Some topics in Extremal Combinatorics

Hong Liu

August 9, 2019

The purpose of this note is to give a touch on some topics in extremal combinatorics. These topics are chosen semi-randomly :) When possible, I try to present the proofs "backwards", or with some intuitions here and there. The proofs would be a bit longer than usual, but hopefully they look more natural this way.

The material covered are based on various notes/books/papers, see main texts for references. Comments are welcome, if you spot mistakes, please let me know :)

Contents

Chapter 1

Extremal graph theory

In this chapter, we will discuss the classical Turán's theorem in extremal graph theory and present some standard techniques such as Zykov's symmetrisation, stability method.

1.1 Turán theorem

Before we start, let us consider the following puzzle. Suppose we have to choose n irrational numbers $x_1, \dots, x_n$. How can we maximise the number of pairs (x_i, x_j) such that $x_i + x_j$ is rational?

One of the most classical extremal problems, nowadays so-called *Turán-type* problem, is:

Problem 1.1.1 (Turán-type)**.** How dense a graph can be without containing another (usually small) graph as a subgraph?

More specifically, given a graph H, we say a graph G *contains a copy* of H, or H is a *subgraph* of G, or $H \subseteq G$, if there is an injective map $\varphi : V(H) \to V(G)$ that preserves adjacencies, i.e. for any $uv \in E(H)$, we have $\varphi(u)\varphi(v) \in E(G)$. We call such a map an *embedding* of H in G. We say G is *H-free* if it does not contain H as a subgraph. If in addition, the map preserves also non-adjacencies, then H is an *induced* subgraph of G.

The main parameter we study for Problem 1.1.1 is the *extremal number* of H,

$$\mathsf{ex}(n, H) = \max\{e(G) : |G| = n \text{ and } G \text{ is } H\text{-free}\},$$

is the maximum size of an n-vertex H-free graph. We call an n-vertex graph G an *extremal graph* for H, if G is H-free of maximum size, i.e. $e(G) = \mathsf{ex}(n, H)$. One of the earliest applications of extremal graph theory, by Erdős, is to construct dense multiplicative Sidon set of integers using a graph without 4-cycles.

The first result in extremal graph theory is the following theorem of Mantel, which answers Problem 1.1.1 when forbidding triangles as subgraphs.

Theorem 1.1.2 (Mantel 1907)**.** *Let G be an n-vertex graph. If G is triangle-free, then*

$$e(G) \le \mathsf{ex}(n, K_3) = \lfloor n^2/4 \rfloor.$$

Exercise **1.1.3.** Solve the puzzle at the beginning of this section, i.e. find the maximum number of pairs of irrationals (x_i, x_j) with $x_i + x_j$ being rational.[1]

Exercise **1.1.4.** Prove that for any tree T, $\mathsf{ex}(n, T) = O(n)$, that is, there exists a constant $C = C(T)$ such that $\mathsf{ex}(n, T) \leq Cn$.[2]

Mantel's result in fact shows that extremal graph for triangle is $K_{\lfloor n/2\rfloor,\lceil n/2\rceil}$. This answers also the following natural question for triangles.

Problem 1.1.5 (Extremal structure/Stability)**.** How do H-extremal graphs look like? What about almost extremal graphs[3], do they look like extremal ones?

Theorem 1.1.2 was later generalised by Turán to forbidding larger cliques. To state his result, we need to define a special family of graphs. Let $r \in \mathbb{N}$, the *r-partite Turán graph* on n vertices, denoted by $T_r(n)$, is the balanced complete r-partite n-vertex graph, i.e. each partite set is of size either $\lfloor n/r \rfloor$ or $\lceil n/r \rceil$. Clearly, $T_r(n)$ is K_{r+1}-free.

Theorem 1.1.6 (Turán 1941)**.** *Let $r \geq 2$ be an integer and G be an n-vertex graph. If G is K_{r+1}-free, then*

$$e(G) \leq \mathsf{ex}(n, K_{r+1}) = e(T_r(n)) = \left(1 - \frac{1}{r}\right)\frac{n^2}{2} - O(r).$$

Furthermore, the Turán graph $T_r(n)$ is the unique extremal graph.

We see from Turán theorem that there is a unique extremal graph $T_r(n)$. The following theorem of Erdős and Simonovits shows that this problem is stable in the sense that every almost extremal graph must be close in structure to the extremal Turán graph, answering Problem 1.1.5 for cliques.

Theorem 1.1.7 (Erdős-Simonovits stability 1966)**.** *Let $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds. Let G be an n-vertex K_{r+1}-free graph. If*

$$e(G) \geq \mathsf{ex}(n, K_{r+1}) - \delta n^2,$$

then G can be changed to $T_r(n)$ by altering at most εn^2 adjacencies.

1.2 Zykov's symmetrisation

There are many proofs for Turán theorem. Here we present one using Zykov's symmetrisation. Zykov's symmetrisation is a process in which we alter the graph, one vertex at a time,

[1]Hint: Build an auxiliary graph and apply Mantel's theorem.

[2]Prove first that every graph with average degree d contains a subgraph with minimum degree at least $d/2$.

[3]We say G is *almost* extremal for H if G is H-free and close to maximum size, i.e. $e(G) \geq ex(n, H) - o(n^2)$.

- without decreasing the number of edges, and
- without increasing the clique number $\omega(G)$.[4]

At the end of the process, we arrive to a complete partite graph, which has a much simpler structure to deal with. In particular, if the original graph is K_{r+1}-free, then all the graphs during symmetrisation will be K_{r+1}-free.

Proof of Turán theorem via Zykov's Symmetrisation. Let G be an n-vertex K_{r+1}-extremal graph. Pick $v_1 \in V(G)$ with maximum degree and symmetrise all of its non-neighbours to v_1. That is, for each u not adjacent to v_1, set $N(u) := N(v_1)$. This operation keeps K_{r+1}-freeness and the resulting graph G_1 has at least as many edges as G. Note that in G_1, $V_1 := V \setminus N(v_1)$ is an independent set and completely joined to $N(v_1)$.

We now repeat this operation as follows. Pick $v_2 \in G_1[N(v_1)]$ and symmetrise all its non-neighbours to v_2. Let G_2 be the resulting graph, then again G_2 is K_{r+1}-free and $e(G_2) \geq e(G_1) \geq e(G)$. Note that in G_2, $V_2 := V(G_2) \setminus N(v_2)$ is an independent set and completely joined to $N(v_2)$.

Continue this process, we will get a complete partite graph, say G', that is also K_{r+1}-free at the end. As the original graph G is an extremal K_{r+1}-free graph, together with $e(G') \geq e(G)$, we see that G' must be also extremal, i.e. $e(G') = e(G)$. We leave the uniqueness of extremal graph as exercise. □

Exercise **1.2.1.** Among all K_{r+1}-free complete partite graphs, the Turán graph $T_r(n)$ is the unique extremal graph.

Further readings. Symmetrisation trick has been used in various extremal problems. To begin, one can read the linear algebraic version, and a recent generalisation due to Füredi and Maleki that can be applied to multiple graphs simultaneously. See also Pikhurko-Staden-Yilma for another application on Erdős-Rothschild problem.

- Motzkin and Straus, *Maxima for graphs and a new proof of a theorem of Turán*, Canad. J. Math., (1965).
- Füredi and Maleki, *The minimum number of triangular edges and a symmetrization method for multiple graphs*, Combin. Probab. Comput., (2017).
- Pikhurko, Staden, and Yilma, *The Erdős-Rothschild problem on edge-colourings with forbidden monochromatic cliques*, Math. Proc. Cambridge Phil. Soc. (2017).

1.3 Erdős-Stone theorem

We have seen that Turán theorem determines the extremal number for cliques and describes the unique extremal structure. The natural next step is what if we forbid general graphs

[4]The clique number $\omega(G)$ of a graph G is the order of the largest clique contained in G.

other than cliques? We shall present in this section a satisfying answer for all non-bipartite graphs.

The seminal result of Erdős and Stone shows that the extremal function of a general graph is completely determined by another important graph parameter: the chromatic number. Recall that the *chromatic number* of a graph H, denoted by $\chi(H)$, is the minimum number of colours needed to colour $V(H)$ so that adjacent vertices do not receive the same colour.

Theorem 1.3.1 (Erdős-Stone 1946)**.** *Let H be an arbitrary graph, then*[5]

$$\mathsf{ex}(n,H)=\left(1-\frac{1}{\chi(H)-1}+o(1)\right)\frac{n^2}{2}.$$

Note that by the definition of chromatic number, the $(\chi(H)-1)$-partite Turán graph is H-free, yielding the lower bound above. The proof of Erdős-Stone theorem proceeds by building a large $(\chi(H)-1)$-partite subgraph. We will not present this proof, instead, we shall give a more "modern" proof later on, which is conceptually simpler, using Szemerédi regularity lemma.

We remark that Erdős-Stone theorem gives the asymptotics of the extremal number for all non-biparite H; while for bipartite H, it only implies that $\mathsf{ex}(n,H)=o(n^2)$. In fact, it is known that extremal number for bipartite graphs is polynomially smaller, i.e. for any bipartite H, there exists $c=c_H$ such that

$$\mathsf{ex}(n,H)=O(n^{2-c}).$$

Further readings. Many bipartite Turán problems are open, we refer the readers to the comprehensive survey of Füredi and Simonovits:

- Füredi and Simonovits, *The history of degenerate (bipartite) extremal graph problems*, arXiv:1306.5167, (2013).

1.4 Stability method

One standard technique in attacking an extremal problem is the so-called *stability method.* We have seen in Section 1.1 the Erdős-Simonovits stability theorem. Such kind of stability statements are not only interesting on its own, but also helpful in obtaining exact results in extremal combinatorics. For instance, the recent result of Liu, Pikhurko and Staden, *The exact minimum number of triangles in graphs of given order and size*, uses, among others, the stability approach.

