MATH 340

Clean ver. by Zisheng Ye

Some remarks would appear in Chinese.
They are not that important and ignoring them is totally fine.
This note, however, does not cover all materials in class,
so attending classes is still the best option.

Class Eval: 20% Assignment, 4 out of 5
(20% midterm) Feb. 16
60% (80%) final

Monty Hall Problem

Q: Should we keep our choice or switch?
A: We should switch thus the prob would rise to $\frac{1}{2}$.

Secretary Problem

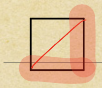
Exist strategy guarantees prob $> \frac{1}{3}$. $\frac{1}{e}$
First pass $\frac{1}{3}$, pick the best from now on.

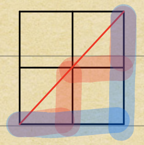
Catalon numbers

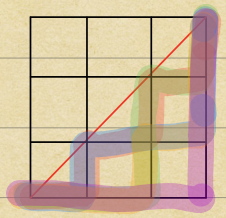
C_n - number of ways of walking from $(0,0)$ to (n,n) in R^2.
Only "one up" and "one right" allowed.
Never go above $x=y$

Examples:

$C_1 = 1$ $\quad$ $C_2 = 2$ $\quad$ $C_3 = 5$ $\quad$ $C_n = \frac{1}{n+1}\binom{2n}{n}$

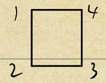




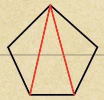


C_n also counts as triangulations of a labelled $(n+2)$gon.

$C_1 = 1$ $\quad$ $C_2 = 2$ $\quad$ $C_3 = 5$



13, 24



each time there's one way meeting at a unique vertex. Hence $5 \times 1 = 5$.

Finding the binomial formula:

$C_1, C_2, \cdots C_n$

$$C(x) = \sum_{n=1}^{n} C_n x^n \xrightarrow{\text{some relations}} C(x) = \frac{1-\sqrt{1-4x}}{2x} \xrightarrow{\text{Taylor series}} C_n = \frac{1}{n+1}\binom{2n}{n}$$

Stable Matchings

graphs: model pairwise relations between objects

n Boys $\{B_1, \cdots, B_n\}$ and n girls $\{G_1, \cdots, G_n\}$.

A perfect matching is a collection of n pairs, each consisting of one boy and one girl, s.t. each boy and girl are in exactly one pair.

There are possibly $n!$ perfect matchings without restrictions.

A perfect matching is a stable matching if there does not exist two pairs (B_1, G_1) (B_2, G_2) matched s.t.

$B_1 - G_2$ $\quad$ $B_2 - G_1$

Find stable matchings quickly: Boy Proposal Algorithm

- An arbitrary single boy proposes to a girl he likes most and has not proposed.
- The girl accepts if single, or likes him better.
- The algorithm terminates if all boys engaged or every single boy proposed to all the girls.

Thm. 1. BPA always terminates in a stable matching.

- Potentially there are many stable matchings.

2. In BPA no girl ever rejects a valid partner.

- Every boy is matched with his best choice.

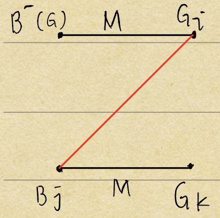
Denote boy's best choice as $G^+(B)$.

So every boy is matched to $G^+(B)$. reasonable.

Similarly denote girl's partner she likes least as $B^-(G)$.

Lemma. Every girl G is matched to $B^-(G)$.

Proof: Suppose not. i.e. $G_i \to B_j \neq B^-(G_i)$



Stable matchings M as shown.

Now G_i prefers B_j.

Only when $G_i > G_k$ would occur a new matching.

Then $G_k \neq G^+(B_j)$. Contradiction.

Basic Graph Theory

1. Definition: $G = (V, E)$

V: vertices (vertex) 节点

E: edges 节点组合. 可理解为连线.

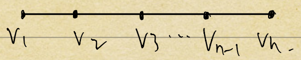
edge $e = ab$, a & b are ends of e

a, b incident to e

if a, b form an edge, a, b are adjacent / neighbors

2. Standard graph classes:

K_n: complete graph on n vertices

P_n: path. 

v_1 & v_n are ends

C_n: cycle Similar to P_n. $E(C_n) = \{v_1v_2, v_2v_3, \cdots v_{n-1}v_n, v_nv_1\}$

3. Subgraph:

A graph H is a subgraph of a graph G if $V(H) \subseteq V(G)$, $E(H) \subseteq E(G)$

A path/cycle in G is a subgraph of G.

