Probabilistic Method

18.204 Fall 2020 term paper

Zhezheng Luo
ezzluo@mit.edu

Pachara Sawettamalya
pachara@mit.edu

Yiwei Zhu
yiweizhu@mit.edu

December 10, 2020

1 Introduction

To show the existence of a combinatorial object, there are two conventional approaches: the constrictive and non-constructive method. Constructive methods show the existence of the object by giving an actual construction, e.g. the proof of the four-color theorem [AH76], the construction of a large independent set [Car79; Wei81], etc. On the other hand, non-constructive methods use external tools to prove existence without giving a concrete example. For example, the Pigeonhole Principle shows that there must be two people sharing the same birthday among a group of 367 people. In this survey, we introduce the probabilistic method, which is a non-constructive method pioneered by Paul Erdős.

The key idea of the probabilistic method is that in order to show some combinatorial object $\mathcal{C}$ with a desired property $\mathcal{P}$ exists, we can show that under an appropriate randomized process $\mathcal{R}$ for constructing $\mathcal{C}$, the property $\mathcal{P}$ is satisfied with positive probability. In other words, $\Pr(\mathcal{C}$ constructed by process $\mathcal{R}$ has property $\mathcal{P}) > 0$ guarantees that there exists a combinatorial object $\mathcal{C}$ that has property $\mathcal{P}$. An intuitive explanation to the validity of the approach is to think of the probability that $\mathcal{P}$ occurs as the ratio of the number of objects with the desired property $\mathcal{P}$ over the total number of randomized constructions. If this ratio is positive, one is guaranteed that there must be at least one construction that exhibits property $\mathcal{P}$.

In this survey, we first introduce the common tools and techniques that are frequently used in the probabilistic method. In Sections 3 to 8, we introduce the Union Bound, the Moment Method, the Chernoff Bound, and the Local Lovász Lemma. These probabilistic analysis tools are powerful in proving probabilistic bounds in combinatorial problems, which enables the probabilistic method to have a wide range of applications. We explore some of its applications to hypergraph coloring, Ramsey Theory, the crossing number, etc.

One of the main advantages of the probabilistic method is that it can prove the existence of a combinatorial object that is computationally hard to find, such as a large clique, a large cut, or a proper coloring of a graph. Therefore, following the existence proof, the question of whether it is possible to find such a construction within reasonable time complexity arises. It turns out that oftentimes, under some additional constraints or looser conditions, we can indeed give an efficient algorithmic procedure, typically within polynomial time, to find such a construction. In Sections 9 to 11, we discuss such constructive procedure.

2 Probability Review

We start with a short introduction to the very basics of probabilistic concepts. We expect the readers to be familiar with these concepts, and will not provide rigorous proofs for these concepts in this survey.

Definition 1. *A* probability space *is a triple* $(\Omega, \Sigma, \Pr)$ *where*

1. Ω *is a non-empty set, which represents all possible outcomes;*
2. Σ *is the* σ*-algebra of* Ω*, that is, a collection of subsets of* Ω *that is closed undertaking complements and countable unions and intersections;*
3. $\Pr$ *is a function mapping* $\Sigma \to [0,1]$ *with* $\Pr(\Omega) = 1$. $\Pr$ *is countably additive i.e. for any two disjoint events* $\sigma_1, \sigma_2 \in \Sigma$*, we have* $\Pr(\sigma_1) + \Pr(\sigma_2) = \Pr(\sigma_1 \cup \sigma_2)$*. Specially, for each outcome* $\omega \in \Omega$*, we define* $\Pr(\omega) = \Pr(\{\omega\})$.

In this survey, we focus on *finite probability space*, which means that Ω is finite, and $\Sigma = 2^\Omega$ is the collection of all subsets of Ω. An element of Σ is called an *event*. For an event σ, $\Pr(\sigma) = \sum_{\omega \in \sigma} \Pr(\omega)$ represents the probability of σ.

Following such definition, we notice that,

$$\Pr(\sigma_1 \cup \sigma_2) = \Pr(\sigma_1) + \Pr(\sigma_2) - \Pr(\sigma_1 \cap \sigma_2) \leq \Pr(\sigma_1) + \Pr(\sigma_2), \tag{1}$$

and in fact, more generally,

Proposition 2. *(Subaddidivity of probabilities, the Union Bound)*

$$\Pr(\sigma_1 \cup \sigma_2 ... \cup \sigma_n) \leq \sum_{i \in [n]} \Pr(\sigma_i). \tag{2}$$

The formal proof of this proposition will be given shortly in Section 3. Note that the Union Bound requires no independence among events.

Definition 3. *(Conditional probability) Given events* A *and* B*, the event* $A \mid B$ *is the event that* A *occurs given that* B *have already occurred. Furthermore, it follows that*

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} \tag{3}$$

Definition 4. *Two events* $\sigma_1, \sigma_2 \in \Sigma$ *are* independent *if* $\Pr(\sigma_1 \cap \sigma_2) = \Pr(\sigma_1) \cdot \Pr(\sigma_2)$.

Definition 5. *(Random variables) Given a probability space with sample space* Ω*, A random variable* X *is a measurable function* $X : \Omega \to S$ *from the sample space* Ω *to another measurable space* S.

Definition 6. *For an event* E*, the associated* indicator variable I_σ *is defined by*

$$I_E(\omega) = \begin{cases} 1 & if \omega \in \sigma \\ 0 & if \omega \notin \sigma. \end{cases} \tag{4}$$

Definition 7. *In a finite probability space, the* expectation *of a random variable X is*

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

Proposition 8. *(Linearity of expectation) For any two random variables X and Y, real numbers a and b,*

$$\mathbb{E}(a \cdot X + b \cdot Y) = a \cdot \mathbb{E}(X) + b \cdot \mathbb{E}(Y). \tag{5}$$

Note that, for any indicator variable I_σ, by definition, $\mathbb{E}(I_\sigma) = \Pr(\sigma)$. Often times, we calculate the expected value of a random variable X by expressing it as a sum of indicator variables, that is $X = \sum_{i=1}^{n} I_{\sigma_i}$. Then by linearity of expectation,

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{E}(I_{\sigma_i}) = \sum_{i=1}^{n} \Pr(\sigma_i) \tag{6}$$

Throughout the survey, we will present several applications of the probabilistic method to hypergraph coloring. Therefore, we now give the definition of hypergraph.

Definition 9. *A* hypergraph *is a pair (V, E) where*

- *V is a set, whose elements will be called* hypernodes *(sometimes referred as* nodes *or* vertices*)*
- *E is a set, whose elements will be called* hyperedges *(sometimes referred as* edges*.) Each hyperedges is a collection of hypernodes in V.*

Furthermore, we call a hypergraph k-uniform if and only of every of its hyperedge consists of exactly k hypernodes.

3 Union Bound

Oftentimes when we want to show an existence of a property (e.g. a proper coloring, a perfect matching, etc.), we can formulate the desired property as an intersection of events. For example, the proper coloring of a graph $G = (V, E)$ can be viewed as the intersection of $|E|$ events – each of which corresponds to the event that an edge is proper colored. The perfect matching of a graph $G = (V, E)$ can also be viewed as the intersection of $|V|$ events – each of them corresponds to the event that a vertex is involved in exactly 1 edge of the matching.

To show an existence of a property, we can define a randomized process and show that the property is satisfied with positive probability. And since we can view the property as an intersection of events, we can show instead that these events occur simultaneously with positive probability. One way to show this is to use probabilistic bound tools which allow us to bound the probability of events in different settings.

Perhaps one of the most straightforward probabilistic bound tools is the Subadditivity of Probabilities, as known as the Union Bound (Proposition 2). For any set of events, the union bound says that the probability that at least one event occurs is upper bounded by the sum of the probability of each individual event.

Proposition 2. *(Subaddidivity of probabilities, the Union Bound)*

$$\Pr(\sigma_1 \cup \sigma_2 ... \cup \sigma_n) \leq \sum_{i \in [n]} \Pr(\sigma_i). \tag{2}$$

Proof. We can prove the union bound by induction. For the case $n = 1$, it is clear that $\Pr(\sigma_1) \leq \Pr(\sigma_1)$. Then for $n = 2, 3, \cdots$, say we have already proved the union bound holds for smaller n, given that

$$\Pr(\sigma_1 \cup \sigma_2) = \Pr(\sigma_1) + \Pr(\sigma_2) - \Pr(\sigma_1 \cap \sigma_2) \tag{7}$$

we have

$$\Pr\left(\bigcup_{i=1}^{n} \sigma_i\right) = \Pr\left(\bigcup_{i=1}^{n-1} \sigma_i\right) + \Pr(\sigma_n) - \Pr\left(\sigma_n \cap \bigcup_{i=1}^{n-1} \sigma_i\right) \tag{8}$$

$$\leq \Pr\left(\bigcup_{i=1}^{n-1} \sigma_i\right) + \Pr(\sigma_n) \tag{9}$$

$$\leq \left(\sum_{i=1}^{n-1} \Pr(\sigma_i)\right) + \Pr(\sigma_n) \tag{10}$$

$$= \sum_{i=1}^{n} \Pr(\sigma_i) \tag{11}$$

so the union bound holds for any positive n. □

Corollary 10. *For any n "bad events" $\sigma_1, \cdots, \sigma_n$, if the sum of their probability is strictly less than 1, then with positive probability all of them can be avoided.*

Proof. For events $\sigma_1, \cdots, \sigma_n$, if $\sum_{i=1}^{n} \Pr(\sigma_i) < 1$, by the union bound, we have $\Pr(\bigcup_{i=1}^{n} \sigma_i) \leq \sum_{i=1}^{n} \Pr(\sigma_i) < 1$, and by taking complement we have $\Pr(\bigcap_{i=1}^{n} \overline{\sigma_i}) > 0$. □

In practical, the bad events usually have some underlying dependency which is difficult to observe, let alone compute the probability. The main advantage of the union bound is that it does not concern the dependency of bad events; however, there is a trade-off. The Union Bound is in fact a *weak* bound, meaning that it often does not give the best result. Using more complicated bounds or more analytical skills often lead to stronger results. Nevertheless, the Union Bound be used to derive some decent results as we will discuss shortly.

3.1 Hypergraph 2-Coloring

Recall the definition of hypergraphs in Definition 9. The Hypergraph 2-Coloring Problem asks that, given a hypergraph, whether it is possible to color each hypernode with Red or Blue which results in no monochromatic hyperedge.

The problem of deciding whether a k-uniform hyperpergraph is 2-colorable has been heavily researched, and is known to be NP-Complete [Lov73]. However, under certain scenarios, we can

use probabilistic method to conclude 2-colorability without knowing much about the underling structure of a hypergraph. In particular, the first example we give in this section is that if a uniform hypergraph does not have too many edges, it is always 2-colorable.

Theorem 11. *Let k be a positive integer and H be a $k-$uniform hypergraph with m edges. If $m < 2^{k-1}$, then H is 2-colorable. [Erd47]*

Proof. Let H's hyperedges be $e_1, e_2, ..., e_m$. Define a randomized process as follows. For every hypernode of H, color it independently and uniformly with Red or Blue – with probability $1/2$ each. In other words, we can think of this process as independently tossing a fair coin once for each hypernode. If the coin lands Heads, color the node Red and if the coin lands Tails, color the node Blue.

