

APM461

Combinatorial Methods

by Brandon Papandrea

The following is based on lecture notes taken during the Winter 2026 offering of APM461: Combinatorial Methods, at the University of Toronto. No set textbook was used, but reference books included *Probabilistic Methods* by Allen & Spencer, *The Linear Algebra Methods in Combinatorics* by Babai and Frankl, and *A Course in Combinatorics* by van Lint and Wilson. The notes are broken up in sections based on the week they were taught, and not necessarily broken up based on the textbook chapters. The intention is for these notes to be a polished version of my own lecture notes that allows me to revise and look over the material multiple times, and thus should not be considered a primary source for learning about combinatorial methods

# Contents

---

<b>1</b>	<b>Week 1</b>	<b>1</b>
1.1	Matchings in Bipartite Graphs . . . . .	1
1.1.1	Hall's Theorem . . . . .	1
<b>2</b>	<b>Week 2</b>	<b>7</b>
2.1	Flows & Cuts in Graphs . . . . .	7
2.1.1	Basics Definitions and Properties . . . . .	7
2.1.2	Max Flow and Cuts . . . . .	8
2.1.3	Hall's Theorem, Revisited . . . . .	11
<b>3</b>	<b>Week 3</b>	<b>13</b>
3.1	Posets . . . . .	13
3.2	Dilworth's Theorem . . . . .	14
3.3	Sperner's Theorem & LYM Inequality . . . . .	15
3.3.1	Sperner's Theorem . . . . .	15
3.3.2	The LYM Inequality . . . . .	16
<b>4</b>	<b>Week 4</b>	<b>18</b>
4.1	A Magic Trick Using Matches . . . . .	18
4.2	The Erdős-Ko-Rado (EKR) Theorem . . . . .	19
4.2.1	Intersecting Families . . . . .	19
4.2.2	Necklaces . . . . .	20
4.3	Fermat's Little Theorem . . . . .	21
<b>5</b>	<b>Week 5</b>	<b>23</b>
5.1	Counting Necklaces Using Burnside's Lemma . . . . .	23
5.1.1	Burnside's Lemma . . . . .	23
5.1.2	Counting Necklaces and FLT . . . . .	24

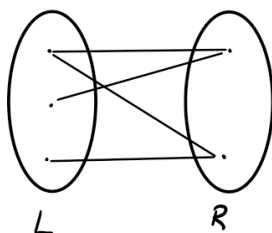
5.2	Catalan Numbers and The Ballot Theorem . . . . .	25
5.3	Probabilistic Results . . . . .	26
5.3.1	Markov's Inequality . . . . .	26
5.3.2	Chebyshev's Inequality . . . . .	26
<b>6</b>	<b>Week 6</b>	<b>28</b>
6.1	The Chernoff/Bernstein/Hoeffding Bounds . . . . .	28
6.2	The Erdős-Renyi Graph . . . . .	29

# Week 1

---

## 1.1 MATCHINGS IN BIPARTITE GRAPHS

We let  $G = (L, R, E)$  be a bipartite graph, where  $E \subseteq L \times R$ .

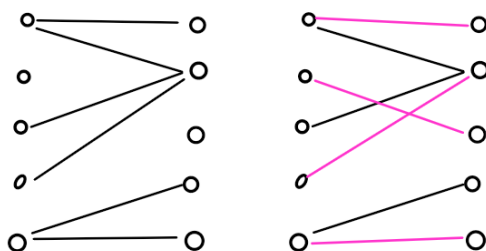


A **matching** in  $G$  is a subset  $M \subseteq E$  such that no two edges in  $M$  share a vertex. If  $|L| = |R| = n$ , then we can consider a **perfect matching**, which is a matching  $M$  such that  $|M| = n$ .

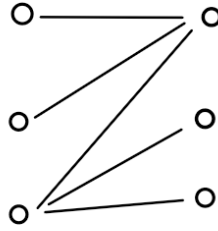
### 1.1.1 HALL'S THEOREM

Given an bipartite graph  $G$  with  $|L| = |R|$ , when does  $G$  contain a perfect matching?

Obviously, one condition is that each vertex must have a non-zero degree. The graph on the left has vertices of the degree 0, so no perfect matching exists, where as the graph on the right has a perfect matching, shown in pink.



Even if this condition is met, we might still have problems. Consider the graph below:



Notice that there are bottlenecks, areas where vertices have degree 1, and their respective edges meet at the same vertex. No matter which edge we choose, we will have to exclude one of these vertices from our matching, hence no perfect matching exists.

Surprisingly, finding one of these bottlenecks is necessary and sufficient in showing that a graph has no perfect matching. This is known as **Hall's Theorem**, and while it is an if and only if condition, it is not efficient if we do not know where the bottleneck is; this is called an NP Problem.

To formulate the theorem, we first state a definition. For a subset of vertices  $S \subseteq L$  in a bipartite graph  $G$ , we define the neighbourhood of  $S$  in  $G$  as

$$N_G(S) = \{v \in R : \exists u \in S, (u, v) \in E\}$$

**Theorem 1.1** (Hall's Theorem). *If  $G$  has no perfect matching, then there exists a set  $S \subseteq L$  such that  $|N_G(S)| < |S|$ .*

*Proof.* We proceed by induction on  $n$ , the number of vertices in  $L$  and  $R$ . We seek to prove the contrapositive, that is, if  $G$  is such that for all  $S \subseteq L$ ,  $|N_G(S)| \geq |S|$ , then  $G$  has a perfect matching. The claim is obvious if  $n = 1$ , so suppose  $n > 1$  and the theorem is true for all values less than  $n$ . We split our proof into two cases:

First, suppose that for all subsets  $S$  with  $0 < |S| < n$ ,  $|N_G(S)| \geq |S| + 1$ . Then let  $(u, v) \in E$  be an edge and consider the graph

$$G' = (L \setminus \{u\}, R \setminus \{v\}, E \setminus \{(u, v)\})$$

For all  $S \subseteq L \setminus \{u\}$ ,  $|N_{G'}(S)| \geq |S|$ , so by the induction hypothesis, we have a perfect matching  $M'$  in  $G'$ . Then the matching

$$M' \cup \{(u, v)\}$$

is perfect in  $G$ , as desired.

Now, suppose that there is a subset  $S \subseteq L$  with  $0 < |S| < n$  such that  $|N_G(S)| = |S|$ . We define two subgraphs,

$$G' = (S, N_G(S), E \cap (S \times N_G(S))) \quad G'' = (L \setminus S, R \setminus N_G(S), E \cap (L \setminus S \times R \setminus N_G(S)))$$

By the induction hypothesis,  $G'$  has a perfect matching,  $M'$ . Does it hold for  $G''$ . Let  $T \subseteq L \setminus S$ . We know by assumption that

$$|N_G(S \cup T)| \geq |S \cup T| = |S| + |T|$$

But we also know that

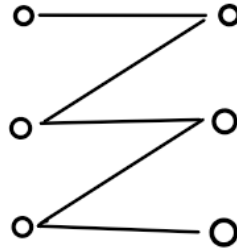
$$|N_G(S \cup T)| = |N_G(S)| + |N_{G''}(T)| = |S| + |N_{G''}(T)|$$

Thus,  $|N_{G''}(T)| \geq |T|$ , and the induction hypothesis applies, giving us a perfect matching  $M''$  in  $G''$ . Combining  $M'$  and  $M''$  gives the desired perfect matching in  $G$ .  $\square$

Second Half of Lecture Scribed by Brandon Papandrea.

### AN ALGORITHM TO CHECK FOR MATCHINGS

Let  $G = (L, R, E)$  be a bipartite graph with  $|L| = |R| = n$ . How can we check if there is a perfect matching in  $G$ ? There's actually a really easy way to compute algorithm that can allows us to check with near perfect accuracy if there is one. For now, let  $G$  be the graph shown below:



Let  $A$  be the  $n \times n$  adjacency matrix corresponding to  $G$ . The rows will represent vertices in  $L$ , and the columns will represent those in  $R$ . We let  $A_{ij} = 1$  if  $(i, j) \in E$ , and 0 otherwise. For the above graph, it is

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

Now let's replace every non-zero value in  $A$  with some random, non-zero real number. This will give us a new, modified adjacency matrix  $\tilde{A}$ :

$$\begin{pmatrix} 7 & 0 & 0 \\ e & 13 & 0 \\ 0 & 1 & \pi \end{pmatrix}$$

Now, we compute the determinant of  $\tilde{A}$ , which is  $7 \cdot 13 \cdot \pi = 91\pi \neq 0$ . Because the determinant is non-zero, we conclude that  $G$  has a perfect matching.