Paths in K_n,

$K_1 = 1$ for $n \geq 2$, $K_n = \frac{n!}{2}$ 握手+不重复

Cycles in K_n: $\frac{(n-1)!}{2}$

Degree of a vertex: number of its neighbors

Handshaking Lemma: $\sum_{v \in V(G)} \deg(v) = 2|E(G)|$.

A connected graph: $\forall u, v \in V(G)$, there's a path with ends u, v.

完全连通

A component C of graph G:

A maximal connected subgraph of G.

最大连通子图

Every vertex of a graph is in a unique connected component.

Let comp(G) be number of components of G.

A forest is a graph with no cycles.

A tree is a connected forest.

Theorem: If G is a forest, then

$$|V(G)| - |E(G)| = comp(G)$$

由tree向上推更易理解

未连接单点也属于 $comp(G)$.

If G is a tree, then

$$|V(G)| - |E(G)| = 1.$$

每个vertex与上一层edge对应,

所以root没有对应到.

We say that $G\backslash e$ is obtained from deleting $e \in E(G)$.

If $V(G\backslash v) = V(G) - \{v\}$,

$E(G\backslash v)$ = consists of all edges of G not containing

Theorem:

If T is a tree with $|V(T)| \geq 2$,

then T has at least two leaves.

If T has exactly two leaves then it is a path.

Corollary:

Let G be a connected path s.t.

$\deg(v) \leq 2 \quad \forall v \in V(G)$

Then G is a cycle or a path.

Bipartition:

A partition (A, B) of $V(G)$ is a bipartition of G if every edge of G has one end in A and another in B.

$A \cap B = \phi$, $A \cup B = V(G)$

G is bipartite if it has a bipartition.

Theorem:

A graph is bipartite iff it contains no odd cycle.

Matchings & vertex covers

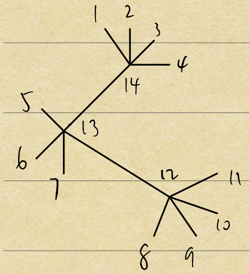
A matching M in G is a subset of $E(G)$ s.t.

every vertex is an end of at most one edge in M.

$\{ab, ac\}$ ✗ $\{ab, cd\}$ ✓

The matching number $\nu(G)$ of G is the max number of edges in a matching in G.

Obviously $\nu(G) \le \lfloor \frac{v(G)}{2} \rfloor$ 所有点尽可能对应,若odd则留下一个.



$\nu(G) \le 3$ since $X = \{12, 13, 14\}$,

then every edge of G has an end in X.

So $\nu(G) \le |X| = 3$.

Such $X \subseteq V(G)$ is a vertex cover in G.

Observation:

Let M be a matching in G, X be a vertex cover in G.

Then $|M| \le |X|$.

Obviously $|M| \le \nu(G) \le |X|$.

The vertex cover number $\tau(G)$ is the min number of vertices in a vertex cover of G.

Thus $\nu(G) \le \tau(G)$

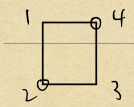
For K_n:

$\nu(K_5) = 2$ $\lfloor \frac{5}{2} \rfloor$ $\tau(K_5) = 4$ 完全连通.

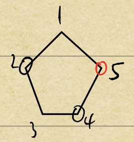
$\nu(K_n) = \lfloor \frac{n}{2} \rfloor$, $\tau(K_n) = n-1$.

For C_n:

$\nu(C_n) = \lfloor \frac{n}{2} \rfloor$, $\tau(C_n) = \lceil \frac{n}{2} \rceil$



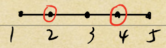
$\tau(C_4) = 2$



$\tau(C_5) = 3$

For P_n:

$\nu(P_n) = \lfloor \frac{n}{2} \rfloor$, $\tau(P_n) = \lfloor \frac{n}{2} \rfloor$

e.g. 

$\tau(P_5) = 2$.

Summary: $\nu(G) \le \tau(G)$

G	$\nu(G)$	$\tau(G)$
K_n	$\lfloor \frac{n}{2} \rfloor$	$n-1$
C_n	$\lfloor \frac{n}{2} \rfloor$	$\lceil \frac{n}{2} \rceil$
P_n	$\lfloor \frac{n}{2} \rfloor$	$\lfloor \frac{n}{2} \rfloor$

Proposition: $\tau(G) \le 2\nu(G)$

When it's max matching, $\tau(G) \le |X| = 2\nu(G)$

König's Theorem

If G is bipartite, then $\nu(G) = \tau(G)$.

证明略.

