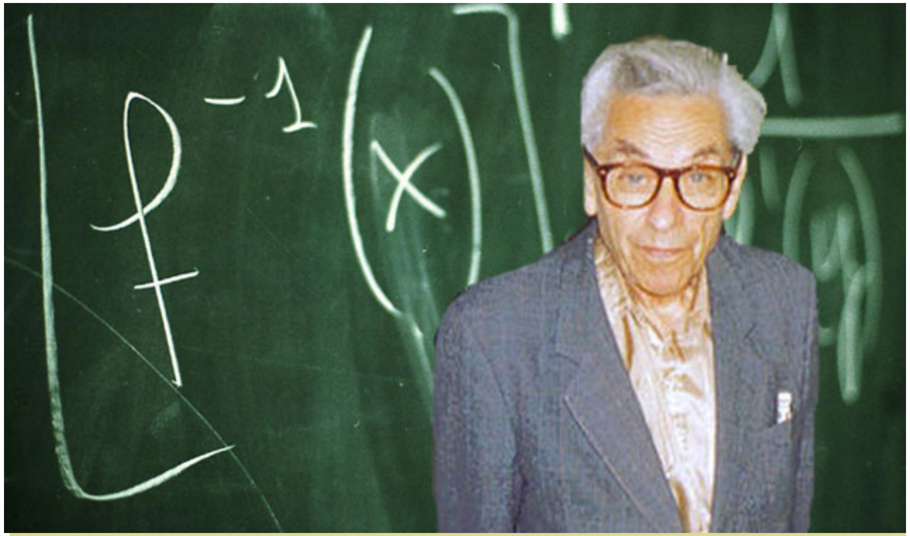
The Probabilistic Method

longhuan@sjtu.edu.cn

The probabilistic method



Paul **Erdős** (26 March 1913 – 20 September 1996)
Hungarian mathematician. Erdős published more papers than any other mathematician in history, working with hundreds of collaborators. He worked on problems in combinatorics, graph theory, number theory, classical analysis, approximation theory, set theory, and probability theory.

- The probabilistic method is a nonconstructive method, primarily used in combinatorics and pioneered by Paul Erdős.
- *For proving the existence of a prescribed kind of mathematical object. It works by showing that if one randomly chooses objects from a specified class, the probability that the result is of the prescribed kind is more than zero.*

Basic Counting Argument

The Expectation Argument

Lovasz Local Lemma

1. Cards Shuffling

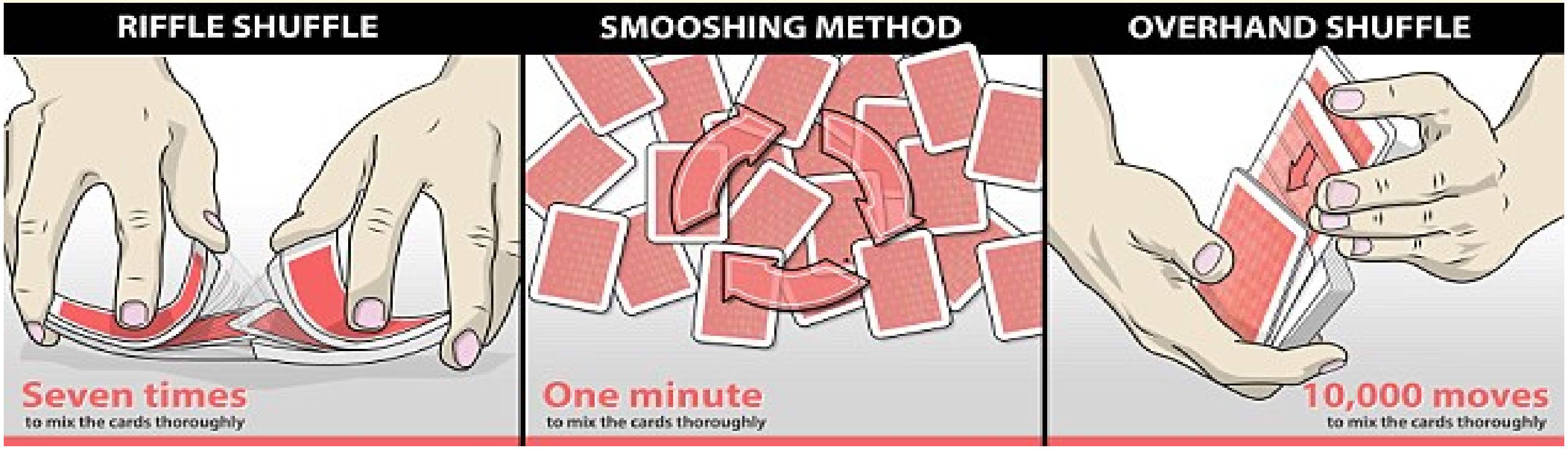
- Consider a new deck of 52 cards. We will shuffle the cards by so-called dovetail shuffling (a.k.a. ‘riffle’).
- Is 4 rounds of dovetail shuffling enough to yield a random order of the cards?