Often time (but not always), we can tackle an extremal problem with the following three steps:

- Step 1. Obtain asymptotic result;

[5] The term $o(1)$ throughout should be understood as a quantity tending to zero as n, the order of the graph, tends to infinity.

- Step 2. Obtain stability statement;
- Step 3. Use the stability statement to get exact result.

The stability method is refered to Steps 2 and 3. Sometimes, the stability statement in Step 2 can be derived by a more careful analysis of the proof for asymptotic result in Step 1. We shall illustrate Step 3 via a baby application: determining the extremal number of pentagon C_5.

Theorem 1.4.1. *For large n, we have $\mathsf{ex}(n, C_5) = \lfloor n^2/4 \rfloor$.*

Note that Step 1 follows from Erdős-Stone theorem: $\mathsf{ex}(n, C_5) = n^2/4 + o(n^2)$. The stability statement in Step 2 in this case reads as follows.

Lemma 1.4.2. *Let $\varepsilon > 0$, there exists $\delta > 0$ such that the following holds for large n. Let G be an n-vertex C_5-free graph. If $e(G) \geq n^2/4 - \delta n^2$, then G can be made bipartite by deleting at most εn^2 edges.*

Exercise **1.4.3.** Prove the weaker version of Lemma 1.4.2 assuming a stronger condition that $\delta(G) \geq (1/2 - \delta)n$.[6]

Exercise **1.4.4.** Deduce Lemma 1.4.2 from the above weaker version.[7]

We now complete Step 3. The idea in stability method is the following. From Step 2, we already know the asymptotic structure of the extremal configuration. Suppose some unwanted imperfection shows up in the configuration, then we can derive a contradiction as we have good control of the structure thanks to Step 2, proving that there is no imperfection to begin with.

Proof of Theorem 1.4.1. Let G be an n-vertex extremal C_5-free graph. As $T_2(n)$ is also C_5-free, extremality of G implies $e(G) \geq e(T_2(n)) = \lfloor n^2/4 \rfloor$. We can use the trick of removing low degree vertices in Exercise 1.4.4 and enlarge n to assume additionally that

$$\delta(G) \geq (1/2 - \varepsilon)n.$$

Let $V(G) = X \cup Y$ be a max-cut of G.[8] By Lemma 1.4.2, we have

$$e(G[X]) + e(G[Y]) \leq \varepsilon n^2.$$

[6] Hint: Take for granted that $\mathsf{ex}(n, C_4) = O(n^{3/2})$. So there are 4-cycles in G. Consider the neighbourhoods of two adjacent vertices in a copy of C_4.

[7] Hint: By deleting vertices of low degree, we can find a subgraph with high minimum degree. Consequently, if the weaker version holds for large n, Lemma 1.4.2 holds for *larger* n.

[8] A max-cut of a graph G is a bipartition $V(G) = X \cup Y$ that maximises the number of *cross edges*, i.e. edges between the two partite sets X and Y. An important property of a max-cut, which we shall use shortly, is that every vertex in one part, say X, has as many neighbours in the other part Y than in its own part X, since otherwise moving this vertex from X to Y would increase the number of cross edges, contradicting to the fact that $X \cup Y$ is a max-cut.

Consequently, this max-cut is almost balanced, i.e.

$$|X|, |Y| = n/2 \pm 2\sqrt{\varepsilon}n.$$

Indeed, otherwise $e(G) \le |X||Y| + e(G[X]) + e(G[Y]) < n^2/4$, a contradiction. We shall show that there is no edge inside X or Y, and so G is bipartite, which together with the extremality of G implies that G has to be $T_2(n)$, as $T_2(n)$ has the maximum size among all bipartite graphs, yielding the desired.

To get rid of the imperfections (edges in X and Y), we first show that the inner degree is $o(n)$, i.e.

$$\Delta(G[X]), \Delta(G[Y]) \le 2\sqrt{\varepsilon}n.$$

Suppose otherwise that there is some $v \in X$ with $d(v, X) \ge 2\sqrt{\varepsilon}n$.[9] As $X \cup Y$ is a max-cut, $d(v, Y) \ge d(v, X) \ge 2\sqrt{\varepsilon}n$. Note that as G is C_5-free, the bipartite graph induced between $X_H := N(v, X)$ and $Y_H := N(v, Y)$ in G is P_4-free, thus having only $O(n)$ edges (by Exercise 1.1.4). Then for large n, the number of missing edges in $G[X, Y]$[10] is at least $|X_H||Y_H| - O(n) \ge 3\varepsilon n^2$. So again $e(G) \le |X||Y| - 3\varepsilon n^2 + e(G[X]) + e(G[Y]) < n^2/4$, a contradiction. With the additional information that inner degree is sublinear, we are now ready to show that there is not even a tiny bit of imperfection, i.e. not a single edge is allowed in X or Y.

Suppose uv is an edge in X. Let w be a third vertex in X. Using that $\Delta(G[X]) \le 2\sqrt{\varepsilon}n$, $|X|, |Y| = n/2 \pm 2\sqrt{\varepsilon}n$ and $\delta(G) \ge (1/2 - \varepsilon)n$, we see that the common neighbourhood of u, v, w contains almost the entire set Y: $d(u, v, w, Y) \ge (1 - 10\sqrt{\varepsilon})|Y|$. Then two such common neighbours in Y together with u, v, w induces a copy of C_5, a contradiction. This completes the proof. □

---

[9] We write $N(v, X) := N(v) \cap X$ for the set of neighbours of v in X, and $d(v, X) = |N(v, X)|$ for the degree of v in X.

[10] We write $G[X, Y]$ for the bipartite graph induced between X and Y in G.

Chapter 2

Szemerédi's regularity lemma and its applications

Szemerédi's regularity lemma is one of the most important tools in extremal graph theory dealing with dense graphs (positive edge-density). Here we give a gentle introduction to this powerful lemma and see some of its applications and other classical results related to it.

Roughly speaking, the regularity lemma states that every large graph admits a partition into bounded number of parts such that between almost all pairs of parts, the induced bipartite subgraphs behave pseudorandomly. The essence of the regularity lemma is:

Approximating large structures by small structures with low complexity.

It usually offers conceptually simple proofs for asymptotic results. For instance, the regularity lemma and its counting lemma together imply that, in terms of subgraph densities, any graph can be approximated by one of the few (weighted) graphs with bounded order (reduced graphs on $O_\varepsilon(1)$ vertices).

2.1 Taster session

To state the regularity lemma rigorously, we need to set up several notions. Before we do so, let us informally describe a common way of applying the regularity lemma:

- Step 1. Reduce an extremal problem A on large graphs to a problem B on small weighted graphs (using the random behaviour of the regular partition, embedding lemma, counting lemma etc.);
- Step 2. Solve problem B (using e.g. classical results in graph theory).

To be more illuminating, let us sketch a proof of Erdős-Stone theorem, Theorem 1.3.1, to get a taste of how one can carry out this approach. We need some definitions. A map $\varphi : V(H) \to V(G)$ is called a *homomorphism* if it is adjacency preserving, i.e. for any $uv \in E(H)$, we have $\varphi(u)\varphi(v) \in E(G)$. When there is a surjective homomorphism from

H to F, we say that F is a *homomorphic image* of H. We also say that G contains a *homomorphic copy* of H if G contains a homomorphic image of H as subgraph. Step 1 above in this case can be done using the following consequence of the regularity lemma and counting lemma: for any graph G, there is a (weighted reduced) graph R on $O(1)$ vertices such that

P.1 for any fixed H, the subgraph density of H in R is roughly the same as that in G;[1]

P.2 if R contains a homomorphic copy of H, then G contains a copy of H.

By subgraphs densities, we mean the following. Denote by $\mathsf{Inj}(H, G)$ the number of labelled (not necessarily induced) copy of H in G, and

$$t(H, G) = \frac{\mathsf{Inj}(H, G)}{|G|^{|H|}}$$

be the H-*density* in G.[2]

For readers convenience, let us recall the upper bound in Erdős-Stone theorem:

$$\mathsf{ex}(n, H) \le \left(1 - \frac{1}{\chi(H) - 1} + o(1)\right) \frac{n^2}{2}$$

.

Informal proof of Erdős-Stone theorem. *Step 1*. Let $r := \chi(H) - 1$. By **P.1** with $H = K_2$, we just need to bound the edge-density of R, i.e. $t(K_2, R) \le 1 - \frac{1}{r}$.
Step 2. Note that K_{r+1} is a homomorphic image of H. Then by **P.2**, R is K_{r+1}-free. The desired bound on edge-density then follows from Turán's theorem. □

Further readings. Before we dive into the details, let us point out a comprehensive survey of Komlós-Simonovits on regularity lemma:

- Komlós and Simonovits, *Szemerédi's regularity lemma and its applications to graph theory*, Bolyai Math. Soc., (1996).

2.2 Formal setup

The basic notion in regularity lemma is that of an ε-*regular pair* which measures the pseudo-randomness/regularity of the induced bipartite subgraph between the pair. The parameter ε is the *precision* of the regularity; the smaller ε is, the more random like the pair is.

[1]As we shall see in counting lemma, more precisely, here by subgraph density in R, we mean the weighted subgraph homomorphism density.

[2]This notation is not standard. More commonly, $t(H, G)$ denotes the homomorphism density. Though, these two versions differ by a lower order term (homomorphisms that are not injective).

