* denote the monochromatic subgraph of H_k having the *most* edges. By the choice of the lower bound $\kappa = R_t$ in the application of the Regularity Lemma,

$$e(H_k^*) > \frac{\tau}{t}k^2 > \left(\frac{k}{t}\right)^2.$$

So we have a monochromatic $K(R_t, R_t) \subseteq H_k$, i.e., two sets of classes, $A_1, \dots, A_{R_t}$ and $B_1, \dots, B_{R_t}$ so that all the pairs (A_i, B_j) are from the pairs in (24) and of the same color, say $s = s_n$.

(α_2) Here the edges $(A_i, A_{i'})$ and $(B_j, B_{j'})$ of H_k, $(1 \le i, i', j, j' \le R_t)$, may have many colors. We apply Ramsey's theorem to H_k. The subgraph of H_k spanned by $A_1, \dots, A_{R_t}$ contains a monochromatic K_t spanned by some classes $\{A_i : i \in I\}$, $(|I| = t)$ and the subgraph spanned by $B_1, \dots, B_{R_t}$ contains a monochromatic subgraph, spanned by some classes $\{B_j : j \in J\}$, $(|J| = t)$. Let the color used for $(A_i, A_{i'})$ be $u = u_n$ and for $(B_j, B_{j'})$ $v = v_n$ Since the colors encode densities, we used at most 3 (rounded) densities. These define a "structure" $G(V_1, V_2, u_n, v_n, s_n)$. Clearly,

$$|d(A_i, A_{i'}) - u_n| \le \frac{1}{t} = \varepsilon, \quad |d(B_j, B_{j'}) - v_n| \le \varepsilon,$$

and

$$|d(A_i, B_j) - s_n| \le \varepsilon.$$

Here we have infinitely many graphs G_n, and each of them corresponds to smaller and smaller $\varepsilon = \varepsilon_t = \frac{1}{t} \to 0$. Take for each of them the corresponding "structure" $G(V_1, V_2, u_n, v_n, s_n)$. If by any chance, (u_n, v_n, s_n) is not convergent, then we take a convergent subsequence:

$$(u_n, v_n, s_n) \to (u_0, v_0, s_0).$$

[15]Generally we would call this graph the colored reduced graph or colored cluster graph.

(Typically, $u_0, v_0 \in \{p, \overline{p}\}$, but we do not know this yet.) By (24),

$$|s_0 - p| \geq \tau - \varepsilon, \quad \text{and} \quad |s_0 - \overline{p}| \geq \tau - \varepsilon.$$

We assert that $G(V_1, V_2, u_0, v_0, s_0)$ is a strong counterexample.

We may assume that the classes U_i have size $h \sim \frac{n}{k} \geq \frac{1}{\Omega_t} n$ where Ω_t was fixed when we applied the Regularity Lemma. Denote by D_{2th} the subgraph of G_n spanned by these A_i's and B_j's.

(α_3) We wish to prove that $G(V_1, V_2, u_0, v_0, s_0)$ is a strong counterexample.[16] Therefore we select a subgraph $G' = G(V_1', V_2', u_0, v_0, s_0)$ of it, with at least $h \geq n\frac{1}{\Omega_t}$ vertices, and count the L_ν's in G'. Of course, the basic idea is that G' is the randomization of $G_n[V_1' \cup V_2']$ and therefore, by Lemma 5.4,

$$\mathbf{N}^*(L_\nu \subseteq G') = \beta_L(p) v(G')^\nu + o(v(G')^\nu). \tag{27}$$

Yet we need several technical steps to prove (27):

(1) To apply Lemma 5.4, we should use the actual probabilities $d(A_i, A_j)$, $d(A_i, B_j)$, and $d(B_i, B_j)$, instead of (u_0, v_0, s_0). Replacing the actual densities by the rounded densities (u_n, v_n, s_n) yields an error at most $\frac{1}{t} v(G')^\nu$; then replacing (u_n, v_n, s_n) by (u_0, v_0, s_0) results an error of at most

$$2(\max\{|u_n - u_0|, |v_n - v_0|, |s_n - s_0|\}) v(G')^\nu \tag{28}$$

L_ν's if $v(G') > h$. (We used the factor 2 to compensate the randomness.)