$$\binom{52}{26}^4 < 52!$$

$$\frac{3\log_2 n}{2}$$

How to shuffle cards like a pro: Mathematician shows why the 'riffle' technique is more effective than the flashy 'overhand'

- A Stanford University mathematician compared shuffling techniques
- Dealers using a 'riffle' shuffle need to repeat the process seven times to get a random pack of cards, said Peri Diaconis
- This technique involves cutting a deck and shuffling the halves together
- Whereas 'overhand' needs to be repeated 10,000 times to get same results
- The 'smooshing' or wash method takes one minute to randomise cards



2. Difficult Boolean Functions

- n variable Boolean functions:

$$f: \{0,1\}^n \to \{0,1\}.$$

- Logical formula in n variables:
  - Symbols: $x_1, x_2, \dots, x_n$;
  - Parenthesis： $($, $)$;
  - Logical connectives: $\wedge$, $\vee$, $\Rightarrow$, $\Leftrightarrow$, $\neg$;

Proposition. There exists a Boolean function of n variables that cannot be defined by any formula with fewer than $2^n / \log_2(n+8)$ symbols.

Proposition. There exists a Boolean of n variables that cannot be defined by any formula with fewer than $2^n / \log_2(n+8)$ symbols.

- Proof:

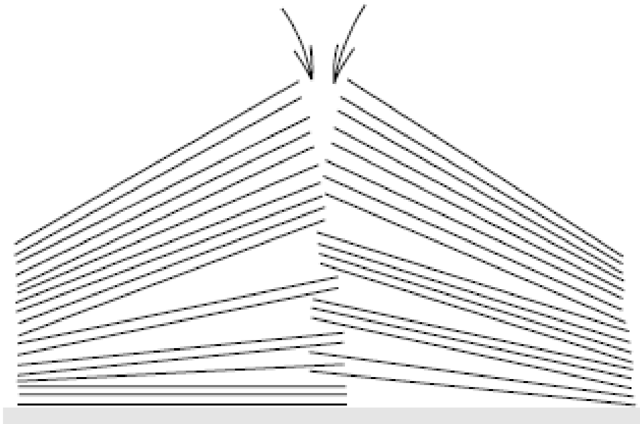
The number of all Boolean functions of n variables: $= 2^{2^n}$