For each $i \in [m]$, denote M_i to be the *bad* event where edge e_i is monochromatic (being all red or all blue). Notice that $\Pr(M_i) = 2^{1-k}$ because the probability of e_i being all red or all blue is 2^{-k} each.

Now we apply the Union Bound,

$$\Pr\left(\bigcup_{i\in[m]} M_i\right) \leq \sum_{i\in[m]} \Pr(M_i) = m \cdot 2^{1-k} < 1$$

which is equivalent to

$$\Pr\left(\bigcap_{i\in[m]} \overline{M_i}\right) = 1 - \Pr\left(\bigcup_{i\in[m]} M_i\right) > 0.$$

Therefore, with positive probability, none of the bad events M_i happens i.e. none of the edge is monochromatic. This implies that there is an color assignment that creates no bad events, i.e., all edge is not monochromatic. This is equivalent to saying that H is 2-colorable. □

3.2 Ramsey Number

Another topic that we will explore throughout the survey is the Ramsey Number.

Definition 12. *For any positive integers m, n, the* Ramsey Number *of (m, n), denoted by $R(m, n)$, is the smallest positive integer k such that every red-blue edge-coloring of K_k is guaranteed a red K_m subgraph or a blue K_n subgraph.*

Example 13. *R(3,3) = 6.*

Proof. First of all, we notice that the coloring of K_5 shown in figure 1 does not contain any monochromatic K_3 subgraph.

Next, we need to show that any red-blue edge coloring of K_6 contains a monochromatic K_3. Let's say the K_6 has vertices $\{A, B, C, D, E, F\}$. Consider all five edges involving A. By Pigeonhole Principle, at least three edges must have the same color. Without loss of generality, we can assume AB, AC, AD are red. If at least one of BC, CD, DB is red, this will create a red K_3. On the other hand, if all BC, CD, DB are blue, this will create a blue K_3. In any case, a monochromatic K_3 is guaranteed. See figure 1 as coloring examples. □

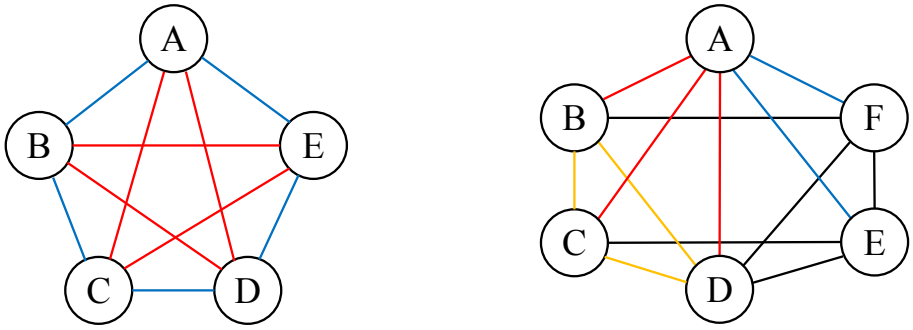

Figure 1: Shown on the left is a coloring of K_5 avoiding any monochromatic K_3 subgraph, and shown on the right is the demostration of our proof on K_6. Note that when AB, AC, AD are all red, any red coloring of BC, CD, DB creates a monochromatic K_3 but coloring them all blue also creates a monochromatic K_3.

The close form of $R(m,n)$ has not been exactly determined. In fact, even the small case of $R(5,5)$ and $R(6,6)$ is not known at the moment. For this reason, mathematicians now focus on bounding the Ramsey Number instead of trying to determine the exact value. It turns out that with a very simple tool like the Union Bound, we can achieve a decent lower bound to the Ramsey Number.

Theorem 14 (Ramsey, [Spe75])**.** *Let n, k be positive integers such that $\binom{n}{k} \cdot 2^{1-\binom{k}{2}} < 1$. Then $R(k,k) > n$.*

Proof. To show that $R(k,k) > n$, it is enough to show that there exists a Red-Blue edge-coloring of K_n for which there is no monochromatic subgraph K_k.

Define a randomized process as follows. We independently and uniformly color each edge with Red or Blue – with probability 1/2 each. For any particular K_k in G, the probability of it being monochromatic is $2^{1-\binom{k}{2}}$ because the probability of all $\binom{k}{2}$ edges being all red or all blue is $2^{\binom{k}{2}}$ each. Furthermore, there are exactly $\binom{n}{k}$ distinct K_k in G. The Union Bound implies that the probability of at least one of K_k is monochromatic is at most $\binom{n}{k} \cdot 2^{1-\binom{k}{2}} < 1$. This implies that the probability of having no monochromatic K_k is > 0, which means that there exists a Red-Blue edge-coloring of K_n with no monochromatic K_k. This proves $R(k,k) > n$. □

With Theorem 14, we can determine a lower bound of $R(k,k)$ by fixing k and determining a positive integer n such that $\binom{n}{k} \cdot 2^{1-\binom{k}{2}} < 1$. For example, if $k = 3$, we can determine that $n = 3$, yielding lower bound $R(k,k) > 3$. Similarly, if $k = 4$, we can determine $n = 6$, yielding the lower bound $R(4,4) > 6$.

3.3 Hitting Set

Definition 15. *Given sets $S_1, S_2, ..., S_m$. We call a set H a hitting set if $H \cap S_i \neq \phi$ for every $i \in [m]$. We say that H hits every S_i.*

With the above setting, we shall consider the scenario that the universe is $[n]$. In other words, we assume that $S_1, S_2, ..., S_m$ are k-element subsets of $[n]$ and the hitting set is also a subset of $[n]$.

Without any prior knowledge, it is trivial that a set $H_1 = [n]$ of size n is a hitting set because each S_i is a subset of $H_1 = [n]$. Another less trivial hitting set is any arbitrary $H_2 \subseteq [n]$ with size $n - k + 1$. This is because for any set S_i, we have $|S_i| + |H_2| = n + 1 > n$. By the Pigeonhole Principle, S_i and H_2 must share an element. Finally, another example of a hitting set is H_3 of size m which consists of one element from each of $S_1, ..., S_m$.

The key observation is that it is easy to find a large hitting set such as H_1, H_2, H_3 as discussed a moment earlier. However, the difficult problem is whether it exists a small hitting set. Using the union bound, we can give an upper bound to the size of the smallest hitting set in the logarithmic of m.

Theorem 16. *Given $k-$element sets $S_1, S_2, ..., S_m \subseteq [n]$. There exists $H \subseteq [n]$ such that $H \cap S_i \neq \phi$ for every $i \in [m]$ and $|H| \leq \lceil \frac{n \log m}{k} \rceil$.*

Proof. Let $l = \lceil \frac{n \log m}{k} \rceil$. Construct a multiset $H \subseteq [n]$ with l elements by the following randomized process: picking l elements one by one uniformly and independently from $[n]$ with repetition.

For each $j \in [m]$, denote B_j to be the event that $H \cap S_j = \phi$. Notice that each element of H is not in B_j with probability $1 - \frac{|S_j|}{n}$. It follows that

$$\Pr(B_j) = \Pr\left(H \cap S_j = \phi\right) = \left(1 - \frac{k}{n}\right)^l < e^{-kl/n} \leq e^{-\log m} = \frac{1}{m}$$

where we use the fact that $(1 - \frac{1}{x})^x < e^{-1}$. By the Union Bound, we have

$$\Pr\left(\bigcup_{j \in [m]} B_j\right) \leq \sum_{j \in [m]} \Pr(B_j) < m \cdot \frac{1}{m} = 1$$

It follows that with positive probability, none of the bad B_j occur which corresponds to $H \cap S_j \neq \phi$ for every $j \in [m]$. Thus, there exists a construction of sized-l multiset H that hits every S_j's. We can remove the repeating elements of H to obtain a hitting set of size at most l. □

We will circle back to the hitting set problem in the Section 9, where we give a procedure of finding the actual hitting that is smaller than $\lceil \frac{n \log m}{k} \rceil$.

4 The First Moment Method

The moment method uses *moments* (i.e. expected values of powers) of a random variable to bound the probability that the random variable is significantly more, or less, than its expected value. In other words, it bridges the expected value and the probability of a random variable. Often times, the expected value of a random variable is easier to calculate given the linearity of expectation, which, unlike probability, exerts no restriction on the dependence relationship between events.

For the purpose of this survey, we are only concerned with finite probability space, which is usually the context for application of the probabilistic methods in combinatorial problems. In this section, we introduce the First Moment Method, which is a simple, yet powerful and fundamental tool of the probabilistic method that concerns $\mathbb{E}(X)$. And in the next section, we introduce the Second Moment Method, which concerns second-order moments $\mathbb{E}(X^2)$.

Theorem 17. *(First Moment Principle) Let X be a random variable with finite expectation. Then for any $\mu \in \mathbb{R}$, if $\mathbb{E}(X) \leq \mu$, then $Pr(X \leq \mu) > 0$.*

Proof. As mentioned, in this survey we only prove this result for finite probability space. Let S denote a finite set of all possible values of X. By definition, $\mathbb{E}(X) = \sum_{i \in S} i \times \Pr(X = i)$. In other words, $\mathbb{E}(X)$ can be expressed as the weighted average of all possible values of X.

We proceed by contradiction. We assume, for the contrary, $\mathbb{E}(X) \leq \mu$ and $\Pr(X \leq \mu) = 0$. That is, X only takes on values strictly greater than μ. Then, it follows that

$$\mathbb{E}(X) = \sum_{i \in S} i \cdot \Pr(X = i) = \sum_{i > \mu} i \cdot \Pr(X = i) > \sum_{i > \mu} \mu \cdot \Pr(X = i) = \mu \sum_{i > \mu} \Pr(X = i) = \mu, \quad (12)$$

which contradicts our assumption. □

We have proved that X takes some value that is at most $\mathbb{E}(X)$ with positive probability, and symmetrically X also takes some value that is at least $\mathbb{E}(X)$ with positive probability. Using a similar proof idea, one can show the following three corollaries.

Corollary 18. *If $\mathbb{E}(X) \geq \mu$, then $\Pr(X \geq \mu) > 0$.*

Corollary 19. *If $\mathbb{E}(X) < \mu$, then $\Pr(X < \mu) > 0$.*

Corollary 20. *If $\mathbb{E}(X) > \mu$, then $\Pr(X < \mu) > 0$.*

The key result from the First Moment Method is that: if we show that the expected value of a random variable is small or large, then the First Moment Method guarantees that there exists an object in the probability space on which this variable is small or large. This is very useful to help us prove existence when using probabilistic methods.

Sometimes, in addition to the existence of an object, we also want to know how unlikely is it that the random variable significantly exceeds its expected value:

Theorem 21. *(Markov's Inequality) Let X be a non-negative random variable and $a > 0$, then*

$$\Pr(X \geq a) \leq \frac{\mathbb{E}(X)}{a}. \quad (13)$$

Proof. For the purpose of this survey, we only prove this result for finite probability space. Let S denote a finite set of all possible values of X. Combining the definition of $\mathbb{E}(X)$ and the fact that X is non-negative,

$$\mathbb{E}(X) = \sum_{i \in S} i \cdot \Pr(X = i) \geq \sum_{i \geq a} i \cdot \Pr(X = i) \geq a \cdot \Pr(X \geq a). \quad (14)$$

□

Notice that Markov's inequality specifically concerns non-negative random variables, which is often the case in the combinatorial context, especially when we want to count the number of certain structures.

Next, we present two applications of the First Moment Method in the hypergraph 2-coloring problem and the crossing number problem to showcase its power in probabilistic methods.

