The Regularity Lemma

Alex Xing

June 9, 2019

Abstract

This is the term paper for MATH336. In this paper, we want to show that every dense graph has a density template that samples constantly many nodes, and we could use that template to show that we could distinguish whether that graph is triangle-free or ϵ-far from a triangle-free graph. This is a strong argument, since the only access we have about a graph G is that we can sample a constant-size subset U and study the correspond subgraph $G[U]$, and the graph G might only contain one triangle, which might not be contained in $G[U]$. [1]

1 Introduction

Consider an undirected graph $G = (V, E)$, testing the property of whether there is a triangle in G is hard, since:

1. the only access we have about a graph G is that we can sample a constant-size subset U and study the correspond subgraph $G[U]$.

2. the graph G might only contain one triangle, which might not be contained in $G[U]$.

But we could first find if there exists a tester, sampling only constantly many nodes, that could distinguish:

1. G is a triangle-free graph.

2. G is ϵ-far from any triangle-free graph.

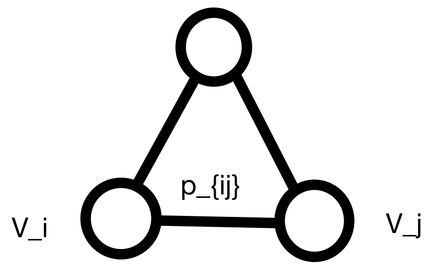
And by ϵ-far, we mean that G is within ϵn^2 many edges different with another graph.

We are still not sure whether we could make this distinction within $O_\epsilon(1)$-many samples. But as an introduction, we could consider an algorithn for a particular type of random graphs and make some observations on it:

Suppose we have a distribution D, from which we generate $G = (V, E)$ such that we have the partition

$$V = V_1 \biguplus \dots \biguplus V_k \qquad \text{for } \frac{1}{\epsilon} \ll k \leq f(\epsilon) \text{ and } |V_1| = \dots = |V_\epsilon|.$$

Suppose for each pair of $i, j \in [k]$, we have probability p_{ij}. Then every edge (u, v) such that $u \in V_i$ and $v \in V_j$ materializes independently with p_{ij}.



For this graph, we can make the observation that if there is a triangle in the graph

$$H = ([k], E(H)) \text{ such that } (i, j) \in E(H) :\Leftrightarrow p_{ij} \geq \frac{\epsilon}{2},$$

then, with high probability, G will contain $\Omega_\epsilon(n^3)$ many triangles. If H does not contain any triangle, by deleting all edges coming from the low density pairs $V_i - V_j$ with $p_{ij} < \frac{\epsilon}{2}$, all triangles in G will vanish. And the number of such edges will likely not be greater than ϵn^2.

Moreover, if there are $\Omega_\epsilon(n^3)$ many triangles, we only need a positive test to sample $O_\epsilon(1)$ many nodes.

The aim for this paper is prove that for any dense graph, we could find such a "density template".

2 The Szemerédi Regularity Lemma

In this section, we introduce and prove the Szemerédi Regularity Lemma.
(probability some background here)

Firstly, we want to give a precise definition of the "density template" that we discussed above. Fix an undirected graph $G = (V, E)$ for $A, B \subseteq V$, and $A \cup B$. Let $e(A, B)$ be the number of edges between A and B sich that

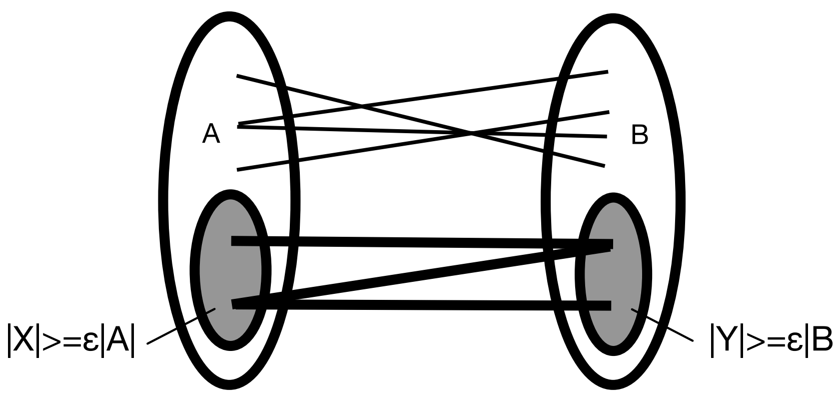
$$e(A,B) := \left|\{e \in E : |e \cap A| = |e \cap B| = 1\}\right|.$$