(2) Some negligible error comes from the fact that we take an *arbitrary* subgraph G' and not only subgraphs spanned by whole classes U_i of the Regular Partition. However, the vertices of $G(V_1, V_2, u_0, v_0, s_0)$ are "symmetric" in the sense that we can replace the vertices of the original G' by other vertices spanning some complete classes A_i or B_j and two "remainder" classes of at most h vertices each. This results in an error at most

$$2hv(G')^{\nu-1} \leq \varepsilon_t v(G')^\nu. \tag{29}$$

(3) To apply Lemma 5.4, we take an integer $\lambda \leq 2t$ and select an arbitrary $\Lambda \subseteq I \cup J$ with $|\Lambda| = \lambda > \sqrt{t}$. Let $D_{\lambda h}$ be the subgraph of G_n spanned by

[16] We remind the reader that $G(V_1, V_2, u_0, v_0, s_0)$ is a generalized random graph and we should have written that "the graph sequence $G_n = G(V_1, V_2, u_0, v_0, s_0)$ is almost surely a strong counterexample if $|V_i| \approx n/2$" but we agreed to say, instead, that $G(V_1, V_2, u_0, v_0, s_0)$ is a strong counterexample, meaning that the "generalized random structure" defined by it provides a strong sequence of counterexamples, see Remark 3.8.

$\bigcup_{i\in\Lambda} U_i$. The Z_M of Lemma 5.4, be $Z_M = D_{\lambda h}$. Its randomization (described in Lemma 5.4) is just $W_M := G(V_1', V_2', u_n, v_n, s_n)$, where

$$V_1' = \bigcup_{i\in I\cap\Lambda} U_i \quad \text{and} \quad V_2' = \bigcup_{j\in J\cap\Lambda} U_j.$$

By Lemma 5.4,

$$|\mathbf{N}^*(L_\nu \subseteq G(V_1', V_1', u_n, v_n, s_n) - \beta(p)(\lambda h)^\nu| < \left(f_\nu(\varepsilon_n) + \frac{\nu^2}{\lambda}\right)(\lambda h)^\nu. \quad (30)$$

Since $G_n \in \mathcal{G}[t]$, it satisfies (26). So we get

$$|\mathbf{N}^*(L_\nu \subseteq D_{\lambda h}) - \beta_L(p)(\lambda h)^\nu| \le \varepsilon_t(\lambda h)^\nu \quad \text{as} \quad n\to\infty. \quad (31)$$

Using (30) and taking into account all the above error estimates, (28), (29),(31) we get (27). Since $s_0 \ne p, \overline{p}$, this concludes the proof of that $G(V_1', V_2', u_0, v_0, s_0)$ is really a strong counterexample sequence.

Details of Case (β)

We have to modify the previous argument just a little bit. We know that there are many pairs (U_i, U_j) with densities around p and also many with densities around $\overline{p}$ and that $0 < \tau < \frac{|p-\overline{p}|}{10}$.

We first consider the pairs of density approximately p and repeat the argument of (α). We get a graph D_{2th} as above. Then, taking the limit, we get a 2-class graph $G(V_1, V_2, p, q, s)$ which is a counterexample, unless it is a $G(V_1, V_2, p, p, p)$. Then we can consider the pairs of density $\sim \overline{p}$ and either get a counterexample or a $G(V_1, V_2, \overline{p}, \overline{p}, \overline{p})$.