More explicitly, the algorithm to check for the existence of a perfect matching is as follows:

1. Compute the adjacency matrix  $A$
2. For each edge  $(i, j)$ , pick a random number  $\tilde{A}_{ij} \in \{1, 2, \dots, M\}$ . For all other  $(i, j)$ , set  $\tilde{A}_{ij} = 0$
3. If  $\det \tilde{A} = 0$ , we have no perfect matching. If it is not zero, we do have a perfect matching

One may look at this algorithm and ask if the choice of  $\tilde{A}_{ij}$  for edges  $(i, j)$  matters. Indeed it does. If  $G$  has a perfect matching, not all choices will lead to a matrix with non-zero determinant, however, most choices will. If  $G$  has no perfect matching, no matter what choices we make, the determinant will be 0. We formalize this in the following theorem:

**Theorem 1.2.** *If  $G$ , with  $|L| = |R| = n$ , has a perfect matching, and we let  $\tilde{A}_{ij} \in \{1, \dots, M\}$  for all edges  $(i, j)$ , then the probability that the above algorithm tells us  $G$  has a perfect matching is at least  $1 - \frac{n}{M}$ .*

*If  $G$  has no perfect matching, the probability the algorithm tell us  $G$  has no perfect matching is 1.*

To prove this, we need to understand what the determinant function is representing when it is applied to an adjacency matrix. Given an  $n \times n$  matrix  $M$ , the formula for the determinant is

$$\det M = \sum_{\sigma \in S_n} (-1)^{\text{sgn}(\sigma)} \prod_{i=1}^n M_{i, \sigma(i)}$$

**Example 1.** *Taking*

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

*Then as the only permutations in  $S_2$  are the identity permutation and the transposition  $(1\ 2)$ , we get that*

$$\det M = M_{11}M_{22} + (-1)M_{12}M_{21} = ad - bc$$

*which we know to be the classic formula for the determinant of a  $2 \times 2$  matrix.*

What happens when  $M$  is an adjacency matrix? Well, each permutation in the sum will correspond to a hypothetical pairing of vertices in  $L$  and  $R$ , meaning each term in the corresponding product is a hypothetical edge. Edges that do not exist in  $G$  will be ommitted, as the corresponding entry in the matrix is 0.

Notice that if a perfect matching exists, there will be a permutation  $\sigma$  that corresponds to it, meaning that the pairings  $(i, \sigma(i))$  correspond to the edges in the matching. If  $G$  has no perfect matching, then for all  $\sigma \in S_n$ , there is an  $i$  such that  $(i, \sigma(i)) \notin E$ . Thus,  $\prod_i \tilde{A}_{i, \sigma(i)} = 0$ , and so  $\det \tilde{A} = 0$ .

Now suppose that  $G$  has a perfect matching. We define a matrix  $M(x_{11}, x_{12}, x_{13}, \dots, x_{nn})$  by

$$M_{ij} = \begin{cases} x_{ij} & \text{if } (i, j) \in E \\ 0 & \text{otherwise} \end{cases}$$

This is technically a function of  $n^2$  variables that maps to a modified adjacency matrix. We then define  $P(X_{11}, \dots, X_{nn})$  to be the polynomial  $\det M$ . It suffices to show that  $P(x_{11}, \dots, x_{nn})$  is a non-zero polynomial. The rough proof of this is as follows: take the permutation  $\sigma$  corresponding to the perfect matching. We know this gives a non-zero term in  $P$ . Moreover, as this permutation does not appear in another term in the sum, it cannot be cancelled out by another term, hence we get a non-zero output to  $P$ . This is formalized as follows:

**Lemma 1.3** (Schwarz-Zippel Lemma). *Let  $|A| = M$  be a subset of  $\mathbb{R}$ . If  $Q(y_1, \dots, y_k)$  is a non-zero polynomial of degree at most  $d$ , then*

$$\Pr[Q(\vec{y}) = 0] \leq \frac{d}{M}$$

where  $\vec{y} \in A^k$  is chosen uniformly.

*Proof.* We proceed by induction on  $k$ , the number of variables in  $Q$ . First suppose that  $k = 1$ . Then we know that  $Q$  has at most  $d$  roots as it is of degree at most  $d$ . The probability of one of  $M$  chosen numbers being one of these roots is at most  $\frac{d}{M}$ , as desired.

Now let  $k > 1$  and suppose the claim holds for  $k - 1$ . We may write  $Q$  as

$$Q(y_1, \dots, y_{k-1}, z) = Q_0(y_1, \dots, y_{k-1}) + Q_1(y_1, \dots, y_{k-1})z + \dots + Q_t(y_1, \dots, y_{k-1})z^t$$

where  $Q_i(y_1, \dots, y_{k-1})$  is non-zero with degree at most  $d - i$ . Now, observe that if all of  $Q$  is non-zero, so too must  $Q_t$ . Using the induction hypothesis and conditional probability, we conclude that

$$\begin{aligned} \Pr_{\substack{\vec{y} \in A^{k-1} \\ z \in A}}[Q(\vec{y}, z) = 0] &= \Pr_{\substack{\vec{y} \in A^{k-1} \\ z \in A}}[Q(\vec{y}, z) = 0 \cap Q_t(\vec{y}) = 0] + \Pr_{\substack{\vec{y} \in A^{k-1} \\ z \in A}}[Q(\vec{y}, z) = 0 \cap Q_t(\vec{y}) \neq 0] \\ &= \Pr_{\substack{\vec{y} \in A^{k-1} \\ z \in A}}[Q(\vec{y}, z) = 0 | Q_t(\vec{y}) = 0] \Pr_{\vec{y} \in A^{k-1}}[Q_t(\vec{y}) = 0] \\ &\quad + \Pr_{\substack{\vec{y} \in A^{k-1} \\ z \in A}}[Q(\vec{y}, z) = 0 | Q_t(\vec{y}) \neq 0] \Pr_{\vec{y} \in A^{k-1}}[Q_t(\vec{y}) \neq 0] \\ &\leq \Pr_{\vec{y} \in A^{k-1}}[Q_t(\vec{y}) = 0] + \Pr_{\substack{\vec{y} \in A^{k-1} \\ z \in A}}[Q(\vec{y}, z) = 0 | Q_t(\vec{y}) \neq 0] \\ &= \Pr_{\vec{y} \in A^{k-1}}[Q_t(\vec{y}) = 0] + \Pr_{z \in A}[Q(\vec{y}, z) = 0 | Q_t(\vec{y}) \neq 0] \\ &\leq \frac{d-t}{M} + \Pr_{z \in A}[Q(\vec{y}, z) = 0 | Q_t(\vec{y}) \neq 0] \quad (Q_t \text{ as at most } d-t \text{ roots by I.H.}) \\ &\leq \frac{d-t}{M} + \frac{t}{M} \quad (Q \text{ has at most } t \text{ roots}) \\ &= \frac{d}{M} \end{aligned}$$

□

There is also a combinatorial way of phrasing this lemma, where we say that

$$|\{\vec{y} \in A^k : Q(\vec{y}) = 0\}| \leq dM^{k-1}$$

## Week 2

---

### 2.1 FLOWS & CUTS IN GRAPHS

#### 2.1.1 BASICS DEFINITIONS AND PROPERTIES

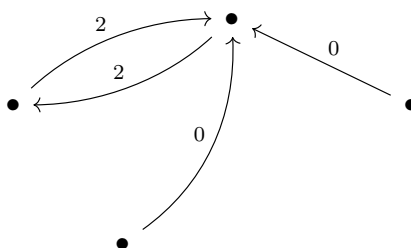
We let  $G = (V, E)$ , with  $E \subseteq V \times V$  be a graph such that  $(a, a) \notin E$  for all  $a \in V$  (no loops).

**Definition 2.1.** A **flow** is a map  $f : E \rightarrow \mathbb{R}_{\geq 0}$  such that for all  $v \in V$ ,

$$\text{NetFlow}(f, v) := \sum_{e \rightarrow v} f(e) - \sum_{v \rightarrow e} f(e) = 0$$

**Remark.** If context makes it evident,  $f$  may be excluded from the argument of  $\text{NetFlow}$ .

Essentially, a flow is an assignment of nonnegative values to the edges of a graph, such that for each vertex, the sum of the flow values coming into the vertex minus the sum of flow values leaving the vertex is 0. An example is shown below:



A specific type of flow occurs when exactly two vertices do not satisfy the above criterion:

**Definition 2.2.** An  $s - t$  **flow** is a map  $f : E \rightarrow \mathbb{R}_{\geq 0}$  such that  $\text{NetFlow}(f, v) = 0$  for all  $v \in V \setminus \{s, t\}$ .

We call  $s$  the **source** and  $t$  the **sink**.

One can think of an  $s - t$  flow as a flow where things go from  $s$  to  $t$ .

Our  $\text{NetFlow}$  function is defined for specific vertices, but can be extended to any subset of vertices in  $V$ . For a subset  $A \subseteq V$ , we define

$$\text{NetFlow}(A) = \sum_{\substack{u \xrightarrow{e} v \\ v \in A \\ u \notin A}} f(e) - \sum_{v \xrightarrow{e} u} f(e)$$

so we taking the sum of flows coming into  $A$  and subtracting flows that exit  $A$ .

