APPLICATIONS OF PROBABILITY THEORY TO GRAPHS

JORDAN LAUNE

Abstract. In this paper, we summarize the probabilistic method and survey some of its iconic proofs. We then go on to draw the connections between ϵ-regular pairs and random graphs, emphasizing their relationship and parallel constructions. We finish with a review of the proof of Szemerédi's Regularity Lemma and an example of its application.

Contents

1. Introduction 1
2. Preliminaries 2
3. The Probabilistic Method 4
4. Random Graphs and ϵ-Regularity 7
5. Szemerédi Regularity Lemma 10
Acknowledgments 13
References 13

1. Introduction

Some results in graph theory could be proven by having a computer check all graphs possible on a specific vertex set. However, this method of brute force is usually completely out of the question: some problems would require a computer to run for longer than the current age of the universe. Other constructive solutions that include algorithms to construct a graph are usually too complicated to find. Thus, probability theory, which gives us an indirect way to prove properties of graphs, is a very powerful tool in graph theory.

Some graph theorists use probability theory as a separate tool to apply to graphs, as in the *probabilistic method*. On the other hand, graph theorists are sometimes interested in the interesection of graph theory and probabilistic models themselves, such as random graphs. While the probabilistic method was interested in using probability merely as a tool for graphs, those studying the random graph are instead interested in its properties for its own sake. As it turns out, the random graph gives us an idea of how a "general" graph behaves.

Roughly speaking, by using Szmederédi's Regularity Lemma, we can treat very large arbitrary graphs with a sense of "regularity." Since random graphs also behave regularly, we are able to think of any large graph as, in some sense, a random graph! This fact has spurred many developments in graph theory, and, as such, it is one of the biggest developments in graph theory to date.

Date: August 29, 2016.

2. Preliminaries

The familiar conception of probability theory includes that probabilities are between zero and one such that, roughly speaking, events with a probability of zero never occur, while those with a probability of one are certain to occur. However, if we treat the colloquial ideas of probability with mathematical rigor, we can prove results in a wide range of unexpected fields. In particular, several problems in graph theory can be proved using probability with more conceptual clarity than other, more direct methods. Thus, we will begin with the axioms from probability theory:

Definition 2.1. Let E be a collection of elements $A, B, C, ...$, called *elementary events*, and let $\mathscr{F}$ be the set of subsets of E. The elements of $\mathscr{F}$ are called *random events.*

I $\mathscr{F}$ is a field of sets.
II $\mathscr{F}$ contains E.
III Every set A in $\mathscr{F}$ is assigned a non-negative real number $\mathbb{P}(A)$ called the *probability* of A.
IV $\mathbb{P}(E) = 1$.
V If $A \cap B = \emptyset$, then $\mathbb{P}(A + B) = \mathbb{P}(A) + \mathbb{P}(B)$.

A system $\mathscr{F}$ along with the assignment of numbers $\mathbb{P}(A)$ that satisfies Axioms I-V is a *field of probability.*

From here, we will specify that we are concerned with discrete probability. For a discrete probability space, we can condense these axioms to a simpler, yet more informal, definition of probability consisting of two components: a countable set Ω and a function $\mathbb{P} : \Omega \to [0, 1]$. We call Ω the *state space* and say that $\mathbb{P}$ assigns each state $\omega \in \Omega$ a *probability* $\mathbb{P}(\omega)$. These probabilities satisfy:

$$\sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1.$$

We speak of an *event* as a subset $E \subset \Omega$. We write $\mathbb{P}(E) = \sum_{\omega \in E} \mathbb{P}(\omega)$. A *random variable* is a mapping $X : \Omega \to \mathbb{R}$. Mathematical events and random variables can be used to represent happenings in reality. For example, each number one through six on a die constitutes a state space Ω. An event could be $E = \{1, 2\}$, which is the case where the die lands on either one or two. A natural random variable would assign each state with its numerical value, i.e. $X : 1 \mapsto 1$, $X : 2 \mapsto 2$ and so on. Here, a natural question one may ask is what we may "expect" from such a random variable. For the case of a single die, we would ask what value we expect to turn up after a roll. Thus, we come to the following formal definition of expectation:

Definition 2.2. The *expectation* of a random variable X is

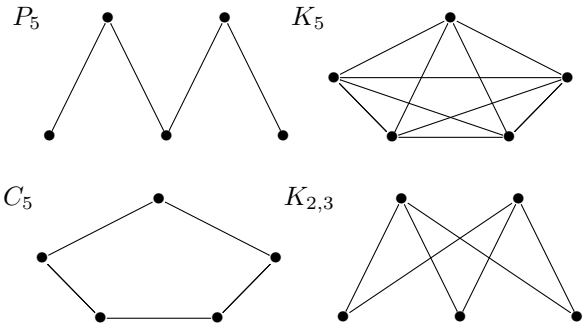
$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega).$$

For the purposes of this paper, the most important characteristic of expectation is its linearity. This allows us to work very easily within probabilistic methods, as the expectation of the sum of random variables is merely the sum of their expectations. This property will become relevant in the following section.