Applications of König's Theorem:

Let M be a matrix with 0 and $*$ entries,

$$\begin{bmatrix} * & 0 & 0 & * & 0 & 0 \\ * & 0 & 0 & 0 & 0 & * \\ 0 & 0 & 0 & * & 0 & * \\ 0 & * & * & 0 & * & * \\ * & 0 & 0 & * & 0 & * \end{bmatrix}$$

$*$ can be replaced by random real numbers

Let $t(M)$ be term-rank of M:

the max rank of a matrix obtained from M

by replacing $*$ with real numbers.

Let $\gamma(M)$ be cover number of M:

the min number of rows and columns

s.t. their deleting results in an all 0 matrix.

$\gamma(M) \geq t(M)$

In the example $4 \geq \gamma(M) \geq t(M)$

Theorem.

$\gamma(M) = t(M) \quad \forall M$

Pf: Let G be a bipartite graph with bipartition (R, C)

row column

A row R is joined by an edge to a column C

iff the edge (intersection) is $*$.

Then removing a set of rows & columns in X. iff X is a

vertex cover of G. So $\tau(G) = \gamma(M)$

By König's Theorem, G contains a matching F s.t.

$|F| = \gamma(M)$

We want to show $t(M) \geq \gamma(M) = |F|$

(In term of a matrix F is a set of stars s.t.

no two of them share a row or a column.)

Replace all entries in F by one, rest by zero.

Thus $t(M) \geq |F|$ $\square$

Perfect Matching

Def. A matching M is perfect in graph G if every vertex of G is an end of an edge in M.

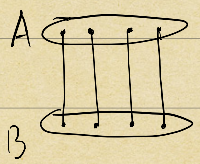
A graph G contains a perfect matching iff $\nu(G) = \frac{|V(G)|}{2}$

A graph is d-regular if every vertex of G has degree d.

e.g. C_n is 2-regular, K_n is $(n-1)$-regular

Theorem: If G is regular for some $d>0$ and bipartite, then G has a perfect matching.

Proof: If G is d-regular, $|E(G)| = d|A|$ as every edge has one end in A. Similarly $|E(G)| = d|B|$.



We need to show $\nu(G) \geq \frac{|V(G)|}{2}$. $\left(\tau(G) \geq \frac{|V(G)|}{2}\right)$

Let X be a vertex cover in G.

Since every edge has an end in X, and every vertex of X is the end of d edges, $|E(G)| \leq d|X|$

$$\downarrow$$

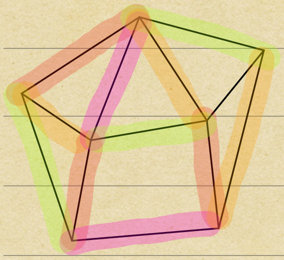
$$\frac{1}{2}\sum_{v \in V(G)} \deg(v) = \frac{1}{2} d |V(G)|. \text{ So } \frac{|V(G)|}{2} \leq |X| \quad \square$$

Edge Coloring.

$C: E(G) \to \mathbb{N}$ is an edge coloring of G

If $C(e) \neq C(f)$ $\forall e \neq f \in E(G)$, e, f share an end.

$\chi'(G)$ - edge-chromatic number of G is the min number of colors needed to color the edges of G.



4-edge coloring

max degree of vertex in G

Observasion: $\chi'(G) \geq \Delta(G)$

In C_3, $\Delta(C_3) = 2$, $\chi'(C_3) = 3$.

Theorem: If G is a d-regular bipartite graph then $\chi'(G) = d$.

Corollary: $\chi'(G) = \Delta(G)$ for every bipartite graph G.

Hall's Theorem

Let G be a bipartite graph with bipartition (A, B) then there exists a matching M in G which uses all vertices of A iff $|N(S)| \geq |S| \quad \forall S \subseteq A$ Hall's condition

Proof: Show if Hall's condition holds then $\nu(G) \geq |A|$

$$\Rightarrow \tau(G) \geq |A|$$

Let X be a vertex cover. Need: $|X| \geq |A|$.

Let $S = A - X$, then $N(S) \subseteq B \cap X$

$$|X| = |A \cap X| + |B \cap X| \geq |A \cap X| + |N(S)|$$

$$\geq |A \cap X| + |S| = |A \cap X| + |A - X| = |A|$$

Systems of distinct representatives

Given a collection $(S_1, S_2, \cdots S_k)$ of finite sets

A system of distinct representatives for this collection $(x_1, x_2, \cdots x_k)$

s.t. $x_i \in S_i$ for $i = 1, 2, \cdots k$

and $x_i \neq x_j$ for $i \neq j$.