**Proposition 2.3.**  $\text{NetFlow}(A) = \sum_{v \in A} \text{NetFlow}(v)$  for all  $A \subseteq V$ .

*Proof.* Unravelling the sums in the definition indicates that for each  $v \in A$ , we sum over all flows entering that vertex and subtract the sum of flows exiting it, which is precisely the sum of  $\text{NetFlow}$  over all vertices in  $A$ .  $\square$

We can also take  $A = V$  and get the net flow of the entire graph. This is always 0 because we are dealing with a flow. Define  $\text{Val}(f) = \text{NetFlow}(t)$ .

**Proposition 2.4.**  $\text{Val}(f) = -\text{NetFlow}(s)$ .

*Proof.* As we have an  $s - t$  flow, we get that

$$\text{NetFlow}(V) = \sum_{v \in V} \text{NetFlow}(v) = \text{NetFlow}(s) + \text{NetFlow}(t) = 0$$

and the claim follows.  $\square$

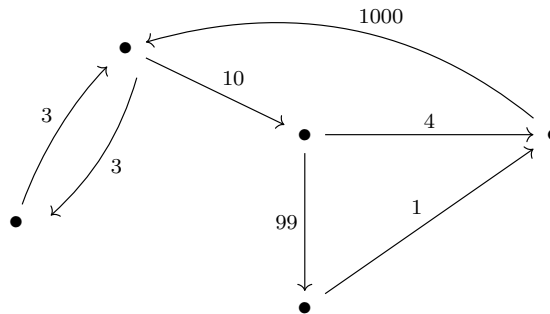
### 2.1.2 MAX FLOW AND CUTS

**Definition 2.5.** A **capacity function** is a map  $C : E \rightarrow \mathbb{R}_{\geq 0}$  that represents the maximum allowable flow in each edge of the graph.

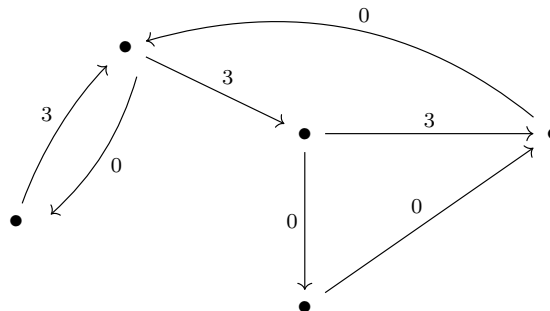
Given a capacity function, we would like to find a flow that maximizes flow while still satisfying the capacity.

**Definition 2.6.** The **Max  $s - t$  Flow Value** for a graph  $G$  and capacity function  $C$  is the maximum value of  $\text{Val}(f)$  over all  $s - t$  flows  $f$  such that for all  $e \in E$ ,  $f(e) \leq C(e)$ .

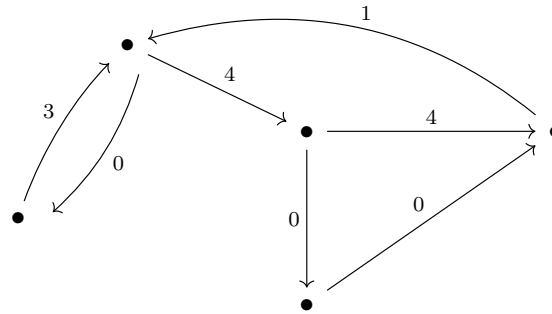
**Example 2.** Below is a graph, with edge labels denoted the capacity of each edge:



An  $s - t$  flow in this graph is the following:



But another flow can be made like this:



In either case, the value of the flow is 3; this is the maximum possible value for the flow.

The key observation is that all of our graphs can be split into two groups of vertices, one containing  $s$ , the other containing  $t$ . To keep our flow within the allowable range, the value of the flow cannot exceed the capacity values for edges between our two vertex groups:

This leads us to a new definition:

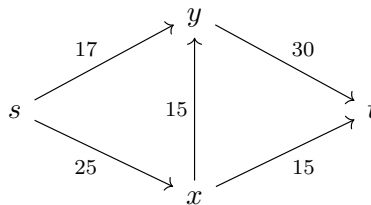
**Definition 2.7.** An  $s - t$  **cut** is a partition of  $V$ ,  $A \subseteq V$ , such that  $s \in A, t \notin A$ . Given a capacity function  $C$ , we define

$$\text{CutValue}(A) = \sum_{\substack{v \xrightarrow{c} u \\ v \in A \\ u \notin A}} C(e)$$

It should be evident that for any  $s - t$  flow  $f$  obeying the capacity function  $C$  and any  $s - t$  cut  $A$ , it must be the case that

$$\text{Val}(f) \leq \text{CutValue}(A)$$

**Example 3.** Consider the graph and constraint function shown below:



We have that

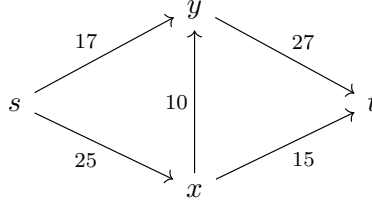
$$\text{CutValue}(\{s, x\}) = 47$$

$$\text{CutValue}(\{s, x, y\}) = 45$$

$$\text{CutValue}(\{s, y\}) = 55$$

$$\text{CutValue}(\{s\}) = 42$$

The best flow we can do is thus shown below, which has value 42, equal to the cut value of  $\{s\}$ .



The above example seems to indicate that the maximum  $s - t$  flow value is given by the smallest CutValue. This is indeed true.

**Theorem 2.8** (Max-Flow Min-Cut Theorem (Ford-Fulkerson Algorithm)). *For all graphs  $G$ , capacity functions  $C$ , and sources and sinks,  $s, t \in V$ , the maximum  $s - t$  flow  $f$  obeying  $C$  is the one such that*

$$\text{Val}(f) = \min_{s-t \text{ cuts } A} \text{CutValue}(A)$$

*Proof.* To find the max flow, we use the following algorithm:

1. Start with a flow  $f$  for which every edge is assigned the value 0.
2. Create a new graph  $G^* = (V, E^*)$ , where

$$E^* = \{(u, v) \in V \times V : ((u, v) \in E \wedge f((u, v)) < C((u, v))) \vee ((v, u) \in E) \wedge f((v, u)) > 0\}$$

The first condition says there is room to increase the value of the flow on that edge, while the second condition says there is room to decrease the value of the flow on that edge.

3. Look for a path from  $s$  to  $t$  in  $G^*$ ; write it as

$$s = v_0, v_1, \dots, v_k = t$$

4. For each  $i = 0, 1, \dots, k - 1$ , increase  $f((v_i, v_{i+1}))$  by 1 or decrease it by 1.
5. Repeat steps 2-4 until it is impossible to find an  $s - t$  path in  $G^*$ .
6. Define  $A = \{v \in V : v \text{ may be reached from } s \text{ in } G^*\}$ .

We now claim that  $\text{CutValue}(A) = \text{Val}(f)$ . We have that

$$\text{CutValue}(A) = \sum_{\substack{v \xrightarrow{e} u \\ v \in A \\ u \notin A}} C(e)$$

Moreover,

$$\begin{aligned} \text{Val}(f) &= -\text{NetFlow}(s) \\ &= -\text{NetFlow}(A) \\ &= \sum_{\substack{v \xrightarrow{e} u \\ v \in A \\ u \notin A}} f(e) - \sum_{\substack{u \xrightarrow{e} v \\ u \notin A \\ v \in A}} f(e) \end{aligned}$$

As there are no edges in  $G^*$  from  $A$  to  $A^c$ , for all  $v \in A, u \in A^c$ ,  $f((v, u)) = C((v, u))$  and  $f((u, v)) = 0$ . Thus,

$$\begin{aligned} &= \sum_{\substack{v \xrightarrow{e} u \\ v \in A \\ u \notin A}} C(e) \\ &= \text{CutValue}(A) \end{aligned}$$

as desired.  $\square$

**Corollary 1** (Menger's Theorem). *Let  $G$  be an undirected graph and let  $s, t \in V$ . Then there are  $k$  edge disjoint path from  $s$  to  $t$  if and only if after deleting  $k - 1$  edges, there is still a path from  $s$  to  $t$ .*

*Proof.* The  $\implies$  direction is obvious.

Suppose if  $k - 1$  edges are removed, a path from  $s$  to  $t$  remains. From  $G$ , we can construct a directed graph  $\tilde{G}$  by taking each undirected edge  $\{u, v\}$  and making two directed edges  $(u, v), (v, u)$ . We create a capacity function  $C$  by setting each edge's capacity to 1.