4.1 Hypergraph 2-Coloring

We revisit the hypergraph 2-coloring problem introduced in Section 3.1. Recall Equation 6, where we expressed the expected value of a random variable as the sum of (often simpler to calculate) indicator random variables. This is particularly powerful because it avoids dealing with the dependence relationship between events, and allows us to apply the First Moment Method. To illustrate this idea, we now provide an alternative proof of Theorem 11 using the First Moment Method and the linearity of expectation.

Theorem 11. *Let k be a positive integer and H be a $k-$uniform hypergraph with m edges. If $m < 2^{k-1}$, then H is 2-colorable. [Erd47]*

Proof. (Alternative proof of Theorem 11) Let H's hyperedges be $e_1, e_2, ..., e_m$. Consider the same randomized process as before: for every hypernode of H, color it independently and uniformly with Red or Blue – with probability 1/2 each.

For each hyperedge e_i, denote by I_i as the indicator variable of the event that e_i is monochromatic, i.e. $I_i = 1$ if e_i is monochromatic, and 0 otherwise. Recall that $\Pr(e_i \text{ is monochromatic}) = \Pr(I_i = 1) = 2^{1-k}$, then it follows that $\mathbb{E}(I_i) = 1 \times \Pr(I_i = 1) + 0 \times \Pr(I_i = 0) = 2^{1-k}$.

Let $I = \sum_{i=1}^{m} I_i$ denote the total number of monochromatic edges in H. By linearity of expectation and given that $m < 2^{k-1}$, $\mathbb{E}(I) = \sum_{i=1}^{m} \mathbb{E}(I_i) = m \times 2^{1-k} < 1$. By Corollary 19 and the fact that I is a non-negative integer, $\Pr(I < 1) = \Pr(I = 0) > 0$.

Therefore, with positive probability, H has no monochromatic hyperedge, i.e., all edges are 2-colored. This is equivalent to saying that H is 2-colarable. □

4.2 Crossing Number

Another application of the probabilistic method is a remarkably simple proof of a lower bound on the crossing number (as we define below) given by Alon [AS16]. This lower bound leads to yield very simple proofs for some theorems in incidence geometry.

Definition 22. *(Immersion, Crossing number, Planner graph) An* immersion *of a graph $G = (V, E)$ into the plane $\mathbb{R}^2$ is a representation of G on the plane in which points are associated with vertices V and simple arcs are associated with edges s.t.*

- *The endpoints of the arc associated with edge (u, v) are the points associated with u and v.*
- *No arcs include points associated with other vertices.*

The crossing number of an immersion of a graph *is the number of crossings in this immersion. That is, the number of pairs of arcs that intersect at non-endpoints.*

The crossing number of G, *denoted by $cr(G)$, is the minimal crossing number over all possible immersion of G into $\mathbb{R}^2$.*

A planar graph *is a graph G such that $cr(G) = 0$.*

Some illustrative examples of the definitions above can be found in Figure 2.

Theorem 23. *Let G be a simple graph that $|E| \geq 4|V|$. Then cr(G)$\geq \frac{|E|^3}{64|V|^2}$.*

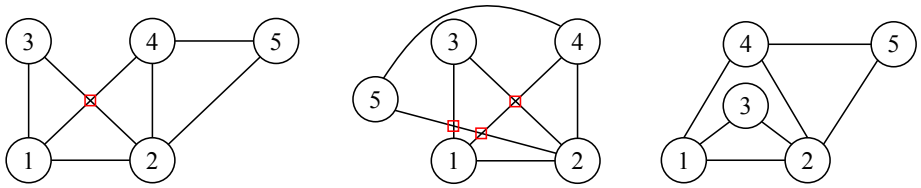

Figure 2: Three different immersions of the same graph G. The first immersion has the crossing number 1 while the second immersion has the crossing number 3 (crossings are highlighted with red squares). The third immersion has no crossings, so $cr(G) = 0$ and G is a planar graph.

Proof. First, we give a trivial lower bound of $cr(G)$. Consider an immersion of G. If for each crossing where e_1 and e_2 intersect, we remove e_1 or e_2 from the immersion, then the crossing is removed. Eventually, we are left with a planar graph where the number of edges is at most $|E| - cr(G)$. Since this is a planar graph, by Euler's formula, $|E| - cr(G) \leq 3|V|$. Therefore, we reach a trivial lower bound that

$$cr(G) \geq |E| - 3|V|. \tag{15}$$

Consider an immersion $\tilde{G}$ of G with $cr(G)$ crossings. Define a randomized process as follows. For every vertex of G, select it uniformly and independently with probability p into S, which is a random subset of V. We will decide the value of p later. Let $G\langle S\rangle$ denote the subgraph of G formed by S, and for simplicity, let $H = (V_H, E_H)$ denote $G\langle S\rangle$, $\tilde{H}$ denote $\tilde{G}\langle S\rangle$. Let $V_{\tilde{H}}, E_{\tilde{H}}, c$ denote the vertices, the edges, and the number of crossings in $\tilde{H}$ respectively.

Applying the trivial bound in Equation 15 and by definition of crossing number, we know

$$c \geq cr(H) \geq |E_{\tilde{H}}| - 3|V_{\tilde{H}}|. \tag{16}$$

Notice that $\mathbb{E}(|V_{\tilde{H}}|) = p \cdot |V|$, $\mathbb{E}(|E_{\tilde{H}}|) = p^2 \cdot |E|$, $\mathbb{E}(c) = p^4 \cdot cr(G)$. This is because, for example, to calculate $\mathbb{E}(|E_{\tilde{H}}|)$, for each edge $e = (u, v) \in E$, let I_e be the indicator variable for the event $e \in E_H$. Then, by linearity of expectation and the fact that each vertex is chosen independently,

$$\mathbb{E}(I_e) = \Pr(u \in S) \cdot \Pr(v \in S) = p^2.$$

Adding indicator variables together across all edges, we have

$$\mathbb{E}(|E_{\tilde{H}}|) = \sum_{e \in E(G)} \mathbb{E}(I_e) = p^2|E|.$$

Similarly, to calculate $\mathbb{E}(c)$, we know that for a crossing to remain in $\tilde{H}$, all 4 vertices of the 2 edges that intersect at the crossings need to be selected. Thus, $\mathbb{E}(c) = p^4 \cdot cr(G)$.

Now, Equation 16 can be expressed as $p^4 \cdot cr(G) \geq p^2 \cdot |E| - 3p \cdot |V|$. Let $p = \frac{4|V|}{|E|}$. By the given condition that $|E| \geq 4|V|$, we know $p \in (0, 1]$ is indeed a valid probability value, then

$$cr(G) \geq \frac{1}{p^4}(p^2|E| - 3p|V|) = \frac{|E|}{p^2} - \frac{3|V|}{p^3} = \frac{|E|^3}{16|V|^2} - \frac{3|E|^3}{64|V|^2} = \frac{|E|^3}{64|V|^2}. \tag{17}$$

□

In fact, the crossing number is on the order of $\Omega(\frac{|E|^3}{|V|^2})$, exactly as we proved above. Notice that in the above proof, we start with a trivial lower bound, and a randomized construction process in which random variables are impacted by the randomized construction parameter p to different orders. Combining this fact and the linearity of expectation, we see that the differences in orders magnify the effect of the trivial lower bound. This is one particularly powerful aspect of the probabilistic method.

5 The Second Moment Method

In this section, we introduce the Second Moment Method, which concerns second-order moments $\mathbb{E}(X^2)$. First, we illustrate the significance of higher-order moments with the following example. Let's say there are two games of gambling. The first one is to win \$1 with probability 0.5 and to lose \$1 with probability 0.5. The second one is to win \$99 with probability 0.01 and to lose \$1 with probability 0.99.

Through the lens of expected values, these two games are indifferent – both have \$0 expected payoff. What distinguishes between these two games is the variance, which loosely measures how much the distribution deviates from the mean. Specifically, the first game has variance 1 and the second game has variance 99. For this reason, the risk-seeking players would prefer the second game, and the risk-averse players would prefer the first game.

The major takeaway from this example is that not only the expected value of a random variable $\mathbb{E}(X)$ is useful in probabilistic bound, but also is the variance, which concerns the second moment $\mathbb{E}(X^2)$.

In particular, the variance, $\text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$, can tell us how *concentrated* the distribution is around its mean. A distribution with low variance will have most values packed around its mean; thus, it is unlikely to have a value far away from the mean. On the other hand, a distribution with high variance means the distribution is more spread-out; thus, it is more likely that some values are far away from the mean. The following inequality formally illustrates this concept – the variance does play a significant role in bounding the probability that a value is far away from the mean.

Theorem 24 (Chebychev's Inequality)**.** *Let X be a random variable with mean μ and variance σ^2. Then for any $k > 0$,*

$$\Pr\left(|X - \mu| \geq k\right) \leq \frac{\sigma^2}{k^2}.$$

Proof. Construct a random variable Y where $Y = (X - \mathbb{E}(X))^2$. By definition $\mathbb{E}(Y) = \text{Var}(X) = \sigma^2$. In addition, notice that $Y \geq 0$.

Recall from Markov's Inequality. We have

$$\Pr\left(|X - \mu| \geq k\right) = \Pr((X - \mathbb{E}(X))^2 \geq k^2) = \Pr\left(Y \geq k^2\right) \leq \frac{\mathbb{E}(Y)}{k^2} = \frac{\sigma^2}{k^2}$$

as wished. □

We will illustrate the use of Chebychev's Inequality through the example of distinct-sum sets.

5.1 Distinct-Sum Sets

Definition 25. *A set $S \subseteq \mathbb{Z}^+$ is* distinct-sum *iff any pair X, Y of distinct subsets of S, it follows that $\sum_{x \in X} x = \sum_{y \in Y} y$*

One example of an n-element distinct-sum set is $\{2^0, 2^1, ..., 2^{n-1}\}$. This is because all subset sums are positive integers from 1 to $2^n - 1$ (can think of this as a binary expression.) This set has the largest element 2^{n-1}. Furthermore, if we consider any n-element distinct-sum set, the largest element must be at least $2^n/n$. This is because otherwise, every subset sum must be less than 2^n, causing at least two distinct subsets to have the same sum by the Pigeonhole Principle. With this observation, the question of determining the tightest lower bound of the largest element of any n-element distinct-sum set arises. Although the problem still remains open today, an approach using Chebychev's Inequality gives a decent lower bound.

Theorem 26. *Let $n \geq 5$ be a positive integers. Let X be a distinct-sum set of positive integers with $|X| = n$. Then*

$$\max_{x \in X} x > \frac{2^{n+1}}{3\sqrt{3n}}.$$

Proof. Let X's elements be $x_1 < x_2 < .. < x_n$. Also let $S(X)$ be the set of all 2^n distinct sums of X. We will prove by contradiction. Assume, for the contrary, that $x_n \leq M$ when $M = \frac{2^{n+1}}{3\sqrt{3n}}$. It follows that $x_i \leq M + 1 - i$ for every $i \in [n]$. Given $n \geq 5$, it follows that

$$\sum_{i \in [n]} x_i^2 \leq \sum_{i \in [n]} (M + 1 - i)^2 < nM^2 - \frac{2^{n+2}}{9} = \frac{2^{2n+2}}{27} - \frac{2^{n+2}}{9}$$

Consider the following randomized process. For each $i \in [n]$, independently and uniformly choose α_i from $\{0, 1\}$. Denote the random variable $S = \sum_{i \in [n]} \alpha_i x_i$. Notice that all 2^n constructions of S happen uniformly and they correspond to 2^n distinct sums of X. Thus, the distribution of S is uniform over $S(X)$.