Now, we will begin our discussion of graph theory with the definition of a graph:

Definition 2.3. A *graph* $G = (V, E)$ consists of a vertex set V and an edge set E. Each edge connects two vertices. We denote the number of vertices $n := |V|$ and the number of edges $m := |E|$.

To list some common graphs, we have a *path* of length n (P_n), which consists of n vertices connected by $n-1$ edges. A *cycle* of length n (C_n) consists of n vertices connected by n edges. The complete graph on n vertices (K_n) is a graph on n vertices with every edge possible. A *bipartite graph* is a graph whose vertex set V can be partitioned into two sets S, T such that $S \cap T = \emptyset$ and for every edge $e \in E$, $e = \{s, t\}$ for some $s \in S$ and $t \in T$. A complete bipartite graph is a bipartite graph with every edge possible between sets S and T. We denote the complete bipartite graph as $K_{s,t}$ such that $s = |S|$ and $t = |T|$. We have drawn these common graphs on five vertices below.



For the rest of this paper, it will be useful to develop a vocabulary for describing graphs and their components. The first definition we have is the *degree* of a vertex, $deg(v)$, which is the number of edges incident with v. Another useful concept is an *independent* set, $S \subset V$, in which there does not exist an edge $\{u, v\} \in E$ for any two vertices $u, v \in S$. Intuitively, an independent set of vertices is one in which none of the vertices in S "touch" each other with an edge. Next, we have k-colorings, a concept closely related to independence:

Definition 2.4. A *k-coloring* of V for a graph G is a function $f : V \to \{1, ..., k\}$. A *valid k-coloring* occurs if for all $\{u, v\} \in E$, $f(u) \neq f(v)$.

Thus, we see that for a valid k-coloring, every subset $S \subset V$ that corresponds to one color must be independent. From here, we define the *chromatic number*, $\chi(G)$, which is the minimum number $k \in \mathbb{N}$ for which G has a valid $k-coloring$. A final property of a graph G that we are interested in is its *girth*, which is smallest cycle C_k for which G includes C_k. A graph with a girth of five cannot include any cycles of length four or less.

3. The Probabilistic Method

The basic outline of the probabilistic method comes in four steps. First, we must choose a random object. For graph theory, the random object is a graph. Next, we must define a useful random variable to analyze. For example, if we are concerned with the girth of the graph, the random variable X could be the smallest cycle included in the graph. The third step is calculating the expectation of the random variable. Then, we make statement of existence based on the calculated expectation. One useful statement worth mentioning is that if a random variable takes on values in $\mathbb{N} \cup \{0\}$, which is common when the random variable is concerned with graphs, if the expectation is lower than one, there *must* exist a graph such that the random variable is equal to zero.

In order to illustrate how the expectation of a random variable can actually enforce the existence of a certain property, let us prove the following short proposition:

Proposition 3.1. *If $X(\omega) \in \mathbb{N} \cup \{0\}$ for all $\omega \in \Omega$, and $\mathbb{E}(X) < 1$, then there must exist $\omega_0 \in \Omega$ such that $X(\omega_0) = 0$.*

Proof. Suppose that $\mathbb{E}(X) < 1$. Now, assume towards a contradiction that there does not exist $\omega_0 \in \Omega$ such that $X(\omega_0) = 0$. Then, for every $\omega \in \Omega$, we have $X(\omega) \geq 1$. So,

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\omega) \geq \sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1$$

from our definitions of probability. Thus, we have reached a contradiction, and so there must exist $\omega_0 \in \Omega$ such that $X(\omega_0) = 0$. □

This statement comes in handy when we are proving lower bounds for properties of graphs on n vertices, like the girth and chromatic number. Roughly speaking, if we can prove that the expectation of a graph containing a cycle of length four or less is less than one, then that means there exists a graph with girth greater than or equal to 5 on n vertices.

A classic example of the power of the probabilistic method comes to us when studying Ramsey's numbers. A $k - clique$ is a subset of a graph G such that it is itself a complete graph. The Ramsey number for natural numbers r and s, $R(r, s)$, is the minimum number of vertices for which a graph with an edge two-coloring must contain a monochromatic clique in one color of size r, or a monochromatic clique in the other color of size s. For this problem, we will label the colors white and blue, with white assigned to r and blue to s in $R(r, s)$. If we examine the case where $k := r = s$, denoted by $R(k)$, we are able to put a lower bound on the Ramsey number by using the probabilistic method:

Theorem 3.2 (Finite Ramsey's Theorem)**.** *For any $k \in \mathbb{N}$,*

$$(\sqrt{2})^k \leq R(k) \leq 4^k.$$

Proof. **Upper Bound.** The upper bound only relies on common combinatorial arguments without using probability, and so we shall skip it.