We claim that the minimum cut value is at least  $k$ . Suppose not. Then let  $A$  be the set with a cut value of  $k - 1$ . Removing the edges connecting  $A$  to  $A^c$  will thus disconnect  $s$  and  $t$ , a contradiction. The maximum flow value is thus at least  $k$ . Start at  $s$  and greedily search for an edge we haven't visited yet until we reach  $t$ ; remove any loops if necessary. Repeating this process  $k$  times will give us  $k$  edge disjoint paths from  $s$  to  $t$ .  $\square$

### 2.1.3 HALL'S THEOREM, REVISITED

We can use flows and cuts to reprove Hall's Theorem on perfect matchings. Before we reprove it we need another result:

**Theorem 2.9.** *Given a bipartite graph  $G = (L, R, E)$  with  $|L| = |R| = n$ ,  $G$  has a perfect matching if and only if there is a flow  $f$  with  $\text{Val}(f) = n$  in the directed graph  $\tilde{G} = (\tilde{V}, \tilde{E})$ , where*

$$\begin{aligned} \tilde{V} &= L \cup R \cup \{s, t\} \\ \tilde{E} &= E \cup \{(s, u) : u \in L\} \cup E \cup \{(v, t) : v \in R\} \end{aligned}$$

with capacity function  $C$  giving each edge a value of 1.

*Proof.* For  $\implies$ , we first show that any matching  $M$  of size  $m$  corresponds to a flow in  $\tilde{G}$  of value  $m$ . Define  $f$  as follows:

$$f((u, v)) = \begin{cases} 1 & \text{if } u = s \vee u = t \vee (u, v) \in M \\ 0 & \text{otherwise} \end{cases}$$

This is indeed a valid flow obeying  $C$  and  $\text{Val}(f) = m$ . Assuming a perfect matching exists, then the above flow will have value  $n$ . There is no flow with value larger than  $n$ , as  $\text{CutValue}(\{s\}) = n$ .

For  $\Leftarrow$ , we show that a flow in  $\tilde{G}$  of value  $m$  corresponds to a matching  $M$  of size  $m$ . Given a flow  $f$  with  $\text{Val}(f) = m$ , without loss of generality we may assume that the image of  $f$  is  $\{0, 1\}$ . There exists  $m$  vertices in  $L$ ,  $u_1, \dots, u_m$ , such that  $f(s, u_i) = 1$  for all  $i$ . Thus for each  $u_i \in L$ , there must be exactly one  $v_j \in R$  such that  $f(u_i, v_j) = 1$ . Furthermore, for each  $v_j \in R$ , there is exactly one  $u_i$  such that  $f(u_i, v_j) = 1$ . A matching  $M$  may then be constructed as

$$M = \{(u_i, v_j) : f(u_i, v_j) = 1\}$$

If the flow  $f$  has value  $n$ , then this matching is of size  $n$ , so it is a perfect matching.  $\square$

This theorem allows us to connect perfect matchings to flows in a neat and simple way, making the proof of Hall's Theorem a lot simpler.

**Theorem 2.10** (Hall's Theorem (Again)). *If  $G$  has no perfect matching, then there exists a set  $S \subseteq L$  such that  $|N_G(S)| < |S|$ .*

*Proof.* We construct a directed graph  $\tilde{G}$  in the same manner as in Corollary 1, except we set our capacity function to be  $\infty$  for all edges  $(u, v) \in E$ . The maximum flow of  $\tilde{G}$  is still at most  $n - 1$ ; if not, a perfect matching would exist given that all other edges have capacity 1. Thus, there is a cut  $A$  such that  $\text{CutValue}(A) \leq n - 1$ . We write

$$A = \{s\} \cup A_L \cup A_R$$

where  $A_L \subseteq L, A_R \subseteq R$ . The cut value is finite, so there is no  $u \in A, v \notin A$  such that  $(u, v) \in E$ . Thus, as  $N(A_L) \subseteq A$ , it must be a subset of  $A_R$ . Moreover,

$$\begin{aligned} n - 1 &\geq \text{CutValue}(A) \\ &\geq |L \setminus A_L| + |A_R| \\ &= n - |A_L| + |A_R| \end{aligned}$$

Thus,  $|A_L| \geq |A_R| + 1$ . Combined with the fact that the neighbours of  $A_L$  are also in  $A_R$ , it must be the case that  $|N(A_L)| \leq |A_L|$ , as desired.  $\square$

## Week 3

---

### 3.1 POSETS

**Definition 3.1.** A **poset**, or *partially ordered set*, is a set  $S$  along with a binary relation  $\leq$  satisfying the following:

- (i) For all  $a \in S$ ,  $a \leq a$  (Reflexive)
- (ii) For all  $a, b, c \in S$ ,  $(a \leq b) \wedge (b \leq c) \implies a \leq c$  (Transitive)
- (iii) For all  $a, b \in S$ ,  $(a \leq b) \wedge (b \leq a) \implies a = b$  (Antisymmetric)

If the condition

- (iv) For all  $a, b \in S$ ,  $a \leq b$  or  $b \leq a$

is also satisfied, then we call it a **totally ordered set**.

**Example 4.** The following are all examples of posets and totally ordered sets

1.  $(\mathbb{R}, \leq)$  is totally ordered
2.  $\mathbb{R}^2$  with coordinate wise  $\leq$  is a poset
3.  $(\{a, b, c, \dots, x, y, z\}, \leq)$  is a totally ordered set
4.  $(\{\text{words in the English language}\}, \text{lexical order})$
5.  $(\mathcal{P}(A), \subseteq)$
6.  $(\mathbb{N} \setminus \{0\}, |)$ .

For our purposes, the poset  $(\mathcal{P}(X), \subseteq)$  is of relevance.

**Definition 3.2.** A **chain** in a poset  $(S, \leq)$  is a subset  $A \subseteq S$  such that for all  $a, a' \in A$ , either  $a \leq a'$  or  $a' \leq a$ .

We can actually draw chains as graphs, as seen here:

The opposite of a chain is a subset where no two elements are comparable. These are also important.

**Definition 3.3.** An **anti-chain** is a subset  $A \subseteq S$  such that for all  $a, a' \in A$ ,  $a \not\leq a'$  and  $a' \not\leq a$ .

### 3.2 DILWORTH'S THEOREM

Dilworth's Theorem is a result that pertains to the largest anti-chain in a poset.

**Definition 3.4.** For a poset  $(S, \leq)$ , a **chain cover** of  $S$  is a set of chains  $C_1, \dots, C_m$  such that

$$\cup_i C_i = S$$

In this case, we say the chain cover has size  $m$ .

**Theorem 3.5** (Dilworth's Theorem). The size of the largest anti-chain is equal to the smallest number of chains that cover the poset.

To prove this, we require some additional results:

**Proposition 3.6.** Given a poset  $(S, \leq)$ , for all anti-chains  $A \subseteq S$ , for all chain cover  $C_1, \dots, C_m$ , we have that  $|A| \leq m$ .

*Proof.* By definition of an anti-chain, at most one element of  $A$  can be in each  $C_i$ . The claim then follows as there are  $m$  chains.  $\square$

Another important result needed for Dilworth is related to vertex covers.

**Definition 3.7.** Given a graph  $G = (V, E)$ , a subset  $U \subseteq V$  is a **vertex cover** of  $G$  if for all  $(u, v) \in E$ , with  $u \in U$  or  $v \in U$ .

**Theorem 3.8** (König's Theorem). In a bipartite graph  $(L, R, E)$ , the maximum size of a matching  $M$  is equal to the minimum size of a vertex cover.

*Proof.* Suppose the maximum matching  $M$  has size  $|M|$ . It is evident that there can be no vertex cover  $U$  with size smaller than  $|M|$ , since a cover must contain a vertex from each edge in  $M$ .

To find a vertex cover of size  $|M|$ , we use the same method as in Theorem 2.9, except we now set the capacity function of all edges in the graph to  $\infty$ .

Now let  $A$  be the cut of the vertex set with smallest cut value. Write

$$A = \{s\} \cup A_L \cup A_R$$

Clearly there is no edges  $(u, v)$  for which  $u \in A_L$  and  $v \in R \setminus A_R$ . Thus,

$$|M| = \text{CutValue}(A) = |L \setminus A_L| + |A_R|$$

From here, it is easy to see that  $(L \setminus A_L) \cup A_R$  is a vertex cover of size  $|M|$ , as for any edge  $(u, v)$ , we either have  $u \in L \setminus A_L$ , or  $v \in A_R$ .  $\square$

We are now prepared to prove Dilworth's Theorem:

*Proof of Dilworth's Theorem.* Suppose  $|S| = n$ , and let  $S^-, S^+$  be two copies of  $S$ . We define a bipartite graph  $G = (L, R, E)$  as

$$\begin{cases} L = S^- \\ R = S^+ \\ E = \{(x^-, y^+) : x \leq y, x \neq y\} \end{cases}$$

By König's Theorem, there is a maximum matching  $M$  and minimum vertex cover  $U$ , where  $|M| = |U| = m$ . We will show that there is a chain cover of  $S$  with size  $n - m$ , and that there is an anti-chain of size  $n - m$ .

For the first claim, we consider the following algorithm: start with  $n$  chains of the form  $\{s\}$ , one for each  $s \in S$ . Then, if  $(x^-, y^+) \in M$ , then we merge the chains containing  $x$  and  $y$ . We repeat this process until we get a chain cover. Observe that because we have  $m$  matchings, we would have to do  $m$  mergings, thus creating a chain cover of size  $n - m$ , as desired.