Then the density of the pair A, B is defined by the quantity:

$$d(A,B) := \frac{e(A,B)}{|A| \cdot |B|}, \quad 0 \leq d(A,B) \leq 1 \text{ for all possible edges between } A \text{ and } B.$$

The pair (A, B) is said to be ϵ-regular if for any $\epsilon > 0$ we have

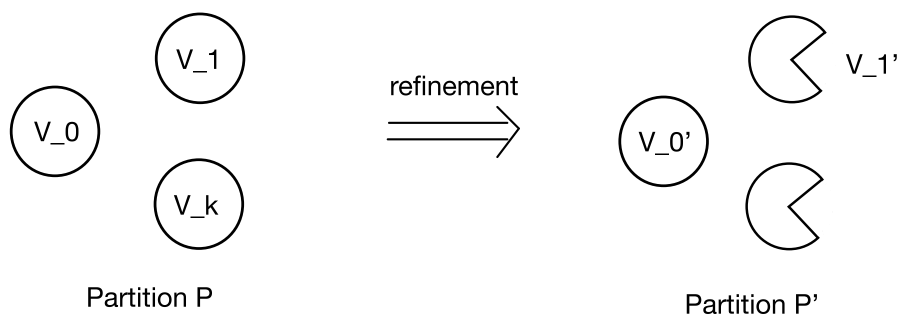
$$\left|d(A,B) - d(X,Y)\right| \leq \epsilon, \text{ for all } X \subseteq A,\ Y \subseteq B,\ |X| \geq \epsilon|A|, \text{ and } |Y| \geq \epsilon|B|.$$



Definition 1 *A partition $V = V_0 \cup V_1 \cup \ldots \cup V_k$ is an equipartition if $|V_1| = \ldots|V_k|$, and we define V_0 to be the exceptional set. If all pairs (V_i, V_j) of an equipartition, except at most ϵk^2's, are regular, and the size of V_0 is bounded by $|V_0| \leq \epsilon|V|$, then that equipartition is ϵ-regular.*

Theorem 2 *(The Szemerédi Regularity Lemma)*
For every $\epsilon > 0$, there is a constant $T = T(\epsilon)$ so that every graph with $|V| \geq T$ vertices has an ϵ-regular partition $P = (V_0, \ldots, V_k)$ with $\frac{1}{\epsilon} \leq k \leq T$.

Proof. Suppose P is a partition with $k := \frac{1}{\epsilon}$ blocks. We can find a refinement of P that is more regular, let it be P'. By refinement of P, we mean that P' is a partition such that every block of P is the disjoint union of some blocks in P'. Consider V_0 as having $|V_0|$ amount of separate singleton nodes, which means, we can always move a small number of nodes into the exceptional set V_0' to obtain such refinement P'.



The question now is to find a measure of the regularity so that it could be improved. In an n-th node graph, consider disjoint sets $U, W \subseteq V$ and M, N to be the partition of u, w, and define the quantity and the weighted average squared density, correspondly, to be

$$q(U,W) := \frac{|U| \cdot |W|}{n^2} \cdot d(U,W)^2 \qquad q(M,N) := \sum_{M \in U, N \in W} q(U', W').$$

Also, define a random variable to represent the average density between blocks in M, N such that those blocks are chosen proportional to their corresponding number of nodes.
Let that variable be $Z \backsim D(M,N)$, formed by picking unirform random elements $u \in U$ and $w \in W$, then set $Z := d(M,N)$. Thus, we could write the following expression

$$q(M,N) = \frac{|U| \cdot |W|}{n^2} \cdot \mathbb{E}[Z^2]. \tag{1}$$

Then, for the partition $P = (V_0, V_1, ..., V_k)$, we can define

$$q(P) := \sum_{\text{blocks } U,W \text{ of } P} q(U,W), \text{ with the sum over } \binom{k + |V_0|}{2} \text{ unordered pairs of blocks.}$$

(We are counting each singleton in V_0 as one single block.)
We call the quantity $q(P)$ to be the index of partition P, which represents the weighted average of squared densities of its partitions.
As the densities are in the range of $[0,1]$, and the sum of the weights is at most $\frac{1}{2}$, we have $0 \leq q(P) \leq \frac{1}{2}$. And as long as the partition is not regular, we could always find such refinements

that increase $q(P)$.