Lower Bound We would like to show that there exists a graph G with $\sqrt{2}^k$ vertices such that it has no monochromatic k-clique. So, let G have $n = \lfloor \sqrt{2}^k \rfloor$. Then, we will color each edge of the graph with a uniform distribution, where the probability of a white edge and the probability of a blue edge both equal $1/2$. Now,

let X be the random variable such that $X(G)$ equals the number of monochromatic k-cliques. If $\mathbb{1}$ is the indicator function, observe that

$$X = \sum_{\text{G' } k\text{-clique}} \mathbb{1}\{\text{G' is monochromatic}\}.$$

Thus,

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{\text{G' } k\text{-clique}} \mathbb{E}(\mathbb{1}\{\text{G' is monochromatic}\} \\
&= \sum_{\text{G' } k\text{-clique}} \mathbb{P}(\text{G' is monochromatic}) \\
&= \sum_{\text{G' } k\text{-clique}} 2\left(\frac{1}{2}\right)^k \\
&= \binom{n}{k}\frac{1}{2^{k-1}}.
\end{aligned}$$

Now, since factorials grow much faster than any exponential, we can see that for a graph on $n = \lfloor\sqrt{2}^k\rfloor$ vertices, for k big enough,

$$\mathbb{E}(X) = \binom{\lfloor\sqrt{2}^k\rfloor}{k}\frac{1}{2^{k-1}} < 1.$$

Now, due to the linearity of expectation, we know that there *must* exist a graph G on $\lfloor\sqrt{2}^k\rfloor$ such that G does not have a monochromatic k-clique. Thus, $R(k) \geq \sqrt{2}^k$. □

In order to put this result into perspective, we shall examine the known Ramsey's numbers. There are only four known Ramsey's numbers, consisting of $R(1) = 1$, $R(2) = 2$, $R(3) = 6$, $R(4) = 18$. $R(5)$ is not known, but we do know that it is somewhere between 43 and 49. Beyond that, the bounds quickly degrade: $R(6)$ is somewhere between 102 and 165; $205 \leq R(7) \leq 540$; and $282 \leq R(8) \leq 1870$. Thus, this is an exciting result. Even though we don't even know exactly what a graph for $k = 5$ looks like, we have still proven $R(k)$'s existence and put bounds on R(k) for all of the values of $k \in \mathbb{N}$ in this short and simple proof!

Now that we have demonstrated the power of the probabilistic method, we will pose a simple yet very difficult question:

Question 3.3. *Can you have a graph with no triangles with an arbitrarily large chromatic number?*

The answer is yes! However, this statement is tricky. In order to have a high chromatic number, we would usually expect a graph to have a more "dense" edge set. K_n has the highest chromatic number possible for a graph on n vertices since every single vertex is adjacent to every other vertex. On the contrary, the empty graph has a chromatic number of one. Thus, a high girth seems to counteract a high chromatic number, and vice versa. Surely, any constructive method that could be used to prove this result would most likely be terribly long and complicated. However, even if a surprisingly simple algorithm existed, it would still be hard to compete with the shockingly simple proof that Paul Erdös gave to prove an even stronger result:

Theorem 3.4. *Fix $k \in \mathbb{N}$. Then, there exists a graph G on n vertices such that $\chi(G), g(G) \geq k$, where $\chi(G)$ is the chromatic number and $g(G)$ is the graph's girth.*

Proof. First, observe that for the independence number, $\alpha(G)$,

$$\alpha(G) \cdot \chi(G) \geq n.$$

In particular, we have that if $\alpha(G) \leq n/k$, then $\chi(G) \geq k$. Now, fix $\epsilon > 0$ so that it is sufficiently small and let $p = 1/n^{1-\epsilon}$. Then, let G be a graph on n vertices such that each edge is chosen independently with probability p. Let X be the number of cycles with size less than or equal to k. Let Y be the number of independent sets of size bigger than or equal to $n/2k$. Then, using methods similar to the previous result to calculate the expectation of X, we have

$$\mathbb{E}(X) = \sum_{\substack{C_i \\ 3 \leq i \leq k}} \mathbb{P}(C_i \subset G) \leq \sum_{3}^{k} n^i p^i = O(n^{\epsilon(k+1)}) << \frac{n}{2}$$

since we chose ϵ to be sufficiently small. Now, we calculate the expectation of Y:

$$\begin{aligned}
\mathbb{E}(Y) &= \sum_{\substack{A \\ \text{size} \\ n/k}} \mathbb{P}(\text{A is independent}) \\
&= \binom{n}{\frac{n}{2k}} \cdot (1-p)^{\binom{\frac{n}{2k}}{2}} \\
&\leq n^{\frac{2}{k}} \cdot (1 - n^{-(1-\epsilon)})^{\frac{n^2}{8k^2}} \\
&\overset{n \to \infty}{=} 0.
\end{aligned}$$

Now, from Markov's inequality, we know that

$$\mathbb{P}\left(X > \frac{n}{2}\right) \leq \frac{\mathbb{E}(X)}{\frac{n}{2}} = \frac{\mathbb{E}(X) \cdot 2}{n} << 1.$$