For the second claim, we let  $U = U^- \cup U^+$ , such that  $U^- \subseteq S^-, U^+ \subseteq S^+$ . Define

$$A = \{x \in S : x^- \notin U^- \wedge x^+ \notin U^+\}$$

we claim  $A$  is an anti-chain. Suppose  $x < y$ . Note that  $(x^-, y^+) \in E$ . By the definition of our vertex cover,  $x^- \in U$  or  $y^+ \in U$ . If  $x^- \in U$ , then  $x \notin A$ . Similarly, if  $y^+ \in U$ , then  $y \notin A$ . Thus, both  $x, y$  cannot be in  $A$ , so  $A$  is indeed an anti-chain.

Finally, we show that  $|A| \geq n - m$ . Indeed,

$$|A| = |\{x \in S : x^- \notin U^- \wedge x^+ \notin U^+\}| \geq |S| - |U^-| - |U^+| = n - m$$

□

### 3.3 SPERNER'S THEOREM & LYM INEQUALITY

#### 3.3.1 SPERNER'S THEOREM

Let  $X = [n] = \{1, \dots, n\}$  and consider the poset  $(\mathcal{P}(X), \subseteq)$ . Sperner's Theorem is a result about the largest anti-chain in this poset:

**Theorem 3.9** (Sperner's Theorem). *The longest anti-chain in this poset has size  $\binom{n}{\lfloor \frac{n}{2} \rfloor}$*

To prove this, we first need to define some new objects. Let  $\sigma \in S_n$  be a uniformly random permutation. For  $B \subseteq [n]$ , we define an event

$$E_B = \{\sigma(1), \sigma(2), \dots, \sigma(|B|)\} = B$$

which means that the first  $|B|$  elements of the permutation are equal to the elements of  $B$ , but not necessarily in the same order.

**Example 5.**  $E_{\{1\}}$  is the event that  $\sigma(1) = 1$ , which has probability  $\frac{1}{n}$ .

$E_{\{1,3\}}$  is the event that either  $\sigma(1) = 1$  and  $\sigma(2) = 3$ , or the other way around. The probability of this event is  $\frac{2}{n(n-1)}$ .

In general,

$$\Pr(E_B) = \frac{1}{\binom{n}{|B|}} = \frac{|B|!}{n(n-1)\cdots(n-|B|+1)}$$

Now let  $A$  be an anti-chain. Consider the events  $\{E_B : B \in A\}$ . As the  $B$ 's are not subsets of each other, the  $E_B$ 's are disjoint events:

**Proposition 3.10.** *If  $B, B'$  are incomparable, then  $E_B, E_{B'}$  are disjoint events.*

*Proof.* If  $\sigma \in E_B, E_{B'}$ , then

$$\begin{aligned} \{\sigma(1), \sigma(2), \dots, \sigma(|B|)\} &= B \\ \{\sigma(1), \sigma(2), \dots, \sigma(|B'|)\} &= B' \end{aligned}$$

so  $B, B'$  are comparable. □

With this in mind, we can now prove our theorem:

*Proof of Sperner's Theorem.* Let  $A$  be an anti-chain and consider  $\{E_B : B \in A\}$ . As the events are disjoint, we have that

$$S = \sum_{B \in A} \frac{1}{\binom{n}{|B|}} \leq 1$$

As  $\binom{n}{\lfloor \frac{n}{2} \rfloor}$  is the largest binomial coefficient, we get that

$$\begin{aligned} \sum_{B \in A} \frac{1}{\binom{n}{\lfloor \frac{n}{2} \rfloor}} &\leq S \leq 1 \\ \implies |A| \frac{1}{\binom{n}{\lfloor \frac{n}{2} \rfloor}} &\leq 1 \\ \implies |A| &\leq \binom{n}{\lfloor \frac{n}{2} \rfloor} \end{aligned}$$

□

### 3.3.2 THE LYM INEQUALITY

Another relevant result about posets of  $\mathcal{P}([n])$  revolves around elements of an anti-chain of a given size  $k$ . For an anti-chain  $A$ , we define

$$A_k = A \cap \binom{[n]}{k}$$

where  $\binom{[n]}{k}$  is the set of subsets of  $[n]$  of size  $k$ .

**Theorem 3.11** (Lubell-Yamamoto-Meshalkin Inequality).

$$\sum_{k=1}^n \frac{|A_k|}{\binom{n}{k}} \leq 1$$

This means that the sum of the ratios of the elements of  $A$  of size  $k$ , to all subsets of  $[n]$  of size  $k$ , can never exceed 1.

We will prove LYM using a stronger statement, that for any anti-chain  $A$ ,

$$\sum_{B \in A} \frac{1}{\binom{n}{|B|}} \leq 1$$

*Proof.* let  $\sigma \in S_n$  be chosen uniformly. Then let

$$C_\sigma : \emptyset \subset \{\sigma(1)\} \subset \{\sigma(1), \sigma(2)\} \subset \cdots \subset \{\sigma(1), \dots, \sigma(n)\} = [n]$$

be the maximal chain built from  $\sigma$ . Now for each  $B \subseteq [n]$ , we define  $E_B$  to be the event in which  $B$  appears in this chain. In other words,

$$E_B = \{B = \{\sigma(1), \dots, \sigma(|B|)\}\}$$

We now compute the probability of  $E_B$ . There are  $n!$  permutations of  $[n]$ . Of these,  $|B|!(n - |B|)!$  of them send the first  $|B|$  entries to elements in  $B$ , and the remaining entries are elements of  $[n] \setminus B$ . Thus,

$$\Pr(E_B) = \frac{|B|!(n - |B|)!}{n!} = \frac{1}{\binom{n}{|B|}}$$

Now note that these events are pairwise disjoint. Indeed if  $B \neq B'$  and  $E_B, E_{B'}$  occur for the same permutation  $\sigma$ , then  $B, B'$  would appear as initial segments of the chain  $C_\sigma$ . It follows that either  $B \subset B'$  or  $B' \subset B$ , which contradicts that  $A$  is an anti-chain. Thus, For any  $B \in A$ , we have that

$$1 \geq \Pr\left(\bigcup_{B \in A} E_B\right) = \sum_{B \in A} \Pr(E_B) = \sum_{B \in A} \frac{1}{\binom{n}{|B|}}$$

□

## Week 4

---

### 4.1 A MAGIC TRICK USING MATCHES

To begin this lecture, we present a fun application of matchings that may prove useful at a party: some math magic.

This trick requires two magicians,  $M_1$  and  $M_2$ , and a volunteer. The volunteer will choose 5 cards at random from a deck of cards, and give them to  $M_2$ .  $M_2$  will then read out exactly 4 cards of the cards they were given in some order.  $M_1$ , who never directly sees the cards, is then able to correctly identify the missing 5th card. How are they able to do this?

Let's simplify this trick a little to get a better idea of what's going on. Suppose that the volunteer only picks up two cards, and also tosses a fair coin.  $M_2$  will read out both cards, and  $M_1$  just needs to guess which side the coin has landed on. This may seem impossible, but note that  $M_2$  has seen the coin flip, and has had time to converse with  $M_1$ . In conversing ahead of time, they could come up with some order that  $M_2$  could read out the cards in, so that  $M_1$  knows the outcome of the coin flip. For example, if the coin lands on heads, then  $M_2$  will read out the cards in ascending order by rank (or if they are the same rank, using CHaSeD order for suit), and if the coin lands on tails, they read them in descending order by rank (or reverse CHaSeD order).

The key observation to make here is that the order of the input is a set (the subset of cards drawn and outcome of the coin flip), but the output given to  $M_1$  is a sequence (a sequence of cards). The ordering of that sequence is what allows  $M_2$  to encode information about the fifth card.

To encode it, we will use a bipartite graph. Identify the deck with  $[52]$ , and let  $G = (L, R, E)$  be bipartite, with

$$L = \binom{[52]}{5}, \text{ the set of possible 5-cards received by } M_2$$

$R = \{(x_1, x_2, x_3, x_4) \in [52]^4 : x_i \neq x_j \ \forall i \neq j\}$ , the ordered sequences of cards  $M_2$  can read out

and  $E$  is defined so that  $A \in L$  is joined to  $b = (x_1, x_2, x_3, x_4) \in R$  if and only if  $b \subset A$ . Notice that for each  $A \in L$ ,  $\deg(A) = \binom{5}{4} \cdot 4! = 120$ , and for each  $b \in R$ ,  $\deg(b) = 48$ . We can use a perfect matching in  $G$  to find an encoding for  $M_2$  to read out the four cards so that  $M_1$  knows exactly what the fifth card is.