Definition 2.2.1 (Regular pair)**.** Given $G = (V, E)$ and disjoint vertex subsets $X, Y \subseteq V$, let $e(X, Y) := e(G[X, Y])$ and denote by

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

the *density* of the pair (X, Y). For $\varepsilon > 0$, the pair (X, Y) is ε*-regular* if for any $A \subseteq X, B \subseteq Y$ with $|A| \geq \varepsilon|X|$, $|B| \geq \varepsilon|Y|$, satisfy

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

Additionally, if $d(X, Y) \geq \delta$, for some $\delta > 0$, we say that (X, Y) is (ε, δ)*-regular*.

In other words, a regular pair (X, Y) has "uniform" edge distribution in the sense that the density of any pair of large (ε-proportion) subsets (A, B) is roughly the same as that of (X, Y).

Definition 2.2.2 (Regular partition)**.** A partition $V = V_0 \cup V_1 \cup \cdots \cup V_r$ is ε*-regular*, if
(i) $|V_0| \leq \varepsilon|V|$; (called *exceptional set*)
(ii) $|V_1| = |V_2| = \cdots = |V_r|$;
(iii) all but εr^2 pairs (V_i, V_j) with $1 \leq i < j \leq r$ are ε-regular.

It is worth making a quick remark that we do not assume that V_i, $i \in [r]$, is larger than the exceptional set V_0. In fact, quite the contrary, most of the time, we take $r \geq m \geq 1/\varepsilon$ to make the edges in V_i negligible.

Another remark is that in the definition of regular partition, we can also have no exceptional set (by distributing V_0 equally to other parts) and instead have $||V_i| - |V_j|| \leq 1$ for all $1 \leq i \leq j \leq r$.

We will use mostly the version of regular partition with no exceptional set V_0, unless otherwise specified.

We can now state the lemma.

Theorem 2.2.3 (Szemerédi regularity lemma 1976)**.** *Given $\varepsilon > 0$ and $m \in \mathbb{N}$, there exists $M = M(\varepsilon, m)$, such that any graph G admits an ε-regular partition $V = V_0 \cup V_1 \cup \cdots \cup V_r$ with $m \leq r \leq M$.*

Remark 2.2.4. Let us make some remarks about the parameters in the regularity lemma.

- We usually think of ε in the regularity lemma as a very small constant, i.e. $o(1)$.
- Both the lower and upper bounds $m \leq r \leq M$ on the number of parts of the partition are meaningful. If there is no lower bound, then the trivial partition $V = V$ consisting of just one part is vacuously a regular partition and clearly this partition is of no use for us. The upper bound on r is also needed as we shall see shortly, the proof of the counting lemma relies crucially on the fact that the reduced graph R we use to approximate the original graph G is of bounded order.

- If the graph G does not have positive edge-density, then the regularity lemma does not say much about G.
- The εr^2 exceptional irregular pairs are needed. Consider the following example:

  Half graph. $G = (A \cup B, E)$, where $A = B = [n]$. For any $a \in A$ and $b \in B$, put $ab \in E(G)$ if and only if $a \geq b$. Notice that $d(A, B) = 1/2$. Let the top half of A be X and bottom half of B be Y, then $d(X, Y) = 0$, while $d(A - X, B - Y) = 1$. There are εr irregular pairs in any partition.
- The upper bound on the size of the partition M coming from the proof of regularity lemma is rather large, it is a tower of 2s with height $2\varepsilon^{-5}$. Gowers gave a construction showing that a tower of 2s with height $\varepsilon^{-1/16}$ is needed.

We end this section with two simple lemmas. The first one states that between a regular dense pair, almost every vertex has the "correct" degree to any large subset of the other side.

Lemma 2.2.5. *Let (X, Y) be an ε-regular pair with density d, and $B \subseteq Y$ with $|B| \geq \varepsilon|Y|$, then all but $2\varepsilon|X|$ vertices in X have degree $(d \pm \varepsilon)|B|$ in B.*

Proof. Let $A \subseteq X$ be the set of vertices with "small" degree in B, i.e.

$$\frac{d(v, B)}{|B|} < d - \varepsilon.$$

Suppose that $|A| > \varepsilon|X|$, consider the pair (A, B). By the choice of A, we have

$$d(A, B) = \frac{e(A, B)}{|A||B|} < \frac{|A| \cdot (d - \varepsilon)|B|}{|A||B|} = d - \varepsilon,$$

contradicting (X, Y) being ε-regular. Thus, $|A| \leq \varepsilon|X|$. Similarly, the same bound holds for the set of vertices of "large" degree, i.e. $d(v, B)/|B| > d + \varepsilon$ in B. □

Given a regular pair (X, Y), one can also show that almost all pairs from one part, say X, have the "correct" codegree to large subsets of the other side.

Exercise **2.2.6.** Formulate the above codegree statement rigorously and prove it.

The second lemma states that regularity is inherited by large subsets of pairs (with a slightly worse precision/regularity). This lemma is useful as it implies that we can further refine a regular partition to get additional properties without losing regularity.

Lemma 2.2.7 (Slicing lemma)**.** *Let $V_0 \cup V_1 \cup \cdots \cup V_r$ be an ε-regular partition. Further refine each part into s equal parts: $V_i = V_i^1 \cup \cdots \cup V_i^s$. The new partition (with $sr + 1$ parts) is $O(s\varepsilon)$-regular.*[3]

Exercise **2.2.8.** Prove the slicing lemma.

[3] Note that $O(s\varepsilon)$-regular implicitly requires that in the slicing lemma, $s \ll 1/\varepsilon$.

2.3 Key lemmas

In the taster session, Section 2.1, we have seen how we can use **P.1** and **P.2** to carry out Step 1. In this section, we shall formally state and prove these two properties. They follows from two key consequences of the regularity lemma: the embedding lemma and the counting lemma.

Roughly speaking, the embedding lemma says that we can embed any (appropriate) bounded degree graphs (up to linear-size); and the counting lemma says for any fixed (small) graph H, we can count accurately the number of copies of H in G. We remark that there is a stronger version of embedding lemma, the blow up lemma, due to Komlós, Sárközy and Szemerédi, which we will not cover for now. The blow up lemma states that we can embed any (appropriate) spanning bounded degree graphs.

2.3.1 Reduced graph

We first define a notion of reduced graphs that appear in **P.1** and **P.2** formally.

Definition 2.3.1 (Reduced graph)**.** Given an ε-regular partition $V(G) = V_0 \cup V_1 \cup \cdots \cup V_r$ of G, and $\delta > 0$, the *reduced/cluster graph* $R = R(\varepsilon, \delta)$ of G is defined as follows:

- $V(R) = [r]$;
- $ij \in E(R)$ if and only if (V_i, V_j) is ε-regular with density at least δ.

We can also think of the reduced graph as a weighted graph, assigning weight $d_{ij} := d(V_i, V_j)$ to the edge ij, and define the weighted degree of vertex $i \in V(R)$ to be $\sum_{j\sim i} d_{ij}$. We will specify it when we treat R as a weighted graph.

Exercise **2.3.2.** Normalised minimum degree is inherited by the reduced graph $R = R(\varepsilon, \delta)$, i.e.[4]

$$\frac{\delta(R)+1}{r} \geq \frac{\delta(G)}{n} - \delta - \varepsilon \left(= \frac{\delta(G)}{n} - o(1) \right).$$

Exercise **2.3.3.** Bound the edge-density of G by that of R's:

$$t(K_2, G) \leq t(K_2, R) + o(1).$$

As we shall soon see in the counting lemma, the notion of reduced graph $R(\varepsilon, \delta)$ captures essentially the whole (asymptotic) information of G in terms of subgraphs densities.

[4] Here we need a slightly stronger version: when we get a regular parition from the regularity lemma, we can assume that each part V_i is in at most εr irregular pairs.

2.3.2 Embedding lemma

Given a graph F, denote by $F(s)$ the *blow-up* of F obtained from replacing each vertex $u \in V(F)$ by an independent set I_u of size s and make (I_u, I_v) complete bipartite in $F(s)$ if and only if $uv \in E(F)$. Observe that the blow-up $F(s)$ contains H as a subgraph is the same as saying that F contains a homomorphic copy of H.

We now present the embedding lemma, which is a formal statement of **P.2**. One thing to notice here is that we can embed appropriate bounded degree graphs of order up to linear size $\Omega(n)$ (so think of d, Δ below as constants and $|G| = \Theta(\ell)$, $|H| = \Theta(s)$ and $s = \Omega(\ell)$), which is what we use in proving Chvátal-Rödl-Szemerédi-Trotter theorem, see Section 2.7.

Lemma 2.3.4 (Embedding lemma)**.** *For any $d \in (0,1]$, $\Delta \geq 1$, there exists $\varepsilon_0 > 0$, such that for any G and H with $\Delta(H) \leq \Delta$ and for any $s \in \mathbb{N}$ and $R = R(\varepsilon, d)$ the reduced graph of G with $\varepsilon \leq \varepsilon_0$. Suppose the corresponding regular partition of G has each of its part of size $\ell \geq 2s/d^{\Delta}$. Then*

$$H \subseteq R(s) \quad \Rightarrow \quad H \subseteq G.$$

Sketch of proof. Given d, Δ, choose $\varepsilon_0 < d$, such that

$$(d - \varepsilon_0)^{\Delta} - \varepsilon_0 \Delta \geq \frac{1}{2} d^{\Delta} \geq \varepsilon_0.$$

Let $\varphi : V(H) \to V(R)$ be a homomorphism (exists as $H \subseteq R(s)$). Order vertices in H as $u_1, \dots, u_h$. Initially, set $Y_j = V_j$, for $j \in [r]$. Embed vertices $u_1, \dots, u_{i-1}$ one by one, and update the sets of eligible vertices $Y_{\varphi(u_j)} \subseteq V_{\varphi(u_j)}$ for each u_j, $j \geq i$ and $u_{i-1}u_j \in E(H)$, to embed by intersecting it with $N(u_{i-1})$, maintaining always $|Y_{\varphi(u_j)}| \geq \varepsilon |V_{\varphi(u_j)}|$. When embedding u_i in $Y_{\varphi(u_i)} \subseteq V_{\varphi(u_i)}$, note that for each $j > i$ with $u_i u_j \in E(H)$, in $V_{\varphi(u_i)}$, all but $\varepsilon |V_{\varphi(u_i)}|$ vertices u, by Lemma 2.2.5, satisfy $d(u, Y_{\varphi(u_j)}) \geq (d - \varepsilon)|Y_{\varphi(u_j)}|$. Since

$$|V_i|(d - \varepsilon)^{\Delta} - \varepsilon \Delta |V_i| \geq \max\{s, \varepsilon |V_i|\},$$

we never get stuck. □

2.3.3 Counting lemma

The formal statement of **P.1** is the following counting lemma.