Example:

$S_1 = \{2, 4, 5\}$ $\quad x_1 = 5$

$S_2 = \{1, 2\}$ $\quad x_2 = 2$ $\quad$ s.d.r

$S_3 = \{1, 3\}$ $\quad x_3 = 1$

$S_4 = \{2, 3\}$ $\quad x_4 = 3$

$S_5 = \{1, 2, 3\}$ $\quad$ No s.d.r. since $|S_2 \cup S_3 \cup S_4 \cup S_5| = 3$.

Theorem:

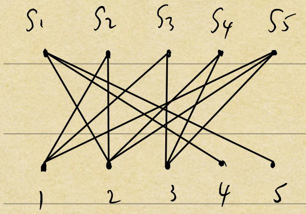
A collection $(S_1, S_2, \cdots S_k)$ has an s.d.r.

iff $\left| \bigcup_{i \in I} S_i \right| \geq |I| \quad \forall I \subseteq \{1, 2, \cdots k\}$

Proof:

Form a bipartite graph with bipartition (A, B),

$A = \{S_1, S_2, \cdots S_k\}$, $B = \bigcup_{i=1}^{k} S_i$



$x \subseteq B$ is joined to S_i by an edge in G

iff $x \in S_i$

A matching covering A is a collection $\{x_1, S_1\}, \{x_2, S_2\}, \cdots \{x_k, S_k\}$

s.t. $x_1, \cdots x_k$ are pairwise distinct.

It exists iff s.d.r. exists.

By Hall's Theorem, it exists iff $\forall I \subseteq \{1, 2, \cdots k\}$,

$$\left| \bigcup_{i \in I} S_i \right| \geq |I|$$

Graph Coloring.

A (proper, vertex) coloring of a graph G is a map assigns to every vertex v of G a color $C(v)$.
s.t. $C(u) \neq C(v)$, u, v adjacent

A coloring is a k-coloring if it uses k colors.
The chromatic number $\chi(G)$ of graph G is the min k s.t. G admits a k-coloring.

e.g. $\chi(G) \leq 1 \quad \Leftrightarrow \quad G$ edgeless

$\chi(G) \leq 2 \quad \Leftrightarrow \quad G$ bipartite

$\chi(K_n) = n$

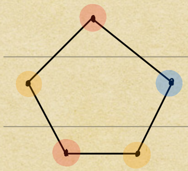
$\Downarrow$

Let the clique number $\omega(G)$ denote the max n s.t. G contains K_n as a subgraph

Then $\chi(G) \geq \omega(G)$.

$(\chi(G) \geq \chi(K_n) = n = \omega(G))$

C_5:



Odd cycle:

$\omega(C_{2k+1}) = 2 \qquad \forall k \geq 2$

$\chi(C_{2k+1}) = 3$

Erdö's Theorem:

$\forall\, k, L > 0$ there exists a graph $G(k, L)$

s.t. $\chi(G) \geq k$ and every cycle in G has length $\geq L$

$(\omega(G) = 2$ if $L \geq 4)$ $L = 3$ 时互相不同.

Greedy Coloring Algorithm

Input: graph G and ordering $(v_1, v_2, \cdots v_n)$ of $V(G)$

Output: A proper vertex coloring of G using colors $\{1, 2, \cdots k\}$ for some k.

Algorithm: $\forall\, i = 1, 2, \cdots k$, use color c not used on neighbors.

Choosing incorrect ordering can produce a suboptimal coloring.

(For C_6, $\begin{matrix} & 3 & 4 & \\ 2 & & & 1 \\ & 6 & 5 & \end{matrix}$ vs $\begin{matrix} & 3 & 2 & \\ 4 & & & 1 \\ & 5 & 6 & \end{matrix}$)

Exercise:

But there's always an ordering which produces a coloring using $\chi(G)$ colors.

Sol: Induction on $|V(G)|$.

Base case: $(|V(G)| = 1)$ trivial.

Induction step: Let e be an edge of G with vertices u and v, and coloring $G \setminus e$ would use at most $\chi(G\setminus e)$ colors.

So when adding back e, i.e. connecting u and v, degree of u or v would increase by 1, thus the potentially used color would gain by 1.

Therefore there's a coloring using $\chi(G\setminus e) + 1 = \chi(G)$ colors as desired.

Theorem. $\chi(G) \leq \Delta(G)+1$

There would be at most $\Delta(G)$ colors forbidden.

