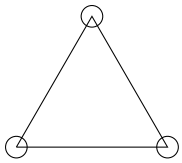
PUMaC 2019 Power Round
“There is an explicit way to define what explicit is.” — Noga Alon

November 9, 2019



Rules and Reminders

1. Your solutions may be turned in in one of two ways:
   - You may email them to us at **pumac2019@gmail.com** by 8AM Eastern Standard Time on the morning of PUMaC, November 16, 2019 with the subject line “PUMaC 2019 Power Round.”
   - You may hand them in to us when your team checks in on the morning of PUMaC. Please staple your solutions together, including the cover sheet.

   The cover sheet (the last page of this document) should be the first page of your submission. Each page should have on it the **team number** (not team name) and **problem number**. This number can be found by logging in to the coach portal and selecting the corresponding team. Solutions to problems may span multiple pages, but include them in continuing order of proof.

2. You are encouraged, but not required, to use LaTeX to write your solutions. If you submit your power round electronically, **you may not submit multiple times**. The first version of the power round solutions that we receive from your team will be graded. If submitting electronically, **you must submit a PDF**. No other file type will be graded.

3. Do not include identifying information aside from your team number in your solutions.

4. Please collate the solutions in order in your solution packet. Each problem should start on a new page, and solutions should be written on one side of the paper only (there is a point deduction for not following this formatting).

5. On any problem, you may use without proof any result that is stated earlier in the test, as well as any problem from earlier in the test, even if it is a problem that your team has not solved. These are the **only** results you may use. You may not cite parts of your proof of other problems: if you wish to use a lemma in multiple problems, please reproduce it in each one.

6. When a problem asks you to "find with proof," "show," "prove," "demonstrate," or "ascertain" a result, a formal proof is expected, in which you justify each step you take, either by using a method from earlier or by *proving* that *everything* you do is correct. When a problem instead uses the word "explain," an informal explanation suffices. When a problem asks you to "find" or "list" something, no justification is required.

7. All problems are numbered as "Problem x.y.z" where x is the section number and y is the subsection. Each problem's point distribution can be found in the cover sheet.

8. **You may NOT use any references, such as books or electronic resources, unless otherwise specified. You may NOT use computer programs, calculators, or any other computational aids.**

9. Teams whose members use English as a foreign language may use dictionaries for reference.

10. **Communication with humans outside your team of 8 students about the content of these problems is prohibited.**

11. There are two places where you may ask questions about the test. The first is Piazza. Please ask your coach for instructions to access our Piazza forum. On Piazza, you may ask any question **so long as it does not give away any part of your solution to any problem**. If you ask a question on Piazza, all other teams will be able to see it. If such a question reveals all or part of your solution to a power round question, your team's power round score will be penalized severely. For any questions you have that might reveal part of your solution, or if you are not sure if your question is appropriate for Piazza, please email us at pumac@math.princeton.edu. We will email coaches with important clarifications that are posted on Piazza.

Introduction and Advice

This year's power round is about **extremal combinatorics**, and more specifically extremal combinatorics in Graph Theory. extremal combinatorics studies the maximum and/or minimum possible cardinalities of combinatorial structures with some desired properties. It has applications in theoretical computer science, information theory, number theory, geometry and others. For example, a standard question in extremal combinatorics is of the form

If we look at structures with specific properties, how big or small can they be?

We will answer this question for a few examples with graphs.

The power round is structured such that it will walk you through proofs of some of the most important theorems. Afterwards there will be some auxiliary problems.

Here is some further advice with regard to the Power Round:

- **Read the text of every problem!** Many important ideas are included in problems and may be referenced later on. In addition, some of the theorems you are asked to prove are useful or even necessary for later problems.
- **Make sure you understand the definitions**. A lot of the definitions are not easy to grasp; don't worry if it takes you a while to fully understand them. If you don't, then you will not be able to do the problems. Feel free to ask clarifying questions about the definitions on Piazza (or email us).
- **Don't make stuff up**: on problems that ask for proofs, you will receive more points if you demonstrate legitimate and correct intuition than if you fabricate something that *looks* rigorous just for the sake of having "rigor."
- **Check Piazza often!** Clarifications will be posted there, and if you have a question it is possible that it has already been asked and answered in a Piazza thread (and if not, you can ask it, assuming it does not reveal any part of your solution to a question). **If in doubt about whether a question is appropriate for Piazza, please email us at pumac@math.princeton.edu.**
- **Don't cheat**: as stated in Rules and Reminders, you may **NOT** use any references such as books or electronic resources. If you do cheat, you will be disqualified and banned from PUMaC, your school may be disqualified, and relevant external institutions may be notified of any misconduct.

Good luck, and have fun!

– Marko Medvedev

I would like to acknowledge and thank many individuals and organizations for their support; without their help, this Power Round (and the entire competition) could not exist. Please refer to the solutions of the power round for full acknowledgments and references.

Contents

1 Introduction **6**

1.1 Graph theory 6

1.2 Asymptotic notation 9

1.3 Probability 9

1.4 Definitions 9

1.5 A few applied problems 13

2 Collegiate partition theorem **15**

2.1 Introduction 15

2.2 The Coupon theorem 21

2.3 The generalized coupon theorem 24

3 Graph discount chances **26**

3.1 Introduction 26

3.2 Connection between graph discount chances testing and the Collegiate partition theorem 30

Notation

- $\forall$: for all. *ex.: $\forall x \in \{1, 2, 3\}$ means "for all x in the set $\{1, 2, 3\}$"*
- $A \subset B$: proper subset. *ex.: $\{1, 2\} \subset \{1, 2, 3\}$, but $\{1, 2\} \not\subset \{1, 2\}$*
- $A \subseteq B$: subset, possibly improper. *ex.: $\{1\}, \{1, 2\} \subseteq \{1, 2\}$*
- $f : x \mapsto y$: f maps x to y. *ex.: if $f(n) = n - 3$ then $f : 20 \mapsto 17$ and $f : n \mapsto n - 3$ are both true.*
- $\{x \in S : C(x)\}$: the set of all x in the set S satisfying the condition $C(x)$. *ex.: $\{n \in \mathbb{N} : \sqrt{n} \in \mathbb{N}\}$ is the set of perfect squares.*
- $\mathbb{N}$: the natural numbers, $\{1, 2, 3, \ldots\}$.
- $[n] = \{1, 2, 3, ..., n\}$.
- $\mathbb{Z}$: the integers.
- $\mathbb{R}$: the real numbers.
- $|S|$: the cardinality of set S.

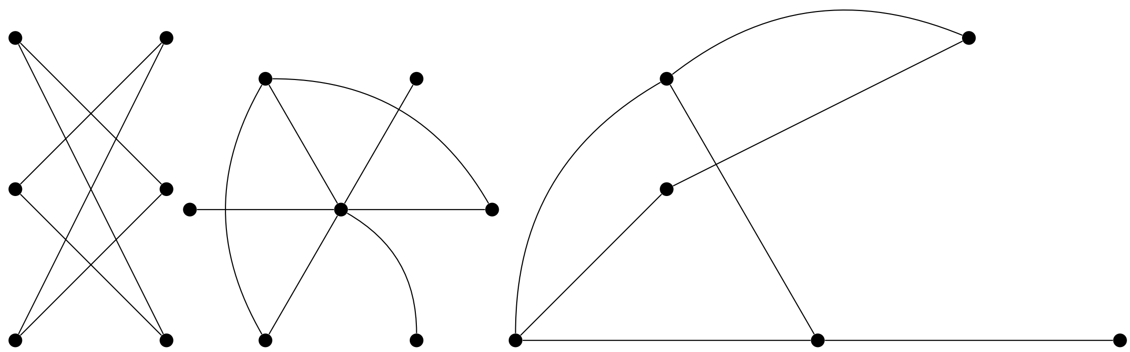


Figure 1: Three examples of graphs.

1 Introduction

1.1 Graph theory

Graphs are objects used to model relations between objects. For example, in a room of people, we can imagine that two people are connected by a line if and only if they are friends. This actually defines a graph. It turns that graphs are important object. They appear in many different areas of combinatorics and mathematics, as well as computer science. To study property of graphs we need following definitions:

Definition 1.1.A. A *graph* is an ordered pair $G = (V, E)$ of where:

1. V is a set of *vertices*,
2. $E \subseteq \{\{u, v\} | (u, v) \in V \times V, u \neq v\}$ is a set of *edges*, i.e. a subset of the set of unordered pairs of vertices. If $(u, v) \in E$, we say that uv is an edge.