Since we choose each α_i's independently, we have

$$\mathrm{Var}(S) = \sum_{i \in [n]} \mathrm{Var}(\alpha_i x_i) = \sum_{i \in [n]} \frac{x_i^2}{4} \leq \frac{2^{2n}}{27} - \frac{2^n}{9}.$$

Now we apply Chebychev's Inequality with $k = \frac{2^n}{3}$.

$$\Pr\left(|S - \mathbb{E}(S)| < k\right) = 1 - \Pr\left(|S - \mathbb{E}(S)| \geq k\right) \geq 1 - \frac{\mathrm{Var}(S)}{k^2} \geq \frac{2}{3} + \frac{1}{2^n}.$$

We notice further that

$$\Pr\left(|S - \mathbb{E}(S)| < k\right) < \frac{2k+1}{2^n} = \frac{2}{3} + \frac{1}{2^n}$$

because there are at most $2k + 1$ possible values of S that lies within the range of k from $\mathbb{E}(S)$ and the probability that S being those values is at most 2^{-n} each. Combining two inequalities, we have reached the contradiction. □

6 Chernoff Bound

In many situations, we need to deal with the sum of random variables, such as the total number of heads in coin flip sequences and the total number of red edges in randomly colored graphs. Sometimes, we can calculate the expected value of the sum rather easily using linearity of expectation, but what is more helpful is a tool to show how unlikely that the actual value deviates much from the expected value.

The Chernoff bound is such a tool. It is a strong and useful probability bound that is commonly used in bounding tail probabilities of the sum of independent Bernoulli random variables. In this section, we introduce and prove the Chernoff bound, and present its application on balanced graph coloring.

Theorem 27. *(Chernoff Bound) Let $X_1, X_2, \cdots, x_n$ be independent Bernoulli random variables i.e. $X_i \sim$ Bernoulli(p_i), and $X = X_1 + X_2 + \cdots + X_n$. Let $\mu = \mathbb{E}[X] = \sum_i p_i$, we have*

$$\Pr\left(X \geq (1+\delta)\mu\right) \leq \left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right]^{\mu}$$

and

$$\Pr\left(X \leq (1-\delta)\mu\right) \leq \left[\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right]^{\mu}$$

Proof. Recall Theorem 21, the Markov's inequality, and apply it to e^{tX}, for real number a we have

$$\Pr(X \geq a) = \Pr\left(e^{tX} > e^{ta}\right) \leq \frac{\mathbb{E}[e^{tX}]}{e^{ta}}$$

And by optimizing the probability over $t > 0$, we have

$$\Pr(X \geq a) \leq \min_{t>0} e^{-ta} \prod_i \mathbb{E}[e^{tX_i}]$$

Similarly,

$$\Pr(X \leq a) \leq \min_{t>0} e^{ta} \prod_i \mathbb{E}[e^{-tX_i}]$$

Note that $\mathbb{E}[e^{tX_i}] = (1-p_i)e^0 + p_i e^t = 1 + p_i(e^t - 1) \leq e^{p(e^t-1)}$, so

$$\mathbb{E}[e^{tX}] = \prod_i \mathbb{E}[e^{tX_i}] \leq e^{\left(\sum_i p\right)(e^t-1)} = e^{\mu(e^t-1)}$$

Then for $\delta > 0$ taking $t = \ln(1+\delta)$ and $a = (1+\delta)\mu$ in equation 6 yields

$$\Pr(X \geq (1+\delta)\mu) \leq \frac{1}{(1+\delta)^{(1+\delta)\mu}} \cdot e^{\delta\mu} = \left[\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right]^{\mu}$$

And similarly, by taking $t = \ln(1-\delta)$ and $a = (1-\delta)\mu$ in equation 6 yields

$$\Pr(X \leq (1-\delta)\mu) \leq \frac{1}{(1-\delta)^{(1-\delta)\mu}} \cdot e^{-\delta\mu} = \left[\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right]^{\mu}$$

□

The Chernoff bound proven in Theorem 27 is tight, but too complex to use in many scenarios. To simplify, we bring in the inequality $\frac{2\delta}{2+\delta} \leq \ln(1+\delta)$, and obtain the following looser but simpler bounds that are often used:

Corollary 28. *(Variant of the Chernoff Bound)*

$$\Pr(X \leq (1-\delta)\mu) \leq e^{-\frac{\delta^2\mu}{2}}, \text{ for } 0 \leq \delta \leq 1 \tag{18}$$

$$\Pr(X \geq (1+\delta)\mu) \leq e^{-\frac{\delta^2\mu}{2+\delta}}, \text{ for } 0 \leq \delta \tag{19}$$

In fact, by limiting δ not to exceed 1 and the union bound, we can have an even simpler form:

$$\Pr(|X-\mu| \geq \delta\mu) \leq 2e^{-\frac{\delta^2\mu}{3}}, \text{ for } 0 \leq \delta \leq 1 \tag{20}$$

6.1 Hypergraph 2-coloring

In addition to studying whether a hypergraph is 2-colorable, another interesting feature to investigate in hypergraph 2-coloring is how *balanced* a coloring is, which means that in each edge, we want the number of hypernodes colored Red not differ much from the number of hypernodes colored Blue. In this section, we apply the Chernoff bound to show how balanced the coloring of some hypergraphs can be.

Definition 29. *In a hypergraph H whose vertices are colored with two colors, for all hyperedge e, the* discrepancy *of e is defined as the absolute value of the difference between the numbers of vertices of e in each color.*

For example, if a hyperedge has five vertices colored with red, blue, red, red and blue, then its discrepancy is $|3-2| = 1$.

Then, we define $disc(H)$ to measure how balanced we can construct a 2-coloring for H.

Definition 30. *Say the* discrepancy *of a hypergraph H under one of its 2-colorings is the maximum discrepancy over all hyperedges. We define $disc(H)$ to be the minimum discrepancy of H over all of its 2-colorings.*

Clearly, for k-uniform hypergraph H, we have $disc(H) \leq k$ all the time, and $disc(H) < k$ if and only if H is 2-colorable i.e. we can color the vertices so that none of the hyperedges are monochromatic.

Theorem 31. *Let H be a k-uniform hypergraph with k hyperedges, and then $disc(H) \leq \sqrt{8k\ln(k)}$.*

Proof. First, note that when $k \leq 8$, we have $disc(H) \leq k < \sqrt{8k\ln(k)}$, so considering all $k \geq 9$ is sufficient.

We can color all vertices of H randomly with red and blue each with probability $\frac{1}{2}$. For each hyperedge, N_r, the number of red vertices it contains is the sum of k Bernoulli random variables each with probability $\frac{1}{2}$. Clearly, $\mu = \mathbb{E}[N_r] = \frac{k}{2}$.

Then, we apply the Chernoff bound (Equation20) with $\delta = \sqrt{\frac{8\ln(k)}{k}}$, and we have

$$\Pr\left(|N_r - \mu| \geq \delta\mu\right) \leq 2e^{-\frac{\delta^2\mu}{3}} = 2e^{-\frac{\frac{8\ln(k)}{k}\cdot\frac{k}{2}}{3}} = 2k^{-\frac{4}{3}}$$

Given that $k > 8$, we have

$$\Pr\left(\left|N_r - \frac{k}{2}\right| \geq \sqrt{2k\ln(k)}\right) \leq 2k^{-\frac{4}{3}} < \frac{1}{k}$$

and the expected number of hyperedges that have $|N_r - \frac{k}{2}| \geq \sqrt{2k\ln(k)}$, by the linearity of expectation, is

$$\Pr\left(\left|N_r - \frac{k}{2}\right| \geq \sqrt{2k\ln(k)}\right) \cdot k < 1$$

Therefore, with positive probability, the number of such edges is 0 i.e. the number of red vertices is in $\frac{k}{2} \pm \sqrt{2k\ln(k)}$ for all hyperedges of H, so the discrepancy of H is at most $\sqrt{8k\ln(k)}$.

Hence, $disc(H) \leq \sqrt{8k\ln(k)}$ for all $k \geq 1$.

□

7 Alteration

While the Union Bound, the First Moment Method, the Second Moment Method, and the Chernoff Bound are all useful tools to help us calculate the probability of the existence of an object, yet sometimes, the random construction only guarantees the existence of an object that *almost* satisfies our demands, but not *completely*. In such cases, we may conduct some alteration to modify the object in a judicious way to get what we exactly need.

7.1 Stability Number

The idea of the stable set and the stability number (as defined below) are important graph properties, and are closely related to other graph property parameters such as clique, vertex coloring, etc. Problems related to stable set are widely studied since they have practical applications in areas like information retrieval, scheduling, and computer vision among many others. Currently, the maximal stable set problem is considered NP-hard and thus hard to approximate.

We now give a lower bound on the stability number, and hope to use this example to epitomize first, how alteration works, and second, the idea introduced in Section 4.2 where a weak bound gets amplified through probabilistic method to achieve a tighter bound.

Definition 32. *A* stable set *of a graph $G(V, E)$ is a set of pairwise non-adjacent vertices. That is, a subset $S \subseteq V$ that no two vertices $v_1, v_2 \in S$ are connected in G.*

The stability number *of a graph $G(V, E)$, denoted by $\alpha(G)$, represents the largest possible size of a stable set of G. In other words, $\alpha(G)$ equals $max\{|S||S$ is a stable set$\}$.*

Notice that by definition, the statement that S is a stable set of G is equivalent to saying that S is a clique of $\overline{G}$, the complement of G.

Theorem 33. *Let $d \geq 1$ be a real number, and G be a graph that $|V| = n$, $|E| = \frac{nd}{2}$. Then the stability number $\alpha(G) \geq \frac{n}{2d}$.*

Proof. Instead of directly proving that a subset S of V of size at least $\frac{n}{2d}$ is a stable set with positive probability, we prove an equivalent statement. We prove that there exists a set S where in the subgraph $G\langle S\rangle$, $|V(G\langle S\rangle)|$ is greater than $|E(G\langle S\rangle)|$ by at least $\frac{n}{2d}$. Then, with an alteration step in which we remove one endvertex of each edge in $G\langle S\rangle$, we reduce $|E(G\langle S\rangle)|$ to be 0, and $|V(G\langle S\rangle)|$ to be at least $\frac{n}{2d}$, which is a valid stable set.

Define a randomized process as follows. For every vertex of G, choose it uniformly and independently with probability $p = \frac{1}{d}$ into $S \subseteq V$. We determine the value of p later.