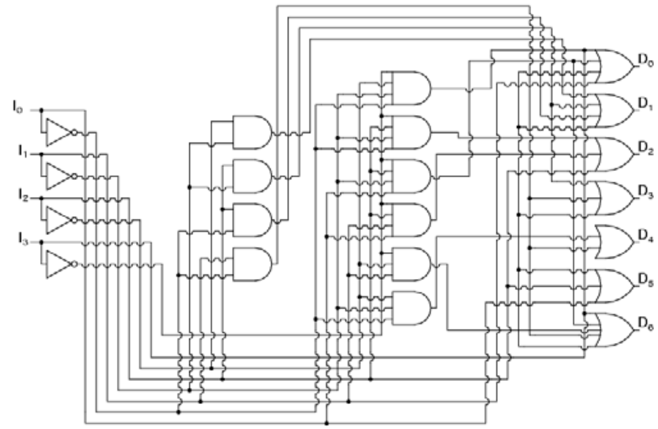
The number of formulas in n variables written by at most m symbols is: $\leq (n+8)^m$

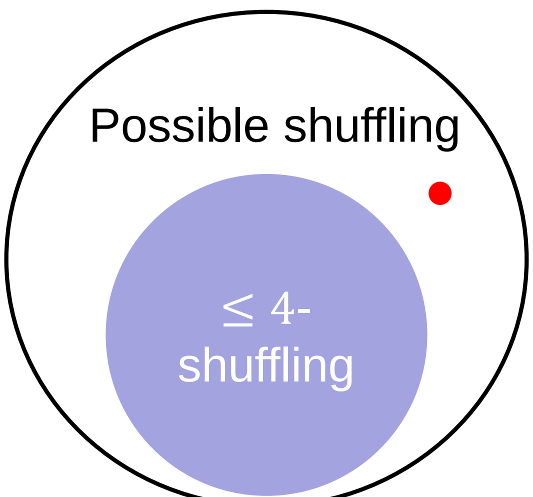
Complications will emerge when: $2^{2^n} > (n+8)^m$