Definition 1.1.B. If $\{u, v\}$ is an edge in a graph we say that u and v are *connected* (by an edge) or *adjacent*.

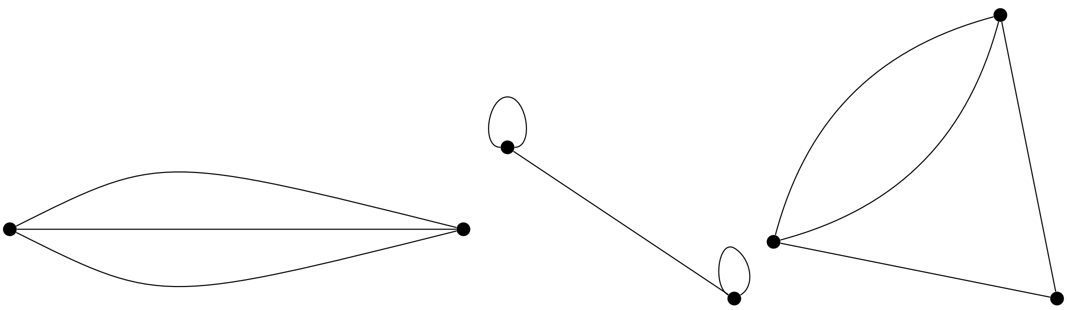


Figure 2: Three examples of non-graphs.

Definition 1.1.C. A cycle of length n, C_n is the sequence of distinct vertices $v_1, v_2, ..., v_n$, $n \geq 3$, in a graph G such that $v_i v_{i+1}$ are edges for every $i = 1, ..., n$ and indices are taken modulo n.

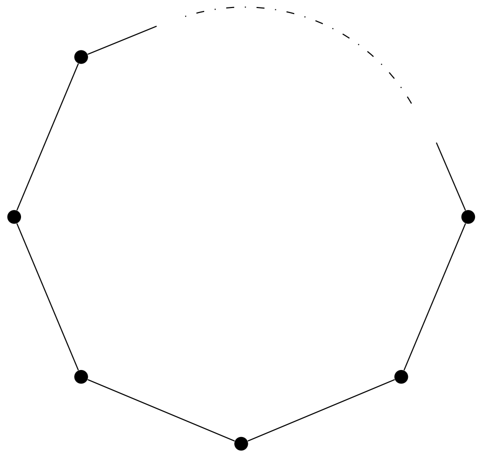


Figure 3: A cycle C_n on n vertices.

Definition 1.1.D. A *complete graph* on n vertices, denoted K_n is a graph on n vertices with all possible edges, i.e. $K_n = (V, E)$ where $V = [n]$ and $E = \{\{u, v\} | (u, v) \in V \times V, u \neq v\}$

Definition 1.1.E. For a vertex $v \in V$ of a graph $G = (V, E)$, *the degree* of vertex v is the number of vertices $u \in V$ such that $\{u, v\}$ is an edge.

Definition 1.1.F. A graph $G = (V, E)$ is said to be d-regular if every vertex in G has degree d.

Definition 1.1.G. A graph $G = (V, E)$ is *bipartite* if the set of vertices V can be split into two sets X, Y such that every edge in E connects two vertices, one in X and the other in Y.

Definition 1.1.H. A *complete bipartite graph on n and t vertices* $K_{n,t}$ is a bipartite graph $G = (X \cup Y, E)$ for $X \cap Y = \emptyset, |X| = n, |Y| = t$ and the set of edges E are all the edges between a point in X and a point in Y.

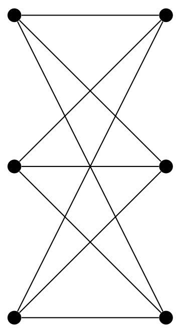


Figure 4: A bipartite graph, $K_{3,3}$

Example: The graph in Figure 4 is a complete bipartite graph $K_{3,3}$; This graph is 3-regular. The leftmost graph in Figure 1 is also 2-regular.

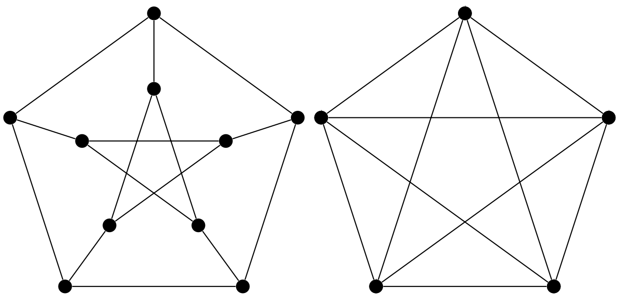


Figure 5: An example of a graph homomorphism from the left graph to K_5. What does the homomorphism look like, symbolically?

Definition 1.1.I. A graph $H = (V', E')$ is said to be a *subgraph* of a graph G if $V' \subseteq V$, $E' \subseteq E$.

Definition 1.1.J. A subgraph $H = (V', E')$ of a graph $G = (V, E)$ is said to be *spanning* if $V' = V$.

Definition 1.1.K. A subgraph $H = (V', E')$ of a graph $G = (V, E)$ is said to be *induced* if E' are all the edges in E between the vertices in V'.

For example, both K_3 and C_4 are subgraphs of K_4 but only K_3 is an induced subgraph.

Definition 1.1.L. An *independent set* in a graph $G = (V, E)$ is a subset of vertices such that no two are connected by an edge.

In C_5, for example, the maximal independent set contains two vertices.

Definition 1.1.M. The *r-blowup* of a graph G is a graph obtained from G by replacing every vertex by an independent set of size r and every edge by a complete bipartite subgraph between corresponding independent sets.

For example, $K_{r,r}$ is an r-blowup of a graph with two vertices and an edge between them.

Definition 1.1.N. Let H and G be graphs. A *homomorphism* from H to G is a function $f : V(H) \to V(G)$ (not necessarily one-to-one) such that for all vertices u, v which form an edge in H, there is also an edge joining $f(u)$ to $f(v)$ in G (two vertices that are connected in$H$ map to sifferent vertices in G, but if they are not connected they can map to the same vertex).

For example, there is a homomorphism between between C_5 and C_3.

Definition 1.1.O. An isomorphism between H and G that is a homomorphism between H and G which is bijective and whose inverse is also a homomorphism.

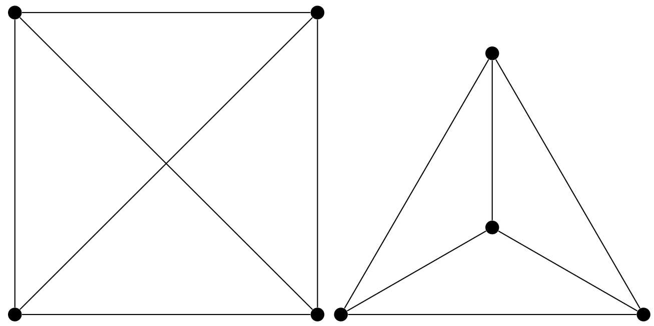


Figure 6: Two isomorphic, hence homomorphic, copies of K_4.

See Figures 3–5 for examples of homomorphisms and isomorphisms.

Definition 1.1.P. The *core* of a graph H is a subgraph H' with the minimum possible number of edges such that there exists a homomorphism from H to H'.

For example, the core of every bipartite graph is a single edge.

1.2 Asymptotic notation

Little o and big O notations are used to compare the order of growth of two real functions.

Definition 1.2.A. For two real functions $f, g : \mathbb{R}^+ \to \mathbb{R}$ we say that $f(x)$ is $o(g(x))$ (or in words f is in small o of g), written as $f(x) = o(g(x))$ as $x \to \infty$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$.

Example: $1/x$ is $o(1)$, x is $o(x^2)$, and any polynomial of finite degree in x is $o(e^x)$.

Definition 1.2.B. For two real function $f, g : \mathbb{R}^+ \to \mathbb{R}$ we say that $f(x)$ is $O(g(x))$ (or in words f is in big O of g), written as $f(x) = O(g(x))$, if there exists a constant $c \in \mathbb{R}$ and $M \in \mathbb{R}^+$, such that for all $x > M$, $f(x) < cg(x)$.