Lemma 2.3.5 (Counting lemma)**.** *Given H, $V_1, ..., V_h$ with $h = |H|$ and $|V_i| = n$, all pairs (V_i, V_j) are ε-regular and $d(V_i, V_j) = d_{ij} \gg \varepsilon$. Then the number of canonical copies*[5] *of H in $V_1, ..., V_h$ is at least*

$$\prod_{ij \in E(H)} (d_{ij} - \sqrt{\varepsilon}) n^h.$$

We skip the proof for the counting lemma, instead leaving the baby case of triangle counting as exercise.

Exercise **2.3.6.** Prove counting lemma for the special case $H = K_3$.

Exercise **2.3.7.** Make the proof of the upper bound of Erdős-Stone theorem rigorous.

[5] By canonical copy, we mean a copy of H with exactly one vertex in each V_i.

2.4 Ruzsa-Szemerédi triangle removal lemma

In this section, we will present, yet, another important consequence of the regularity lemma, the removal lemma, due to Ruzsa and Szemerédi, which states that an almost triangle-free graph ($o(n^3)$ triangles) can be made genuinely triangle-free by removing a negligible amount of edges ($o(n^2)$ edges).

Lemma 2.4.1 (Ruzsa-Szemerédi triangle removal lemma 1976)**.** *Given $c > 0$, there exists $a = a(c) > 0$, such that for sufficiently large n the following holds. Let G be an n-vertex graph. If G has at most an^3 triangles, then it can be made triangle-free by removing at most cn^2 edges.*

The contrapositive says if one cannot make a graph triangle-free by removing few edges, then the graph contains lots (positive proportion) of triangles. The removal lemma has many applications, e.g. $(6, 3)$-theorem and Roth's theorem.

2.4.1 Cleaning the graph G

Before proving the removal lemma, It is convenient to define the subgraph G_R of G corresponding to a reduced graph $R = R(\varepsilon, \delta)$, obtained by keeping only edges between (regular and dense) pairs (V_i, V_j) for which $ij \in E(R)$. We can obtain the subgraph G_R via the following standard cleaning process, showing that only a negligible amount of edges are deleted:

$$e(G_R) = e(G) - o(n^2).$$

- Remove inner edges, i.e. edges in V_i, $i \in [r]$. By choosing $m \geq 1/\varepsilon$ when applying the regularity lemma to obtain the regular partition corresponding to $R(\varepsilon, \delta)$, we can guarantee the number of parts satisfies $r \geq m \geq 1/\varepsilon$. Then the number of inner edges is at most
$$\binom{n/r}{2} \cdot r \leq \frac{n^2}{2r} \leq \frac{n^2}{2m} = \frac{1}{2}\varepsilon n^2.$$

- Remove edges between irregular pairs. As there are at most εr^2 irregular pairs, the number of edges of this kind is at most
$$\varepsilon r^2 \cdot (n/r)^2 = \varepsilon n^2.$$

- Remove edges between sparse pairs with density at most δ, i.e. (V_i, V_j) with $ij \notin E(R)$. The number of such edges is at most
$$\delta \left(\frac{n}{r}\right)^2 \binom{r}{2} \leq \frac{1}{2}\delta n^2.$$

Thus, in forming G_R, we delete in total at most

$$\frac{1}{2}(3\varepsilon+\delta)n^2 = O(\varepsilon+\delta)n^2$$

edges, which is negligible as we usually choose ε, δ sufficiently small.

The cleaning graph process above is exactly the non-essential information we discard when forming the reduced graph. Edges in G_R all lie in regular and dense pairs and so we can employ e.g. the counting lemma, which is how we shall prove the triangle removal lemma.

2.4.2 Proof of triangle removal lemma

Suppose the statement is not true. That is, there is some $c > 0$ such that for any a there exists a counterexample G, i.e. G has at most an^3 triangles, but the removal of any cn^2 edges does not make it triangle-free.

Apply Szemerédi's regularity lemma with $\varepsilon = c/8$ and $m = 1/\varepsilon$ to G to get an ε-regular partition $V(G) = V_1 \cup ... \cup V_r$, where $M \geq r \geq m$ and $||V_i| - |V_j|| \leq 1$, for $1 \leq i, j \leq r$. Let $R = R(\varepsilon, c/4)$ be the reduced graph, and $G_R \subseteq G$ be the cleaned subgraph, as in Section 2.4.1. Then the number of edges deleted is at most $cn^2/2$.

By the choice of G, there is still triangles in G_R, which can only be in three sets, say X, Y, Z, that are pairwise regular with density larger than $c/4$. We can then apply the counting lemma to the tripartite graph $G_R[X, Y, Z]$ to see that there are at least

$$\left(\frac{c}{4}-\varepsilon\right)^3 \cdot \left(\frac{n}{r}\right)^3 \geq \left(\frac{c}{8r}\right)^3 n^3 \geq \left(\frac{c}{8M}\right)^3 n^3$$

triangles. Note that $M = M(\varepsilon, m)$ depends in fact only on c. Then choosing $a = a(c) < \left(\frac{c}{8M}\right)^3$, we get that G has more than an^3 triangles, a contradiction.

Remark 2.4.2. To get more familiar, let us write a streamlined proof without all the calculations. Let G be an almost triangle-free graph. Then its reduced graph R must be triangle-free, as otherwise, by the counting lemma, G would contain too many triangles. Thus, G can be made triangle-free by removing few edges not corresponding to R.

2.5 $(6,3)$-theorem and Roth's theorem

In this section, we present Ruzsa-Szemerédi $(6,3)$-theorem and see how it implies Roth's theorem on 3-term arithmetic progression (3AP).

Throughout this section, we will work with 3-uniform hypergraphs $\mathcal{H} = (V, E)$, where the edge set $E \subseteq \binom{V}{3}$ consists of triples in V. We say a hypergraph is *linear* if any two of its edges share at most one vertex in common. For $s, t \in \mathbb{N}$, an (s,t)-configuration (or simply (s,t)) in a hypergraph is a set of s vertices inducing at least t edges. A hypergraph is (s,t)-*free* if it does not contain any (s,t)-configuration.

Theorem 2.5.1 ((6,3)-theorem)**.** *If a 3-uniform hypergraph $\mathcal{H}$ is $(6,3)$-free, then*

$$e(\mathcal{H}) = o(n^2).$$

We remark that this upper bound is not very far from optimal: there exists 3-uniform $\mathcal{H}$ that is $(6,3)$-free and have $e(\mathcal{H}) > n^2 \cdot e^{-c\sqrt{\ln n}}$, which is larger than $n^{2-\varepsilon}$ for any constant $\varepsilon > 0$. We shall give this lower bound construction after Theorem 2.5.3.

Proof of (6,3) theorem. Suppose to the contrary that there exists $c > 0$ such that for infinitely many n, there is a $(6,3)$-free 3-uniform n-vertex $\mathcal{H}$ with $e(\mathcal{H}) > cn^2$. By zooming into a subgraph with higher average degree (which is still a counterexample), we may in addition assume that $\mathcal{H}$ is maximal in the sense that no subgraph of $\mathcal{H}$ has larger average degree than $\mathcal{H}$.

The maximality of $\mathcal{H}$ implies that it is linear. Indeed, if there are two edges intersecting at two points, we have a $(4,2)$-configuration, then no other edges intersect at these four points, as otherwise we get a $(6,3)$. Thus these two edges form a component themselves. Then deleting this component results in a subgraph with higher average degree than $\mathcal{H}$, contradicting the maximality of $\mathcal{H}$. Note also that, since $\mathcal{H}$ is linear, $\mathcal{H}$ is a steiner triple system on n vertices, which is known to have at most $(1/6 + o(1))n^2$ hyperedges.

Let G be the *shadow graph* of $\mathcal{H}$, obtained by setting $V(G) = V(\mathcal{H})$ and turning every hyperedge in $\mathcal{H}$ into a triangle. We say a triangle in G is an *$\mathcal{H}$-triangle* if it corresponds to a hyperedge in $\mathcal{H}$. Since $\mathcal{H}$ is linear, no two $\mathcal{H}$-triangles in G share an edge. Thus, there are at least $e(\mathcal{H}) > cn^2$ edge-disjoint $\mathcal{H}$-triangles in G. Consequently, G cannot be made triangle-free by removing at most cn^2 edges, and so the removal lemma implies that G contains at least an^3 triangles in G, where $a = a(c)$. For large n, $an^3 > n^2 > e(\mathcal{H})$, meaning that there are triangles in G that does not come from a hyperedge in H. Such a triangle in G corresponds to a $(6,3)$ in $\mathcal{H}$ as $\mathcal{H}$ is linear, a contradiction. □

Let us now see how $(6,3)$-theorem implies Roth's theorem on 3APs.