If we have not obtained the desired 2-class counterexample as yet, then we proceed as follows. In the original graph G_n we have the classes $A_1, \dots, A_r$ and $B_1, \dots, B_r$ corresponding to $G(V_1, V_2, p, p, p)$, and $C_1, \dots, C_r$ and $D_1, \dots, D_r$ corresponding to $G(V_1, V_2, \overline{p}, \overline{p}, \overline{p})$. To simplify the notation, change it by using $A_{i+r} := B_i$ and $C_{j+r} := D_j$. All the densities $d(A_i, A_j) \sim p$ and all the densities $d(C_i, C_j) \sim \overline{p}$. Since $p \ne \overline{p}$, after deleting at most one group we may assume that these two sets of groups are disjoint. Let the corresponding groups be $A_1, \dots, A_{2r-1}$ and $C_1, \dots, C_{2r-1}$. The densities $d(A_i, C_j)$ could be arbitrary, but applying (22) to them we can get (apart from the indexing) $A_1, \dots, A_t$ and $C_1, \dots, C_t$, where $d(A_i, C_j) \sim s$ for some s. So, taking the limit, we get the same type of $G(V_1, V_2, u, v, s)$ as in (α), except that there we knew that s_0 is far from p and $\overline{p}$, now we know that the densities $d(A_i, A_j)$ are far from $d(C_i, C_j)$. That is equally good for us. ∎

5.4. The algorithm

We have a system of polynomials, which have some trivial zeros. We shall eliminate some trivial zeros and then the question is reduced to finding an algorithm which decides if some polynomials have zeros at all in the 3-dimensional unit cube. Such an algorithm was given by Tarski [11].

One part of Tarski's theorem is trivial here. It is trivial that if there are no counterexamples, then that can be algorithmically proven: We have to decide if the copy-polynomials $\mathbb{P}^k_{u,v}(s)$ have zeros for some (u, v, s) where not all the coordinates are equal. Assume there are no such zeros and we wish to prove this.

There are two cases: either $u = p$ and $v = p$ or $u = p$ and $v = \overline{p} \neq p$. In the first case we have polynomials depending on one parameter p and having the root $s = p$. So we can factor out $s - p$, (maybe more than once) and we can often factor out some powers of p, q, $(1-p)$ and $(1-q)$. (Often $p = 0, 1$ are also roots and we have to get rid of them, too.) Then we have to show that the resulting system of polynomials has no zeros at all in $[0, 1]$.

Indeed, we have the polynomials, can estimate their gradients, say by a constant M, and therefore if we check the values of the polynomials on a sufficiently fine grid, in the (u, v, s)-space, then we have a lower bound in the grid points and also know that the polynomials are positive in all the other points of $[0, 1]^3$ as well. ∎

(b) The other case is when there exist strong counterexamples for (L_ν, p). According to a theorem of Tarski [11] this can algorithmically be proven.

(c) Clearly, the two parts can be combined in one algorithm which decides in finitely many steps if case (a) or case (b) holds.

6. PROOF OF THEOREM 3.12

Below we show that if L_ν is d-regular and $u, v, s \in (0, 1)$, and

$$\begin{aligned} \mathbb{P}^1_{u,v}(s) &= 0 \\ \mathbb{P}^2_{u,v}(s) &= 0 \end{aligned} \tag{32}$$

then $u = v = s = p$ or $u = v = s = \overline{p}$.

Using $\mathbb{P}^{\nu-1}_{u,v}(s) = 0$

We know that for any L_ν, if $G(V_1, V_2, u, v, s)$ is a counterexample, then $u, v \in \{p, \overline{p}\}$. For d-regular graphs we can easily see that even $s \in \{p, \overline{p}\}$. Indeed, we can use (15) from the proof of Lemma 4.3:

$$0 = \nu(1-u)^{\nu-1} - \sum_{i=1}^{\nu} u^{-d_i}(1-u)^{d_i} s^{d_i}(1-s)^{\nu-d_i-1}$$

yielding

$$u^d(1-u)^{\nu-1-d} = s^d(1-s)^{\nu-d-1} \tag{33}$$

Observe that (by Lemma 4.4) $s = u$ or $s = \overline{u}$. Further, (14) reduces to

$$1 = \left(\frac{s}{u}\right)^d \left(\frac{1-s}{1-u}\right)^{\nu-1-d}, \tag{34}$$