1.3 Probability

Probability theory is the mathematical study of randomness. It has broad applications across natural sciences, engineering, social sciences and pure mathematics. As we will see later, probability has also applications in combinatorics. First we will introduce the main definitions of probability that are needed to use it in combinatorics, and then we will see a few examples of applications of probability in general.

1.4 Definitions

A *random experiment* is an experiment whose outcome cannot be predicted before the experiment is performed, but the possible outcomes are known.

Definition 1.4.A. The *sample space*, usually denoted Ω, is the set of all possible outcomes of a random experiment.

Example: Let the random experiment be throwing one red and one blue die. Denote the outcome as (i, j) if the red one comes up i and the blue one comes up j. Define the sample space

$$\Omega = \{(i, j) : 1 \leq i, j \leq 6\}$$

Thus, here there are $6^2 = 36$ possible outcomes.

Now we can discuss what types of question we can ask about outcomes of a random experiment. Informally an *event* is a statement for which we can determine whether it is true or false after the experiment has been performed. Formally, we define an event in the following way:

Definition 1.4.B. An *event* is a subset A of the sample space Ω.

Example: From the previous example of two dice, an example of an event might be

"The sum of numbers on the dice is 8."

Consider now two events A and B. The intersection $A \cap B$ is the event that A *and* B both occur. The union $A \cup B$ is the event that A *or* B occurs (meaning either A, B or both occur). The complement $A^c = \Omega - A$ is the event that A *does not occur*.

Definition 1.4.C. A probability measure is an assignment of a nonnegative real number $\mathbb{P}(A) \in [0, 1]$ to every event A such that the following rules are satisfied.

1. $\mathbb{P}(\Omega) = 1$ (something certainly happens);
2. If events A and B are such that $A \cap B = \emptyset$, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$. More generally, if events $E_1, E_2, \ldots$ satisfy $E_i \cap E_j = \emptyset$ for all $i \neq j$ then

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i).$$

From this we can prove following two things:

Problem 1.4.1. Prove that for any events A, B:

1. $\mathbb{P}(A) + \mathbb{P}(A^c) = 1$.
2. If $A \subset B$ then $\mathbb{P}(B) \geq \mathbb{P}(A)$.

Example: Say that we want to model the random experiment of throwing a die. The natural sample space is

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

To assign the probabilities we have to make some modeling assumptions. We make the assumption that each outcome is equally likely to occur. This can be therefore be written as

$$\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \cdots = \mathbb{P}(\{6\}) \implies$$
$$\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \cdots = \mathbb{P}(\{6\}) = \frac{1}{6}.$$

Many other random processes can be modeled in this way. For another example, let's model a biased coin that always when tossed has twice as much probability to turn heads than tails. The natural sample space in this example is

$$\Omega = \{H, T\}.$$

To assign the probabilities, we know that the coin will turn either heads or tails. Let $\mathbb{P}(\{H\})$ be the probability that the coin ends up as heads and $\mathbb{P}(\{T\})$ as tails. For simplicity, we write $\mathbb{P}(\{H\}) = \mathbb{P}(H)$. By what we know about the probability measure,

$$\mathbb{P}(H) + \mathbb{P}(T) = 1$$
$$\mathbb{P}(H) = 2\mathbb{P}(T) \implies$$
$$\mathbb{P}(H) = \frac{2}{3}$$
$$\mathbb{P}(T) = \frac{1}{3}.$$

Next, we will define independent events. Imagine that you are throwing a fair six sided die. It is reasonable to assume that different throws don't depend on each other, i.e. that if you throw a six on the first try, you are not more or less likely to throw a six on the second throw. Gaining information about the first throw doesn't gives us any information about the second. Another example we might consider is that assume that we know that the probability that it is going to snow for every day of the year. The probability that it will snow tomorrow might increase if we know that it is snowing today. To reason rigorously in problems of this kind, we introduce the formal notation of *conditional probability.*

Definition 1.4.D. The *conditional probability* $P(A|B)$ of an event A given that the event B occurs is defined as

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

provided $\mathbb{P}(B) > 0$. We don't define conditional probability for an event B that happens with probability 0.

Now we can define independent events.

Definition 1.4.E. Events A, B are *independent* if $\mathbb{P}(A|B) = \mathbb{P}(A)$, or equivalently $\mathbb{P}(A \cap B) = \mathbb{P}(A)P(B)$.

In most situations, we will be interested in random quantities that are not necessarily represented by a yes/no question. Image that we are throwing a coin 5 times. The number of heads is a random number between 0 and 5. Such random quantities are called *random variables*. Let Ω be the sample space as usual.

Definition 1.4.F. A *random variable* is a function X assigning a real number $X(\omega) \in \mathbb{R}$ to each possible outcome $\omega \in \Omega$. We write the probability that X takes value t as $\mathbb{P}(X = t)$.

Example: We flip a coin two times. A good choice for sample space for this problem is

$$\Omega = \{HH, HT, TH, TT\}.$$

Let X be the total number of head that are flipped. Then X is a random variable, and we can define it explicitly as follows:

$$\mathbb{P}(X = 0) = \frac{1}{4}, \quad \mathbb{P}(X = 1) = \frac{1}{2}, \quad \mathbb{P}(X = 2) = \frac{1}{4}.$$

We can interpret $\mathbb{P}(X = i)$ as the fraction of repeated experiments in which the random variable X takes value i. In many cases we would like to know what X is "on average". If f is a function from $\mathbb{R} \to \mathbb{R}$ we can compute the average value of $f(X)$:

Two random variables X and Y are said to be *independent* if for all $x, y \in \mathbb{R}$, we have that $\mathbb{P}(X = x \text{ and } Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$. To be more precise the event $X = x$ and $Y = y$ is the intersection of events $X = x$ and $Y = y$.

Definition 1.4.G. Let $X : \Omega \to \mathbb{R}$ be a random variable taking finitely many values from a set $S \subset \mathbb{R}$. Let $f : \mathbb{R} \to \mathbb{R}$ be a function. The *expectation* of $f(X)$ is defined as

$$\mathbb{E}(f(X)) = \sum_{t \in S} f(t)\mathbb{P}(X = t).$$

In particular $\mathbb{E}(X)$ is defined by taking f to be the identity function.

A very important property of random variable is the following;

Theorem 1.4.I. For *any* two random variables $X, Y : \Omega \to \mathbb{R}$ we have that

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

This is non trivial and requires proof but we do not present it here because it is not particularly illuminating in this context. One of the most useful things about this formula is that it holds even if X and Y are not independent. While this summarizes the variable X in some sense, we still can't know much about X only from its expectation. For example X' could take the value $\mathbb{E}(X)$ with probability 1 (i.e. not be random) and have the same expectation as X. Thus we think of variance as a "measure of randomness"(but still, it doesn't capture all the details about a distribution), formally:

Definition 1.4.H. The *variance* of a random variable X is defined as

$$\mathrm{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2).$$

That is, $\mathrm{Var}(X)$ is the expected value of the square difference between X and $\mathbb{E}(X)$.

Variance has a lot of nice properties.

Problem 1.4.2. Prove that the following properties of variance hold:

1. $\text{Var}(X) \geq 0$.
2. $\text{Var}(X) = 0$ if and only if X is nonrandom, meaning that X takes a single value with probability 1.
3. $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.
4. $\text{Var}(X+Y) = \text{Var}(X)+\text{Var}(Y)$ if X and Y are independent random variables
5. $\text{Var}(aX) = a^2 \text{Var}(X)$ for $a \in \mathbb{R}$.

Some other useful properties of the probability measure, variance and random variables are given as exercises in the next problems.

Problem 1.4.3. Let $A_1, ..., A_n$ be events. Prove that $\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) \leq \sum_{i=1}^{n} \mathbb{P}(A_i)$.

Problem 1.4.4. Let X be a non-negative random variable and let $t > 0$. Prove that $\mathbb{P}(X \geq t) \leq \frac{E(X)}{t}$.

1.5 A few applied problems

The first few problems are general problems in probability. The last few problems are combinatorics problems that can be solved using some of the probability notions defined above in a clever way.

Problem 1.5.1. Find the expectation and variance of a 6 sided die roll. Find the expectation and variance of the sum of 100 rolls of a 6 sided die.