Then,

$$\frac{\mathbb{E}(X)}{\frac{n}{2}} + \mathbb{E}(Y) < 1.$$

Now, per the probabilistic method, we know that there must exist a graph on n vertices such that there are at most $\frac{n}{2}$ cycles with length less than or equal to k and there is no independent set larger than $\frac{n}{2k}$. Thus, removing a vertex from each cycle of length less than or equal to k to form G', we still have a graph on $\frac{n}{2}$ vertices such that there is no independent set of size greater than $\frac{n}{2k}$. Furthermore, G' will have no cycle of length less than or equal to k. Due to our previous observation about the relationship between $\chi(G')$ and $\alpha(G')$, we have shown that there exists a graph G' such that $\chi(G'), g(G') \geq k$. □

This concludes our discussion of the probabilistic method. Using only basic probability, we have proven an incredible result in graph theory. However, this entire time we have been looking at probability as a *tool* to apply to a graph. In this next section, we will move on to viewing probabilistic *objects* in graph theory. We got a little bit of a taste of what these objects are in the previous theorem, when we chose each edge to the graph G independently with probability p. These kinds of objects are called random graphs and will be an invaluable concept to intuit as we move forward.

4. Random Graphs and ϵ-Regularity

We are interested in a random graph with a specified number of vertices, n, and a fixed, independent probability of there being an edge between any two vertices, p. Let us denote this model of the random graph as G_p on the vertex set $V = \{1, 2, ..., n\}$. Note that G_p is actually a *probability distribution* of graphs on n vertices. To be more precise, we will use the Bollobás and Erdös random graph:

Definition 4.1. A random graph $G \in \mathscr{G}(\mathbb{N}, p)$ is a collection $(X_{ij}) = X_{ij} : 1 \leq i < j$ of independent random variables with $P(X_{ij} = 1) = p$ and $P(X_{ij} = 0) = q$ such that a pair ij is an edge of G if and only if $X_{ij} = 1$. $G_n = G[1, 2, ..., n]$, the subgraph of G that is spanned by $[n]$, is exactly G_p on $V = \{1, 2, ..., n\}$.

Even though this precise definition is quite a bit more complicated than the first one given, for the purposes of this paper, it will suffice to picture a random graph in the simpler way. Furthermore, even though G_p is a probability distribution, it will be convenient to picture it as a single graph. When we talk of G_p, we will implicitly mean a graph randomly selected from the distribution G_p.

Since the random graph is somewhat of a mixture between both probability theory and graph theory, we can expect it to have properties from both fields. Morally, since we are dealing with randomness, we would expect a smaller sample out of the entire vertex set to behave like the entire vertex set itself. For graphs, the characteristic we are concerned with "sampling" is the density:

Definition 4.2. For disjoint vertex sets A and B, we define the *edge density* to be

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

where $e(A, B)$ is the number of edges $\{a, b\}$ for some $a \in A$ and $b \in B$.

For a random graph G_p, we would expect that for any two (reasonably large) vertex subsets $A, B \subset V$, subsets $X \subset A$ and $Y \subset B$ would have $d(X, Y)$ that does not deviate very far from $d(A, B)$. This expectation is based on the same intuition that a sample from a larger, well-behaved population will have a mean that does not vary far from the population's mean. To put this statement into precise terms, we will give the useful definition of ϵ-regularity based on this expectation and then prove that subsets of a random graph G_p will satisfy ϵ-regularity with high probability.

Definition 4.3. Let $\epsilon > 0$. Given a graph G and two disjoint vertex subsets $A, B \subset V$, the pair (A, B) is *ϵ-regular* if for any two sets $X \subset A$ and $Y \subset B$ such that

$$|X| > \epsilon|A| \text{ and } |Y| > \epsilon|B|,$$

we have

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

For the following, we will use both Boole's inequality and Chernoff's bound. However, their proofs are beyond the scope of this paper and so we will give their statements while skipping their proofs:

Theorem 4.4 (Boole's Inequality)**.** *For a countable number of events $A_1, A_2, A_3, ...$, we have*

$$\mathbb{P}\left(\bigcup_i A_i\right) \leq \sum_i \mathbb{P}(A_i).$$

Theorem 4.5 (Multiplicative Chernoff's Bound)**.** *Let $X_1, ..., X_n$ be random variables with $X_i \in \{0, 1\}$. Let X denote their sum and define $\mu = \mathbb{E}(X)$. Then we have*

$$\mathbb{P}(X > (1+\delta)\mu) < \left(\frac{e^\delta}{(1-\delta)^{(1-\delta)}}\right)^\mu.$$

Now, we will move on to prove that any two subsets of the random graph G_p are *ϵ-regular* with high probability.