A graph G is **k-degenerate** if every subgraph H of G contains a vertex v s.t. $\deg_H(v) \leq k$

$\Rightarrow$ · G is 0-degenerate iff edgeless.

· 1-degenerate iff it is forest.

· Every graph G is $\Delta(G)$-degenerate.

Theorem. Let G be a k-degenerate graph, then $\chi(G) \leq k+1$.

$\forall v \in V(G)$ $\deg(v) \leq k$, applying greedy coloring algorithm.

Planar Graphs

In a planar drawing of a graph G, vertices of G are represented by points in the plane, and edges are represented by (simple) curves joining the points corresponding to their ends,

s.t. curves corresponding are disjoint except for their common ends.

☆可以做同构变换.

A graph G is planar if it admits a planar drawing.

Cutting the plane along the edges of a drawing of a graph partitions the plane into pieces called regions or faces of the drawing.

There will always be an unbounded region. → infinite / outer space

Reg(G) denotes the number of regions or faces in the planar drawing of G.

Euler's Formula

If G is a connected planar graph,

$|V(G)| - |E(G)| + Reg(G) = 2$ $\qquad (V - E + F = 2)$

Prove by induction.

- Induction is the most important method!

Proof Sketch: Induction on $|E(G)|$

Base case: ($|E(G)|=0$) $|V(G)|=1$ $Reg(G)=1$ ✓

Induction step:

Case 1. G is a tree:

$Reg(G)=1$, $|V(G)|-|E(G)|=1$. ✓

Case 2. G contains a cycle C

Choose $e \in E(C)$, let $G' = G \setminus e$, G' connected.

By the induction hypothesis,

$$|V(G')| - |E(G')| + Reg(G') = 2$$

$$\downarrow \qquad \downarrow \qquad \downarrow$$

$$|V(G)| - |E(G)+1| + (Reg(G)+1) = 2$$

- It remains to show that $Reg(G') = Reg(G)+1$

Jordan Curve Theorem:

Cutting the plane along a closed simple curve seperates the plane into two regions.

Def: For a region R in the planar drawing of G, let **length(R)** be the number of edges of G on the boundary of R with edges that only belong to R counted twice.

Something really hard to understand.

Check 7 Feb note if necessary.

Fact 1: $\sum_{R \text{ regions of drawing of } G} \text{length}(R) = 2|E(G)|$

Fact 2: If G is a connected graph with $|V(G)| \geq 3$ drawn in the plane then $\text{length}(R) \geq 3$.

For every region R of the drawing and if $\text{leng}(R) = 3$ then R is bounded by a cycle of length 3.

Theorem. Let G be a planar graph with $|V(G)| \geq 3$

then $|E(G)| \leq 3|V(G)| - 6$

Proof: We can finally make G connected.

So we assume G is connected.

$$2|E(G)| \underset{①}{=} \sum_{R \text{ regions of } G} \text{length}(R) \underset{②}{\geq} 3\,\text{Reg}(G)$$

$$|V(G)| - |E(G)| + \text{Reg}(G) = 2$$

$$6 = 3|V(G)| - 3|E(G)| + 3\,\text{Reg}(G)$$

$$\leq 3|V(G)| - 3|E(G)| + 2|E(G)|$$

$\Rightarrow |E(G)| \leq 3|V(G)| - 6$, as desired.

max number of edges in a planar graph with n vertices is $3n-6$

Corollary: K_5 is not planar.

$|E(K_5)| = 10$, $|V(K_5)| = 5$. $10 > 3 \cdot 5 - 6$

Corollary: Every planar graph contains a vertex with degree ≤ 5.

Applying handshaking, $|E(G)| \geq 3|V(G)|$ when ≥ 6, contradiction.

The Four Color Theorem

$\chi(G) \le 4$ for every planar graph G.

Six color: Every planar graph is 5-degenerate.

So $\chi(G) = 5+1 = 6$.

Theorem: Let G be a planar graph s.t. $|V(G)| \ge 3$

and G contains no cycles of length 3.

then $|E(G)| \le 2|V(G)| - 4$ Leave as exercise

Sol: $|V(G)| - |E(G)| + Reg(G) = 2$

$2|V(G)| - 4 = 2|E(G)| - 2Reg(G)$

from last theorem we know $2|E(G)| = \sum\limits_{R \text{ regions of } G} length(R) \ge 3Reg(G)$

Since G contains no cycles of length 3,

$2|E(G)| \ge 4Reg(G)$, plug in, $2|V(G)| - 4 \ge |E(G)|$ as desired.