Problem 1.5.2. You are given a standard deck of 52 cards. You draw cards, and after each draw you return the card in the deck and shuffle it. Let R be the random variable representing the minimal number of draws needed to draw two red cards in a row, let B be a random variable representing the minimal number of moves needed to draw a red and a black card in that order, and let K be a random variable representing minimal number of moves needed to draw a two red kings in a row. Find the expectations of R, B, K. Let M be a random variable defined as $M = \min(R, B)$. Find the expected value of M.

Problem 1.5.3. Let X_{10} be the random variable representing an outcome of a roll of a ten sided die and let X_{20} be the random variable representing an outcome of a roll of a twenty sided die. The two rolls are independent. Find $\mathbb{E}(X_{10}X_{20})$, $\mathrm{Var}(X_{10}X_{20})$ and $\mathrm{Var}(X_{10}(X_{20} - X_{10}))$.

Problem 1.5.4. You have a bag with 3 coins. One of them is fair, second has heads on both sides and the third has tails on both sides. You take one coin from the bag and toss and it shows head. Then you take another coin without returning the first one and toss it. What is the probability that it shows head? What if you return the first coin?

As an example of how proability can be used to solve combinatorics problems cosnider the following two problems.

Theorem 1.5.I. Every graph on n edges contains a bipartite subgraph with at least $\frac{n}{2}$ edges.

Proof 1.5.1. Take a subset of vertices, X, by randomly including any vertex with probability $\frac{1}{2}$, i.e. for every vertex flip a coin and if the coin comes up Head you include the vertex otherwise you don't include it. Now for every edge e let X_e be the *indicator* of the event E_e="edge exactly one vertex of e is in X", meaning that $X_e = 1$ if E_e happens, and 0 otherwise. Now we have that $\mathbb{E}(X_e) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$, now let R_X be the random variable representing the number of edges having exactly one vertex in X, thus now by linearity of expectation and since by design $\mathbb{E}(R_X) = \sum_{e \in E} \mathbb{E}(X_e) = \frac{n}{2}$. Thus for some X there is at least $\frac{n}{2}$ edges between X $V - X$ where V is the set of all vertices of G.

Theorem 1.5.II. Suppose that every vertex of a bipartite graph on n vertices is given a personalized list of $> \log_2 n$ possible colors. Then, it is possible to give each vertex a color from its list such that no two adjacent vertices receive the same color.

Proof 1.5.2. Let the partition of the vertex set be $V_1 \cup V_2$. For any color in the set of colors, C, flip a fair coin - so we assign either Head or Tail for any color in C. If, for any vertex in V_1 we can choose a color that corresponds to Head, and to any vertex in V_2 a color that corresponds to Tail we would be done. Let X be a random variable representing the number of vertices for which this fails, i.e. we cannot assign it a color. Then we can prove that the probability for this is $< \frac{1}{n}$ for any vertex (think about how would you prove this). Thus we have that $\mathbb{E}(X) < 1$, thus we cannot always have $X \geq 1$ since then we would have $\mathbb{E}(X)$, the weighted average of all of them be $\mathbb{E}(X) \geq 1$. Then, there exists an assignment of Head and Tail to colors in C such that X(of this assignment of colors) $= 0$.

Problem 1.5.5. Let S be a finite set of nonzero integers. Prove that there a constant $c > 0$ such that for every S, there exists a subset of S, X, such that for any two a, b in X, $a + b$ is not in X and $|X| \geq c|S|$.

Problem 1.5.6. From a complete graph G on n vertices, we choose a subset of its edges by including every edge with probability p in the subset. What is the expected number of triangles?

Problem 1.5.7. Prove that for every integer n there exists a bipartite graph on two classes of vertices A, B such that $|A| = |B| = n$ with less than $32n$ edges so that for any $X \subseteq A, Y \subseteq B, |X| = |Y| = \lfloor \frac{n}{4} \rfloor$, there is an edge from X to Y.

2 Collegiate partition theorem

"The following lemma is just combinatorics." — Jose D. Edelstein

2.1 Introduction

Consider the following (hypothetical) scenario, which motivates the mathematics that follow.

All students at Princeton University are separated into k colleges so called *residental colleges*. All residental buildings are separated into residental colleges, in a way such that any two residental colleges are disjoint. Currently, there are six: Rockefeller, Mathey, Butler, Whitman, Wilson, and Forbes (but there are plans to build more). Before freshman year, each person is assigned to one college. For the first two years of your time at Princeton University, your room is in your college and there are many different college specific events intended to make people hang out with each other. Therefore, in some sense, people from one college will spend more time with other people from their college, and are more likely to get to know each other. Thus in order to have the most homogeneous distribution of friendships on campus, we want to distribute people into colleges in a pseudo-random way such that there is no preference whether people that were friends before coming to Princeton are put in the same or different college. Because of this we want friendships between people of different colleges to be close to random.

We will prove that this is always possible under certain restrictions on number of students and friendships between them prior to coming to Princeton. We will call this theorem the *collegiate splitting theorem*. Informally, it says that the vertices of every large graph can be partitioned into a bounded number of nearly equal parts so that nearly all bipartite graphs between pieces are "random-like." We model this problem with graphs. Each person will represent a vertex in the *friendship graph* and two vertices will be connected if and only if the two people were friends before coming to Princeton. To formally formulate the theorem, we will need a few definitions first.

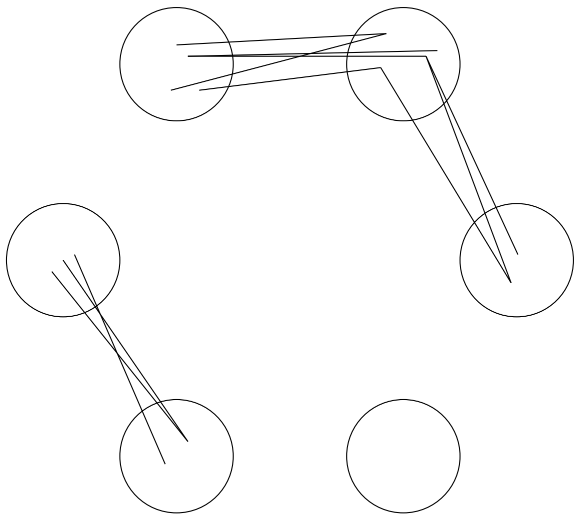


Figure 7: An example of what a random like split of people into colleges might look like

Definition 2.1.A. Let $G = (V, E)$ be a graph with $|V| = n$. For two disjoint sets $U, W \subset V$, let

$$e(U, W) = \text{number of edges between } U \text{ and } W,$$

meaning that we take only edges with one vertex in U and the other in W.

Definition 2.1.B. The *density* of the pair (U, W) is defined as

$$d(U, W) = \frac{e(U, W)}{|U||W|}.$$

Definition 2.1.C. For $\varepsilon > 0$ the pair (U, W) is called ε-*regular* if for every $X \subseteq U$ and $Y \subseteq W$ satisfying $|X| \geq \varepsilon|U|$ and $|Y| \geq \varepsilon|W|$ we have

$$|d(X, Y) - d(U, W)| \leq \varepsilon.$$

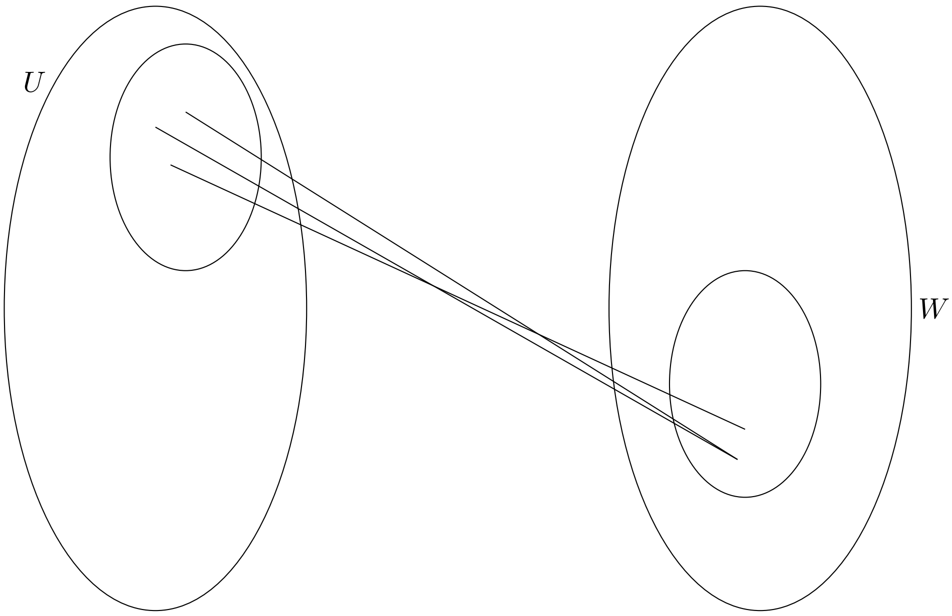


Figure 8: An ε-regular pair (U, W)

For example, if U and W are two classes of vertices of $K_{n,n}$ then they are ε-regular since $d(X, Y)$ is a constant function on $X \in U, Y \in W$.