Let X and Y denote the number of vertices and the number of edges in $G\langle S\rangle$ respectively. Then from the given conditions, $\mathbb{E}(X) = p|V| = pn$, and $\mathbb{E}(Y) = p^2|E| = p^2\frac{nd}{2}$, since for an edge (u, v) to be in $G\langle S\rangle$, both u and v need to selected into S. By linearity of expectation,

$$\mathbb{E}(X - Y) = \mathbb{E}(X) - \mathbb{E}(Y) = pn - \frac{nd}{2}p^2.$$

This quadratic function of p reaches its maximum value $\frac{n}{2d}$ at $p = \frac{1}{d}$. Given $d \geq 1$, $p \in (0, 1]$ is a valid probability. Set $p = 1/d$, then $\mathbb{E}(X - Y) = \frac{n}{2d}$. By the First Moment Method, there exists a subset $S \subseteq V$ such that $X - Y \geq \frac{n}{2d}$. As discussed, in the alteration step, for each edge $(u, v) \in E(G\langle S\rangle)$, we delete u or v from S. Then the remaining set of vertices $S' \subseteq S$ contains no edges, and has size $|S'| \geq X - Y \geq \frac{n}{2d}$. This is equivalent to saying that S' is a stable set, and the stability number $\alpha(G)$ is at least $|S'|$, which is at least $\frac{n}{2d}$. □

In fact, if for each $v \in V(G)$, denote by $d(v)$ be the degree of v, then by constructing a randomized process on $|V|!$ orderings of V and applying the First Moment Method, we can achieve a tighter lower bound on stability number:

$$\alpha(G) \geq \sum_{v \in V} \frac{1}{d(v) + 1}.$$

And, the relationship between stable set and clique indicates that, G contains a clique of size at least $\sum_{v \in V} \frac{1}{|V| - d(v)}$.

7.2 Ramsey Number

To reinforce the idea of alteration, we return to the Ramsey Number problem introduced in Section 3.2, and present a better lower bound for the diagonal Ramsey numbers $R(k, k)$ using alteration.

Theorem 34. *Let n, k be positive integers, then $R(k, k) > n - \binom{n}{k} \cdot 2^{1-\binom{k}{2}}$.*

Proof. We define the alteration procedure as follows. In an uncolored K_n called G, color each edge independently and uniformly with Red or Blue - with probability $1/2$ each. Initialize a set $D = \emptyset$. We sequentially go over every K_k of G: if it is monochromatic, choose any vertex from that K_k and add it to D. Let $G' = G \backslash D$, which represents deleting vertices in D from G. It follows that $|V(G')| = |V(G)| - |D| = n - |D|$. Since for each K_k in G, it must have at least one vertex added to D and deleted from G, we know that G' contains no monochromatic K_k. By definition of Ramsey number, $R(k, k) > |V(G')| = n - |D|$.

To complete the proof, we next show that there exists some random coloring that yields $|D| \leq \binom{n}{k} \cdot 2^{1-\binom{k}{2}}$. For each C being a K_k in G, denote by I_C as the indicator variable of the event that

C is monochromatic, i.e. $I_C = 1$ if C is monochromatic, and 0 if not. Let $X = \sum_{C \text{ being } K_k \text{ of } G} I_C$ represent the total number of monochromatic K_k in G. Then $|D| \leq X$, because we add 1 vertex to D for each monochromatic K_k.

Recall from the proof for Theorem 14, $\mathbb{E}(I_C) = \Pr(C \text{ is monochromatic}) = 2^{1-\binom{k}{2}}$, and the number of K_k in G is $\binom{n}{k}$ by definition. Therefore,

$$\mathbb{E}(X) = \text{number of } K_k \text{ in } G \cdot I_C = \binom{n}{k} \cdot 2^{1-\binom{k}{2}}.$$

By the First Moment Method, some random coloring yields $X \leq \binom{n}{k} \cdot 2^{1-\binom{k}{2}}$. It follows that in this coloring, $|D| \leq X \leq \binom{n}{k} \cdot 2^{1-\binom{k}{2}}$, which completes the proof. □

In fact, we can further optimize the lower bound of $R(k,k)$ by determining the largest positive integer n (as a function of k) such that $\binom{n}{k} \cdot 2^{1-\binom{k}{2}} < 1$. With this optimization, it follows that

$$R(k,k) > \left(\frac{1}{e\sqrt{2}} + o(1)\right) \cdot k \cdot 2^{k/2}$$

8 Lovász Local Lemma

So far, we have made no assumption about the independence of events. We can notice that the Union Bound is effective mostly when events are disjoint or almost disjoint, so that the sum of probability over all events is less than 1. On the other hand, when computing expected values, we often make use of the linearity of expectation which does not concern the independence among random variables. In this section, we introduce the Lovász Local Lemma [EL74] which allows us to deal with *local dependency*, meaning that each depends on few other events. There are two majors variations of the Lovász Local Lemma: the general and symmetric form. We will discuss both variations along with their applications.

8.1 General Lovász Local Lemma

Definition 35. *Given events $A_1, A_2, ..., A_n$. We call $G = (V, E)$ a dependency graph iff*

- $V = \{A_1, A_2, ..., A_n\}$
- *For each $i \in [n]$, A_i is totally independent of all events but those $N^+(A_i)$. In other words, A_i is independent of S for any $S \subseteq V \setminus N^+(A_i)$*

where $N^+(A_i)$ is the set of A_i and all neighbors of A_i in G.

Theorem 36. (General Lovász Local Lemma) *Let $A_1, ..., A_n$ be events with dependency graph G. Suppose that there exists $p_1, ..., p_n \in (0,1)$ which for any $i \in [n]$,*

$$\Pr(A_i) \leq p_i \cdot \prod_{(i,j) \in E(G)} (1 - p_j). \tag{21}$$

Then

$$\Pr\left(\bigwedge_{i \in [n]} \overline{A_i}\right) \geq \prod_{i \in [n]} (1 - p_i) > 0. \tag{22}$$

Proof. For any $S \subseteq [n]$, denote $\overline{A_S} = \bigwedge_{s \in S} \overline{A_s}$. Also denote $N(i) = \{j \in [n] \mid (A_i, A_j) \in E(G)\}$. In other words, $N(i)$ is the set of indices of neighbors of A_i in G. The proof proceeds through the following claim.

Claim 37. *For any $i \in [n]$ and $S \subset [n] \setminus \{i\}$, we have $\Pr\left(A_i \mid \overline{A_S}\right) \leq p_i$.*

Proof. We will prove the claim by strong induction. Let $H(k)$ represent the following statement: for any $i \in [n]$ and $S \subset [n] \setminus \{i\}$ such that $|S| = k$, we have $\Pr\left(A_i \mid \overline{A_S}\right) \leq p_i$.

The base case $k = 0$ is trivially true since for any $i \in [n]$, $\Pr(A_i) \leq p_i \cdot \prod_{(i,j) \in E(G)} (1 - p_j) \leq p_i$.

For the induction step, assume that $H(0), ..., H(k-1)$ are true. Let $i \in [n]$ and $S \subset [n] \setminus \{i\}$ such that $|S| = k$. Partition S into $R \cup T$ for $R = S \cap N(i)$ and $T = S \setminus N(i)$. Also denote $R = \{r_1, ..., r_m\}$ for some $m \leq |S| = k$.

Our goal is to show that $\Pr\left(A_i \mid \overline{A_S}\right) = \Pr\left(A_i \mid \overline{A_R}\ \overline{A_T}\right) \leq p_i$. Notice that

$$\Pr\left(A_i \mid \overline{A_R}\ \overline{A_T}\right) = \frac{\Pr\left(A_i \overline{A_R} \mid \overline{A_T}\right)}{\Pr\left(\overline{A_R} \mid \overline{A_T}\right)} \leq \frac{\Pr\left(A_i \mid \overline{A_T}\right)}{\Pr\left(\overline{A_R} \mid \overline{A_T}\right)} = \frac{\Pr(A_i)}{\Pr\left(\overline{A_R} \mid \overline{A_T}\right)}$$

where the last equality follows from the fact that A_i is independent from $\overline{A_T}$.

The numerator is

$$\Pr(A_i) \leq p_i \cdot \prod_{(i,j) \in E(G)} (1 - p_j) = p_i \cdot \prod_{j \in N(i)} (1 - p_j)$$

and the denominator is

$$\Pr\left(\overline{A_R} \mid \overline{A_T}\right) = \prod_{i \in [m]} \Pr\left(\overline{A_{r_i}} \mid \overline{A_{T \cup \{r_1, ..., r_{i-1}\}}}\right) \geq \prod_{i \in [m]} (1 - p_{r_i}) = \prod_{j \in R} (1 - p_j) \geq \prod_{j \in N(i)} (1 - p_j)$$

where the first equality follows from a chain of Bayes Rule, and the first inequality follows from the assumption the induction hypotheses $H(|T|+1), ..., H(|T|+m-1)$. This can be done since $|T| + m - 1 = |T| + |R| - 1 \leq k - 1$.

It follows that $\Pr\left(A_i \mid \overline{A_S}\right) = \Pr\left(A_i \mid \overline{A_R}\ \overline{A_T}\right) \leq p_i$. This completes the induction step and the proof of Claim 37. □

Finally, we now can conclude that

$$\Pr\left(\bigwedge_{i \in [n]} \overline{A_i}\right) = \Pr\left(\overline{A_{[n]}}\right) = \prod_{i \in [n]} \Pr\left(\overline{A_i} \mid \overline{A_{[i-1]}}\right) \geq \prod_{i \in [n]} (1 - p_i) > 0$$

where the second equality follows from the Bayes Rule, and the first inequality follows directly from the Claim 37. □

Now that we have proved the General Lovász Local Lemma, we present its applications. We start with a fairly straightforward example, followed by revisiting the Hypergraph 2-Coloring problem and strengthening our previous result.

Theorem 38. *Let $A_1, ..., A_n$ be events with dependency graph G. Suppose that for every $i \in [n]$, we have $\Pr(A_i) < 1/2$ and $\sum_{(i,j)\in E(G)} \Pr(A_j) \leq 1/4$. Then with positive probability, no event A_i occurs.*

Proof. We use the General Lovász Local Lemma. We assign $p_i = 2 \cdot \Pr(A_i)$. First, we need to verify that p_i lies in $[0, 1)$ which is true since $\Pr(A_i) < 1/2$. Then, we need to show that this assignment satisfies the algebraic condition for the General Lovász Local Lemma as in Equation 21. We verify this by showing that

$$\begin{aligned} p_i \cdot \prod_{(i,j)\in E(G)} (1 - p_j) &= 2\Pr(A_i) \cdot \prod_{(i,j)\in E(G)} (1 - 2\Pr(A_j)) \\ &\geq 2\Pr(A_i) \cdot \left(1 - 2 \cdot \sum_{(i,j)\in E(G)} \Pr(A_j)\right) \\ &\geq 2\Pr(A_i) \cdot \left(1 - 2 \cdot \frac{1}{4}\right) = \Pr(A_i). \end{aligned}$$

Therefore, directly following the General Lovász Local Lemma, we conclude that with positive probability, no event A_i occurs. □

Recall Theorem 11 where we showed that a hypergraph is always 2-colorable if the number of hyperedges is not too large. However, the number of hyperedges can be massive despite having small number of hypernodes. The following theorem presents another aspect of the hypergraph 2-colorability under constraints on parameters including the size and the number of neighbors of hyperedges.

Theorem 39. *Given positive integers k, l. Let H be a hypergraph which each edge contains at least k hypernodes. In addition, for each hyperedge, the sum of sizes of its neighbors is at most l. If $e\left(1 + \frac{l}{k}\right) \leq 2^{k-1}$, then H is 2-colorable.*

It also is worth noting that if H is k-unifrom and every hyperedge is adjacent to at most d others, we will find that $l = dk$ is a valid choice. Thus, the condition $e\left(1 + \frac{l}{k}\right) \leq 2^{k-1}$ is equivalent to $e(1 + d) \leq 2^{k-1}$, which is the result we got from Theorem 41 using Symmetric Lovász Local Lemma. Thus, this result is stronger than Theorem 41

Proof. For every hyperedge A, denote $N(A)$ to be the set of A's neighbor, i.e. hyperedges that intersect with A. Also denote $|A|$ to be the number of hypernodes contained in A. Now the second condition has become

$$\forall A \in E(H), \sum_{B\in N(A)} |B| \leq l.$$

Denote a randomized process as follows. For each hypernode, uniformly and independently color it with Red or Blue – with probability 1/2 each. For each edge $A \in E(H)$, we denote M_A to be the bad event that edge A is monochromatic. It follows that $\Pr(M_A) = 2^{1-|A|} \leq 2^{1-k}$.