Theorem 2.5.2 (Roth's theorem)**.** *For any $\delta > 0$, there exists n_0 such that for $n \geq n_0$, any subset $S \subseteq [n]$ with size δn contains a three-term arithmetic progression.*

Theorem 2.5.3. *$(6,3)$-theorem $\Rightarrow$ Roth's Theorem.*

Proof. Suppose there is a 3AP-free set $A \subseteq [n]$ with $|A| \geq \delta n$. Define a 3-partite 3-uniform $\mathcal{H}$ as follows: $V(\mathcal{H}) = V_1 \cup V_2 \cup V_3$, where $V_1 = [n]$, $V_2 = [2n]$ and $V_3 = [3n]$; for the edge set, for each $x \in [n]$ and $a \in A$, add the hyperedge $(x, x+a, x+2a)$. So

$$e(\mathcal{H}) = |A||V_1| \geq \delta n^2.$$

Thus, Theorem 2.5.1 implies that there is a $(6,3)$ in $\mathcal{H}$. Say the three hyperedges in this $(6,3)$-configuration are $(x, x+a, x+2a), (y, y+b, y+2b)$ and $(z, z+c, z+2c)$. Since two points completely determine an edge in $\mathcal{H}$, this $(6,3)$ has to have two points from each V_i, $1 \leq i \leq 3$. Without loss of generality, say $x = z \neq y$, then $x + a \neq z + c$ and $x + 2a \neq z + 2c$, otherwise two edges coinside. Again without loss of generality, say $y + 2b = x + 2a$, then similarly $y+b \neq x+a$ and so $y+b = z+c$. Then a simple calculation shows that $b+c = 2a$. Note that $a, b, c \in A$ and $a \neq c$ since $x = z$ and $x + a \neq z + c$. Thus $\{b, a, c\} \subseteq A$ forms a 3AP, a contradiction. □

Dense (6,3)-free $\mathcal{H}$. In the above proof, in fact A is 3AP-free if and only if $\mathcal{H}$ is (6,3)-free. Behrend constructed a 3AP-free subset of $[n]$ of size $n \cdot e^{-c\sqrt{\log n}}$. The corresponding hypergraph $\mathcal{H}$ then is (6,3)-free and has $n^2 \cdot e^{-c\sqrt{\log n}}$ edges.

We end this section with an old conjecture.

Conjecture 2.5.4 (Brown-Erdős-Sós 1973)**.** *If a 3-uniform $\mathcal{H}$ is $(s+3, s)$-free, then*

$$e(\mathcal{H}) = o(n^2).$$

The simplest open case is $(7, 4)$.

2.6 Ramsey-Turán problem for K_4

In this section, we present an appication of the regularity lemma in Ramsey-Turán problem. Recall that Turán's theorem states that among all n-vertex K_{r+1}-free graphs, the Turán graph $T_r(n)$ has the largest size. Notice that these Turán graphs have rigid structures, in particular, there are independent sets of size linear in n. It is then natural to ask what happens when there is no such big holes. Such problems, first introduced by Sós in 1969, are the substance of the Ramsey-Turán theory.

Given a graph H and natural numbers $m, n \in \mathbb{N}$, the *Ramsey-Turán number* for H is:

$$\mathsf{RT}(n, H, m) := \max\{e(G) : \ |G| = n, \ \alpha(G) \leq m, \text{ and } G \text{ is } H\text{-free}\}.$$

The most classical case is when m is sublinear in n, i.e. $m = o(n)$. Formally,

Definition 2.6.1. Given a graph H and $\delta \in (0, 1)$, let

$$\varrho(H, \delta) := \lim_{n\to\infty} \frac{\mathsf{RT}(n, H, \delta n)}{n^2} \quad \text{and} \quad \varrho(H) := \lim_{\delta\to 0} \varrho(H, \delta).$$

Define

$$\mathsf{RT}(n, H, o(n)) = \varrho(H) \cdot n^2 + o(n^2).$$

Exercise **2.6.2.** Prove that $\mathsf{RT}(n, K_3, o(n)) = o(n^2)$.

When there is no restriction on the independence number, recall that $\mathsf{ex}(n, K_4) = n^2/3 \pm O(1)$. In comparison,

Theorem 2.6.3 (Szemerédi)**.** $\mathsf{RT}(n, K_4, o(n)) \leq n^2/8 + o(n^2)$.

Sketch of proof. Let G be an n-vertex K_4-free graph with $\alpha(G) = o(n)$. Let R be a weighted reduced graph of G. It suffices to show that R is triangle-free and no edge in R has density larger than $1/2$. Indeed, K_3-free implies that, as a graph, R has at most $r^2/4$ edges; each edge having weight at most $1/2+o(1)$ implies that, as a weighted graph, $e(R) \leq r^2/8+o(r^2)$, and hence $e(G) \leq n^2/8 + o(n^2)$ as desired.

Suppose R has a triangle ijk. Consider the corresponding pairwise dense regular triple V_i, V_j, V_k in G. We can find two typical adjacent vertices $v_iv_j \in E(G)$ with $v_i \in V_i$ and $v_j \in V_j$, having linear codegree in V_k: $d(v_i, v_j, V_k) = \Omega(n)$. As $\alpha(G) = o(n)$, there is an edge in $N(v_i, v_j, V_k)$, yielding a copy of K_4, a contradiction.

Suppose R has a chubby edge ij, and so $d(V_i, V_j) \geq 1/2 + \Omega(1)$. Then any two typical vertices in V_i has codegree $2(n/2 + \Omega(n)) - n = \Omega(n)$ linear in V_j. This also yields a K_4, as almost all vertices (hence linear many) in V_i are typical, we can find two adjacent ones and pick an edge in their coneighbourhood in V_j, again reaching a contradiction. □

An ingenious geometric construction of Bollobás and Erdős later yields a matching lower bound:

$$\mathsf{RT}(n, K_4, o(n)) = \frac{n^2}{8} + o(n^2).$$

The following question is open.

Question 2.6.4. *Is* $\mathsf{RT}(n, K_{2,2,2}, o(n)) = o(n^2)$*?*

2.7 Chvátal-Rödl-Szemerédi-Trotter theorem

In this section, we present an application of the regularity lemma in graph Ramsey theory. Recall that the Ramsey number $r(G, G)$ for a graph G is the minimum integer N such that any 2-edge-colouring of K_N contains a monochromatic copy of G. The Ramsey number for cliques is exponential. A theorem of Chvátal-Rödl-Szemerédi-Trotter states that bounded degree graphs have linear Ramsey number.

Theorem 2.7.1. *Let* $d \in \mathbb{N}$ *and* G *be a graph with* $\Delta(G) \leq d$*, then*

$$r(G, G) = O_d(|V(G)|).$$

We will make use of the multicolour version of the Szemerédi Regularity Lemma. For a k-edge-coloured graph G, a partition $V(G) = V_1 \cup \ldots \cup V_r$ is an *ε-regular partition* if

- for all $ij \in \binom{[r]}{2}$, $\big||V_i| - |V_j|\big| \leq 1$;
- for all but at most $\varepsilon\binom{r}{2}$ choices of $ij \in \binom{[r]}{2}$, the pair (V_i, V_j) is ε-regular in *every* colour.

Lemma 2.7.2 (Multicolour regularity lemma)**.** *For every real* $\varepsilon > 0$ *and integers* $k \geq 1$ *and* m*, there exists* $M = M(\varepsilon, m, k)$ *such that every* k*-edge-coloured graph* G *with* $n \geq m$ *vertices admits an* ε*-regular partition* $V(G) = V_1 \cup \ldots \cup V_r$ *with* $m \leq r \leq M$*.*

We can similarly define a reduced graph corresponding to a regular partition, with the only difference that $ij \in E(R)$ if and only if (V_i, V_j) is regular with respect to every colour. The reduced graph inherits a (multi)edge-colouring from G: we can assign each edge $ij \in E(R)$, the set of all colours that is dense in $G[V_i, V_j]$. For the application here, it suffices to just assign the *majority* colour, i.e. if $ij \in E(R)$ is red, then (V_i, V_j) has red-density at least $1/k - o(1) = \Omega(1)$.

Let us recall Brook's theorem, which will be needed in the proof.

Theorem 2.7.3 (Brook's theorem)**.** *Every graph G can be properly vertex-coloured using $\Delta(G) + 1$ colours, i.e. $\chi(G) \leq \Delta(G) + 1$.*

Proof of Theorem 2.7.1. Let $m \geq 5r(K_{d+1}, K_{d+1})$ be sufficiently large, $\varepsilon = 1/m$,

$$C := 2M/(1/2 - \varepsilon)^d,$$

where $M = M(\varepsilon, m, 2)$ returned from Lemma 2.7.2. Let $N \geq C|V(G)|$ and fix an arbitrary 2-edge-colouring of K_N. We shall find a monochromatic copy of G.

Apply multicolour regularity lemma to the given 2-edge-coloured K_N and let R be the corresponding reduced graph. Recall that R is almost complete:

$$e(R) \geq (1 - 2\varepsilon)\binom{r}{2} > \left(1 - \frac{1}{m/3 - 1}\right)\frac{r^2}{2},$$

with a 2-edge-colouring indicating the majority colour.

By Turán theorem, it R contains a clique $K_{m/3}$. As $m \geq 5r(K_{d+1}, K_{d+1})$, in this 2-edge-coloured clique $K_{m/3}$, there is a monochromatic K_{d+1}. By Brook's theorem, K_{d+1} is a homomorphic image of G, as $\Delta(G) \leq d$. Then by the embedding lemma, the original 2-edge-coloured K_N contains a monochromatic copy of G as desired. □

2.8 Spectral proof of regularity lemma

Finally, we give a proof of the regularity lemma. The original proof proceeds by refining partition and energy increment strategy. Here, we shall give a proof based on the spectral decomposition of the adjacency matrix given by Tao, and independently by Szegedy. This idea originates from Frieze-Kannon's proof of the weak regularity lemma.