Taking the logarithms, and dividing by $\nu - 1$,

$$\frac{d}{\nu-1} \log\left(\frac{s}{u}\right) + \left(1 - \frac{d}{\nu-1}\right) \log\left(\frac{1-s}{1-u}\right) = 0.$$

Put

$$\alpha := 1 - \frac{d}{\nu-1}, \quad \beta := \frac{d}{\nu-1}, \quad x := \left(\frac{1-u}{1-s}\right)^2, \quad y := \left(\frac{u}{s}\right)^2, \tag{35}$$

Then

$$\alpha \log x + \beta \log y = 0. \tag{36}$$

Using $\mathbb{P}^{\nu-2}_{u,v}(s) = 0$

Let us calculate $\mathbb{P}^{\nu-2}_{u,v}(s) = 0$. By the *$d$-regularity* (!) $\mathbb{P}^{\nu-2}_{u,v}(s)$ has (at most) three distinct terms:

- A fixed term:
$$\binom{\nu}{2} u^{\frac{1}{2}\nu d}(1-u)^{\binom{\nu}{2} - \frac{1}{2}\nu d};$$

- A term corresponding to the case when the two points in B are independent: $e(B) = 0$, $e(A, B) = 2d$, $e(A) = \frac{1}{2}\nu d - 2d$; and

- A term corresponding to when B contains an edge: $e(B) = 1$, $e(A, B) = 2d - 2$, $e(A) = \frac{1}{2}\nu d - 2d + 1$;

Hence

$$\begin{aligned}\mathbb{P}^{\nu-2}_{u,v}(s) &= \binom{\nu}{2} u^{\frac{1}{2}\nu d}(1-u)^{\binom{\nu}{2}-\frac{1}{2}\nu d} \\ &- \left(\binom{\nu}{2} - \frac{1}{2}\nu d\right)(1-v)u^{\frac{1}{2}\nu d - 2d}(1-u)^{\binom{\nu-2}{2}-\frac{1}{2}\nu d + 2d} s^{2d}(1-s)^{2\nu-2d-4} \\ &- \frac{1}{2}\nu d v u^{\frac{1}{2}\nu d - 2d + 1}(1-u)^{\binom{\nu-2}{2}-\frac{1}{2}\nu d + 2d - 1} s^{2d-2}(1-s)^{2\nu-2d-2}\end{aligned}$$

Plugging in the conjugacy for s,

$$s^{2d}(1-s)^{2\nu-2d-2} = u^{2d}(1-u)^{2\nu-2d-2},$$

into the equation $\mathbb{P}^{\nu-2}_{u,v}(s) = 0$ and then simplifying we get

$$0 = (\nu - 1) - \Big(\nu - 1 - d\Big)(1-v)(1-u)(1-s)^{-2} - dvus^{-2} \tag{37}$$

From here on, we distinguish two cases:

- $u = v$, $s = \overline{u}$ and
- $u \neq v$.

The symmetric case: $v = u$

Now (37) gives

$$0 = (\nu - 1) - \Big(\nu - 1 - d\Big)(1-u)^2(1-s)^{-2} - du^2s^{-2} \tag{38}$$

Rearranging, we get

$$\left(1 - \frac{d}{\nu - 1}\right)\left(\frac{1-u}{1-s}\right)^2 + \frac{d}{\nu - 1}\left(\frac{u}{s}\right)^2 = 1.$$

Hence we get

$$\alpha x + \beta y = 1,$$

but this and the concavity of $\log t$ contradicts (36), unless $x = y$, implying that $u = s$ and completing this part of the proof.