Problem 2.1.1. Let $\varepsilon > 0$ be a real number. Let U and W be two disjoint sets of vertices such that $|U| = |W| = m$. Let $0 \leq p \leq 1$ be a fixed real number. We generate a random bipartite graph G on $U \cup W$ by declaring that each $(u, w) \in U \times W$ is an edge of G with probability p (and there are no other edges).

1. Prove that for infinitely many m the probability of having (U, W) be ε-regular is larger than $\frac{1}{2}$.
2. Prove that the probability of having (U, W) be ε-regular tends to 1 as m tends to infinity, i.e. prove that for any $\delta > 0$ there is $m_0 \in \mathbb{N}$ such that for any $m > m_0$, $\mathbb{P}((U, W) \text{ is } \varepsilon\text{-regular}) > 1 - \delta$.

Definition 2.1.D. For us, a *partition* P of V is a union of disjoint sets V_i for $0 \leq i \leq k$

$$V = V_0 \cup V_1 \cup \cdots \cup V_k.$$

The sets V_i for $i \geq 1$ are called the k *parts* of the partition. The set V_0 behaves differently; it is called the *exceptional set* of the partition. In a few cases we may also consider each individual vertex of V_0 to also be a *part* (from the definition above, just with only one vertex); we will state so explicitly when such a case arises.

The partition is *equitable* if $|V_1| = |V_2| = \cdots = |V_k|$.

Definition 2.1.E. If P, P' are two partitions of V, P' is a *refinement* of P if each part of P is a union of parts of P'. In particular, if we obtain P' from P by shifting vertices to V_0, then P' is a refinement of P.

Definition 2.1.F. A partition $V = V_0 \cup \cdots \cup V_k$ is ε*-regular* if

- $|V_0| \leq \varepsilon n$; and
- for all but at most εk^2 of the pairs (i, j) with $1 \leq i < j \leq k$, the pair (V_i, V_j) is ε-regular.

An example of this looks like the following:

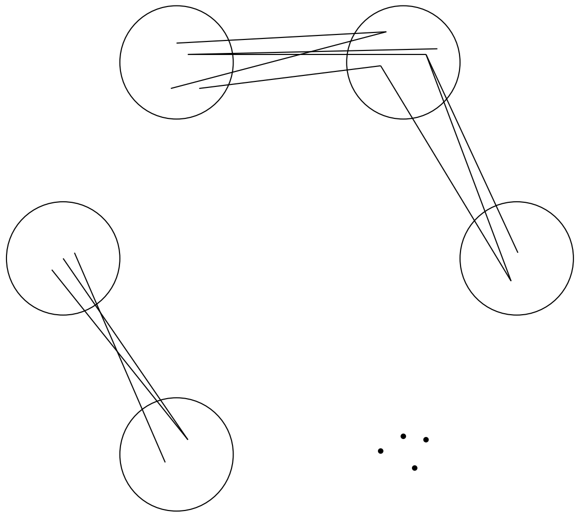


Figure 9: ε-regular partition with some edges drawn.

Now we can state the theorem.

To-be-proved 2.1.1 (Collegiate splitting theorem)**.** For any integer t and any real number $\varepsilon > 0$, there exists an integer $T = T(\varepsilon, t)$ such that any graph $G = (V, E)$ with $|V| \geq T$ has an ε-regular partition with k parts, where $t \leq k \leq T$.

You will now prove the theorem by proving a few auxiliary lemmas. The idea will be the following.

- Start with any partition into t equal parts and put the remainder in the exceptional set.
- Show that as long as the partition is not ε-regular we can refine it in a controlled way, such that the number of parts does not increase too much, the size of an exceptional set does not increase too much, and the weighted average of square of density between parts increases by some $f(\varepsilon)$.
- As this average is always between 0 and $O(1)$, the number of steps is bounded by $O(\frac{1}{f(\varepsilon)})$.

Before we see details we need a bit more notation.

Definition 2.1.G. Let $G = (V, E)$ be a graph with $|V| = n$. For disjoint $U, W \subset V$ let

$$q(U, W) = \frac{|U||W|}{n^2} d(U, W)^2.$$

Definition 2.1.H. For partitions $\mathcal{U}$ of U and $\mathcal{W}$ of W define

$$q(\mathcal{U}, \mathcal{W}) = \sum_{u \in \mathcal{U}, w \in \mathcal{W}} q(u, w).$$

Definition 2.1.I. for a partition $P = (V_0, V_1, \ldots, V_k)$ with $V_0 = \{v_1, \ldots, v_{|V_0|}\}$, the *index* of P is defined as

$$q(P) = \sum q(U, W),$$

where the sum is over distinct $U, W \in \{\{v_1\}, \ldots, \{v_{|V_0|}\}, V_1, \ldots, V_k\}$ (i.e. here we consider V_0 to be $|V_0|$ parts of size 1). This means that $q(P)$ is a sum of $\binom{k+|V_0|}{2}$ terms $q(U, W)$.

The problems 2.1.2–2.1.6 will guide you through the proof of the collegiate splitting theorem.

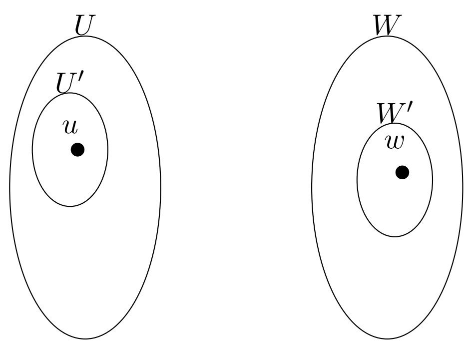
Problem 2.1.2. Prove that for every partition P, we have $0 \leq q(P) \leq \frac{1}{2}$.

Next we will first prove a useful lemma.

Lemma 2.1.1. Let U, W be disjoint sets of vertices. If $\mathcal{U}$ is a partition of U and $\mathcal{W}$ is a partition of W, show that

$$q(\mathcal{U}, \mathcal{W}) \geq q(U, W)$$

Proof 2.1.1. Let Z be a random variable defined on pairs $(u, w) \in U \times W$ in the following way: Let $U' \in \mathcal{U}$ and $W' \in \mathcal{W}$ such that $u \in U', w \in W'$, then $Z(u, w) = d(U', W')$.



Now we have that

$$\begin{aligned}\mathbb{E}(Z) &= \sum_{U' \in \mathcal{U}, W' \in \mathcal{W}} \frac{|U'||W'|}{|U||W|} \frac{e(U', W')}{|U'||W'|} \\ &= \frac{e(U,W)}{|U||W|} = d(U,W),\end{aligned}$$

and similarly

$$\begin{aligned}\mathbb{E}(Z^2) &= \sum_{U' \in \mathcal{U}, W' \in \mathcal{W}} \frac{|U'||W'|}{|U||W|} d^2(U', W') \\ &= \frac{n^2}{|U||W|} q(\mathcal{U}, \mathcal{W}),\end{aligned}$$

thus we have that

$$\mathrm{Var}(Z) = \frac{n^2}{|U||W|}(q(\mathcal{U}, \mathcal{W}) - q(U,W)) \geq 0,$$

which implies the result of the lemma.