We need to consider a lemma that demonstrates the following:

1. the refinement could only increase the calue of $q(P)$.

2. an irregular pair can be used to get a refinement that strictly increases $q(P)$.

■

Lemma 3 *The following are ture:*

1. *Let $U, W \supset V$ be disjoint. Let M, N be partitions of U, W. Then $q(M,N) \geq q(U,W)$.*

2. *If P' is a refinement of P, then $q(P') \geq q(P)$.*

3. *Suppose a disjoint pair (U,W) is not ϵ-regular, due (U_1, W_1) such that $U_1 \supset U$ and $W_1 \supset W$, then*

$$M := \{U_1, \frac{U}{U_1}\} \quad N := \{W_1, \frac{W}{W_1}\} \text{ satisfy } q(M,N) > Q(U,W) + \epsilon^4 \cdot \frac{|U| \cdot |W|}{n^2}.$$

Proof.

For 1:

First, note that the overall edge density is $\mathbb{E}[z] = d(U,W)$. Consider the random variable $Z \backsim D(M,N)$, which gives the density of the random pair (U', W') of the partitions. By applying the regularity lemma, the Jensen inequality and the definition of $q(U,W)$, we have

$$\frac{n^2}{|U||W|} q(M,N) = \mathbb{E}[Z^2] \geq \mathbb{E}[Z^2] = d(U,W)^2 = \frac{n^2}{|U||W|} q(U,W).$$

For 2:

The conclusion follows directly from 1.

For 3:

To prove what we want, the only thing we need to do is to lower bound the variance of Z:

$$Var[Z] = \frac{n^2}{|U||W|} \left[q(M,N) - q(U,W) \right].$$

By choosing an irregular pair (U_1, W_1), we could rewrite the variance of Z as:

$$Var[Z] = \mathbb{E}[(Z - \mathbb{E})^2] = Pr_{u \backsim U, w \backsim W}[u \in U_1, w \in W_1] \cdot [d(U_1, W_1) - d(U,W)]^2 \geq \epsilon^4,$$

following from the fact that $Pr_{u \sim U, w \sim W}[u \in U_1, w \in W_1] \geq \epsilon^2$, and $[d(U_1, W_1) - d(U, W)]^2 \geq \epsilon^2$. And the conclusion follows easily from here. ■

In the central part of the proof for the Szemerédi Regularity Lemma, we want to show that if the partition is not ϵ-regular, then there is a refinement P' such that

$$q(P') \geq q(P') + \frac{\epsilon^5}{2},$$

and that the size of the exceptional set only increases marginally.

Lemma 4 *Suppose $0 < \epsilon < \frac{1}{4}$, let $P = \{V_0, \cdots, V_k\}$ be an equipartition of V, where V_0 is the exceptional set, $|V_0| \leq \epsilon n$, and $|V_i| = c$ for all $1 \leq i \leq k$.*
If P is not ϵ-regular, then there exists a refinement $P' = \{V_0', \cdots, V_l'\}$ of P, where $k \leq l \leq k4^k$, $|V_0'| \leq |V_0| + \frac{n}{2^k}$, and all other sets V_i are of the same size and satisfy

$$q(P') \geq q(P) + \frac{\epsilon^5}{2}.$$

Proof.

Consider a pair (V_i, V_j) with $1 \leq i < j \leq k$. If the pair is ϵ-regular, then $V_{ij} := \{V_i\}$, and $V_{ji} := \{V_j\}$. Otherwise, if the pair is not regular, $U \subseteq V_i$ and $W \subseteq V_j$ are the two parts contained in each partition, chosen according to Lemma 3, 3. For each $1 \leq i \leq k$, let $\mathcal{V}_i$ be the partition of V_i obtained by the Venn diagram of all $(k-1)$-partitions $\mathcal{V}_{ij}$. Therefore, we know that each pair has at most 2^{k-1} parts. Let $\tilde{P}$ be the partition containing $\mathcal{V}, \cdots, \mathcal{V}$ together with the exceptional set V_0. Since P is not ϵ-regular, there will be ϵk^2 many pairs that are irregular and from Lemma 3, 3, we know that every single pair will increase the function q. The estimate follows:

$$q(\tilde{P}) \geq q(P) + \sum_{\text{irregular pairs } (V_i, V_j)} \epsilon^4 \frac{|V_i||V_j|}{n^2} \geq q(P) + \epsilon^4 \cdot \epsilon k^2 \cdot \frac{1}{k^2} \cdot \frac{3}{4} \geq q(P) + \frac{\epsilon^5}{2},$$

where we use the fact that $kc \geq (1-\epsilon)n \geq \frac{3n}{4}$.
Note that P has at most $k \cdot 2^{k-1}$ parts except the exceptional set, but they are not necessarily all equal sizes. If we define $b = \frac{c}{4^k}$ and split every part of P argitrarily into disjoint sets of size b and throw the remaining vertices in each part, we will get a partition P' with

1. at most $k \cdot 4^k$ non-exceptional parts of equal size,

2. a new exceptional set of size smaller that

$$|V_0| + k \cdot 2^{k-1} \cdot b < |V_0| + \frac{kc}{2^k} \leq |V_0| + \frac{n}{2^k}.$$

Moreover, from Lemma 3,2, we know that the index $q(P')$ of P' is at least

$$q(\tilde{P}) > q(P) + \frac{\epsilon^5}{2}.$$

Therefore, we've completed the proof. ■

Proof for Regularity Lemma.

We begin with an argitrary partition of the n vertices into $k_0 = \frac{1}{\epsilon}$ many equal size blocks, and this requires to move at most $\frac{1}{\epsilon} \ll \frac{\epsilon n}{2}$ many nodes into the exceptional set.

In the ith iteration, as long as the current partition is not ϵ-regular, we can use Lemma 4 to show that the number of partitions increases from k_i to $k_{i+1} \leq k_i 4^{k_i}$.

As $q(P)$ increases by at least $\frac{\epsilon^5}{2}$, we terminate after getting at most $\frac{2}{\epsilon^5}$ calls. And in each of these calls the size of the exceptional set increases by the amount of $\frac{n}{2^{k_i}}$, but the total increase in size is bounded by $\frac{\epsilon n}{2}$ as $k_i \geq \frac{1}{\epsilon}$.

The argument works as long as n stays bigger than the bounde on k_i. And we've finished the proof.

Pessimistically, it could happen that for $\Theta(\frac{1}{\epsilon^5})$ times the number of partitions increases exponentially. In particular, the bound on $T(\epsilon)$ is a *tower of exponents* with height of $\Theta(\frac{1}{\epsilon^5})$. From the result of Gowers, every ϵ-regular partition in come graphs requires a number of partitions that is a tower of height polynomial in $\frac{1}{\epsilon}$.

3 Testing Triangle-Freeness

In the original application, we want to distinguish the triangle-free graph from a graph that is ϵ-far from being triangle-free. Using the proof of the Regularity Lemma, the property of containing no triangle is testable with one-sided error. The required combinatorial lemma here is the fact that if we want to delete ϵn^2 edges of an n-vertex graph, in order to destroy all the triangle in it, there must be at least δn^3 triangles in the graph, where $\delta = \delta(\epsilon) > 0$.

From Ruzsa and Szemerédi, the fact mentioned above implies that any set of integers with positive upper density contains a three-term arithmetic progression.

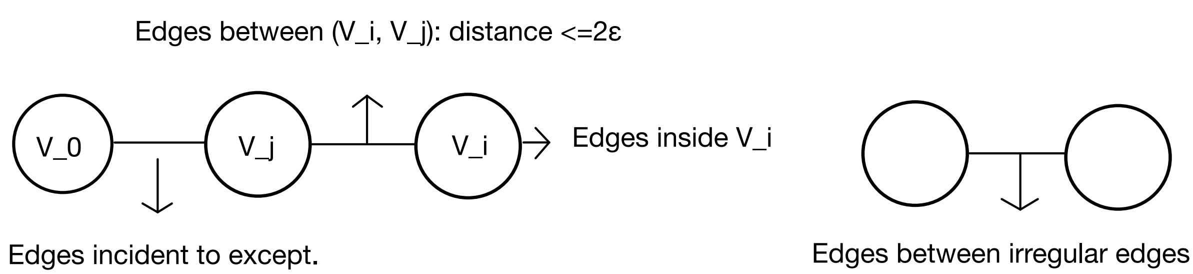
Lemma 5 *Let $G = (V, E)$ be a graph so that for every $H \subseteq E$ with $|H| \le \epsilon n^2$, $(V, E \backslash H)$ still contains at least one triangle. Then G itself contains δn^3 many triangle with $\delta := \delta(\epsilon) > 0$.*

Proof.