Collary: $K_{m,n}$ is a bipartite graph with bipartition (A, B)

$|A| = m$, $|B| = n$. every vertex of A is adjacent to every vertex of B.

Prove $K_{3,3}$ not planar.

$|V(K_{3,3})| = 6$, $|E(K_{3,3})| = 9$. no length 3 cycle;

$9 > 2 \cdot 6 - 4$, so $K_{3,3}$ not planar.

Characterising Planar Graph

Is it true that every non-planar graph has a subgraph which is K_5 or $K_{3,3}$?

No. We can add extra vertices onto existing edges.

We say that a graph H is a subdivision of a graph G
if H is obtained from G by replacing some edges by paths with same ends, which otherwise don't share vertices.

Observation:

Let H be a subdivision of G, then H is planar iff G is planar.

- If G contains a subdivision of K_5 or $K_{3,3}$, then G is not planar.

Kuratowski's Theorem:

A graph G is planar iff G does not contain a subgraph which is a subdivision of K_5 or $K_{3,3}$.

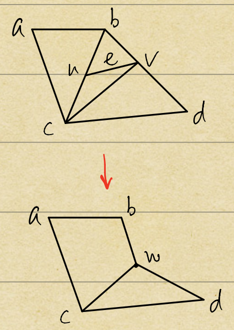
H is a subgraph of G if H can be obtained by repeatedly deleting vertices and/or edges.

Contracting an edge e with ends u and v in G:

delete e, set new vertex w,

Connect w with all vertices adjacent to u and v.

The resulting graph is denoted by G/e.



Minor Def:

H is a minor of G if H can be obtained by (Y/N) contracting edges in a subgraph of G.

If H is a minor of G, we will write $H \leq G$ or $H \leq_m G$.

Properties of minors:

- If H is a subgraph of G, $H \leq G$.
- $G \leq G$
- If $F \leq H$, $H \leq G$, then $F \leq G$
- If G is planar, $H \leq G$, then H is planar.
- If G is a subdivision of H then $H \leq G$.

Converse does not hold.

Five color theorem:

Proof Sketch: Induction on $|V(G)|$.

Base case: $(|V(G)|=1)$ trivial.

Induction step: G planar, $v \in V(G)$.

Case 1: $\deg(v) \leq 4$ trivial

Case 2: $\deg(v) = 5$

must not all connected otherwise K_6.

Select two not adjacent and contract.

We obtain a 5-coloring by using the color of post-vertex on two original vertices, so 4 colors in 5 adjacent vertices.

Then the center be the fifth color.

Four Color Theorem:

Hadwiger's Conjecture.

If G does not have a K_t minor, then $\chi(G) \leq t-1$.

$K_2 \npreceq G \longrightarrow \chi(G) \leq 1$.

(no edges)

$K_3 \npreceq G \longrightarrow \chi(G) \leq 2$

(forest)

$K_4 \npreceq G \longrightarrow \chi(G) \leq 3$

(G is 2-degenerate) , theorem of Dirac

$K_5 \npreceq G \longrightarrow \chi(G) \leq 4$. 4 - Color !

Kuratowski's Theorem (cont.)

Statement:

A graph G is planar iff G contains neither a subdivision of K_5 nor $K_{3,3}$ as a subgraph.

Thm: (minor variant of the theorem)

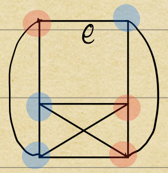
A graph G is planar iff it contains neither K_5 nor $K_{3,3}$ as a minor.

(Proof ignored, check original note).

Deduction: Every non-planar graph G contains a subdivision of K_5 or $K_{3,3}$ as a subgraph.

! It remains to show that if G contains K_5 or $K_{3,3}$ as a **minor**, then G contains a subdivision of K_5 or $K_{3,3}$ as a **subgraph**.

Reverse ? FALSE !

G  $K_{3,3}$.

G/e is K_5, but no K_5 minor in G.

Discrete Probability

Applications:

Al-Kindi: Permutation code

Brute Force: 26!

Optimized by analyzing letter frequencies.

Hypothesis Testing: Drug Testing

CS: Random Algorithms

Physics: Models of particle behaviour

Finance: Calculating prices of options

Def:

Prob space will be finite or countably infinite.

Sample space: Set of outcomes

An event $A \subseteq S$ is a subset of the sample space.

Prob distribution:

A function $P: S \to \mathbb{R}$ s.t.

1. $P(x) \geq 0 \quad \forall x \in S$
2. $\sum_{x \in S} P(x) = 1$
3. For event $A \subseteq S$, $P(A) = \sum_{x \in A} P(x)$

Conditional Prob.