Problem 2.1.3. If P, P' are partitions of G and P' is a refinement of P then show that

$$q(P') \geq q(P).$$

Problem 2.1.4. Let U, W be disjoint sets of vertices. If the pair (U, W) is not ε-regular then show that there exists $U_1 \subseteq U$ and $W_1 \subseteq W$ so that for $\mathcal{U} = (U_1, U - U_1), \mathcal{W} = (W_1, W - W_1)$

$$q(\mathcal{U}, \mathcal{W}) \geq q(U, W) + \frac{|U||W|}{n^2} \varepsilon^4.$$

Hint: Look at Z again.

Now we need one final corollary to prove the theorem.

Problem 2.1.5. Let $G = (V, E)$, $|V| = n$ be a graph. Let $P = (V_0, V_1, \ldots, V_k)$ be an equitable partition, where $|V_0| \leq \varepsilon n$ and $\varepsilon < \frac{1}{4}$. Suppose P is not ε-regular. Show that there exists a refinement P' of P, $P' = (V_0', V_1', \ldots, V_l')$, such that P' is equitable such that $|V_0'| \leq |V_0| + \frac{n}{2^k}$, $k \leq l \leq k4^k$, and $q(P') \geq q(P) + \frac{1}{2}\varepsilon^5$.

Hint: Partition the given sets V_i to get a new partition that satisfies the required conditions. Using the last problem we can prove the collegiate splitting theorem:

Problem 2.1.6. Without loss of generality assume that $\varepsilon < \frac{1}{4}$ and $2^{t-2} > \frac{1}{\varepsilon^6}$. Define $k_0 = t$ and $k_{i+1} = k_i 4^{k_i}$ for $i \geq 0$. Prove the collegiate splitting theorem for $T = k_s$, where $s = \lceil \frac{1}{\varepsilon^5} \rceil$.

With this statement proven, we have finished the proof of the collegiate splitting theorem. Through the next chapter, we will see how useful it can be in extremal graph theory. We will use similar ideas that we used to prove the Collegiate splitting theorem to solve another related problem in combinatorics.

To motivate the discussion in the next chapter consider the following scenario. The first two weeks at Princeton are the so called *orientation*, meant to introduce first year students to life at Princeton and to one another. There are many events organized with the goal of making people meet each other such as orientation trips, club and organization meetings, Lawnparites etc. Unfortunately, after this period ends, classes start and people get more busy with work.

Since the orientation week was fun and everyone met so many new friends, at some point each person will want to catch up with many other people. If there is a group of at least 3 people such that every two of them want to catch up, they will organise a party. However, loud music disturbs PUMaC problem writers. Because of this, the Office of the Dean of Undergraduate Students (ODUS) wants to prevent the parties. To this end, ODUS is giving out coupons for an ice cream shop meant to enable a pair of students who want to catch up with each other to do so over ice cream. ODUS can choose the two students, give them a coupon for ice cream, and after that the two students don't want to catch up any more. ODUS focus on preventing all the parties, hence they realize it is enough to prevent the parties of exactly 3 people.

Again we will model this problem with graphs, the vertices of the graph will be people and two vertices will be connected if and only if the people represented with these vertices want to catch up with each other. The number of coupons needed to prevent the parties can be bounded by the *coupon theorem*.

2.2 The Coupon theorem

We begin with a few definitions.

Definition 2.2.A. A *triangle* in a graph G is a set of three vertices in G such that any two of them are connected with an edge. To *destroy* a triangle means to remove *one of* the three edges forming a triangle from graph G.

Definition 2.2.B. We say that a graph is *triangle-free* if there is no subgraph that is a triangle.

Definition 2.2.C. We say that a graph G is ε-far from being triangle-free if one has to remove at least εn^2 edges from G to destroy all triangles.

Problem 2.2.1 (Coupon theorem)**.** Let $\varepsilon > 0$ be a real number. Prove that there exists $\delta = \delta(\varepsilon) > 0$, such that for every graph $G = (V, E)$ on n vertices, if G is ε-far from being triangle free, then G contains at least δn^3 triangles.

Hint: Take a regular partition of G. After taking a particular partition remove all edges incident to V_0, all edges inside sets V_i, all edges between irregular pairs and, all edges between pairs of sets of too small density. Look at how many edges were removed and how many triangles are left.

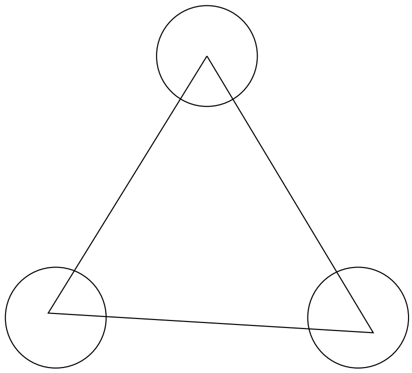


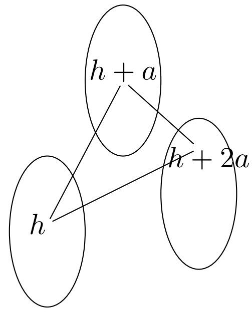
Figure 10: Removing a triangle between parts of a regular partition of G.

Now we will see a few problems that are applications of the collegiate partition theorem and/or the coupon theorem. A useful advice is to use one of the previous theorems in a clever way.

Two players, Alice and Bob, are playing two versions of a game.

Problem 2.2.2. In the first version of the game Alice goes first and she picks any number $p, 0 < p < 1$. Then Bob goes, and after hearing which number Alice picked, he picks a positive integer N, which defines Bob's set $X = \{1, 2, \ldots N\}$. Then Alice can remove at most $p|X|$ elements of X. Let the set Bob is left with after Alice removed the elements she wanted be Y . Alice wins if Y contains **no** three distinct elements such that one is the mean of the other two (elements need not be distinct). Otherwise Bob wins. Who has a winning strategy in the first version of the game and why?

Hint: Consider a graph whose vertex set are three copies of X and choose edges in a way such that you can use The Coupon theorem, i.e. consider the following setup:



Problem 2.2.3. In the second version of the game Bob and Alice decide to solve a problem. They want to prove that there exists a function $N(p)$ such that for every real number $0 < p < 1$ and any positive integer $N > N(p)$ the following holds: The set $X = \{1, \dots, N\}$ has a proper subset $Y \subset X$ such that Y contains no three distinct numbers such that one of them is the mean of the other two, and $|Y| > |X|^{1-p}$. Prove this statement.

Problem 2.2.4. Let $A \subset \{1, \dots, n\} \times \{1, \dots, n\}$. Define an (x, y)-shift of A as $A_{x,y} = \{(x, y) - a | a \in A\}$. Prove that there exists (x, y) such that $|A \cap A_{x,y}| \geq \frac{|A|^2}{(2n)^2}$.

Problem 2.2.5. Prove that for any $\varepsilon > 0$ there exists an integer $n_0 = n_0(\varepsilon)$ such that whenever $n > n_0$ and $A \subset \{1, \dots, n\} \times \{1, \dots, n\}$ with $|A| \geq \varepsilon n^2$ then A contains a corner; that is, three vertices of a square (x, y), $(x, y + d)$, $(x + d, y)$ for some integers $x, y, d \geq 1$. Is it true that the problem holds if we replace (x, y), $(x, y + d)$, $(x + d, y)$ with (x, y), $(x, y + 2d)$, $(x + 3d, y)$? What if we replace it with (x, y), $(x, y + 2d)$, $(x + 4d, y)$?

Problem 2.2.6. Prove or provide a counterexample to the following claim: For every $\varepsilon > 0$ there exists an integer n_0 such that for every $n > n_0$ and every subset $X \subset \{1, 2, \ldots, n\}$ with $|X| \geq \varepsilon n$, there exist *distinct* $a, b, c \in X$ such that

$$2019a + 2020b = 4039c.$$

2.3 The generalized coupon theorem

After some time has passed after the orientation, people became less interested in parties. Now students will organize a party only if there are k of them such that each pair wants to catch up. Now ODUS seeks to solve the generalized version of the previous problem.

Definition 2.3.A. An H-copy of a graph H in G is a subgraph of G that is isomorphic to H.

Two H copies are the same if and only if they are associated with the same set of vertices and set of edges.

Theorem 2.3.I (Generalized coupon theorem)**.** For every integer k and for every $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon, k) > 0$ such that for every graph H with k vertices and for every graph $G = (V, E)$ with $|V| = n$, if we have to remove more than εn^2 edges from G to destroy all H-copies, then the number of copies of H in G is at least δn^k.