Suppose that at least $C\epsilon n^2$ edges can be delated without destroying all the triangles, with $C > 0$, a constant that's large enough. Consider the Regularity Lemma and the partition $P = (V_0, \cdots, V_k)$ that is ϵ-regular. Consider the new graph by deleting the following edges:

1. Edges that are incident to the exceptional set V_0
2. Edges between irregular pairs
3. Edges between regular pairs where the density is less that 2ϵ
4. Edges inside some block

The visualization looks like the following:



The common thing shares between those four types is that we delete at most $O(\epsilon n^2)$ many edges in each case. Assume that the remaining graph still has at least one single triangle, and we could construct this triangle to be running between partitions where all pairs are regular, and the densities are bigger than or equal to 2ϵ:

For V_1, V_2, V_3, all pairs (V_i, V_j) are regular, and the densities $d(V_i, V_j) \ge 2\epsilon$ for $1 \le i < j \le 3$.

Recall that $s := |V_1| = |V_2| = |V_3|$ and $s \ge (3/4) \cdot (n/k)$. Define $X_i := \{u \in V_1 : |N(u) \cap V_i| \le \epsilon |V_i|\}$, $i = 2$ or $i = 3$, to be the nodes with rather few neighborhoods.

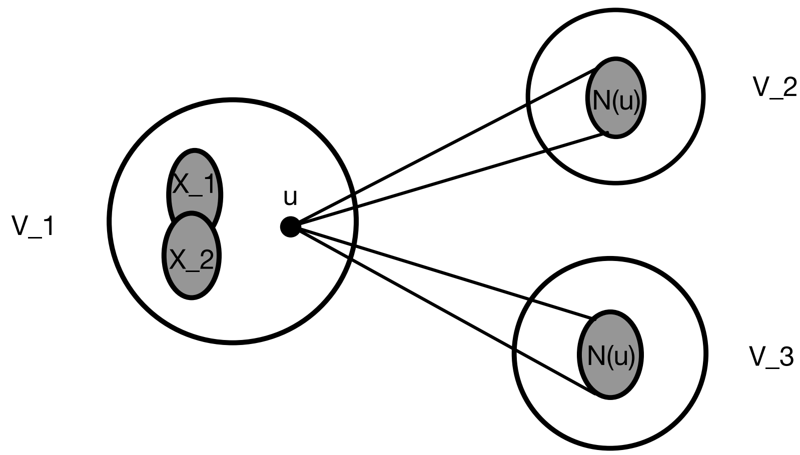
If $|X_i| \geq \epsilon|V_1|$, then we say that (X_i, V_i) is an ϵ-irregular part of (V_1, V_i), and we call the nodes $u \in V_1 \backslash (X_2 \cup X_3)$ *typical.*

By regularity we know that the densities between neighborhoods of a typical node $u \in V_1$ is

$$d(N(u) \cap V_2), N(u) \cap V_3) \geq \epsilon.$$

And every edge between $N(u) \cap V_2$ and $N(u) \cap V_3$ forms a triangle together with u. And the number of triangles between V_1, V_2, V_3 is at least

$$(1 - 2\epsilon)|V_1| \cdot \epsilon \cdot |V_2| \cdot \epsilon |V_3| = \delta(\epsilon) \cdot n^3.$$



■

4 Characterizing the Testable Graph Properties

[2] In this section, we will describe and prove several results using the variant of the Regularity Lemma.

A *monotone* graph is the type of graphs that is closed under removing vertives and edges. From the definition, we know that being triangle free is a monotone property.

A *hereditary* graph is the type of graphs that is closed under removal of vertices (not necessarily under removal of edges). From the definition, we know that every monotone graph is also hereditary, but not in the opposite. Esamples are perfect graphs, chordal graphs, adn interval graphs, etc.