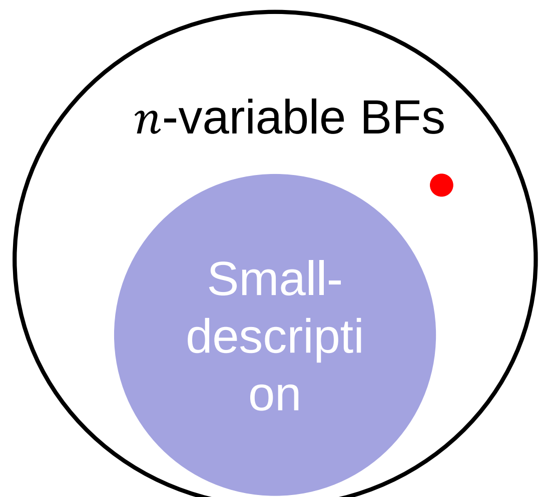
$$m < 2^n / \log_2(n+8)$$

The existence of certain objects

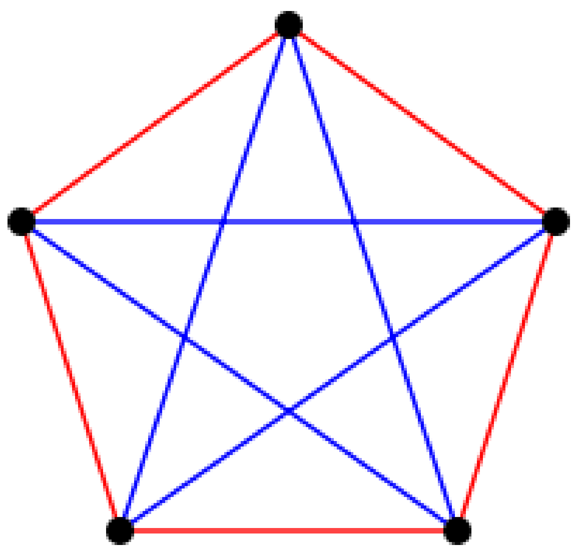


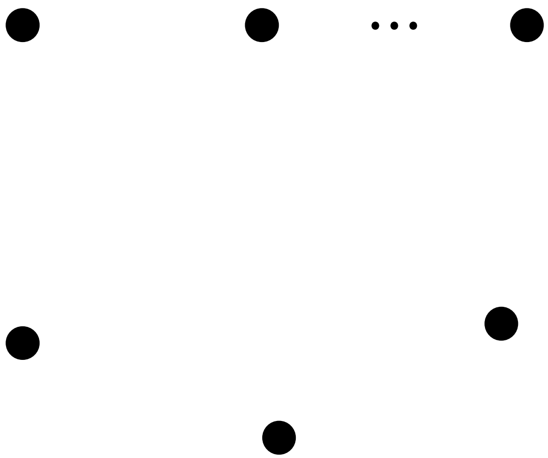




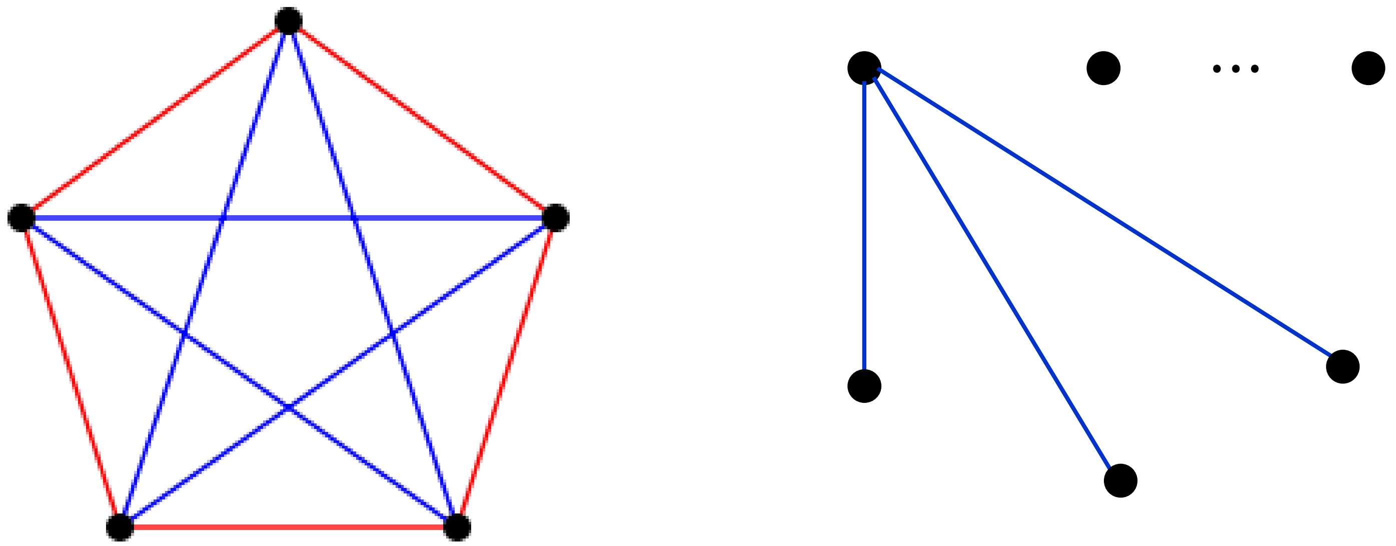


3. Edge Coloring
(a.k.a. Ramsey number $R(k,k)$)