We shall only prove a weaker version in which we do not require equipartition. One can refine this regular partition further to get a equipartition.

Lemma 2.8.1. *Let G be an n-vertex graph and let $\varepsilon > 0$. Then there exists a partition $V = V_1 \cup \cdots \cup V_M$, $M \leq M(\varepsilon)$, such that apart from an exceptional set $\Sigma \subseteq \binom{[M]}{2}$ with*

$$\sum_{(i,j)\in\Sigma} |V_i||V_j| = O(\varepsilon|V|^2),$$

we have for every $(i,j) \notin \Sigma$, $A \subseteq V_i$ and $B \subseteq V_j$ that

$$|e(A, B) - d_{ij}|A||B|| = O(\varepsilon|V_i||V_j|).$$

Before we dive into the details of the proof, let us sketch briefly how it goes. We write the adjacency matrix T as the sum of the rank-1 matrices from eigenvectors of T with weights being the associated eigenvalues. Then the structure of T is dictated mostly by the part with large eigenvalues (main term); while the part with small eigenvalues is more like noise (error term). Thus, to capture the behaviour of T, we shall get a partition in which each eigenvector with large eigenvalue is approximately constant in each part.

We need Cauchy-Schwarz inequality for the proof, let us recall it here.

Lemma 2.8.2 (Cauchy-Schwarz inequality)**.** *Let $u, v \in \mathbb{C}^n$, then*

$$\sum_{i=1}^{n} u_i \overline{v_i} = \langle u, v \rangle \leq \|u\|_2 \|v\|_2 = \sqrt{\sum_{i=1}^{n} |u_i|^2 \sum_{i=1}^{n} |v_i|^2}.$$

Furthermore, equality holds if and only if u and v are linearly dependent.

Proof of the regularity lemma, Lemma 2.8.1. Let T be the adjacency matrix of G. As T is a real symmetric matrix, it is self-adjoint and has eigenvalue decomposition:[6]

$$T = \sum_{i=1}^{n} \lambda_i u_i u_i^*,$$

where $u_1, \ldots, u_n$ form an orthonormal basis of $\mathbb{C}$ with eigenvalues $|\lambda_1| \geq \ldots \geq |\lambda_n| \in \mathbb{R}$.

Splitting T. As outlined above, we shall splits $T = T_1 + T_2 + T_3$ into main term T_1 and error terms T_2, T_3. To do so, we need a bound on the eigenvalues. Note that the (i, j)-th entry in T^k records the number of v_i, v_j-walks with length k in G. In particular, each diagonal entry of T^2 is the degree of the corresponding vertex. Thus, the trace

$$\mathsf{tr}(T^2) = \sum_i d_i = 2e(G) \leq n^2,$$

and for each $i \in [n]$, we have

$$i \cdot |\lambda_i|^2 \leq \sum_{i=1}^{n} |\lambda_i|^2 \leq n^2 \quad \Rightarrow \quad |\lambda_i| \leq \frac{n}{\sqrt{i}}. \tag{2.1}$$

Let $F = F(\varepsilon) : \mathbb{N} \to \mathbb{N}$ be a function to be chosen later with $F(i) \geq i$. By averaging, for some $J \leq F^{1/\varepsilon^3}(1)$[7], we can take out a piece in the middle with small weight[8]:

$$\sum_{i \in [J, F(J)]} |\lambda_i|^2 \leq \varepsilon^3 n^2. \tag{2.2}$$

We can now write $T = T_1 + T_2 + T_3$, where

[6] We treat all vectors here as column vectors.

[7] We use $F^k = F \circ \cdots \circ F$ for k iteration of F.

[8] Proof: Consider the partition of $[n]$ into intervals $[1, F(1)) \cup [F(1) \cup F^2(1)) \cup [F^2(1), F^3(1)) \cdots$. As $\sum_{i \in [n]} |\lambda_i|^2 \leq n^2$, one of the first $1/\varepsilon^3$ intervals should be at most $\varepsilon^3 n^2$.

- $T_1 = \sum_{i \le J} \lambda_i u_i u_i^*$ is the "*structured*" term;
- $T_2 = \sum_{i \in [J, F(J)]} \lambda_i u_i u_i^*$ is the "*small*" term;
- $T_3 = \sum_{i > F(J)} \lambda_i u_i u_i^*$ is the "*pseudorandom*" term.

Partition for the structured term T_1. We now construct a partition of $V(G)$ such that T_1 is approximately constant in most parts. For each $i \le J$, we partition $V(G)$ into $O_{J,\varepsilon}(1)$ parts in which u_i only fluctuates by $O(\frac{\varepsilon^{3/2}}{J} n^{-1/2})$ apart from an exceptional part of size $O(\frac{\varepsilon}{J} n)$ where $|u_i|$ is excessively large (of value at least $\sqrt{\frac{J}{\varepsilon}} n^{-1/2}$). Let $u = u_i$ and write $u(j)$ for the j-th coordinate of u. Recall that $\|u\|_2^2 = \sum_{j \in [n]} u(j)^2 = 1$, so the number of coordinates with value at least $\sqrt{\frac{J}{\varepsilon}} n^{-1/2}$ is at most $\frac{\varepsilon}{J} n$. Thus, for the rest of the coordinates, we can partition it into at most $\sqrt{\frac{J}{\varepsilon}} n^{-1/2} / (\frac{\varepsilon^{3/2}}{J} n^{-1/2}) = O_{J,\varepsilon}(1)$ parts as claimed.

Combining all of these J partitions together, we get $V(G) = V_1 \cup \dots \cup V_{M-1} \cup V_M$, $M = O_{J,\varepsilon}(1)$, where the exceptional part $|V_M| \le \varepsilon n$, and for any $1 \le i \le M-1$, the eigenvectors $u_1, \dots, u_J$ all fluctuate at most $O(\frac{\varepsilon}{J} n^{-1/2})$. We claim that T_1 fluctuates at most $O(\varepsilon)$ on each block $V_i \times V_j$, for $1 \le i, j \le M-1$, and consequently, writing d_{ij} for the mean value of entries of T_1 on $V_i \times V_j$, we have for any $A \subseteq V_i$, $B \subseteq V_j$, that

$$\mathbb{1}_A^* T_1 \mathbb{1}_B = d_{ij} |A||B| + O(\varepsilon |V_i||V_j|). \tag{2.3}$$

Indeed, recall that $|\lambda_i| \le n/\sqrt{i}$, we see that each $V_i \times V_j$-entry of $T_1 = \sum_{i \le J} \lambda_i u_i u_i^*$ fluctuates by at most

$$\sum_{i \le J} \lambda_i \cdot \sqrt{\frac{J}{\varepsilon}} n^{-1/2} \cdot O\left(\frac{\varepsilon^{3/2}}{J} n^{-1/2}\right) = O\left(\frac{\varepsilon}{\sqrt{J}} n^{-1}\right) \cdot \sum_{i \le J} \frac{n}{\sqrt{i}} = O(\varepsilon).$$

Bounding error term T_2. By the choice of T_2 and (2.2), $\mathsf{tr}(T_2^2) \le \varepsilon^3 n^2$. On the other hand, let x_{ab} be (a,b)-th entry of T_2, as T_2 is self-adjoint, we have $\mathsf{tr}(T_2^2) = \sum_{a,b \in V(G)} |x_{ab}|^2$. Then by Markov inequality, we get

$$\sum_{a \in V_i, b \in V_j} |x_{ab}|^2 \le \varepsilon^2 |V_i||V_j|, \tag{2.4}$$

for all $1 \le i, j \le M-1$ apart from an exceptional set $\Sigma' \subseteq \binom{[M-1]}{2}$ with $\sum_{(i,j) \in \Sigma'} |V_i||V_j| \le \varepsilon n^2$. Hence, for any $(i,j) \notin \Sigma'$ and $A \subseteq V_i, B \subseteq V_j$, by (2.4) and Cauchy-Schwarz, we have

$$\mathbb{1}_A^* T_2 \mathbb{1}_B \le \sum_{a \in V_i, b \in V_j} |x_{ab}| \le \left(\sum_{a \in V_i, b \in V_j} |x_{ab}|^2 \right)^{1/2} (|V_i||V_j|)^{1/2} = O(\varepsilon |V_i||V_j|). \tag{2.5}$$

Bounding error term T_3. By the choice of T_3 and (2.1), the operator norm of T_3 is at most $\|T_3\|_{\mathsf{op}} \le \frac{n}{\sqrt{F(J)}}$. Then by Cauchy-Schwarz, we have

$$\mathbb{1}_A^* T_3 \mathbb{1}_B = \langle \mathbb{1}_A, T_3 \mathbb{1}_B \rangle \le \|\mathbb{1}_A\|_2 \cdot \|T_3 \mathbb{1}_B\|_2 \le \|\mathbb{1}_A\|_2 \cdot \|T_3\|_{\mathsf{op}} \cdot \|\mathbb{1}_B\|_2 = O\left(\frac{n^2}{\sqrt{F(J)}}\right). \quad (2.6)$$

Set $\Sigma := \Sigma' \cup \{(i,j) : \ i \text{ or } j = M\} \cup \{(i,j) : \min\{|V_i|, |V_j|\} \le \varepsilon n/M\}$. Then easy to check that $\sum_{(i,j)\in\Sigma} |V_i||V_j| \le O(\varepsilon n^2)$. By (2.3), (2.5), (2.6), we have

$$e(A,B) = \mathbb{1}_A^* T \mathbb{1}_B = \sum_{i=1}^{3} \mathbb{1}_A^* T_i \mathbb{1}_B = d_{ij}|A||B| + O(\varepsilon |V_i||V_j|) + O\left(\frac{n^2}{\sqrt{F(J)}}\right).$$

As $|V_i|, |V_j| \ge \varepsilon n/M$, we have $\frac{n^2}{\sqrt{F(J)}} \le \frac{M^2 |V_i||V_j|}{\varepsilon^2 \sqrt{F(J)}}$. To absorb the 2nd error term into the first one, we need $\frac{1}{\sqrt{F(J)}} = O(\varepsilon^3/M^2)$. □

Remark 2.8.3. The point of having T_2-term is to have local control on the fluctuation of $e(A,B)$, i.e. $O(\varepsilon|V_i||V_j|)$. For the tail T_3, we only have a global type control $O(\frac{n^2}{\sqrt{F(J)}})$, and we need $\sqrt{F(J)} \ge \frac{M^2}{\varepsilon^3}$, to make it into a local error. Recall that $M \ge J^J$ when we combine the partitions for each u_i, $i \le J$. Thus, we need to create a gap between $F(J)$ and J by splitting out a small term T_2 in the middle.