Theorem 4.6. *Let G_p be a random graph on n vertices, and let $A, B \subset V$. As $n \to \infty$,*

$$\mathbb{P}((A, B) \text{ is } \epsilon\text{-regular}) \to 1.$$

Proof. Let G_p be a random graph on n vertices and $A, B \subset V$ and ϵ be sufficiently small. For simplicity's sake, assume that $|A| = |B| = k \leq \frac{n}{2}$. Then, the probability of an edge existing between $a \in A$ and $b \in B$ is p. We would like to find the probability that, for any two subsets $X \subset A$ and $Y \subset B$ such that $|X| > \epsilon|A|$ and $|Y| > \epsilon|B|$, we have $|d(X, Y) - d(A, B)| < \epsilon$. To find this, we shall prove that the probability of the existence of a pair (X, Y) that has $d(X, Y)$ such that $|d(X, Y) - d(A, B)| > \epsilon$ goes to zero as $k \to \infty$. We will denote the probability of the existence of such a pair as q. Then, $1 - q$ is the probability of (A, B) being ϵ-regular.

We begin by using Boole's inequality:

$$\begin{aligned} q &\leq \sum_{\substack{l \geq \epsilon|A| \\ m \geq \epsilon|B|}} \binom{|A|}{l}\binom{|B|}{m}\mathbb{P}(|d(X, Y) - d(A, B)| > \epsilon) \\ &\leq 4^k[\mathbb{P}(|d(X, Y) - p| > \epsilon/2) + \mathbb{P}(|d(A, B) - p| > \epsilon/2)]. \end{aligned}$$

Let $\mu = \mathbb{E}(e(X, Y)) = p|X||Y|$. Now, using the Chernoff bound,

$$\begin{aligned} \mathbb{P}(|d(X, Y) - p| > \epsilon/2) &= \mathbb{P}(|e(X, Y) - \mu| > \frac{\epsilon}{2}|X||Y|) \\ &= \mathbb{P}(e(X, Y) > (1 + \frac{\epsilon}{2p})\mu) + \mathbb{P}(e(X, Y) < (1 - \frac{\epsilon}{2p})\mu) \\ &\leq 2e^{\frac{-\epsilon^2 k^2}{4p}}. \end{aligned}$$

Similarly, we have

$$\mathbb{P}(|d(A, B) - p| > \epsilon/2) \leq 2e^{\frac{-\epsilon^2|A||B|}{4p}} \leq 2e^{\frac{-\epsilon^2 k^2}{4p}}.$$

Thus, we can now see that $q \to 0$ as $k \to \infty$. Equivalently, $p \to 1$ as $k \to \infty$. Thus, for sufficiently large subsets of G_p, we know that they are ϵ-regular with high probability. □

With this definition we can prove a quick property for ϵ-regular graphs based on our intuition for the density of random graphs that will be useful later:

Proposition 4.7. *Let (A, B) be an ϵ-regular pair of density d. Then, all but at most $\epsilon|A|$ vertices in A have at least $(d - \epsilon)|B|$ neighbors in B.*

Proof. Let

$$X = \{v \in A \mid |N_B(v)| < (d - \epsilon)\}.$$

Then, $d(X, B) < d - \epsilon$. So, by the definition of ϵ-regularity, we know that $|X| < \epsilon|A|$. □

Since we have a precise definition of ϵ-regularity now, and we know about its relationship to random graphs, we will move on to making useful statements with them both. However, first, we must discuss the notion of "blowing up" a graph H on n vertices. This new, blown-up graph $H(t)$ consists of vertex sets $A_1, ..., A_n$ with $|A_1| = \cdots = |A_n| = t$. The edge set of $H(t)$ is

$$E = \{(v, w) \text{ for } v \in A_i \text{ and } w \in A_j \text{ if } (i, j) \in E(H)\}.$$

Now we can turn to random graphs, and blow up H into a graph $R \subset H(t)$, but this time replacing the edges with a random bipartite graph of density d. Then, intuitively, we would expect that if d is high enough, and if G is a subgraph of $H(t)$, then G is also a subgraph of R. This intuition is purely based on informal argumentation: if we keep "enough" randomly selected edges of $H(t)$ in R, we will still expect to find subgraphs of $H(t)$ in R. Since this intuition comes from our concept of random graphs, it turns out that we can prove a formal statement using ϵ-regularity:

Lemma 4.8 (Embedding Lemma)**.** *Let $d > \epsilon > 0$, $m \in \mathbb{Z}^+$, and H be an arbitrary graph. Let R be the graph constructed by replacing each vertex of H with m vertices, and replacing the edges of H with ϵ-regular pairs of density at least d. Then, let G be a subgraph of $H(t)$ with n vertices and maximum degree Δ, and let $\delta = d - \epsilon$. If $\epsilon \leq \delta^\Delta$ and $t - 1 \leq (\delta^\Delta - \Delta\epsilon)m$, then the number of labeled copies of G in R is $||G \to R|| > [(\delta^\Delta - \Delta\epsilon)m - (t-1)]^n$.*

Proof. The vertices $v_1, ..., v_n$ of G will be embedded into R one at a time in an algorithm. For every v_j with $j > i$ that has not been picked by time step i, we denote the possible locations of v_j by the set C_{ij}. At time 0, C_{0j} is restricted to the vertex set A_j of $H(m)$, so $|C_{0j}| = m$. Our algorithm consists of two steps. In the first step, we choose a vertex v_i so that it has "enough" neighbors in each C_{ij} with $j > i$. In the second, we update C_{ij} with $j > i$ after choosing v_i.