**Theorem 4.1.** *In the graph  $G$  described above, there is a matching which saturates all vertices of  $L$ . Equivalently, there is an injective map  $f : L \hookrightarrow R$  such that  $f(A)$  is an ordered 4-tuple of cards from  $A$  for all  $A \in L$ .*

*Proof.* We prove this via Hall's Theorem. Let  $S \subset L$ ,  $N(S)$  the set of neighbours, and  $E_S, E_{N(S)}$  the corresponding edge sets. As every vertex of  $S$  has degree 120,

$$|E_S| = 120|S|$$

and as every vertex of  $N(S)$  has degree 48,

$$|E_{N(S)}| = 48|N(S)|$$

Now,  $E_S \subseteq E_{N(S)}$ , so

$$120|S| \leq 48|N(S)| \implies |N(S)| \geq \frac{120}{48}|S| \geq |S|$$

as required. Hall's Theorem now applies.  $\square$

So when  $M_2$  receives a hand  $A \in L$ , they read out the 4-card ordered tuple  $b \in R$  that  $A$  is matched to.  $M_1$  recognizes this, identifies what  $A$  is using the matching, and reads out the fifth card!

Note that this only guarantees that a matching exists, not how to create it. There is some leeway in how you do this, and there are many methods, such as mnemonics.

## 4.2 THE ERDŐS-KO-RADO (EKR) THEOREM

### 4.2.1 INTERSECTING FAMILIES

**Definition 4.2.** A family of sets  $\mathcal{F}$  is **intersecting** if for all  $A, B \in \mathcal{F}$ ,  $A \cap B \neq \emptyset$ .

An intersecting family of sets is simply a family of pairwise disjoint sets. We ask the following question: If  $\mathcal{F} \subset \mathcal{P}([n])$  is intersecting, then how large can  $|\mathcal{F}|$  be?

**Example 6.** Consider the family of sets

$$\{A \in \mathcal{P}([n]) : \{1, 2\} \subseteq A, \{2, 3\} \subseteq A, \text{ or } \{1, 3\} \subseteq A\}$$

This is an intersecting family of sets, and its size is  $2^{n-1}$

**Theorem 4.3.** If  $\mathcal{F} \subset \mathcal{P}([n])$  is intersecting, then  $|\mathcal{F}| \leq 2^{n-1}$ .

*Proof.* We partition  $\mathcal{P}([n])$  into pairs  $(A, A^c)$ , of which there are  $2^{n-1}$ . Notice that any intersecting family  $\mathcal{F}$  must contain at most one set from each pair, as if it contained both sets in a pair their intersection would be empty. Thus,  $|\mathcal{F}| \leq 2^{n-1}$ .  $\square$

Let's explore a slightly more complicated question: If  $\mathcal{F} \subseteq \binom{[n]}{k}$ , how large can  $|\mathcal{F}|$  be? Notice that if  $k > n/2$ , then  $\binom{[n]}{k}$  is already an intersecting family, so let's assume  $k < n/2$ . Consider the family

$$\{A \in \binom{[n]}{k} : 1 \in A\}$$

Then this is an intersecting family of size  $\binom{n-1}{k-1}$ , and this grows at a rate of  $\Theta(n^{k-1})$ . Another intersecting example is the family

$$\{A \in \binom{[n]}{k} : A \cap \{1, 2, 3\} = \{1, 2\} \text{ or } \{1, 3\} \text{ or } \{1, 3\}\}$$

This is also intersecting, and is of size  $3\binom{n-3}{k-2} \ll \binom{n-1}{k-1}$ , and growing at a rate of  $\Theta(n^{k-2})$  when  $k \ll n$ . The size we found in our first example is actually the maximum size for such a family, an important result known as the Erdős-Ko-Rado (EKR) Theorem.

**Theorem 4.4** (Erdős-Ko-Rado Theorem). *If  $k < \frac{n}{2}$  and  $\mathcal{F} \subseteq \binom{[n]}{k}$  is intersecting, then*

$$|\mathcal{F}| \leq \binom{n-1}{k-1}$$

### 4.2.2 NECKLACES

To prove EKR, we will use a probabilistic method that involves a special object called a **necklace**.

**Definition 4.5.** *Let  $\Sigma$  be an alphabet. Two sequences  $(a_1, \dots, a_n), (b_1, \dots, b_n)$  of elements of  $\Sigma$  are equivalent if there is an  $i \in [n]$  such that*

$$a_1 = b_i, a_2 = b_{i+1}, \dots, a_{n-i+1} = b_n, a_{n-i+2} = b_1, \dots, a_n = b_{i-1}$$

*in other words, one can be made into the other by a shift of indices. An equivalence class of sequences of length  $n$  over  $\Sigma$  is called a **necklace**, and is denoted  $[a_1, \dots, a_n]$ .*

*Proof of EKR.* Let  $\sigma \in S_n$  be chosen uniformly, and let

$$N = [\sigma(1), \dots, \sigma(n)]$$

be the necklace corresponding to  $\sigma$ . For each  $\mathcal{F} \subseteq \binom{[n]}{k}$ , let  $\mathcal{F}_N$  be the set of all  $A \in \mathcal{F}$  such that  $A$  is also a contiguous subsequence of  $N$ . We then define the random variable  $X = |\mathcal{F}_N|$ . For each  $A \in \mathcal{F}$ , we define an indicator function

$$X_A = \begin{cases} 1 & A \in \mathcal{F}_N \\ 0 & A \notin \mathcal{F}_N \end{cases}$$

It follows that  $X = \sum_{A \in \mathcal{F}} X_A$ . What is the expectation of  $X$ ? Notice that there are  $n$  cyclic starting positions in  $N$  and each position gives a uniformly random size  $k$ -subset of  $N$ . Thus,

$$\mathbb{E}[X_A] = \Pr[A \in \mathcal{F}_N] = \frac{n}{\binom{n}{k}}$$

Thus,

$$\mathbb{E}[X] = \sum_{A \in \mathcal{F}} \mathbb{E}[X_A] = \sum_{A \in \mathcal{F}} \frac{n}{\binom{n}{k}} = |\mathcal{F}| \cdot \frac{n}{\binom{n}{k}}$$

We now make a key observation: As  $k < \frac{n}{2}$ , then  $X \leq k$ . Indeed, notice that two contiguous segments of  $N$  must intersect to both be in  $\mathcal{F}$ . If we had  $k + 1$  contiguous segments of length  $k$ , we remove a point not in any of these segments (which is possible as  $k < \frac{n}{2}$ ). We get a line with  $k + 1$  distinct length  $k$  intervals on this line. The left and rightmost intervals are disjoint, as required. Thus as  $X \leq k$ , so too is its expectation. We conclude that

$$\begin{aligned} |\mathcal{F}| \cdot \frac{n}{\binom{n}{k}} &\leq k \\ \implies |\mathcal{F}| &\leq \frac{\binom{n}{k} \cdot k}{n} \\ &= \frac{n! \cdot k}{k!(n-k)!n} \\ &= \frac{(n-1)!}{(k-1)!(n-k)!} \\ &= \binom{n-1}{k-1} \end{aligned}$$

□

### 4.3 FERMAT'S LITTLE THEOREM

Recall this basic theorem of number theory:

**Theorem 4.6** (Fermat's Little Theorem). *Let  $p$  be prime. Then for any integer  $a$ ,*

$$a^n \equiv a \pmod{n}$$

*Equivalently, if  $n \nmid a$ , then*

$$a^{n-1} \equiv 1 \pmod{n}$$

We can actually prove this theorem using necklaces. Consider necklaces over  $\Sigma = [a]$  of length  $n$ . For now we will let  $a = 2$ .

**Example 7.** *If  $a = 2$ , then*

- *If  $n = 2$ , there are 3 unique necklaces:  $[0, 0], [0, 1], [1, 1]$*
- *If  $n = 3$ , there are 4 unique necklaces:  $[0, 0, 0, 0], [0, 0, 1], [0, 1, 1], [1, 1, 1]$*
- *If  $n = 4$ , there are 6 unique necklaces:  $[0, 0, 0, 0], [0, 0, 0, 1], [0, 0, 1, 1], [0, 1, 0, 1], [0, 1, 1, 1], [1, 1, 1, 1]$ .*

*In general, there are  $2^n$  strings of length  $n$ , and each necklace is equivalent to  $n$  of them by rotation.*

For a string  $x$ , we will define  $R^i x$  to be that string rotated/shifted by  $i$  places; we let  $\mathbb{Z}_n$  act on the string by shifting. We let

$$\text{Period}(x) : \{i \in \mathbb{Z}_n : R^i x = x\}$$

be the values of  $i$  that fix  $x$ . This is a group, a subgroup of  $\mathbb{Z}_n$ . We have that

$$|\mathcal{O}(x)| = \frac{n}{|\text{Period}(x)|}$$

If we let  $n$  be prime, then the only way for  $x$  to be fixed is if all its elements are the same. Thus,

$$|\text{Period}(x)| = \begin{cases} n & x \in \{(0, 0, \dots, 0), (1, 1, \dots, 1)\} \\ 1 & \text{otherwise} \end{cases}$$

and so we have that the orbit is 1 in the first case, and  $n$  in the second case. Partitioning  $\Sigma^n$  into its orbits (which we know to be disjoint), we get 2 orbits of size 1 (one for  $\{0, 0, \dots, 0\}$ , and one for  $\{1, 1, \dots, 1\}$ ), and some number of orbits of size  $n$ , which we denote by  $\alpha$ . It follows that

$$|\{0, 1\}^n| = 2^n = 1 + 1 + n\alpha \implies \alpha = \frac{2^n - 2}{n}$$

Therefore, the total number of necklaces of  $\Sigma$  of length  $n$  is

$$\alpha + 2 = \frac{2^n - 2}{n} + 2$$

because we have a one-to-one correspondence between a necklace and its orbit. The fraction  $\alpha$  must be an integer, so we have that

$$2^n - 2 \equiv 0 \pmod{n} \iff 2^n \equiv 2 \pmod{n}$$

which completes the proof of Fermat's Little Theorem for  $a = 2$ . This generalizes for any  $a$ .