The asymmetric case: $v \neq u$

(ii) We start again with (37)

$$0 = (\nu - 1) - \Big(\nu - 1 - d\Big)(1 - v)(1 - u)(1 - s)^{-2} - dvus^{-2}$$

By $v \neq u$, either $s = u$ or $s = v$. By symmetry, we may assume that $s = v$.[17]

$$(\nu - 1) = \Big(\nu - 1 - d\Big)(1 - u)(1 - s)^{-1} + dus^{-1}$$

$$\left(1 - \frac{d}{\nu - 1}\right)\left(\frac{1 - u}{1 - s}\right) + \left(\frac{d}{\nu - 1}\right)\left(\frac{u}{s}\right) = 1. \tag{39}$$

Here we can use the same convexity argument used in the previous subsection, with

$$x := \left(\frac{1 - u}{1 - s}\right), \quad y := \left(\frac{u}{s}\right).$$

(The squaring is missing!) This completes the whole proof. ∎

7. PROOF OF THEOREM 3.9 ON INDUCED P_3'S

First we calculate the Copy-polynomials of P_3, (using (12)) and then solve the corresponding system of equations.

Clearly,

$$\mathbb{P}^0_{u,v}(s) := u^2(1 - u) - v^2(1 - v). \tag{40}$$

To get $\mathbb{P}^1_{u,v}(s)$ use $k = 1$ in (12): $e(A) = 0$ and either $e(B) = 0$ or $e(B) = 1$. If $e(B) = 0$, then $e(A, B) = 2$. So we get $(1 - v)s^2$. In the other case, when $e(B) = 1$, then $e(A, B) = 1$: we get $vs(1 - s)$, but we get this term twice:

$$\mathbb{P}^1_{u,v}(s) = s^2(1 - v) + 2vs(1 - s) - 3u^2(1 - v). \tag{41}$$

Exchanging in the first two terms u and v we get

$$\mathbb{P}^2_{u,v}(s) = s^2(1 - v) + 2us(1 - s) - 3u^2(1 - v). \tag{42}$$

[17] Below we shall use (34) which uses u and this may seem to create some asymmetry. However, (34) remains valid if u is replaced bu v.

(i) $\mathbb{P}^2_{u,v}(s)$ does not contain v. Since $s = u$ is a trivial root of $\mathbb{P}^2_{u,v}(s) = 0$, we can decompose $\mathbb{P}^2_{u,v}(s)$:

$$\mathbb{P}^2_{u,v}(s) = (s - u)\Big((1 - 3u)(s + u) + 2u\Big).$$

For each value of u, the equation $\mathbb{P}^2_{u,v}(s) = 0$ yields two values of s. One of them is $s = u$, (of course!) but we are interested in the other one:

$$s = 3u\frac{1 - u}{3u - 1},$$

which is negative in $(0, \frac{1}{3})$ and exceeding 1 in $(\frac{1}{3}, \frac{1}{\sqrt{3}})$. Observe that $3u\frac{1-u}{3u-1} = u$, iff $u = 2/3$.

(ii) To verify Theorem 3.9, observe that $G(V_1, V_2, u, v, s)$ is a strong counterexample (for $u = p$ or $u = \overline{p}$) iff the corresponding copy-polynomials vanish and $s \in [0, 1]$ and $u = v = s$ does not hold. Now we assumed that $u = v$, so (40) is automatically satisfied, (41) and (42) coincide. So we have to satisfy only that (42) vanishes and ensure that $u, s \in [0, 1]$. But the formula of Theorem 3.9 (b) is just the solution of $(1 - 3p)(s + p) + 2p = 0$, providing an $s \in (0, 1)$ for every $p > \frac{1}{\sqrt{3}}$. This proves (b).