Step 1. Pick v_i in $C_{i-1,i}$ such that

$$|N(v_i) \cap C_{i-1,j}| > \delta |C_{i-1,j}|$$

for all $j > i$ such that $\{v_i, v_j\} \in E(G)$.

Step 2. For $j > i$, let

$$C_{ij} = \begin{cases} N(v_i) \cap C_{i-1,j} & \text{if } \{v_i, v_j\} \in E(G) \\ C_{i-1,j} & \text{otherwise} \end{cases}$$

For $i < j$, let $d_{ij} = |\{l \in [i] \mid \{v_l, v_j\} \in E(G)\}|$. Then, if $d_{ij} > 0$, we know from Step 1 that $|C_{ij}| > \delta^{d_{ij}} m$. So, for all $i < j$, we have $|C_{ij}| > \delta^\Delta m \geq \epsilon m = \epsilon |A_j|$. From 4.7, we know that at most $\Delta\epsilon m$ do not satisfy the conditions for Step 1. Thus, for each i, we have at least

$$|C_{i-1,i}| - \Delta\epsilon m - (t-1) > (\delta^\Delta - \Delta\epsilon)m - (t-1)$$

choices. We subtract $(t-1)$ to take into account the maximum number of vertices that G could have in the blown-up vertex of $H(t)$. Thus, this proves our claim that there are at least $(\delta^\Delta - \Delta\epsilon)m - (t-1)$ copies of G in R. □

In other words, if we have a blown-up graph with edges consisting of ϵ-regular pairs of density d, then we can treat these edges as complete bipartite graphs when embedding a subgraph, as long as the subgraph isn't too "complicated." By

complicated, we mean that the subgraph must fulfill the requirements at the end of the statement of the Embedding Lemma.

5. Szemerédi Regularity Lemma

To begin with, we will extend our sense of ϵ-regular pairs to ϵ-regular partitions of vertex sets:

Definition 5.1. An *ϵ-regular partition*, $\mathscr{P}$, of a graph's vertex set is a collection of pairwise disjoint sets $V_0, V_1, ..., V_k$ such that $V_0 \cup V_1 \cup \cdots \cup V_k = V$ and $|V_1| = \cdots = |V_k|$. Furthermore, $|V_0| \leq \epsilon|V|$, and all pairs (V_i, V_j) with $1 \leq i < j \leq k$, except at most ϵk^2 of them, are ϵ-regular.

In addition to ϵ-regular partitions, we also have the notion of a *refinement*. If we let each point in V_0 and V_0' be a separate "part" of the partitions, a partition $\mathscr{P}'$ is a refinement of $\mathscr{P}$ if and only if $\mathscr{P}$ can be created through the union of some of the parts of $\mathscr{P}'$, namely some of the sets V_i' and points from V_0'.

We must now introduce the concept of an index of a partition, which will be useful in the following proof:

Definition 5.2. For disjoint subsets $U, W \subset V$, define their index

$$q(U, W) = (|U||W|/n^2)d^2(U, W).$$

Next, for two partitions $\mathscr{U}, \mathscr{W}$ of U and W, define their index

$$q(\mathscr{U}, \mathscr{W}) = \sum_{\substack{U' \in \mathscr{U} \\ W' \in \mathscr{W}}} q(U', W').$$

Finally, for a partition $\mathscr{P}$ of V,

$$q(\mathscr{P}) = \sum q(U, W),$$

where U and W are the distinct parts of $\mathscr{P}$, including each individual vertex of the exceptional set $V_0 \in \mathscr{P}$.

Now that we have the proper vocabulary, will move on to proving Szemerédi's Regularity Lemma. This result is one of of the most important results in graph theory. Roughly speaking, it states that, as long as a graph is large enough, we can approximate it well through the union of ϵ-regular pairs of k-sets. Coupling this with the parallels we drew between ϵ-regular pairs and random graphs, the Szemerédi Regularity Lemma tells us that we can approximate very large *arbitrary* graphs as the collection of random looking bipartite graphs. This statement is indispensible for dealing with extremal graph properties.

Theorem 5.3 (Szemerédi's Regularity Lemma)**.** *Let $\epsilon > 0$ and $t \in \mathbb{Z}$. Then, there exists an integer $T = T(\epsilon, t)$ such that every graph G with $|V| = n \geq T$ has an ϵ-regular partition $(V_0, V_1, ..., V_k)$, where $t \leq k \leq T$.*

Proof. **Part 1.** Suppose that P is not ϵ-regular. Consider a non ϵ-regular pair (A, B). Then, there exists (X, Y) with $|X| \geq \epsilon|A|$ and $|Y| \geq \epsilon|B|$ such that $|d(X, Y) - d(A, B)| > \epsilon$. Let $X' = A \setminus X$ and $Y' = B \setminus Y$. Let $\mathscr{X}$ and $\mathscr{Y}$ be the

partitions $\{X, X'\}$ and $\{Y, Y'\}$, respectively. Pick $a \in A$ and $b \in B$ randomly. Let $S \in \mathscr{X}$ and $T \in \mathscr{Y}$ such that $a \in S$ and $b \in T$. Finally, let $Z = d(S,T)$. Then,

$$\mathbb{E}(Z) = \sum_{\substack{S\in\mathscr{X}\\T\in\mathscr{Y}}} \frac{|S||T|}{|A||B|} d(S,T) = \sum_{\substack{S\in\mathscr{X}\\T\in\mathscr{Y}}} \frac{e(S,T)}{|A||B|} = d(A,B).$$