For any prob distribution $P: S \to \mathbb{R}$ and events A, B,

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Baye's Theorem:

$P(A|B) = \dfrac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(A|B)P(B)$

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

Independence

Event A, B are **independent** if the outcome of one has no impact on the other.

We want $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = P(A)$.

So $P(B|A) = P(B)$, $P(A \cap B) = P(A)P(B)$

Events $A_1, \cdots A_n$ are independent if $\forall I \subseteq \{1, \cdots n\}$

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i)$$

Pairwise indep doesn't mean all of them indep!

The rate of total prob:

$S = B_1 \cup B_2 \cup \cdots B_n$, B_i pairwise disjoint:

$P(A) = P(A|B_1)P(B_1) + \cdots + P(A|B_n)P(B_n)$

Random Variable and Expectation

r.v.:

Let S be a prob space.

A random variable is a function $X: S \to R$

Expectation of X (avg. value): $E[X] = \sum_{s \in S} X(s)P(s)$

Thm: (Linearity of Expectation)

X, Y be r.v., $a, b \in R$.

$E[aX + bY] = aE[X] + bE[Y]$

Thm: If X and Y are indep r.v., $E[XY] = E[X]E[Y]$

Applications

1. Secretary Problem.

Suppose n is even, we can look at first $\frac{n}{2}$ cards without stopping. For the remaining cards we say stop once we find a larger one.

Lemma: winning prob $= P(A \cap B) = P(A) \cdot P(B|A) = \frac{1}{2} \cdot \frac{\frac{n}{2}}{n-1} > \frac{1}{4}$.

(A: second largest in first half, B: largest in second half)

- Prob of winning is larger if 3rd largest number is in the first half, but 1st and 2nd are in the second half.

  We win with prob $\geq \frac{1}{4} + \left(\frac{1}{2}\right)^4 = \frac{5}{16}$

Generalized version.

Look at first k cards.

Thm. We win $\sim \frac{k}{n} \log \frac{n}{k}$

Best k: $\frac{n}{e} \sim \frac{1}{e}$

- (?) If we want to select 2nd best, n is even, then we can't do better than $p = \frac{1}{k-1} \cdot \frac{n}{n-1}$.

2. Birthday Problem.

Given k people in a room, we want p(some people share a birthday).

How large must k be s.t. $p \geq \frac{1}{2}$?

Sol: Let $x_1, x_2, \cdots x_k$ be birthdays. $n = 365$ days / yr

$$P(\text{all different}) = \left(1-\frac{1}{n}\right)\cdot\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)$$

$$\leq e^{-\frac{1}{n}} e^{-\frac{2}{n}} \cdots e^{-\frac{k-1}{n}} \quad \text{Taylor's series}$$

$$= e^{-\frac{1}{n}\frac{k(k-1)}{2}} \leq \frac{1}{2}$$

$$\Rightarrow \log 2 \leq \frac{1}{n} \cdot \frac{k(k-1)}{2} \quad \text{配方法即可.}$$

$$\Rightarrow k \geq \frac{1}{2} + \sqrt{\frac{1}{4} + 2n \log 2} = 22.99994$$

So 23.

Balls & Bins

[original notes are too messy, just put some results and useful claims here]

Hashing: map $h: U \to T$ $\quad |T| \approx |S|$

What is $\max |X_i| = M$?

$h(i) = ai + b \mod p$, (a,b) random from $(1,p)$.

"birthday paradox": $h[m] \to h[n]$,

with prob ≈ 0.5 is not injective when $n \approx m^2$.

Thm: For N balls & N bins:

$$E(M) = E(\max_i |B_i|) \le \lceil 2\log n \rceil + 1$$

Claim in the proof: $(2k)! \ge k^k$ (ez proved by induction)

Markov Inequality:

$P(X \ge c) \le \frac{E[X]}{c}$, $c > 0$

$P(X \ge c) = P(f(x) \ge f(c)) \le \frac{E[f(x)]}{f(c)}$, f be strict increasing.

Chernoff Bound

- $p(X \geq (1+\delta)\mu) \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}$, $\mu = E[X]$, $\delta > 0$

For large enough n. $p\left(X \geq \frac{3\log n}{\log\log n}\right) \leq \frac{1}{n^2}$

- $p(X \geq (1+\delta)\mu) \leq e^{-\frac{1}{3}\mu\delta^2}$, $\delta \in (0,1)$

Proofs → Wiki page.