This result is special: **it can only be used in problems after 2.3.3**.

Its proof is rather complicated in the general case, but in some special cases it is just a small variation of the proof of the triangle removal lemma. To understand the special case, we define the *chromatic number of a graph.*

Definition 2.3.B. The *chromatic number* of a graph G, denoted $\chi(G)$, is the minimal number of colors needed to color its vertices such that any two adjacent vertices are of different color.

For example, the chromatic number of any bipartite graph is 2.
The case $\chi(H) = 3$ can be proved using things we know so far. We can prove that the conclusion holds even if we assume that we have to remove $\geq \varepsilon n^2$ edges to destroy all triangles. We can reformulate the previous theorem differently, using the following definition.

Definition 2.3.C. A graph G is H*-free* if it doesn't contain H as a subgraph (not necessarily an induced subgraph).

For example, C_5 is C_3 free, but K_5 is not C_4 free.

Definition 2.3.D. For a fixed graph H, a graph $G = (V, E)$ with $|V| = n$ is ε*-far* from being H-free if one has to remove at least at least εn^2 edges to get an H-free subgraph.

Then the theorem claims that for every fixed H, if G is ε-far from being H-free, then G contains at least δn^k copies of H. But we can actually prove a stronger result.

Problem 2.3.1. If $H = K(h_1, h_2, h_3)$, meaning that H is a complete 3-partite graph with vertex classes of sizes h_1, h_2, h_3 (meaning that all edges have vertices in different vertex classes), with $h = h_1 + h_2 + h_3$, then if $G = (V, E)$, $|V| = n$, and G is ε-far from being triangle free then G contains $\geq \delta(\varepsilon, h)n^h$ copies of H.

Hint: Proceed similarly as in the coupon theorem.

Using this problem we will be able to prove a useful bound on the following quantity in a graph.

Definition 2.3.E. For a graph H, let

$$\text{ex}(n, H) = \text{maximum possible number of edges in an } H\text{-free graph on } n \text{ vertices.}$$

Problem 2.3.2. Prove that for a fixed H with $\chi(H) = 3$ we have that

$$\text{ex}(n, H) = \left(\frac{1}{4} + o(1)\right) n^2.$$

A very similar argument proves a bound for general $\chi(H) = k$,

$$\text{ex}(n, H) = \left(1 - \frac{1}{k-1} + o(1)\right) \binom{n}{2}.$$

This determines the asymptotic behavior of $\text{ex}(n, H)$ whenever $\chi(H) \geq 3$. But when $\chi(H) = 2$ we only have $\text{ex}(n, H) = o(n^2)$. In fact, in this case $\text{ex}(n, H) \leq n^{2-\varepsilon(H)}$ for some $\varepsilon(H) > 0$, and in many cases the best possible $\varepsilon(H)$ is not known.

In the next few problems we will see some applications of both the coupon theorem and the generalized coupon theorem.

As a reminder, here is the defintion of *triangle-free.*

Definition 2.3.F. We say that a graph is *triangle-free* if there is no subgraph that is a triangle.

Problem 2.3.3. a) Prove that the number of triangle-free graphs on a set of n labeled vertices is

1. exponential in n^2 (at least a^{cn^2} for some real $c > 0, a > 1$)
2. less than $2^{\frac{1}{2}n^2}$ for large enough n.

b) Let $T(n)$ be the number of triangle free graphs on n vertices. Prove that the limit $\lim_{n\to\infty} \frac{\log_2 T(n)}{n^2}$ exists and find its value.

Problem 2.3.4. Prove that for every $c > 0$ there is a $\delta = \delta(c) > 0$ so that every graph G with n vertices and at least cn^2 edges contains a copy of every bipartite graph H with at most δn vertices and maximum degree 3 (i.e. every such H is a subgraph of G).

Problem 2.3.5. Show that for every $\varepsilon > 0$ there is a $\delta = \delta(\varepsilon) > 0$ and $n_0(\varepsilon)$ such that if Princeton admits more than n students to their Class of 2024, and ODUS guarantees that orientation events will make at least εn^2 people want to catch up with each other, then ODUS will be able to guarantee they can select a subset of students in which every student wants to catch up with exactly k others for some $k \geq \delta n$.

3 Graph discount chances

3.1 Introduction

ODUS is now interested in finding which party sizes are possible. Hence, ODUS has to determine for which k there are no k students such that they all know each other. So, given the friendship graph of students, they need to see if the graph has a K_k as a subgraph. Thus we want to find an algorithm for determining whether a graph representing the friendship in the student body G has a given trait, which in the case of the ODUS is "is graph G K_k-free." This trait will be called a *discount chance* as it allows ODUS to save money by exploiting it. Since there are many students on campus, making the graph we consider relatively large, finding exact solution would take too much time. So we consider the relaxation of this: we want to find an efficient *randomized* algorithm that will distinguish with high probability whether a graph has a discount chance or is close to having it. This is only one example of a discount chance we might want to see if the graph has. So this motivates the study of *graph discount chances*, which we will formulate using the following definitions.

Definition 3.1.A. A *graph discount chances* is a family of graphs closed under isomorphism. Examples include being H-free, being k-colorable, etc.

Definition 3.1.B. For a graph discount chances P, a graph $G = (V, E)$ with $|V| = n$ is *ε-far from P* if one has to add or delete $\geq \varepsilon n^2$ edges to or from G to get a graph satisfying P.

We want to find an efficient (randomized) algorithm that distinguishes with high probability between G satisfying P and G that is ε-far from satisfying it.

Definition 3.1.C. An ε*-tester* is a (randomized) algorithm that knows n, and can access an input $G = (V, E)$ on n vertices by asking queries of the form "are the vertices i, j adjacent ?". The algorithm produces either an output "accepts" or "rejects" (corresponding to whether the algorithm thinks G satisfies P or not). The algorithm should distinguish between G satisfying P and G being ε-far from P by satisfying the following requirements:

1. for every G satisfying P, the tester accepts it with probability greater than 0.9. (If it accepts G with probability 1, we say it is a *one-sided tester.*)

2. for every G which is ε-far from P, the tester rejects it with probability greater than 0.9.

Remark: 0.9 can be replaced by any $0.9 \leq c < 1$ by repeating the algorithm and taking the results that appears the most often.

Definition 3.1.D. The *query complexity* of the tester is the maximum number of queries it makes on an input G (of n vertices).

Definition 3.1.E. A discount chance P *is testable* if the query complexity is bounded by a function of ε (independent of n), for every $\epsilon > 0$.

Here is one example: consider the discount chance P which is "G has no edges". A natural ε-tester is to choose randomly and uniformly some number of edges, say $\frac{10}{\varepsilon}$ pairs and check adjacency, and accept if there is no edge found. If G satisfies P, the tester always accepts (so that is one sided tester (defined above)). If G is ε-far from P, i.e. it has at least εn^2 edges,) the probability that the tester will accept is at most $(1 - 2\varepsilon)^{\frac{10}{\varepsilon}} \leq e^{-20}$. It is possible to construct other algorithms for testing different discount chances of graphs.

Problem 3.1.1. Prove that if graph G is ε-far from some discount chance P then for any H such that H is isomorphic to G, H is also ε-far from the discount chance P. Does the same hold if the graph H is only homomorphic to G?

Problem 3.1.2. Show, with proof, whether the following discount chances are testable for a graph G:

- G is connected
- G has a cycle
- G contains a k-clique
- G has a vertex of degree $|V(G)| - 1$

- G is 3-colorable, meaning that you can color vertices of G in three different colors such that any two adjacent vertices are of different color?

Next we will prove that the discount chance of being 3-colorable is testable. Take the ε-tester to be the following: choose randomly uniformly $\frac{100}{\varepsilon^2}$ vertices and accept if and only if the induced subgraph on them is 3-colorable. Here, the tester is also one-sided (if G is 3-colorable it will always accept). In order to prove correctness we have to show the following.

Theorem 3.1.I. If $G = (V, E)$ with $|V| = n$ is ε-far from 3-colorable then the induced subgraph on a random set of $\frac{100}{\varepsilon^2}$ vertices is with high probability *not* 3-colorable.

Remark: It is not obvious that such G contains a subgraph on $\geq f(\varepsilon)$ vertices which is not 3-colorable.