Furthermore,

$$\mathbb{E}(Z^2) = \sum_{\substack{S\in\mathscr{X}\\T\in\mathscr{Y}}} \frac{|S||T|}{|A||B|} d^2(S,T) = \frac{n^2}{|A||B|} \sum_{\substack{S\in\mathscr{X}\\T\in\mathscr{Y}}} \frac{|S||T|}{n^2} d^2(S,T) = \frac{n^2}{|A||B|} q(\mathscr{X},\mathscr{Y}).$$

Then, from an expanded expression for the variance of Z, we have

$$Var(Z) = \mathbb{E}(Z^2) - \mathbb{E}(Z)^2 = \frac{n^2}{|A||B|}(q(\mathscr{X},\mathscr{Y}) - q(A,B)).$$

Now, since $\mathbb{E}(Z) = d(A,B)$ and

$$\mathbb{P}(|Z - \mathbb{E}(Z)| > \epsilon) \geq \frac{|X||Y|}{|A||B|} > \epsilon^2,$$

we know from Chebyshev's inequality that

$$Var(Z) > (\epsilon^2)\epsilon^2 = \epsilon^4$$

and so finally we get

$$q(\mathscr{X},\mathscr{Y}) > q(A,B) + \epsilon^4 \frac{|A||B|}{n^2}.$$

Similarly, if we let $\mathscr{X}$ and $\mathscr{Y}$ be arbitrary partitions of A and B with Z defined as earlier, from Jensen's inequality, we get

$$\mathbb{E}(Z^2) \geq \mathbb{E}(Z)^2$$
$$\frac{n^2}{|A||B|} q(\mathscr{X},\mathscr{Y}) \geq \frac{n^2}{|A||B|} q(A,B).$$

Now, if P' is a refinement of P, then the parts of P' are partitions of P and so it follows that $q(P') \geq q(P)$.

Part 2. Let P be the partition described above, and let $c := |V_i|$ for $V_i \in P$. There are two cases for each pair (V_i, V_j):

1. If (V_i, V_j) is ϵ-regular, let $\mathscr{V}_{ij} = V_i$ and $\mathscr{V}_{ji} = V_j$.

2. If (V_i, V_j) is not ϵ-regular, choose the partitions $\mathscr{V}_{ij}$ of V_i and $\mathscr{V}_{ji}$ of V_j as in Part 1.

Now, after setting all of these partitions, let $\mathscr{V}_i$ be the venn diagram of $\mathscr{V}_{ij}$ for $j \neq i$. There are then $k-1$ partitions eligible, so $\mathscr{V}_i$ has at most 2^{k-1} parts. Then, let P' be the refinement of P formed by $\mathscr{V}_i$ and the original exceptional set V_0. We know that there were at least ϵk^2 non ϵ-regular pairs in P since p was not ϵ-regular. Since $|V_0| \leq \epsilon n$,

$$kc \geq (1-\epsilon)n \geq \frac{3}{4}n.$$

We have essentially just applied the process from Part 1 at least ϵk^2 times on sets of size c. So,

$$\begin{aligned} q(P') &\geq q(P) + \epsilon^4(\epsilon k^2)\left(\frac{c^2}{n^2}\right) \\ &> q(P) + \frac{\epsilon^5}{2}. \end{aligned}$$

Now, we must turn P' into an equipartition Q. P' has at most $k2^{k-1}$ parts besides V_0. So, let $c' = \lfloor c/4^k \rfloor$. Split each part of P' into disjoint sets of size c' with the extra vertices going to V_0. Then, there are at most $(kc)/(c/4^k) = k4^k$ parts in Q. Furthermore, the exceptional set of Q will have size

$$|V_0'| < |V_0| + k2^{k-1}c' \leq |V_0| + \frac{n}{2^k}.$$

From the end of Part 1, since Q is a refinement of P', we know that

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

Part 3. Observe that, for any partition P, $q(P) < 1/2$. Now, let $s = \lceil 2/\epsilon^5 \rceil$ and $k_0 = t$ for t such that

$$2^{t-2} > \frac{1}{\epsilon^6}.$$

Then, let $k_i = k_{i-1}4^{k_{i-1}}$ and $T = k_s$. Now, consider a graph G with $|V| = n \geq T$. Let P_0 be the equipartition into k_0 pairwise disjoint parts, each of size $\lfloor n/t \rfloor$ with the extra vertices comprising V_0. Continually refine it as in Part 2, and, since $q(P_i) \leq 1/2$, this process must terminate after at most s steps. Let P be the resulting partition after this refinement process.