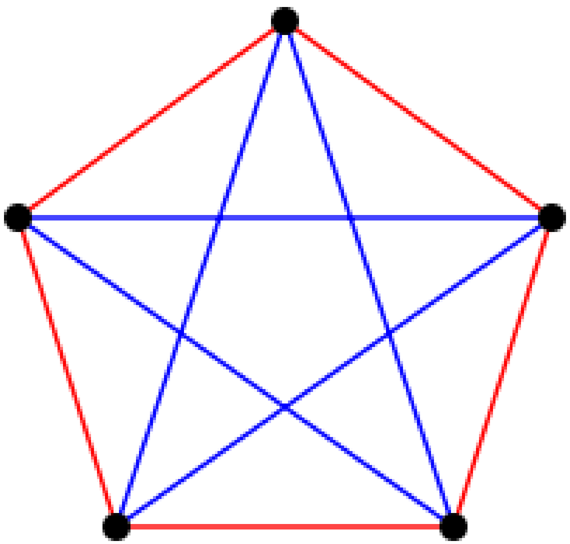


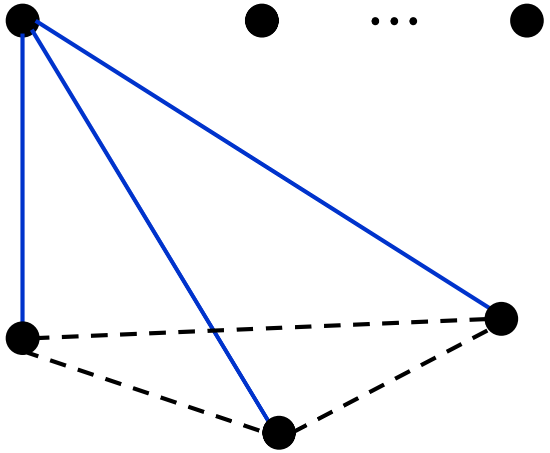


3. Edge Coloring
(a.k.a. Ramsey number $R(k,k)$)

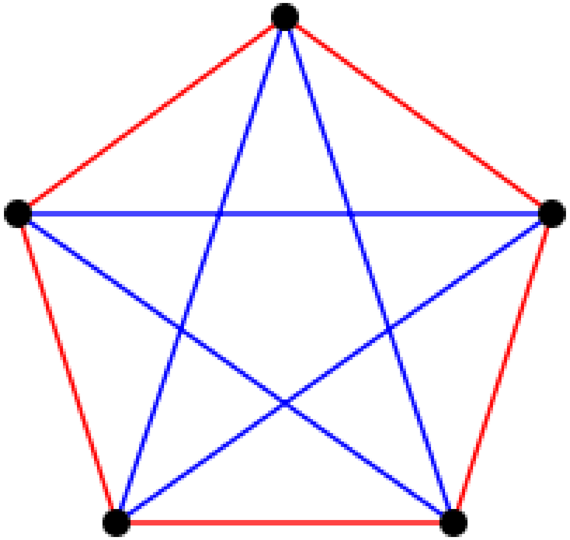


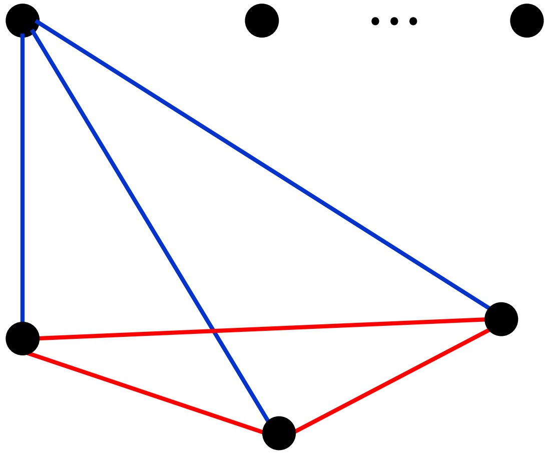
3. Edge Coloring (a.k.a. Ramsey number $R(k,k)$)