(iii) To prove (a) we have to solve the system of equations provided by (40), (41), (42). Subtract $\mathbb{P}^2_{u,v}(s)$ from $\mathbb{P}^1_{u,v}(s)$:

$$\begin{aligned} \mathbb{P}^1_{u,v}(s) - \mathbb{P}^2_{u,v}(s) &= s^2(1 - v) + 2vs(1 - s) - s^2(1 - v) - 2us(1 - s) \\ &= s(3s - 2)(u - v). \end{aligned}$$

This means that $\mathbb{P}^1_{u,v}(s)$ and $\mathbb{P}^2_{u,v}(s)$ can vanish at the same time if and only if $v = u$ or $s = 0$ or $s = 2/3$. Here $s = 0$ is excluded, since then $\mathbb{P}^2_{u,v}(s) = 0$ would imply $u = 0$ or $u = 1$. For $s = 2/3$ (42) yields

$$4 - 27u^2(1 - v) = 0.$$

Then $u^2(1-v) = \frac{4}{27}$. This $\frac{4}{27}$ is the maximum of the conjugacy curve $u^2(1-v)$, at $u = 2/3$. So $u = 2/3$ and $v = u$. Hence case $s = 2/3$ is completely settled.

So $u = v$. Then $\mathbb{P}^0_{u,v}(s) = 0$ automatically holds, $\mathbb{P}^1_{u,v}(s) = \mathbb{P}^2_{u,v}(s) = 0$ follows from (i). This completes the proof of Theorem 3.9. ∎

Corrigendum to [9]. *In [9] we have forgotten to explicitly state that* $e(L_\nu) > 0$. *If* $e(L_\nu) = 0$, *then* $\mathbf{N}(L_\nu \subseteq F_h) = \binom{h}{\nu}$ *is independent of the structure of* G_n, *and of* p: *the theorem trivially does not hold.*

References

[1] F. R. K. Chung, R. L. Graham and R. M. Wilson: Quasi-random graphs, Combinatorica, **9**(4), (1989), 345-362.

[2] F. R. K. Chung, R. and R. L. Graham: Quasi-random hypergraphs, Random Structures and Algorithms, **1**(1), (1990), 105-124.

[3] F. R. K. Chung: Regularity lemmas for hypergraphs and quasi-randomness, Random Structures and Algorithms, **2**(2) (1991) 241-252.

[4] F. R. K. Chung and R. L. Graham: Maximum cuts and quasi-random graphs, *in* Random Graphs, (Poznan Conf, 1989) Wiley-Intersci, Publ. vol 2, 23–33.

[5] P. Erdős and A. H. Stone: On the structure of linear graphs, Bull. Amer. Math. Soc. **52** (1946), 1089-1091. MR**8**, 333b.

[6] P. Frankl, V. Rödl and R. M. Wilson: The number of submatrices of given type in an Hadamard matrix and related results, Journal of Combinatorial Theory, (B) **44**(3) (1988) 317-328.

[7] T. Kővári, Vera T. Sós, P. Turán: On a problem of Zarankiewicz, Colloq. Math. **3** (1954), 50-57; MR**16** #456a.

[8] M. Simonovits and V. T. Sós: Szemerédi's Partition and quasi-randomness, Random Structures and Algorithms, **2**(1) (1991) 1-10.

[9] M. Simonovits and V. T. Sós: Hereditarily extended properties, quasi-random graphs and not necessarily induced subgraphs, Combinatorica, **17**(4) (1997) 577-596.

[10] E. Szemerédi: On regular partitions of graphs, Problemes Combinatoires et Théorie des Graphes (ed. J. Bermond et al.), CNRS Paris, 1978, 399-401.

[11] A. Tarski, A. Mostowski and R. M. Robinson, Undecidable Theories. North Holland, 1953.

[12] A. Thomason: Random graphs, strongly regular graphs and pseudo-random graphs, in Surveys in Combinatorics, 1987 (Whitehead, ed.) LMS Lecture Notes Series 123, Cambridge Univ. Press, Cambridge, 1987, 173–196

[13] A. Thomason: Pseudo-random graphs, in Proceedings of Random graphs, Poznan, 1985, (M. Karonski, ed.), Annals of Discrete Math. **33** (1987), 307-331.