Chapter 3

Pseudorandomness

In the last chapter, we studied Szemerédi's regularity lemma, which partitions any (large) graph into parts such that almost all pairs of parts induces a random-like bipartite graph. This random-like property then enables us (counting lemma, embedding lemma) to use expectation to approximate some graph parameters (subgraph density).

In this chapter, we will take a look at the notion of *pseudorandomness*, also referred to in other contexts as *quasirandomness, regularity, uniformity.* Pseudorandomness has played an important role in not just extremal combinatorics, but also other fields such as number theory, probability, coding theory and theoretical computer science.

3.1 Quasirandom graphs

We will first take a look at *quasirandom graphs*, introduced in the 80s by Thomason and independently by Chung-Graham-Wilson. We shall define several properties that at the first glance seems irrelevant of one another but turns out to be equivalent in the sense of being random-like. One immediate application of this is that we have many different ways of checking whether a graph is quasirandom, as if a graph satisfies any one of the equivalent properties, then it satisfies all of them.

We need some notations before stating the equivalent quasirandom properties. Throughout this section, G will be an n-vertex graph with edge density $p \in (0,1)$, i.e. $e(G) = p\binom{n}{2}$. When reading this section, we should compare G with the Erdős-Rényi random graph $G(n,p)$.[1] We let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of the adjacency matrix T of G, ordered by

$$|\lambda_1| \geq \ldots \geq |\lambda_n|.$$

We will write

$$d(u,v) = |N(u) \cap N(v)|$$

for the *codegree* of u and v.

[1] The Erdős-Rényi random graph $G(n,p)$ is the probability space of all graphs on vertex set $[n]$ with p-biased measure, equivalently, $G(n,p)$ is a random graph on $[n]$ in which every pair forms an edge with probability p independent of all other pairs.

3.1.1 Equivalent definitions of quasirandomness

We can now state the aforementioned properties:

- (Induced Subgraph Count) For every graph H, the number of labelled induced copy of H in G is $p^{e(H)}(1-p)^{e(\overline{H})} + o(1)$.
- (Subgraph Count) For every graph H, $t(H, G) = p^{e(H)} + o(1)$.
- (4-cycle Count) $t(C_4, G) \leq p^4 + o(1)$.
- (Spectral Gap) $|\lambda_2| = o(n)$.
- (Discrepancy) For any $A, B \subseteq V(G)$, $e(A, B) = p|A||B| + o(n^2)$.
- (Codegree) $\sum_{u,v \in V(G)} |d(u,v) - p^2 n| = o(n^3)$.

Notice that these properties hold almost surely in $G(n, p)$.

The result of Thomason and Chung-Graham-Wilson states that all the above properties are equivalent. A graph G is called *quasirandom* if it satisfies any one of the above properties. It is surprising at first that the seemingly weaker property of having the correct C_4 count implies the correct count of all subgraphs densities.

We shall give a proof for regular graphs.

Theorem 3.1.1. *Let $p \in (0,1)$ and G be an n-vertex d-regular graph with $d = pn$, then all the above properties are equivalent.*

Proof. We will prove (Induced Subgraph Count) ⇒ (Subgraph Count) ⇒ (4-cycle Count) ⇒ (Spectral Gap) ⇒ (Discrepancy) ⇒ (Codegree). We defer the proof of (Codegree) ⇒ (Induced Subgraph Count) to the next subsection.

- (Induced Subgraph Count) ⇒ (Subgraph Count) Exercise.
- (Subgraph Count) ⇒ (4-cycle Count) By definitions.
- (4-cycle Count) ⇒ (Spectral Gap) This amounts to write C_4-count using trace of T^4 and the correct count of C_4 means the contribution from the non-trivial eigenvalues λ_i, $i \geq 2$, is negligible.

  More precisely, note that $T^k_{u,v}$, the u,v-th entry of the k-th power of the adjacency matrix T, is the number of u,v-walk of length k in G. Then the trace of T^k

  $$\mathsf{tr}(T^k) = \sum_{i \in [n]} \lambda_i^k$$

  counts the number of closed walks of length k in G. Among these walks, the non-degenerate ones are C_k, while the degenerate ones is easily seen to be negligible,

$O(n^{k-1})$. Recall that for d-regular graphs, $\lambda_1 = d$. Splitting out the first term in $\mathsf{tr}(T^4)$, we see that

$$p^4 n^4 \pm o(n^4) \geq t(C_4, G) n^4 \pm o(n^4) = \mathsf{tr}(T^4) = (pn)^4 + \sum_{i=2}^{n} \lambda_i^4,$$

implying that $|\lambda_i| = o(n)$ for all $i \geq 2$.

- (Spectral Gap) $\Rightarrow$ (Discrepancy) Expander mixing lemma.
- (Discrepancy) $\Rightarrow$ (Codegree) We shall prove a stronger statement that every vertex has small codegree deviation: for any u,

$$\sum_{v: v \neq u} |d(u,v) - p^2 n| = o(n^2).$$

  To see this, we split $V(G) \setminus \{u\} = B \cup B'$, where $B := \{v : d(u,v) > p^2 n\}$. This splitting helps us to get rid of absolute value sign: writing $A := N(u)$ and so $|A| = pn$, we have

$$\begin{aligned} \sum_{v: v \neq u} |d(u,v) - p^2 n| &= (e(A,B) - p^2 n|B|) + (p^2 n|B'| - e(A,B')) \\ &= (e(A,B) - p|A||B|) + (p|A||B'| - e(A,B')). \end{aligned}$$

  Now applying (Discrepancy) to each of the two terms above finishes the proof.

□

Exercise **3.1.2.** Prove that (Induced Subgraph Count) $\Rightarrow$ (Subgraph Count).

3.1.2 (Codegree) $\Rightarrow$ (Induced Subgraph Count).

Proof. Let H be a graph on vertex set $\{v_1, v_2, \ldots, v_s\}$, and for $r \in [s]$, let $H_r := H[\{v_1, \ldots, v_r\}]$. We will use induction on $1 \leq r \leq s$, via building H_{r+1} from H_r, to show that G has the correct count of copies of $H = H_s$. That is, writing N_r for the number of labelled induced copies of H_r in G, we shall show

$$N_r = (1 + o(1)) n^r p^{e(H_r)} (1-p)^{e(\overline{H_r})}. \tag{3.1}$$

The base case $r = 1$ is clearly true. Assume now it holds for $1 \leq r < s$, we will prove it for $r + 1$.

<u>*Extension function.*</u> To count copies of H_{r+1}, we will make use of a function that helps us to count the number of ways to extend a copy of H_r to H_{r+1}. For this purpose, let $\epsilon \in \{0,1\}^r$ be the 0/1-vector encoding the adjacencies of v_{r+1} to $\{v_1, \ldots, v_r\}$ in H_{r+1}, namely,

$$\epsilon = \{\epsilon_1, \ldots, \epsilon_r\} \text{ with } \epsilon_i = 1 \text{ if and only if } v_i \sim v_{r+1} \text{ in } H_{r+1}.$$

For $z = \{z_1, \ldots, z_r\} \in V_{(r)}$[2], an ordered set of r distinct vertices in G, let

$$X(z) := |\{v \in V(G) : v \neq z_1, \ldots, z_r, \text{ and } v \sim z_i \text{ if and only if } \epsilon_i = 1, \text{ for } 1 \leq i \leq r\}|.$$

We write $z \cong H_r$ when $z \in V_{(r)}$ induces a copy of H_r. Note that for any $z \cong H_r$, the function $X(z)$ counts exactly the number of ways to extend z to a labelled induced copy of H_{r+1}.