Claim mentioned in proofs: $\frac{\binom{n}{\frac{n}{2}}}{2^n} \sim \frac{1}{\sqrt{\frac{\pi}{2}n}}$

Randomized Quicksort

For distinct numbers $X_1, X_2, \cdots X_n$

Quicksort $(X_1, \cdots X_n)$ = Output [Quicksort(G^-), X_1, Quicksort(G^+)].

Runtime upperbound $\frac{n(n-1)}{2}$

Thm. Assuming $X_1, \cdots X_r$ are randomly sorted
expected running time is at most $2n\log n$

Thm. Almost surely, recursion depth of Quicksort $< 32\log n$
and always $\geq \log_2 n$

Claim: $\left(\frac{3}{4}\right)^4 < \frac{1}{e}$

Counting

Counting function $F: X \to Y$

A function f from X to Y assigns to each $x \in X$ a unique $y \in Y$ denoted by $f(x)$.

A function $f: X \to Y$ is:

① injective (1-to-1) if $f(x_1) \neq f(x_2)$ $\forall$ $x_1 \neq x_2$, $x_1, x_2 \in X$

② subjective (onto) if $\forall y \in Y$, there is $x \in X$ s.t. $f(x) = y$.

③ bijective: ① + ②

Notation: $[n] = \{1, 2, \cdots n\}$

Thm: If $f: [n] \to [k]$ is injective then $n \leq k$.

Pigeonhole principle:

- If function $f: [n] \to [k]$, $n < k$, then f is not injective.
- If n balls are distributed among k bins, $k < n$, then some bin(s) contains at least two balls.

Corollary: If for same f which is bijective. then $n = k$.

k^n ways to find functions: k ways for $f(i)$ $\forall$ $i \in [1, n]$

Multiplication principle:

The # of sequences of the form $(a_1, a_2, \cdots a_n)$

s.t. $a_i \in A_i$, $|A_i| = k_i$ is $k_1 k_2 \cdots k_n$ (by induction)

More generally, suppose we are interested in counting sequences $(a_1, \cdots a_n)$ s.t. $\forall i = 1, 2, \cdots n-1$, if $(a_1, \cdots a_i)$ are chosen, then there are exactly k_{i+1} ways of choosing a_{i+1} which leads to a valid sequence. Then there are $k_1 k_2 \cdots k_n$ ways.

- How many injections in $f: [k] \to [n]$?

$$\begin{array}{cccc} [f(1), & f(2), & \cdots & f(k)] \\ \downarrow & \downarrow & & \downarrow \\ n & n-1 & \cdots & n-k+1 \end{array} \text{ ways.}$$

So $\frac{n!}{(n-k)!}$ ways. While $k=n$, $0! = 1 \Rightarrow n!$ ways.

- # of k element subsets of $[n]$: $\binom{n}{k} = C_k^n = \frac{n!}{k!(n-k)!}$

- Division principle

Let $f: X \to Y$ be a function s.t. $f^{-1}(y) = \{x \in X, f(x) = y\}$.

$|f^{-1}(y)| = m$ $\forall y \in Y$, then $|Y| = \frac{|X|}{m}$.

Binomial theorem:

$$(x+y)^n = \binom{n}{0} x^0 y^n + \cdots + \binom{n}{n} x^n y^0 = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

Pascal triangle identity

$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$, $k \in [0, n]$, assuming $\binom{n-1}{n} = \binom{n-1}{-1} = 0$

Algebraic manipulation

$$\frac{n!}{k!(n-k)!} = \frac{(n-1)!}{k!(n-1-k)!} + \frac{(n-1)!}{(k-1)!(n-k)!} \qquad n = (n-k)+k.$$

Combinatorial proof.

We want a bijection f:

(k element subsets of $[n]$) $\rightarrow$ (k or $(k-1)$ element subset of $[n-1]$)

$$f: \underset{\substack{X \subseteq [n] \\ |X| = k}}{X} \begin{cases} X & \text{if } n \notin X, \\ X \setminus \{n\} & \text{o.w.} \end{cases}$$

f is bijective as there is inverse function

$$\underset{\substack{|Y| = k-1/k, \\ Y \subseteq [n-1]}}{Y} \xrightarrow{f^{-1}} \begin{cases} Y & \text{if } |Y| = k \\ Y \cup \{n\} & \text{o.w} \end{cases}$$

if $n \geq 1$: $0 = (-1+1)^n = \sum_{k \text{ even}} \binom{n}{k} - \sum_{k \text{ odd}} \binom{n}{k}$

$\binom{n}{k}$: # of subsets, as mentioned before.