We can further generalize this for composite values of  $n$ . Suppose that  $n = pq$  for primes  $p, q$ . We determine the number of necklaces of size  $n$  over  $[a]$ . The orbit sizes of the action of  $\mathbb{Z}_n$  on  $[a]^n$  must divide  $n$ , meaning they are either 1,  $p$ ,  $q$ , or  $pq = n$ . Strings with orbit size  $pq$  are the constant strings, and there are  $a$  of them.

For orbits of size  $q$ , if  $x \in [a]^n$  is in this orbit, then

$$|\text{Period}(x)| = n/q = p \iff \text{Period}(x) = \{0, q, 2q, \dots, (p-1)q\}$$

because  $\text{Period}(x)$  is a subgroup of  $\mathbb{Z}_n$  of size  $p$ .  $x$  has this period if and only if the first  $q$  elements repeat themselves. Such elements may be arbitrarily chosen from  $[a]$  in  $a^q$  ways. In  $a$  of these choices, all elements of the string are the same, so the size of the period is 1. Thus, the number of orbits of size  $q$  is  $a^q - a$ . Similarly, there are  $a^p - a$  orbits of size  $p$ . Continuing with the process done previously gives a generalization of FLT for these values of  $n$ .

## Week 5

---

### 5.1 COUNTING NECKLACKES USING BURNSIDE'S LEMMA

We continue our study of necklaces by deriving a way to count the number of necklaces over  $[a]$  of length  $n$  using Burnside's Lemma.

#### 5.1.1 BURNSIDE'S LEMMA

Burnside's Lemma is a theorem in abstract algebra, and describes a formula for the number of orbits of a given group action in terms of the number of fixed points. For a group  $G$  acting on a set  $S$ , and  $g \in G$ , we let

$$\text{Fix}(G) = |\{s \in S : gs = s\}|$$

Be the number of elements fixed by  $g$ .

**Theorem 5.1** (Burnside's Lemma). *The number of orbits of the given group action is*

$$\frac{1}{|G|} \sum_{g \in G} \text{Fix}(g)$$

*Proof.* We have that

$$\begin{aligned}
 \sum_{g \in G} \text{Fix}(g) &= \sum_{g \in G} \sum_{s \in S} 1_{gs=s} \\
 &= \sum_{s \in S} \sum_{g \in G} 1_{gs=s} \\
 &= \sum_{s \in S} |\text{Stab}_G(s)| \\
 &= \sum_{s \in S} \frac{|G|}{|\mathcal{O}(s)|} \\
 &= |G| \sum_{s \in S} \frac{1}{|\mathcal{O}(s)|} \\
 &= |G| \sum_{\text{orbits } \mathcal{O}} \sum_{s \in \mathcal{O}} \frac{1}{|\mathcal{O}(s)|} \\
 &= |G| \sum_{\text{orbits } \mathcal{O}} 1 \\
 &= |G| \cdot \text{number of orbits}
 \end{aligned}$$

and rearranging completes the proof.  $\square$

### 5.1.2 COUNTING NECKLACES AND FLT

Now, consider necklaces over  $[a]$  of length  $n$ . We let  $\mathbb{Z}_n$  act on  $[a]^n$  by rotation/shifting by  $i$ . Notice that each orbit is a necklace. By Burnside, we have that the number of such orbits is

$$\frac{1}{|G|} \sum_{g \in G} \text{Fix}(g)$$

where  $\text{Fix}(i)$  is the number of strings that are  $i$ -periodic (shifting by  $i$  does not affect the string). In particular, it is the number of strings that are  $\gcd(1, n)$ -periodic. This follows from Bézout's Theorem modulo  $n$ . Thus,

$$\begin{aligned}
 \text{number of necklaces} &= \frac{1}{n} \sum_{i=0}^{n-1} a^{\gcd(i, n)} \\
 &= \frac{1}{n} \sum_{d|n} (\text{number of } i \text{ such that } \gcd(i, n) = d) a^d \\
 &= \frac{1}{n} \sum_{d|n} \phi(n/d) a^d
 \end{aligned}$$

where  $\phi$  is the Euler Totient Function. This also gives us a generalization of FLT: as this value is an integer, we get that

$$n \mid \sum_{d|n} \phi(n/d) a^d$$

## 5.2 CATALAN NUMBERS AND THE BALLOT THEOREM

Consider an  $n \times n$  grid. Starting at  $(0, 0)$ , we may move up or to the right by 1. How many ways can we end up at  $(n, n)$ ? Given that we have to move up  $n$  times, and move right  $n$  times, giving us a total of  $2n$  moves, we get that there are

$$\binom{2n}{n} \approx \Theta\left(\frac{1}{\sqrt{n}} 2^{2n}\right)$$

such ways.

Now let's restrict our paths so that they cannot go above the diagonal  $y = x$ . How many paths are there now? The answer is approximately  $1/n$  of all paths. In particular, we get

$$\frac{1}{n+1} \binom{2n}{n}$$

This value is called the  **$n$ -th Catalan Number**, and they show up in a wide variety of combinatorial problems.

To derive this value, we first need to consider a special theorem that, once again, involves necklaces. Consider a necklaces of beads, each labelled either  $+1$  or  $-1$ . We say a bead is **special** if, starting at that bead, the partial sums in the clockwise direction are positive.

**Example 8.** Suppose we have the necklace

$$[+1, -1, +1, +1, -1, +1, +1, +1, -1, -1, +1, -1, +1]$$

Then the first  $+1$  is not special, since the 2nd partial sum is 0, while the second  $+1$  is special.

Suppose I know that the necklace has a certain number of beads labelled  $+1$ , and a certain number labelled  $-1$ . How can I arrange them to maximize the number of special beads?

**Theorem 5.2** (The Ballot Theorem). *Given  $a$  beads labelled  $+1$ , and  $b$  labelled  $-1$ , the maximum number of special beads is  $\max(a - b, 0)$ .*

*Proof.* We can always find two adjacent beads labelled  $+1$  and  $-1$  (assuming  $a, b \neq 0$ ). Delete these beads, and repeat this process until there are either all positive beads or all negative beads. If there are all positive beads, then there are  $a - b$  of them, and all are special. If all are negative, then none are special.  $\square$

Using this theorem, we prove our above claim about paths which don't cross the diagonal: Consider necklaces with  $(n + 1)$  beads labelled  $+1$  and  $n$  labelled  $-1$ . Each necklace has  $2n + 1$  representations as a string in  $\{+1, -1\}^{2n+1}$  (by rotation) with exactly  $n + 1$  beads labelled as  $+1$ 's. By the Ballot Theorem, exactly one out of each string representation has all partial sums positive. To construct a string with the desired number of beads, we just consider choose where the  $n + 1$  positive beads will go in the  $2n + 1$  positions. Thus, the number of desired walks is

$$\frac{1}{2n+1} \binom{2n+1}{n+1} = \frac{1}{n+1} \binom{2n}{n}$$

### 5.3 PROBABILISTIC RESULTS

We now move to discussing some results in probability. These will become useful in discussing some results in graph theory later on.

#### 5.3.1 MARKOV'S INEQUALITY

We let  $X$  be a nonnegative random variable. For our purposes, we assume that it is discrete.