Definition 3.1.F. Call a graph discount chance P *easily testable* if there exists ε-tester for P with query complexity that is bounded by a polynomial of $\frac{1}{\varepsilon}$.

Some examples are having no edges and being 3-colorable. Example of a testable discount chance that's not easily testable is being triangle free.

To prove that being traingle-free is not easily testable we will need the following theorem.

Theorem 3.1.II. Let P be the discount chances of a graph $G = (V, E)$ be that G has partition $V = V_1 \cup V_2$, $|V_1| = |V_2| = \frac{n}{2}$ and has no edges between V_1, V_2, and induced subgraphs on V_1, V_2 are isomorphic. Then P is not testable.

This result is special: **it cannot be cited without proof in Problem 3.2.1.** This is informally saying that for every $c > 0$, there exists n such that with less than $c\sqrt{n}$ queries we can't distinguish between two random graphs on $\{1, 2, \dots, \frac{n}{2}\}$ and $\{\frac{n}{2}+1, \dots, n\}$ and two copies of the same random graph. Using the previous theorem we can prove the following:

Problem 3.1.3. Show that being triangle-free is testable.

This can be generalized for arbitrary H.

Theorem 3.1.III. Being H-free is testable.

Next we will see what happens when being H-free is easily testable.

To-be-proved 3.1.1. Being H-free is easily testable if and only if H is bipartite.

To prove this we will a few things.

Problem 3.1.4. Let $G=(V,E)$ be a graph with n vertices and at least εn^2 edges. Show that there exist functions $c(H)$ and $n(H)$ such that for every bipartite graph H with h vertices, if $n > n(H)$ then G contains at least $\lfloor c(H)\varepsilon^{\frac{h^2}{4}} n^h \rfloor$ copies of H. In fact, if $H = K_{s,t}$, if we take randomly uniformly $s+t$ ordered vertices in G (and have $n > n_0(s,t)$) then probability that we get a copy of $K_{s,t}$ on these $s+t$ vertices is $\geq (1-o(1))(2\varepsilon)^{st}$.

Note: $c(H)$ means that the constant c may depend on H. Likewise, $n(H)$ means that n may need to be bigger than some value determined by H.

Hint: Look at the number of homomorphisms between graph G and a particular graph that is useful.

Now we look at a useful lemma for the next thing that will help us with the next problem.

Lemma 3.1.1. For every integer m there exists a subset X of $\{1,\dots,m\}$ such that $|X| \geq \frac{m}{2^{10\sqrt{\log m}}}$ and the following holds: if $x_1+x_2+x_3+x_4 = 4x_5$ for $x_i \in X$ then $x_1 = \cdots = x_5$.

Problem 3.1.5. Show that for every large n there exists $G=(V,E)$ which is ε-far from being C_5-free and contains no more than $\varepsilon^{c\log\frac{1}{\varepsilon}} n^5$ copies of C_5, where $c > 0$ is some constant.

Hint: Start from a $5-\textit{partite}$ graph and look at its r-blowup

This can be generalized for any non-bipartite H on h vertices as follows.

Theorem 3.1.IV. For every non-bipartite graph H with h vertices and for every $\varepsilon > 0$, there is $n_0 = n_0(\varepsilon, n)$ such that there exists $G=(V,E), |V| = n$, such that G is ε-far from being H-free and contains less than $\varepsilon^{c_H \log\frac{1}{\varepsilon}} n^h$ copies of H.

For general non-bipartite graph H the core of any non-bipartite graph H we start with a construction of an appropriate variant of G' and show that it contains a small number of copies of core of H.

Now you can prove:

Problem 3.1.6. Being H-free is easily testable if and only if H is bipartite.

Note that if a graph is H-free, then deleting edges preserves the property of a graph being H-free.

3.2 Connection between graph discount chances testing and the Collegiate partition theorem

While many discount chances of graph are indeed testable, there are some that are not.

Problem 3.2.1. Find a discount chance of a graph G that is not testable.

A natural question arises: which graph discount chances are testable and which are not? While examining discount chance by discount chance we can in many cases determine whether a discount chance is testable or not, there are some cases in which it's still an open problem whether a discount chance is testable or not. In order to answer the previous question, a natural thing to do is to somehow characterize all discount chance that are testable. The relaxation of this question is finding the characterization of the discount chances that are testable in constant number of queries. This characterization surprisingly turns out to be purely combinatorial. The combinatorial structure is supplied via the Collegiate partition theorem, i.e. the discount chance that proves crucial is that graph G contains any Collegiate-like partition. To state the theorem we will need the following definition:

An equitable parition is said to be an *equipartition*.

Definition 3.2.A (ε-regular equipartition)**.** An equipartition $E = \{V_i | 1 \leq i \leq k\}$ of the vertex set of a graph is called ε-regular if all but at most $\varepsilon\binom{k}{2}$ of the pairs (V_i, V_j) are ε-regular.

An equivalent of the collegiate partition theorem can be proven for ε-regular equipartitions.

Definition 3.2.B. A *regularity-instance* R is given by error parameter ε, an integer k, a set of $\binom{k}{2}$ densities $0 \leq \eta_{i,j} \leq 1$ indexed by $1 \leq i \leq j \leq k$, and a set $\bar{R}$ of pairs (i, j) of size at most $\varepsilon\binom{k}{2}$. A graph is said to satisfy the regularity-instance if it has an ε-regular equipartition $E = \{V_i | 1 \leq i \leq k\}$ partition such that for all $(i, j) \notin \bar{R}$ the pair (V_i, V_j) is ε-regular and satisfies $d(V_i, V_j) = \eta_{i,j}$. The *complexity* of regularity-instance is $max(k, \frac{1}{\varepsilon})$

We now have the following theorem for a testing regularity-instances:

Theorem 3.2.I. For any regularity-instance R, the discount chance of satsifying R is testable.

And the following crucial definition:

Definition 3.2.C. A graph discount chances P is *regular-reducible* if for any $\delta > 0$ there exists and integer $r = r_P(\delta)$ such that for any n there is a family $\mathcal{R}$ of at most r regularity-instances each of complexity at most r, such that the following holds for any $\varepsilon > 0$ and any graph on n vertices:

1. If G satisfies P, then for some $R \in \mathcal{R}$, G is δ-close to satisfying R.

2. If G is ε-far from satisfying P, then for any $R \in \mathcal{R}$, G is $\varepsilon - \delta$-far from satsifying R.

We say that a graph G is δ-close to satisfying R if we have to remove or add $< \delta n^2$ edges to or from G to get a graph satisfying G.

Now the main result can be stated as follows:

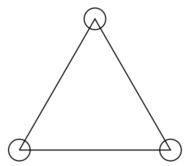
Theorem 3.2.II. A graph discount chance is testable if and only if it is regular-reducible.

Note that many discount chances we have seen so far are regular-reducible. For example it can be shown that the following holds

Problem 3.2.2. Prove, without using theorem 3.2.II that being H-free is regular-reducible.

This means that all the discount chances that we have seen such as being triangle-free or H-free are regular reducible hence testable.

This result connects two seemingly unrelated areas in combinatorics.



Team Number: ______________________________

PUMaC 2019 Power Round Cover Sheet

Remember that this sheet comes first in your stapled solutions. You should submit solutions for the problems in increasing order. Write on one side of the page only. The start of a solution to a problem should start on a new page. Please mark which questions for which you submitted a solution to help us keep track of your solutions.

Problem Number	Points	Attempted?
1.4.1	5	
1.4.2	5	
1.4.3	5	
1.4.4	10	
1.5.1	5	
1.5.2	5	
1.5.3	10	
1.5.4	5	
1.5.5	20	
1.5.6	10	
1.5.7	30	
2.1.1	20	
2.1.2	10	
2.1.3	10	
2.1.4	30	
2.1.5	50	
2.1.6	15	
2.2.1	50	
2.2.2	60	
2.2.3	60	
2.2.4	30	
2.2.5	30	
2.2.6	40	
2.3.1	50	
2.3.2	50	
2.3.3	75	
2.3.4	60	
2.3.5	75	

Problem Number	Points	Attempted?
3.1.1	10	
3.1.2	15	
3.1.3	30	
3.1.4	40	
3.1.5	15	
3.1.6	50	
3.2.1	25	
3.2.2	15	