We can view $X(z)$ probabilistically as follows. Think of $X(z)$ as a random variable X drawn from the space $\Omega := V_{(r)}$ with uniform measure, that is, for any $z \in \Omega$,

$$\Pr[X = X(z)] = \frac{1}{n_{(r)}}.$$

As observed above, writting $\Omega^* = \{z : \ z \cong H_r\} \subseteq \Omega$, we can count copies of H_{r+1} by summing up the number of extensions of each $z \cong H_r$ to H_{r+1}:

$$N_{r+1} = \sum_{z \in \Omega^*} X(z). \tag{3.2}$$

<u>*Inductive step assuming concentration of X.*</u> Later, we will bound the variance of X to show that each $X(z)$ is close to the mean $\mathrm{E}[X]$, in particular:

$$\sum_{z \in \Omega^*} X(z) = |\Omega^*| \cdot \mathrm{E}[X] + o(n^{r+1}). \tag{3.3}$$

Assuming (3.3) for now, let us see how it finishes the proof. Observe first that $\sum_{z \in \Omega} X(z)$ counts the number of ordered $(r+1)$-tuple $\{u_1, \ldots, u_r, u_{r+1}\} \in V_{(r+1)}$ such that the adjacencies of u_{r+1} to $\{u_1, \ldots, u_r\}$ is ϵ. We can count this quantity from u_{r+1}'s point of view as follows. First choose u_{r+1}, for which there are n ways; and then choose an ordered r-tuple, among which exactly $|\epsilon|$ are neighbours of u_{r+1} and $r - |\epsilon|$ are non-neighbours of u_{r+1}, where $|\epsilon|$ denotes the number of 1s in ϵ. The number of such ordered r-tuple is

$$(1 + o(1)) p^{|\epsilon|} (1-p)^{r-|\epsilon|} n^r.$$

For simplicity, let

$$\rho := p^{|\epsilon|} (1-p)^{r-|\epsilon|}.$$

Thus, $\sum_{z \in \Omega} X(z) = (1 + o(1)) \rho n^{r+1}$. We can then compute the mean:

$$\mathrm{E}[X] \ = \ \sum_{z \in \Omega} \Pr[X = X(z)] \cdot X(z) = \frac{1}{n_{(r)}} \sum_{z \in \Omega} X(z) = (1 + o(1)) \rho n.$$

Recall the definition of Ω^*, we see that

$$|\Omega^*| = N_r.$$

[2] We write $V_{(r)}$ for the set of all ordered r-tuples of vertices in V, and $|V_{(r)}| = n_{(r)} = n \cdot (n-1) \cdots (n-r+1)$ for the r-falling factorial of n.

Then, using (3.2) and (3.3), we derive

$$\begin{aligned} N_{r+1} &= |\Omega^*| \cdot \mathrm{E}[X] + o(n^{r+1}) = (1+o(1))N_r \rho n + o(n^{r+1}) \\ &= (1+o(1))n^{r+1} p^{e(H_{r+1})}(1-p)^{e(\overline{H_{r+1}})}, \end{aligned}$$

where the last equality follows from the induction hypothesis, i.e. (3.1), and that $|\epsilon| = e(H_{r+1}) - e(H_r)$ and $r - |\epsilon| = e(\overline{H_{r+1}}) - e(\overline{H_r})$. This concludes the inductive step, hence the proof.

Concentration of f via bounding variance. We are left to prove (3.3). By Lemma 3.1.4 below, it suffices to bound the error term:

$$\sqrt{|\Omega^*| \cdot \sum_{z\in\Omega} (X(z) - \mathrm{E}[X])^2} = \sqrt{|\Omega^*| \cdot \sum_{z\in\Omega} (X(z)^2 - \mathrm{E}[X]^2)} = o(n^{r+1}).$$

As $|\Omega^*| = N_r = O(n^r)$, it suffices to show

$$\sum_{z\in\Omega} X(z)^2 = |\Omega| \cdot \mathrm{E}[X]^2 + o(n^{r+2}) = \rho^2 n^{r+2} + o(n^{r+2}). \tag{3.4}$$

To see this, For the second moment $\sum_{z\in\Omega} X(z)^2$, we can approximate it with

$$T := \sum_{z\in\Omega} X(z)(X(z)-1),$$

which can be computed by double counting, using the following combinatorial meaning. To do so, we need the following claim: for any $u \neq v \in V$ and any integer $k, k' \geq 1$,

$$\sum_{u\neq v} d_G(u,v)^k d_{\overline{G}}(u,v)^{k'} = (1+o(1))p^{2k}(1-p)^{2k'} n^{k+k'+2}. \tag{3.5}$$

We leave this claim as an exercise, see Exercise 3.1.3.

We can now compute T. Note that T counts the number of pairs $(z, \{u,v\})$, with $z \in V_{(r)}$ and $\{u,v\} \in V_{(2)}$, such that $u, v \notin z$ and both u and v have the same adjacency ϵ to z. On the other hand, counting such pairs from $\{u,v\}$'s perspective, for each given $\{u,v\}$, we can pick z according to ϵ by choosing $|\epsilon|$ terms from $N_G(u,v)$ and $r - |\epsilon|$ terms from $N_{\overline{G}}(u,v)$, then by (3.5), we have

$$\begin{aligned} T &= (1+o(1)) \sum_{u\neq v} d_G(u,v)^{|\epsilon|} d_{\overline{G}}(u,v)^{r-|\epsilon|} \\ &= (1+o(1))p^{2|\epsilon|}(1-p)^{2(r-|\epsilon|)} n^{r+2} \\ &= (1+o(1))\rho^2 n^{r+2}. \end{aligned}$$

Consequently, (3.4) follows:

$$\sum_{z\in\Omega} X(z)^2 = T + \sum_{z\in\Omega} X(z) = T + O(n^{r+1}) = \rho^2 n^{r+2} + o(n^{r+2}),$$

as desired. □

Exercise **3.1.3.** Prove (3.5) using (Codegree) property. It is worth noting that the right hand side in (3.5) is what we would expect from a genuine random graph $G(n,p)$.[3]

When proving (Codegree) implies (Induced Subgraph Count), we use the following lemma, which basically says that we can approximate a subset sum by average provided that the variance is small.

Lemma 3.1.4. *Let X be a random variable defined on space Ω with uniform measure. Let $\Omega^* \subseteq \Omega$, then*

$$\begin{aligned}\sum_{\omega \in \Omega^*} X(\omega) &= |\Omega^*| \cdot \mathrm{E}[X] \pm \sqrt{|\Omega^*| \cdot \sum_{\omega \in \Omega} (X(\omega) - \mu)^2} \\ &= |\Omega^*| \cdot \mathrm{E}[X] \pm \sqrt{|\Omega^*| \cdot |\Omega| \operatorname{Var}[X]}.\end{aligned}$$

Proof. Let $\mu = \mathrm{E}[X]$. By Cauchy-Schwarz,

$$\left|\sum_{\omega \in \Omega^*} X(\omega) - |\Omega^*| \cdot \mu\right|^2 = \left|\sum_{\omega \in \Omega^*} (X(\omega) - \mu)\right|^2 \le |\Omega^*| \cdot \sum_{\omega \in \Omega^*} (X(\omega) - \mu)^2 \le |\Omega^*| \cdot \sum_{\omega \in \Omega} (X(\omega) - \mu)^2,$$

implying that

$$\left|\sum_{\omega \in \Omega^*} X(\omega) - |\Omega^*| \cdot \mu\right| = \sqrt{|\Omega^*| \cdot \sum_{\omega \in \Omega} (X(\omega) - \mu)^2}.$$

Thus,

$$\sum_{\omega \in \Omega^*} X(\omega) = |\Omega^*| \cdot \mu \pm \sqrt{|\Omega^*| \cdot \sum_{\omega \in \Omega} (X(\omega) - \mu)^2},$$

as desired. □

3.1.3 (4-cycle Count) ⇒ strongly regular via Cauchy-Schwarz

We can also prove a stronger version of (Codegree) property directly from (4-cycle Count).

Proposition 3.1.5. *Let G be an n-vertex graph with $e(G) = p\binom{n}{2}$. If $t(C_4, G) \le p^4 + o(1)$, then for any $u \neq v \in G$,*

$$d(u) = pn + o(n) \quad \text{and} \quad d(u,v) = p^2 n + o(n).$$

Proof. Note first that, by double counting the cherry $K_{1,2}$ and Cauchy-Schwarz,

$$\begin{aligned}\sum_{u \neq v} d(u,v) &= \sum_{w} d(w)(d(w)-1) = \sum_{w} d(w)^2 - 2e(G) \\ &\ge \frac{1}{n}\left(\sum_{w} d(w)\right)^2 - 2e(G) = (1 + o(1))p^2 n^3.\end{aligned}$$

[3]Hint: Let $\delta_{uv} = d_G(u,v) - p^2 n$. Show first that for any integer $k \ge 1$, $\sum_{u \neq v} |\delta_{uv}|^k = o(n^{k+2})$, using the fact that $|\delta_{uv}| \le n$ for any $u \neq v$.

Consequently, the number of labelled C_4, using Cauchy-Schwarz again, is

$$\begin{aligned} p^4n^4 + o(n^4) &\geq t(C_4, G)n^4 = \sum_{u \neq v} d(u,v)(d(u,v)-1) = \sum_{u \neq v} d(u,v)^2 - \sum_{u \neq v} d(u,v) \\ &\geq \frac{1}{n(n-1)} \left(\sum_{u \neq v} d(u,v) \right)^2 - \sum_{u \neq v} d(u,v) \\ &\geq p^4n^4 + o(n^4). \end{aligned}$$

We thus should have the equalities above throughout. Equalities hold in the two applications of Cauchy-Schwarz implies that both the degree vector $\{d(w)\}_{w \in V(G)}$ and the codegree vector $\{d(u,v)\}_{(u,v) \in V_{(2)}}$ are linearly dependent with all 1s vector. In other words, the graph is strongly regular, i.e. every vertex has degree $pn + o(n)$ and codegree of every pair is $p^2n + o(n)$ as desired. □

Remark 3.1.6. In retrospect, it is not that surprising now that the seemingly weaker property of (4-cycle Count) implies the (Induced Subgraph Count). As we have seen above, (Codegree) follows from (4-cycle Count) with two applications of Cauchy-Schwarz. And in the proof of (Codegree) ⇒ (Induced Subgraph Count), we count H-subgraphs by building it up one vertex at a time, $H_1, \ldots, H_r, \ldots, H_s = H$. Let H^*_{r+1} be the graph obtained from H_{r+1} by adding a new vertex v^*_{r+1}, which is a copy of v_{r+1}. To count H_{r+1}, we need to control its "variance": the number of copies of H^*_{r+1}, which, in point of views of the twins v_{r+1} and v^*_{r+1}, is governed by (Codegree) property.