We know each refinement increased P by at most $k4^k$ parts, and so P ends up having at most T nonexceptional parts. V_0 increased by at most $n/2^k < \epsilon n/2s$ vertices each time, and since it started with at most $t << \epsilon n/2$ points, we know

$$|V_0| < \frac{\epsilon n}{2} + \frac{\epsilon n}{2s}(s) = \epsilon n.$$

Since the process terminates, the resulting P *must* be ϵ-regular. Thus, we have found a partition P that satisfies the requirements of the Regularity Lemma. □

The Regularity Lemma is truly invaluable for dealing with extremal graphs. Since it gives us a statement on how ϵ-regular pairs make up an arbitrary graph, we can couple it with the embedding lemma from last section to prove challenging results. We will conclude this paper with an answer to the following loosely-phrased question:

Question 5.4. *Given a graph G with a certain amount of triangles, how many edges must we remove to make a new graph triangle-free?*

The answer is less than the number of triangles. In fact, it is *much* less than the number of triangles. This question was not answered until 1976, and, until recently, the Regularity Lemma was the only method able to prove the following result:

Theorem 5.5 (Triangle Removal Lemma)**.** *Let $\epsilon > 0$. Then, there exists $\delta > 0$ such that if G has at most δn^3 triangles, then G can be made triangle-free by removing at most ϵn^2 edges.*

Proof. Let us apply the Regularity Lemma with $\epsilon' = \epsilon/4 > 0$ to partition G into $V_0, V_1, ..., V_k$. Then, we will remove the edges between any ϵ-irregular pairs, of which there are at most $\frac{\epsilon}{4}n^2$ edges. We will remove the edges between any pairs with density at most $\epsilon/2$, which will be at most $\frac{\epsilon}{2}n^2$ edges. Finally, we will remove all edges inside pairs, of which there are at most $kn\frac{\epsilon}{4k}n = \frac{\epsilon}{4}n^2$ edges. Thus, we have removed at most ϵn^2 edges from graph G to form G'.

If we were to have a triangle in G', its three vertices would be positioned in pairs that are $\epsilon/4$-regular with density at least $\epsilon/2$. Thus, by the embedding lemma, we would have at least

$$\left(\left(\frac{3}{4}\right)^2 - 2\epsilon\right)\frac{n}{k} - \left(\frac{n}{k} - 1\right)^n$$

copies of the triangle. We know that k is bounded from the Regularity Lemma, and so we reach a contradiction by choosing δ small enough. □

Thinking about ϵ-regularity and its relationship to random graphs, we can very roughly summarize what goes on behind the scenes in the proof of the Triangle Removal Lemma: for an arbitrary graph, we can think of it as the union of a collection of "random" bipartite graphs, and so many of the triangles will have overlapping edges. This way, it takes a smaller number of edges to destroy all of the triangles in a graph than the actual number of triangles. This may seem weird, but that's the magic of the Regularity Lemma!

Acknowledgments. I would like to thank Professor Peter May for organizing and running such a remarkable mathematics REU program. It is also a pleasure to thank my mentor, Anthony Santiago Chaves Aguilar, for guiding me through my paper and going the extra mile to offer deep insight into the mathematics behind it.

References

[1] Noga Alon and Joel H. Spencer. *The Probabilistic Method.* John Wiley & Sons, Inc., 2008.
[2] Béla Bollobás. *Random Graphs: Second Edition.* Cambridge Studies in Advanced Mathematics. Cambrige University Press, 2001.
[3] David Conlon. *Notes on Extremal Graph Theory: Lecture 6.* Wadham College, Oxford. <https://www.dpmms.cam.ac.uk/~dc340/EGT6.pdf>
[4] János Komlós, Ali Shokoufandeh, Miklós Simonovits, Endre Szemerédi. "The Regularity Lemma and Its Applications in Graph Theory." *Theoretical Aspects of Computer Science.* Springer-Verlag, 2002.
[5] Ryan Martin. *Notes on Extremal Graph Theory.* Iowa State University, 5 April 2012. <http://orion.math.iastate.edu/rymartin/ISU608EGT/EGTbook.pdf>
[6] Robert Morris and Roberto Imbuzeiro Oliveira. *Extremal and Probabilistic Combinatorics.* Publicações Matemáticas. Instituto de Matemática Pura e Aplicada, 2011.
[7] A. N. Kolmogorov. *Foundations of the Theory of Probability.* Chelsea Publishing Company, 1956.
[8] Ramsey's Theorem. *Wikipedia.* Wikimedia Foundation, Inc., 5 July 2016. <https://en.wikipedia.org/wiki/Ramsey%27s_theorem>
[9] Boole's Inequality. *Wikipedia.* Wikimedia Foundation, Inc., 28 August 2016. https://en.wikipedia.org/wiki/Boole%27s_inequality
[10] Chernoff Bound. *Wikipedia.* Wikimedia Foundation, Inc., 28 August 2016. https://en.wikipedia.org/wiki/Chernoff_bound