We notice for any hyperedge A, the event M_A is totally independent of all M_B which $A \cap B = \phi$. Thus, in the dependency graph, M_A is only adjacent to those M_B's where $A \cap B \neq \phi$.

We will use the General Lovász Local Lemma. For any bad event M_A, we assign $p(M_A) = \frac{|A|}{l+|A|}$ which correctly lies within $(0,1)$. Also note that $p(M_A) \geq \frac{k}{k+l}$ and $1 - p(M_A) \geq e^{-|A|/l}$. It follows that

$$p(M_A) \cdot \prod_{B \in N(A)} (1 - p(M_B)) \geq \frac{k}{k+l} \cdot \prod_{B \in N(A)} e^{-|B|/l} = \frac{k}{e(k+1)} \geq 2^{1-k} \geq \Pr(M_A).$$

The General Lovász Local Lemma implies that with positive probability, no bad event occurs. This is equivalent to having no monochromatic edge. Thus, H is 2-colorable. □

8.2 Symmetric Lovász Local Lemma

We have already proven the General Lovász Local Lemma, but applying the general version requires us to construct appropriate p's, which is not convenient in most cases. In this section we introduce the symmetric version of Lovász Local Lemma by simply setting all p_i to be equal, and assuming that all $\Pr(A_i)$ can be bound using the same bound p.

Theorem 40. *(Symmetric Lovász Local Lemma) Let $A_1, \cdots, A_n$ be events in a probability space, for $p \in [0,1]$ and integer d, if we have*

1. *For each $i \in [n]$, $\Pr(A_i) \leq p$*
2. *For each $i \in [n]$, A_i is independent of all but at most d other events.*
3. *$ep(d+1) \leq 1$, where $e = \exp(1)$ is the Euler's number.*

then $\Pr\left(\bigcap_{i \in [n]} \bar{A}_i\right) > 0$

Proof. We apply the General Lovász Local Lemma in which we have that $\Pr(A_i) \leq p \leq \frac{1}{e(d+1)}$, so

$$\Pr(A_i) \leq \frac{1}{d+1} \cdot \frac{1}{e} < \frac{1}{d+1}\left(1 - \frac{1}{d+1}\right)^d$$

Then we set $p_i = \frac{1}{d+1}$ for i $in[n]$, and note that $\sum_{j|(i,j) \in E(G)} 1 \leq d$, so we can get

$$\Pr(A_i) \leq \frac{1}{d+1}\left(1 - \frac{1}{d+1}\right)^d \leq p_i \cdot \prod_{(i,j) \in E(G)} (1 - p_j)$$

Then, by the General Lovász Local Lemma, we have

$$\Pr\left(\bigcap_{i \in [n]} \overline{A_i}\right) \geq \prod_{i \in [n]} (1 - p_i) > 0$$

□

8.3 Hypergraph Coloring

We are going to prove a special case of Theorem 39 using Symmetric Lovász Local Lemma on hypergraph color-ability.

Theorem 41. *Let H be a k-uniform hypergraph in which each hyperedge intersects at most d other hyperedges. Then, H is 2-colorable if $e(d+1) \leq 2^{k-1}$ where $e = \exp(1)$ is the Euler number.*

Proof. We can prove the theorem by applying the Symmetric Lovász Local Lemma directly as following

1. For each hyperedge i, we set the random event A_i to be that the hyperedge is monochromatic. Clearly there are only two monochromatic colorings over all 2^k colorings and $\Pr(A_i) \leq p = \frac{1}{2^{k-1}}$.
2. Each random event A_i is independent of all other random events but at most d other events whose hyperedges intersect with i.
3. If $e(d+1) \leq 2^{k-1}$, we have $ep(d+1) \leq 1$.

Then by the Symmetric Lovász Local Lemma, we have the probability that none of the hyperedges are monochromatic, to be positive. Therefore, there must be one such coloring. □

Moving forward to constructive methods. In previous sections, we focus on the non-constructive side of the probabilistic methods, in which we prove the existence of an object without giving an explicit example. For the rest of the paper, we will shift our focus to the question of *how to construct an object* whose existence can be guaranteed by the probabilistic method. The most naive way is to exhaustively enumerate every possible construction, and select the one that has the desired property. But this is unrealistic and unnecessarily complicated in most cases. For example, if we enumerate all possible 2-coloring of a hypergraph on its n vertices, we need to do 2^n rounds of searching and this is very inefficient. It turns out that we can often efficiently (in polynomial time) find such object under some additional constraints or in looser settings. In upcoming sections, we will discuss three ways of construction: the Greedy Algorithm in Section 9, the Beck's Algorithm in Section 10, and the Derandomization Method in Section 11.

9 Greedy Algorithms

The first constructive method we introduce is the Greedy Algorithm. Although there is no formal definition of the term, one can interpret it as trying to grab as much as we can from the opportunity that presents in front of us. The following example illustrates this idea.

Example 42. *Alice is going on a game show. Bob is the host who has a series of integers in his mind. Bob then reveals the numbers one by one to Alice who can choose either to keep the number or pass it. However, there is an additional rule that the sum of Alice's numbers must not exceed 10. At the end of the game, Alice will take home the amount of money equal to the sum of her numbers. How will Alice tackle on this game?*

One strategy that Alice can employ is a greedy approach: always keep the number if the sum does not exceed 10. In fact, this is a decent strategy given that Alice has no prior knowledge of Bob's series of number. Thus, it would be the best for Alice to act upon what is currently offered in front of her.

For example, it Bob's numbers are $(3, 1, 2, 5, 4)$, Alice will keep $3, 1, 2, 4$ – earning \$10. On the other hand, if the numbers are $(3, 9, 3, 6, 9, 3)$, Alice will keep $3, 3, 3$ – earning \$9. In these scenarios, the greedy strategy is *optimal*, meaning that Alice cannot earn more even if she knew the entire series before hand. In the first scenario, Alice can never earn more than \$10 because the rule forbids Alice to keep the numbers whose sum up more than 10. In the second scenario, Alice cannot earn \$10 because the numbers in the sequence are multiple of 3.

On the other hand, consider the scenario when Bob's numbers are $(3, 1, 5, 2, 4)$. Alice will keep $3, 1, 5$ – earning \$9. In this scenario, the greedy strategy is not optimal because Alice can actually earns \$10 from keeping $3, 1$, passing 5, and keeping $2, 4$. Thus, greedy strategy does not always give the optimal solution.

Even worse, consider the unfortunate scenario when Bob's numbers are $(1, 10, 10, 10, 10)$. Employing greedy strategy, Alice can earn only \$1 from keeping 1, while the optimal solution is to earn \$10 by passing the first 1 and keeping any of the following 10's. Thus, greedy strategy can sometimes perform very poorly, though this rarely happens.

The major takeaway from this example is that the Greedy approach does not always give the optimal solution. In fact, most of the time it will not. However, we can expect that most of the time, the greedy approach to give a *good enough* solution while the scenario where it gives a *bad* solution rarely happens.

It also turns out that oftentimes, the greedy approach gives a construction which is as good as what we derive from probabilistic method.

9.1 Hitting Set Problem

Theorem 43. *Given $k-$element sets $S_1, S_2, ..., S_m \subseteq [n]$. There exists an algorithm for finding the hitting set H with $|H| \leq \lceil \frac{n \log m}{k} \rceil$ that runs in in $O(mk \log n)$ time.*

Proof. We claim that the following greedy algorithm finds the desired hitting set H.

Algorithm 1 : Greedy-Hitting-Set$(S_1, S_2, ..., S_m)$

```
S ← {S_1, S_2, ..., S_m}
H ← φ
while S ≠ φ do
    x ← the element of [n] \ H that has the most occurrences in S
    Add x to H
    Remove every S_i ∈ S such that x ∈ S_i
Output H
```

First of all, notice that this algorithm always terminates because for each round of the while loop, we remove at least one element from S, which allows at most m rounds of looping.

Notice that at each round, we add one element to H. Therefore, when the algorithm terminates, the size of H is indeed the number of rounds of the while loop.

For each $r \in \mathbb{Z}^+$, let t_r be the size of S before entering round r, which is the same as the size of S at the end of round $r-1$. With this notation, we have $t_1 = m$ as it represents initial S before entering the while loop. Finally, let the x determined in Line 4 of round r denoted by x_r.

Now we notice that once the while loop terminates, H becomes a hitting set. This is because each $S_i \in S$ removed at round r has x_r in common with H.

For each round r, we begin with $|S| = t_r$ and $|H| = r - 1$. The total number of occurrences of $[n] \setminus H$ in S is exactly $|S| \cdot k = t_r k$. This means the occurrences of x_r in S must be at least

$$\frac{t_r k}{|[n] \setminus H|} = \frac{t_r k}{n - r + 1} \geq \frac{t_r k}{n}.$$

This means in round r, we remove at least $t_r k/n$ elements from S. This yields

$$t_{r+1} \leq t_r - \frac{t_r k}{n} = t_r \cdot \left(1 - \frac{k}{n}\right) < t_r \cdot e^{-k/n}.$$

Therefore, for any round r,

$$t_{r+1} < t_1 \cdot e^{-rk/n} = m \cdot e^{-rk/n}.$$

Notice that when $r = \lceil \frac{n \log m}{k} \rceil$, we have $t_{r+1} < 1$, which means $t_{r+1} = 0$, i.e., $S = \phi$. This means the while loops never runs more than $\lceil \frac{n \log m}{k} \rceil$ rounds before reaching $S = \phi$ and then terminate. Since $|H|$ is the number of rounds of the while loop, the proof is complete. □

Recall that Theorem 16 guarantees that the hitting set H with size at most $\lceil \frac{n \log m}{k} \rceil$ exists. Thus, the above greedy algorithm gives the construction of a hitting set which is as good as guaranteed by the probabilistic method.

10 Beck's Algorithm

After introducing the greedy algorithm as an example of constructive methods, we now return to the hypergraph coloring example, and show how constructive methods shed light on the topic. Recall that in Theorem 41, we proved that if a hypergraph satisfies certain conditions, then a proper 2-coloring exists. Then, the question that naturally arises is, given a hypergraph that satisfies the condition $e(d+1) \leq 2^{k-1}$, how to construct the proper 2-coloring on the hypergraph. It turns out that under a slightly looser constraint, we can find such coloring in polynomial time. The algorithm was first given by [Bec91] and is considered a significant breakthrough in the hypergraph 2-coloring problem. The version of Beck's algorithm that we are going to present is revised in [Alo91].

Beck's algorithm consists of two passes. In the first pass, we independently and uniformly color each hypernode with Red or Blue. There is no constraints imposed at the moment, meaning that it is acceptable in this stage to have monochromatic edges. We will show that with the random coloring, we can guarantee a useful underlying structure with constant probability.