In the previous sections, we discussed the result of the Regularity Lemma related to cases where the graphs are both hereditary and monotone, that is, being triangle-free (and also k-colorable).

Consider a family of graph $mathcalF$, a graph is said to be an *induced $\mathcal{F}$-free* graph if it contains no $F \in \mathcal{F}$ as an induced subgraph. Consider the following lemma:

Lemma 6 *Let $\mathcal{F}$ be a family of graphs (possibly infinite), and suppose there are functions $f_{\mathcal{F}}(\epsilon)$ and $\delta_{\mathcal{F}}(\epsilon)$ such that the following holds for every $\epsilon > 0$:*
Every graph G on n vertices that is ϵ-far from being induced $\mathcal{F} - free$ contains at least $\delta_{\mathcal{F}}(\epsilon)n^f$ induced copies of a graph $F \in \mathcal{F}$ of size $f \leq f_{\mathcal{F}}(\epsilon)$. Then, being induced $\mathcal{F}$-free is testable with one-sided error.

And the result can be generalized as follows:

Theorem 7 *For any family of graphs $\mathcal{F}$ there are functions $f_{\mathcal{F}}(\epsilon)$ and $\delta_{\mathcal{F}}(\epsilon)$ satisfying the conditions of Lemma 6.*

Combine Lemma 6 with Theorem 7, and define for any hereditary property $\mathcal{P}$, a family of graphs $\mathcal{F}_{\mathcal{P}}$ is equivalent to being induced $\mathcal{F}_{\mathcal{P}}$-free, since we can put a graph F in $\mathcal{F}_{\mathcal{P}}$ if and only if F does not satisfy the hereditary property $\mathcal{P}$. The theorem we obtain by combining the previous two results are:

Theorem 8 *Every hereditary graph property is testable with one-sided error.*

From the above theorem, we can obtain a characterization of the *natural* graph properties, which we will discuss later. And we could also obtain the following result from this theorem:

Corollary 9 *For every hereditary graph property $\mathcal{P}$, there is a function $W_{\mathcal{P}}(\epsilon)$ with the following property:*
If G is ϵ-far from satisfying $\mathcal{P}$, then G contains an induced subgraph of size at most $W_{\mathcal{P}}(\epsilon)$ that does not satisfy $\mathcal{P}$.

From Theorem 8, we can obatin a characterization of the natural graph, which are testable with one-sided error.

Consider a tester, one-sided or two-sided, is defined to be *oblivious* if:
Given ϵ, the tester computes $Q = Q(\epsilon)$, and requres an oracle of a subgraph to be induced by

a set of vertices of size Q. It is important to note that the oracle choose the vertices randomly and uniformly from the vertices of the graphs. If the size is larger than that of the inpur graph, then the oracle returns the entire graph. According to the graph induced by S, the tester accepts it.

Some properties of graphs cannot have oblivious tester, however, these properties cannot be natural. An example is the property of not containing an induced cycle of length 4, givern the number of vertices to be even; or, to say, the property of not containing an induced cycle of length 5, given the number of vertices to be odd.

Using Theorem 8, we can show that if we consider only oblivious tester, then it is possible to precisely cnaracterize the graph properties, with the following definition:
A *semi-hereditary* graph $\mathcal{P}$ is the kind of graphs if there exists a hereditary graph property $\mathcal{H}$ such that the following holds:

1. Any graph satisfying $\mathcal{P}$ also satisfies $\mathcal{H}$
2. For any $\epsilon > 0$, there is an $M(\epsilon)$ such that any graph of size at least $M(\epsilon)$ that is ϵ-far from satisfying $\mathcal{P}$ does not satisfy $\mathcal{H}$.

From the definition, we could state the following characterization:

Theorem 10 *A graph property $\mathcal{P}$ has an oblivious one-sided tester if and only if $\mathcal{P}$ is semi-hereditary.*

References

[1] Alon N and Spencer H. *The probabilistic method, Chapter 17, p294-p301.* John Wiley & Sons, 2004.

[2] Alon N and Shapira A. *A characterization of the (natural) graph properties testable with one-sided error, p429-p438.* 46th Annual Symposium on Foundations of Computer Science(FOCS), IEEE, 2005.