**Theorem 5.3** (Markov's Inequality).

$$\Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}$$

*Proof.*

$$\begin{aligned} \mathbb{E}[X] &= \sum_a \Pr[X = a] \cdot a \\ &= \sum_{a \geq t} \Pr[X = a] \cdot a + \sum_{0 \leq a < t} \Pr[X = a] \cdot a \\ &\geq t \cdot \Pr[X \geq t] \end{aligned}$$

□

**Example 9.** Let  $X_1, \dots, X_n$  be independent coin flips, with  $\Pr[X_i = 1] = 0.1, \Pr[X_i = 0] = 0.9$ , and let  $X = \sum X_i$ . Then by Markov's Inequality

$$\Pr[X > 0.2n] \leq \frac{1}{2}$$

#### 5.3.2 CHEBYSHEV'S INEQUALITY

Let  $\mathbb{E}[X] = \mu$ , and let  $\sigma = \text{Var}(X) = \mathbb{E}[(X - \mu)^2]$ . If we apply Markov to  $(X - \mu)^2$ , then

$$\Pr[(X - \mu)^2 \geq t^2] \leq \frac{\mathbb{E}[(X - \mu)^2]}{t^2} = \frac{\sigma}{t^2}$$

More explicitly,

**Theorem 5.4** (Chebyshev's Inequality).

$$\Pr[|X - \mu| \geq t] \leq \frac{\sigma}{t^2}$$

**Example 10.** Using the example above, with  $X = \sum X_i$ , we have that

$$\mathbb{E}[X] = \sum \mathbb{E}[X_i] = 0.1n$$

$$\begin{aligned}
\sigma &= \mathbb{E}[(X - \mu)^2] \\
&= \mathbb{E}[(\sum_i X_i - 0.1n)^2] \\
&= \mathbb{E}[(\sum_i X_i - 0.1)^2] \\
&= \mathbb{E}[\sum_i \sum_j (X_i - 0.1)(X_j - 0.1)] \\
&= \sum_i \sum_j \mathbb{E}[(X_i - 0.1)(X_j - 0.1)] \\
&= \sum_i \mathbb{E}[(X_i - 0.1)^2] + \sum_{i \neq j} \mathbb{E}[X_i - 0.1] \mathbb{E}[X_j - 0.1] \\
&\leq n
\end{aligned}$$

Thus, we have that  $\Pr[|X - 0.1| > t] \leq \frac{n}{t^2}$ , and

$$\Pr[X > 0.2n] \leq \Pr[|X - 0.1n| > 0.1n] \leq \frac{n}{(0.1n)^2} = O\left(\frac{1}{n}\right)$$

## Week 6

---

We continue our discussion of relevant probabilistic results

### 6.1 THE CHERNOFF/BERNSTEIN/HOEFFDING BOUNDS

These results provide upper bounds on the deviation of a sum of i.i.d bounded random variables from its expected value. Let  $X_1, \dots, X_n$  be i.i.d with value either 0 or 1, with distribution so that

$$\Pr[X_i = 1] = p \quad \Pr[X_i = 0] = 1 - p$$

with  $0 \leq p \leq 1$ . Then for  $t \geq 0$ , and setting  $X = \sum X_i$ , we have that

$$\Pr[|X - np| > t] \leq e^{-\frac{t^2}{3n}}$$

Taking  $\varepsilon = \frac{t}{n}$ , we get

$$\Pr[|X - np| > \varepsilon n] \leq e^{-\frac{\varepsilon^2 n}{3}}$$

**Lemma 6.1.** Set  $Z_i = X_i - p$ , and  $Z_n = \sum_{i=1}^n Z_i$ . Then for any  $c > 0$ .

$$\Pr\left[\sum_i Z_i > t\right] \leq \frac{(pe^{c(1-p)} + (1-p)e^{-cp})^n}{e^{ct}}$$

*Proof.* We have

$$\begin{aligned} \Pr[Z_n > t] &= \Pr[cZ_n > ct] \\ &= \Pr[e^{cZ_n} > e^{ct}] \\ &\leq \frac{\mathbb{E}[\prod_i e^{cZ_i}]}{e^{ct}} && \text{(By Markov)} \\ &= \frac{\prod_i \mathbb{E}[e^{cZ_i}]}{e^{ct}} \\ &= \frac{(pe^{c(1-p)} + (1-p)e^{-cp})^n}{e^{ct}} \end{aligned}$$

□

More importantly, we have the following inequalities:

$$\begin{aligned}\Pr[|X - np| > \omega(\sqrt{n})] &= o(1) \\ \Pr[|X - np| > \Omega(n)] &= e^{-\Omega(n)}\end{aligned}$$

## 6.2 THE ERDŐS-RENYI GRAPH

The Erdős-Renyi model is a way of creating a random graph that has applications in probabilistic methods, often used to prove the existence of graphs with various properties.

**Definition 6.2.** Let  $n \in \mathbb{N}, p \in [0, 1]$ . Then the **Erdős-Renyi graph**  $G(n, p)$  has vertex set  $[n]$ , and for each  $\{i, j\} \in \binom{[n]}{2}$ , the edge  $(i, j)$  is in the edge set with probability  $p$ .

Now, we say that vertices  $v_i, v_j, v_k$  form a **triangle** there are edges between each of them. We ask the following: what is the distribution of the number of triangles in  $G(n, p)$ ?

For each  $\{i, j\}$ , we define a random variable that encodes when an edge exists:

$$Z_{ij} = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

Then for each triple  $\{i, j, k\} \in \binom{[n]}{3}$ , we have another random variable encoding when a triangle between the vertices is present:

$$\begin{aligned}X_{i,j,k} &= \begin{cases} 1 & Z_{ij} = Z_{jk} = Z_{ik} = 1 \\ 0 & \text{otherwise} \end{cases} \\ X &= \sum_{(i,j,k) \in \binom{[n]}{3}} X_{i,j,k}\end{aligned}$$

Let's compute the expectation and variance of  $X$ . We have that

$$\mathbb{E}[X] = \sum_{(i,j,k)} \mathbb{E}[X_{i,j,k}] = p^3 \binom{n}{3}$$

The variance computation is much more complicated:

$$\begin{aligned}
 \text{Var}(X) &= \mathbb{E} \left[ \left( X - p^3 \binom{n}{3} \right)^2 \right] \\
 &= \mathbb{E} \left[ \left( \sum_{(i,j,k)} (X_{(i,j,k)} - p^3) \right)^2 \right] \\
 &= \mathbb{E} \left[ \sum_{(i,j,k), (i',j',k') \in \binom{[n]}{3}} (X_{(i,j,k)} - p^3)(X_{(i',j',k')} - p^3) \right] \\
 &= \sum_{(i,j,k), (i',j',k') \in \binom{[n]}{3}} \mathbb{E}[(X_{i,j,k} - p^3)(X_{i',j',k'} - p^3)] \\
 &= \sum_{(i,j,k), (i',j',k') \in \binom{[n]}{3}: |(i,j,k) \cap (i',j',k')| \leq 1} \mathbb{E}[(X_{i,j,k} - p^3)(X_{i',j',k'} - p^3)] \\
 &\quad + \sum_{(i,j,k), (i',j',k') \in \binom{[n]}{3}: |(i,j,k) \cap (i',j',k')| = 2} \mathbb{E}[(X_{i,j,k} - p^3)(X_{i',j',k'} - p^3)] \\
 &\quad + \sum_{(i,j,k), (i',j',k') \in \binom{[n]}{3}: |(i,j,k) \cap (i',j',k')| = 3} \mathbb{E}[(X_{i,j,k} - p^3)(X_{i',j',k'} - p^3)] \\
 &= \sum_{(a,b,c,d) \in \binom{[n]}{4}} \mathbb{E}[(X_{a,b,c} - p^3)(X_{a,b,d} - p^3)] + \sum_{(i,j,k) \in \binom{[n]}{3}} \mathbb{E}[(X_{i,j,k} - p^3)^2] \\
 &= \sum_{(a,b,c,d) \in \binom{[n]}{4}} (\mathbb{E}[X_{a,b,c}X_{a,b,d}] - \mathbb{E}[p^3X_{a,b,c}] - \mathbb{E}[p^3X_{a,b,d}] + \mathbb{E}[p^6]) + \sum_{(i,j,k) \in \binom{[n]}{3}} \mathbb{E}[(X_{i,j,k} - p^3)^2] \\
 &= \Theta(n^4)(p^5 - p^6) + \Theta(n^3)(p^3 - p^6)
 \end{aligned}$$

By Chebyshev, we have

$$\Pr[|X - \binom{n}{3}p^3| > t] \leq O\left(\frac{n^4p^5 + n^3p^3}{t^2}\right)$$

and applying  $p = \frac{1}{2}$ ,

$$\Pr[|X - \frac{1}{8}\binom{n}{3}| > t] \leq O\left(\frac{n^4}{t^2}\right)$$

for  $t = \omega(n^2)$ ,  $O(\frac{n^4}{t^2}) \in o(1)$ . So

$$X \in \left[ \frac{1}{8}\binom{n}{3} - \omega(n^2), \frac{1}{8}\binom{n}{3} + \omega(n^2) \right]$$

with probability  $1 - o(1)$ . What values of  $p = p(n)$  can we say that  $X > 0$  with probability  $1 - o(1)$ ? Set  $t = \binom{n}{3}p^3$ , then

$$\begin{aligned}
 \Pr[X = 0] &\leq \Pr[|X - \binom{n}{3}p^3| \geq \binom{n}{3}p^3] \\
 &\leq O\left(\frac{n^4p^5 + n^3p^3}{n^6p^6}\right) \\
 &= O\left(\frac{1}{n^2p} + \frac{1}{n^3p^3}\right)
 \end{aligned}$$

If  $p = \omega(\frac{1}{n^2})$ , then  $O(\frac{1}{n^2p+n^3p^3}) = o(1)$ . If  $p = \omega(\frac{1}{n})$ , then  $\Pr[X = 0] = o(1)$ , while if  $p = o(\frac{1}{n})$ , then  $\Pr[X = 0] = 1 - o(1)$ .