Since it is possible to have some monochromatic hyperedges as a result from the random coloring, the second pass of the algorithm is dedicated to fixing those edges. Specifically, we will choose not-so-many hypernodes to be recolored in such a way that guarantees that there will be no monochromatic edge after the recoloring.

We define some terminologies that are essential to the proof in Section 10.1, introduce the first pass in Section 10.2, and conclude the proof with the second pass in Section 10.3.

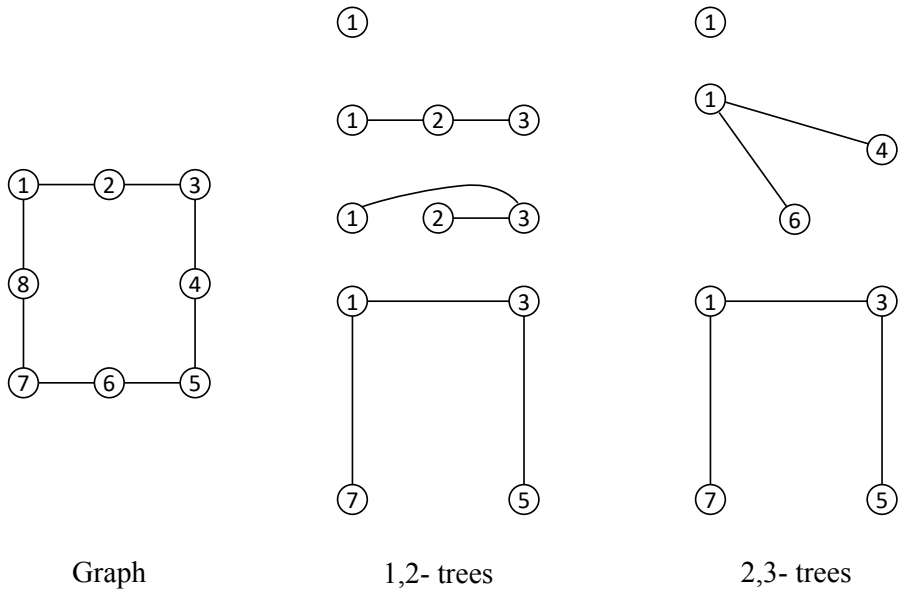

Figure 3: Shown above is an example of a graph G (left) along with its four examples of 1,2-trees (middle) and three examples of 2,3-trees (right). Each 1,2-tree and 2,3-tree satisfies properties specified in Definition 44 and Definition 45 respectively.

10.1 Definitions

Definition 44. *Given an undirected graph G with distance function d_G. We call C a 1,2-tree of G if it satisfies the following three conditions. (See Figure 3)*

1. *$V(C) \subseteq V(G)$*
2. *For any $(a, b) \in E(C)$, we must have $d_G(a, b) = 1$ or 2.*
3. *C is a connected tree*

Definition 45. *Given an undirected graph G with distance function d_G. We call T a 2,3-tree of G if it satisfies the following three conditions. (See Figure 3)*

1. *$V(T) \subseteq V(G)$*
2. *For any two arbitrary vertices $a, b \in V(T)$, we must have $d_G(a, b) \geq 2$. It is worth noting that (a, b) is not necessarily an edge of T.*
3. *For any $(a, b) \in E(T)$, we must have $d_G(a, b) = 2$ or 3*
4. *T is a connected tree*

The main result we present in this section is the following theorem.

Theorem 46. *Given fixed $n \geq 2$ and d, and suppose that some $0 < \alpha < 1$ satisfies*

$$4ed^3 < 2^{n(1-H(\alpha))}, H(\alpha) = -\alpha \log_2 \alpha - (1-\alpha)\log_2(1-\alpha), \tag{23}$$

and

$$2e(d+1) < 2^{\alpha n}, \tag{24}$$

then there is a randomized algorithm that finds a proper 2-coloring of a n-uniform hypergraph H with N edges in which no edge intersects more than d other edges in expected $N^{O(1)}$ time.

10.2 The First Pass

The first step towards the algorithm is to color every hypernode independently and uniformly with Red or Blue. At this stage, we do not impose any constraint on coloring (some edges can be monochromatic) as we will fix that up in the second pass.

After the random coloring, we will call an edge *bad* if it contains at most αn vertices of one color, and let B denote the set of bad edges. We can see that the probability that a particular edge is bad is

$$2 \cdot \sum_{0 \leq i \leq \alpha n} \binom{n}{i} \cdot 2^{-n} \leq 2 \cdot 2^{n(1-H(\alpha))} = p,$$

which follows from the connection between entropy and binomial coefficients that can be proved with the subadditivity of entropy. We start with showing some important properties of 1,2-trees and 2,3-trees.

Lemma 47. *Given a graph G with a 1,2-tree C. Let T be the maximal 2,3-tree whose vertices are within vertices of C, i.e. $V(T) \subseteq V(C)$. Then for every $p \in V(C) \setminus V(T)$, there must exists $q \in V(T)$ such that $(p,q) \in E(G)$.*

Proof. For a vertex u and a real number r, we say $u \geq_G r$-away from T iff $d_G(u,v) \geq r$ for every $v \in V(T)$. To prove this lemma, we first need to introduce two following miniclaims.

Claim 48. *If $p \in V(C) \setminus V(T)$ and $p \geq_G 2$-away from T, then $p \geq_G 4$-away from T.*

Proof. Suppose, for the sake of contradiction, that there exists a vertex $q \in V(T)$ which $d_G(p,q) \leq 3$. In addition, $p \geq_G 2$-away from T implies $d_G(p,q) \geq 2$; therefore, $d_G(p,q) = 2$ or 3. This essentially means we can add edge (p,q) to T to become a larger 2,3-tree, which contradicts the maximality of T. □

Claim 49. *If $p, q \in V(C) \setminus V(T)$ such that $p \geq_G 4$-away from T and $(p,q) \in E(C)$, then $q \geq_G 4$-away from T.*

Proof. Since $(p,q) \in E(C)$, we must have $d_G(p,q) \leq 2$. Take arbitrary $r \in V(T)$. It follows that $d_G(p,r) \geq 4$. From triangle inequality, it follows that $d_G(q,r) \geq d_G(p,r) - d_G(p,q) \geq 2$. Since this holds for arbitrary $r \in V(T)$, this implies $q \geq_G 2$-away from T. Applying Claim 48, we can boost q to be $\geq_G 4$-away from T. □

Now it's time to finish off the proof. Given $p \in V(C) \setminus V(T)$. Assume, for the contrary, that $d_G(p,q) \geq 2$ for every $q \in V(T)$. This is equivalent to saying that $p \geq_G$ 2-away from T. Since $V(T) \subseteq V(C)$ and C is a connected tree, there must me a path $(p, p_1, p_2, ..., p_f)$ in $E(C)$ where $p, p_1, p_2, ..., p_{f-1} \in V(C) \setminus V(T)$ and $p_f \in V(T)$.

With the assumption that $p \geq_G$ 2-away from T, we apply Claim 48 to boost it to $p \geq_G$ 4-away from T. We then iteratively apply Claim 49 with $(p, p_1), (p_1, p_2), ..., (p_{f-2}, p_{f-1}) \in E(C)$ to get $p_{f-1} \geq_G$ 4-away from T. However, $(p_{f-1}, p_f) \in E(C)$ implies $d_G(p_{f-1}, p_f) \leq 2$. Since $p_f \in V(T)$, we reach a contradiction. □

Lemma 50. *With probability at least* $1/2$*, every 2,3-tree of* G *whose vertices are in* B *has size at most* $\log(2N)$

Proof. Build a new graph G' where $V(G') = V(G)$ and $(p,q) \in E(G')$ iff $d_G(p,q) = 2$ or 3. The maximum degree of G' is most d^3. Notice that any 2,3-tree of G must be also be a connected tree in G'.

We now set $u = \log(2n)$. A result from [Knu69] implies that for any vertex $v \in G'$, there are at most $\frac{1}{(d^3-1)u+1}\binom{d^3u}{u} \leq (ed^3)^u$ sized-u trees in G' that contains v. Therefore, there are at most $N(ed^3)^u$ sized-u trees of G'. We also notice that for any sized-u tree T in G', the probability of $V(T) \subseteq B$ is at most p^u. Thus, the expected number of sized-u 2,3-trees of G is at most

$$\sum_{\text{T = sized-}u\text{ tree of }G'} \Pr\left(V(T) \in B\right) \leq p^u \cdot \left(\text{number of sized-}u\text{ tree of }G'\right) \leq N(eDp)^u < 1/2$$

given the condition 1. Finally, we apply Markov's Inequality.

$$\Pr\left(G \text{ has at least 1 sized-}u\text{ 2,3-trees}\right) < 1/2$$

which is equivalent to saying that the probability of having no sized-u 2,3-trees of G is at least $1/2$. □

Our goal in the first pass is to show the following lemma.

Lemma 51. *With probability at least* $1/2$*, every 1,2-tree of* G *whose vertices are in* B *has size at most* $(d+1)\log(2N)$

Proof. Consider the scenario which every 2,3-tree of G whose vertices are in B has size at most $\log(2N)$ – this happens with probability at least $1/2$ according to CLemma 50..

Let C be an arbitrary 1,2-tree of G whose vertices are entirely in B. Also let T be the maximal 2,3-tree whose vertices are entirely in C. So we have $|T| \leq \log(2N)$.

By Lemma 47, every vertex in $V(C) \setminus V(T)$ is a neighbor (in G) to some vertex in $V(T)$. Since each vertex has at most d neighbors in G, we must have $|V(C) \setminus V(T)| \leq d \cdot |V(T)|$. Thus, $|V(C)| = |V(T)| + |V(C) \setminus V(T)| \leq (d+1) \cdot |V(T)| \leq (d+1) \cdot \log(2N)$.

In conclusion, with probability at least $1/2$, every 1,2-tree of G whose vertices are entirely in B must have size at most $(d+1) \cdot \log(2N)$. □

10.3 The Second Pass

Given the coloring from the First Pass, we can make the following assumption.

Assumption 52. *There is no 1,2-tree of size greater than* $(d+1)\cdot\log(2N)$ *all of whose vertices in* B*.*

This is because the probability of this happening is at leas $1/2$. So we can repeat the first pass to achieve this in expected $O(1)$ rounds. In the Second Pass, we fix the coloring by recoloring all the hypernodes that belong to the *bad* edges in such a way that each edge is non-monochromatic.

We call a hyperedge *dangerous* if it contains at least αn vertices that belong to *bad* edges, and let D denote the set of *dangerous* edges. By definition, *bad* edges are also *dangerous*, and thus $B \subseteq D$. In fact, with this definition, we only need to focus on *dangerous* edges. This is because:

Claim 53. *A non-dangerous edge will remain non-monochromatic no matter how we recolor.*

Proof. For an edge $e \notin D$, on the one hand, $e \notin B$, and thus it has at least αn Red vertices, and at least αn Blue vertices. On the other hand, since it is non-dangerous, it has less than αn vertices that belong to *bad* edges. In other words, less than αn vertices gets recolored, and thus after recoloring, there must still be at least 1 Red and 1 Blue vertices in e. That is, e remains non-monochromatic. □

Therefore, we now narrow our attention to recolor in such a way that makes sure all *dangerous* edges are non-monochromatic. We first prove that such a proper 2-coloring exists.

Lemma 54. *Given the Assumption 52, there exists a proper 2-coloring by recoloring all the hypernodes that belong to the* bad *edges.*

Proof. Notice that we need to recolor at least αn hypernodes in each hyperedge. Therefore, the probability that a particular dangerous edge becomes monochromatic is at most $2^{1-\alpha n}$. In addition, each dangerous edge shares a vertex with at most d others. By the Symmetric Lovász Local Lemma, along with the condition in Equation 24, a proper 2-coloring must exist. □

Next, we show how to find such a proper 2-coloring. In fact, we exhaustively enumerate all possible ways to color the hypernodes that belong to the bad edges, and we prove that we can do this in polynomial time. We first divide B into a set of non-overlapping maximal 1,2 tress $C_1, C_2, ...C_m$, such that

$$\bigcup_{i\in[m]} C_i = B, \text{ and } C_i \cap C_j = \emptyset \text{ for } i \in [m], j \in [m], i \neq j$$

To do this, we use a greedy-like approach. Starting with a random vertex in B, we iteratively add new vertices in such a way that it maintains a valid 1,2-tree. Each new vertex can be determined via a Breadth First Search in the dependency graph G. If we cannot add any more vertex, it becomes a maximal 1,2-tree in B. We repeat this process until there is no vertex left in B. The final result is such partition given above.

Lemma 55. *For a* dangerous *edge* $e \in D$*, in the hypergraph it only intersects edges from one such maximal 1,2-tree* C_i*.*

Proof. We prove by contradiction. If, for the contrary, a dangerous edge e intersects edges from two distinct maximal 1,2-tree. That is, in the hypergraph, e intersects $u \in C_1$ and $v \in C_2$. It follows that e is adjacent to u and v in G, $d_G(e,u) = d_G(e,v) = 1$

Then, $d_G(u,v) \leq d_G(e,u) + d_G(e,v) = 1 + 1 = 2$. This means that we can add edge v to C_1, or add u to C_2, to become larger 1,2 trees, which contradicts the maximality of C_i. □

Following Lemma 55, we can now recolor the points that belong to the edges of each C_i separately, and in such a way that makes sure all dangerous edges intersecting edges in C_i are non-monochromatic. Recall that the First Pass guarantees $|C_i| \leq (d+1) \cdot \log(2N)$, it follows that enumerating all possible 2-colorings for each C_i takes

$$O(2^{|C_i|}) = O(2^{O((d+1)\log 2N}) = O(2^{O(\log N)}) = N^{O(1)} \tag{25}$$

time. Therefore, by enumerating all possible 2-colorings for each C_i separately, we can find the proper 2-coloring in polynomial time.

As a final note, recall that we make the Assumption 52 by claiming that we can repeat the first pass until there is no 1,2-tree of size greater than $(d+1) \cdot \log(2N)$ all of whose vertices in B. However, it seems to be impossible to check the size of every 1,2-tree in polynomial time. Luckily, there is a way to circumvent it.

Notice that we only need the size of each C_i to be at most $(d+1) \cdot \log(2N)$. Therefore, we can do the First Pass, generate and check the size of each C_i's. If some C_i is larger than $(d+1) \cdot \log(2N)$, we repeat the first pass; otherwise we start the recoloring process. Each first pass is successful with probability at least $1/2$, thus we only need expected $O(1)$ repetitions of the first pass.

11 Derandomization

As we have mentioned, the naive algorithm of enumerating across the probability space to find one object with the desired property has unbounded time complexity and may not even terminate. In this section, we introduce derandomization, a technique that derandomizes a derived randomized algorithm into a deterministic and fast algorithm. That is, an algorithm that is guaranteed to terminate in polynomial time. This could significantly help with the construction process for an object instance.

Proposition 56. *(Randomized Algorithm) Let A be a randomized algorithm that runs in $t(n)$ time using a random bit sampler. Note that $t(n)$ is an upper bound on the number of random bits used in the algorithm. Therefore, let $x = \{0,1\}^n$ be the input to A, and $r = \{0,1\}^{t(n)}$ be $t(n)$ random bits, we can write $A(x;r)$ for A's output of input x under random bits r. Then we have*

$$\Pr\left(A(x;r) \text{ accepts}\right) = \frac{1}{2^{t(n)}} \sum_{r \in \{0,1\}^{t(n)}} [A(x;r) \text{ accepts}] \tag{26}$$

We can enumerate all possible r and run $A(x;r)$ to simulate the random algorithm and make sure we accept x if A accepts x with positive probability. However, this enumeration technique takes $O(2^{t(n)}t(n))$ time which is exponential.

Proposition 57. *(Derandomization) If A is a probabilistic polynomial-time algorithm that runs in $t(n)$ using $r(n)$ random bits. There is a corresponding deterministic algorithm which runs in*

$2^{r(n)}t(n)$. *Therefore, if* $t(n) = O(poly(n))$ *and* $r(n) = O(\log n)$*, we have a polynomial derandomization version for* A.

11.1 Application on Graph Cut problem

We first define the cut and the size of a cut of undirectional graphs.

Definition 58. *For an undirectional graph* $G = (V, E)$*, a cut of the graph is a division of the vertex set* V *into* $C = (S, T)$ *i.e.* $S \cap T = \emptyset$ *and* $S \cup T = V$*. The size of a cut is the number of edges across* S *and* T*, which is*

$$size(C) = size(S, T) = \sum_{(u,v)\in E|(u\in S\wedge v\in T)\vee(u\in T\wedge v\in S)} 1$$

Then let's consider how large the cut we can obtain from a graph G.

Theorem 59. *For a graph* $G = (V, E)$*, there is a cut with size at least* $\frac{|E|}{2}$.

Proof. We can randomly assign vertices into S and T each with probability $\frac{1}{2}$. Then for each edge e, the probability it is cut is equal to the probability that its two ends are assigned to different set, which is of probability $\frac{1}{2}$. By the linearity of expectation, $\mathbb{E}[size(S, T)] = |E| \cdot \frac{1}{2} = \frac{|E|}{2}$. According to the first moment method, with positive probability, $size(S, T) \geq \frac{|E|}{2}$. Hence such cut must exist. □

The theorem is easily proven using probabilistic method, which also naturally gives a randomized algorithm for construction. That is, to sample $|V|$ random bits indicating each vertex to be assigned to S or T, and then compute the size of the cut, and repeat until we get a cut of size at least $\frac{|E|}{2}$.

The algorithm runs in $O(|V| + |E|)$ and requires $|V|$ random bits, and derandomizing it with enumeration takes $O(2^{|V|}|E|)$ time. To obtain a polynomial derandomized algorithm, we need to reduce the number of random bits used to $O(\log(|V|))$.

Definition 60. *(Pairwise-Independent) A collection of* n *random bits,* $x_1, x_2, \cdots, x_n$ *is called pairwise-independent iff every pair of bits* x_i, x_j *is independent, i.e.* $Pr(x_i) = Pr(x_i|x_j)$ *for any* $i \neq j$.

Recall that for each edge (u, v), if the random bits for u and v are independent with each other, $\Pr((u, v) \text{ is cut}) = \frac{1}{2}$. Thus, if we use x_i to assign node i and $x_1, \cdots, x_n$ are pairwise-independent, we can conclude that $\mathbb{E}[size(S, T)] = \frac{|E|}{2}$.

Hence, it is sufficient to construct n pairwise-independent random bits $x_1, \cdots, x_n$ from $O(\log n)$ bits.

Proposition 61. *Let* $k = \lfloor \log_2 n \rfloor + 1$*, and* $B(i)$ *to be an vector of* i*'s first* k *binary bits i.e.* $i = \sum_{j=0}^{k-1} 2^j B(i)_j$*. Then let* $R = [r_0 r_1 \cdots r_{k-1}]$ *be a vector of* k *random bits. We construct the pairwise-independent random bits as*

$$x_i = R^T B(i) \mod 2$$

Clearly, for all $i \geq 1$, $x_i \sim \text{Bernoulli}(\frac{1}{2})$. This is because $B(i)$ must have at least one none-zero bit, and consider the lowest non-zero bit that flips the outcome with probability $\frac{1}{2}$ (because the corresponding random bit has probability $\frac{1}{2}$ to be 1).

Then note that the event $x_i \neq x_j$ is equivalent to the event $(x_i + x_j) \mod 2 = 1$ and $(x_i + x_j) \mod 2 = [R^T(B(i) - B(j))] \mod 2$.

Further, $(B(i) - B(j)) \mod 2$ must have at least one none-zero bit, similarly, $(x_i + x_j) \mod 2 \sim \text{Bernoulli}(\frac{1}{2})$.

Therefore, no matter what distribution x_j has, the distribution of x_i is still $\text{Bernoulli}(\frac{1}{2})$, which proves the pairwise-independency of $x_1, \cdots, x_n$.

We have got a derandomized algorithm of finding a cut of size at least $\frac{|E|}{2}$.

1. Enumerate all possible value combinations of k random bits $r_0, \cdots, r_{k-1}$.
2. For each assignment, construct $x_1, \cdots, x_n$ using Proposition 61.
3. Check if the resulted cut has size $\geq \frac{|E|}{2}$, and output the cut if so.

This algorithm runs in $O(2^k|E|) = O(|V| \cdot |E|)$, and is a deterministic polynomial construction.

12 Final Remarks

The probabilistic method is a powerful non-constructive method that elegantly proves the existence of some object with desired property. Shown in the early sections, oftentimes the probabilistic method relies on probability bounds as common tools to calculate the probability. In later sections, we also discuss the inspired exciting research on the construction process to find an object with desired property, whose existence is guaranteed by the probabilistic method.

Despite its initial usage in combinatorics and graph theory, the probabilistic method now extend its applications to areas such as number theory, linear algebra, theoretical computer science, information theory, etc. A number of longstanding open problems, such as bounds on the Ramsey Number, stability number, the largest element of distinct-sum sets, etc, continue to motivate ongoing important research.

References

[AH76] K. Appel and W. Haken. "Every map is four colourable". In: *Bulletin of the American Mathematical Society 82* (1976), pp. 711–712.

[Alo91] N. Alon. "A parallel algorithmic version of the Local Lemma". In: *FOCS* (1991), pp. 586–593.

[AS16] Noga Alon and Joel H Spencer. *The probabilistic method*. John Wiley & Sons, 2016.

[Bec91] J. Beck. "An algorithmic approach to the Lovász Local Lemma I". In: *Random Struct. Alg.* 2 (1991), pp. 343–365.

[Car79] Y. Caro. "New Results on the Independence Number". In: *Technical report* (1979).

[EL74] Paul Erdős and László Lovász. "Problems and results on 3-chromatic Hypergraphs and some related questions". In: *Coll Math Soc J Bolyai* 10 (Jan. 1974), pp. 609–627.

[Erd47] P. Erdős. “Some remarks on the theory of graphs”. In: *Bull. Amer. Math. Soc.* 53.4 (Apr. 1947), pp. 292–294.

[Knu69] Donald E. Knuth. *The Art of Computer Programming, Volume II: Seminumerical Algorithms*. Addison-Wesley, 1969.

[Lov73] L. Lovász. “Coverings and coloring of hypergraph”. In: *Proceedings of the Fourth Southeastern Conference on Combinatorics, Graph Theory, and Computing* (1973), pp. 3–12.

[Spe75] Joel Spencer. “Ramsey’s Theorem - A New Lower Bound”. In: *J. Comb. Theory, Ser. A* 18.1 (1975), pp. 108–115.

[Wei81] V. K. Wei. “A Lower Bound on the Stability Number of a Simple Graph”. In: *Technical Report* (1981).