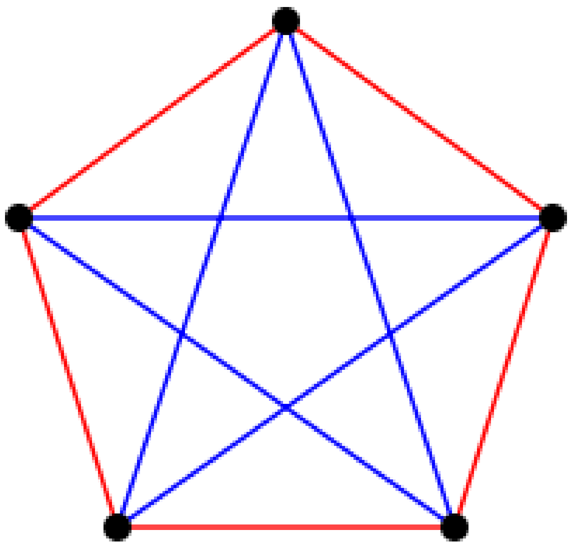


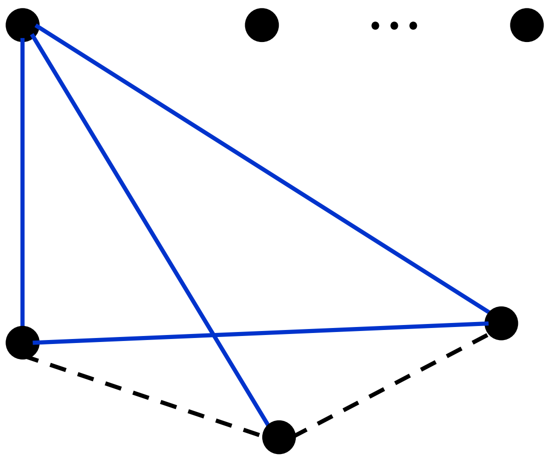
3. Edge Coloring (a.k.a. Ramsey number $R(k,k)$)





3. Edge Coloring (a.k.a. Ramsey number $R(k,k)$)





Theorem. If $\binom{n}{k} 2^{-\binom{k}{2}+1} < 1$, then it is possible to color the edges of K_n with two colors so that it has no single-colored (monochromatic) K_k subgraphs.

- Proof.

For each $e = \{u, v\}$  $\begin{cases} \text{Head:} & f(e) = \text{RED} \\ \text{Tail:} & f(e) = \text{BLUE} \end{cases}$

A certain K_k subgraph is monochromatic: $= 2 \cdot \frac{1}{2^{\binom{k}{2}}}$

The probability that one of K_k subgraph is monochromatic: $\leq \binom{n}{k} \cdot 2 \cdot \frac{1}{2^{\binom{k}{2}}} = \binom{n}{k} 2^{-\binom{k}{2}+1}$

$< \mathbf{1}$

4. Coloring set systems by two colors(*)

- X is a finite set, $M \subseteq P(X)$.
- **Coloring function** $f: X \to \{\mathrm{RED}, \mathrm{BLUE}\}$
- **2-Colorability**. if there is a coloring function such that every $S \in M$ contains points of both colors. Then M is 2-colorable.
- **Example**. $X = \{1,2,3\}, M = \{\{1,2\},\{1,3\},\{2,3\}\}$ then M is not 2-colorable.

- X is a finite set, $M \subseteq P(X)$.
- **Coloring function** $f: X \to \{\mathrm{RED}, \mathrm{BLUE}\}$
- **2-Colorability**. if there is a coloring function such that every $S \in M$ contains points of both colors. Then M is 2-colorable.
- $\forall\, S \in M\ (|S| = k)$
- $s(k)$ is the smallest number of sets in a system M that is not 2-colorable.
- **Example:** $s(2) = 3$.

Theorem. $s(k) \geq 2^{k-1}$, i.e. any system consisting of fewer than 2^{k-1}sets of size k admits a 2-coloring.

Theorem. $s(k) \geq 2^{k-1}$, i.e. any system consisting of fewer than 2^{k-1} sets of size k admits a 2-coloring.

- Proof. $M \subseteq \binom{X}{k}$, $|M| = m$

For each $x \in X$

$$\begin{cases} \text{Head:} & f(x) = \text{RED} \\ \text{Tail:} & f(x) = \text{BLUE} \end{cases}$$

$S \in M$, the probability that S is single-colored is: $\frac{1}{2^k} + \frac{1}{2^k} = 2^{1-k}$

The probability that at least one of the m sets in M is *monochromatic (single-color)* is: $\leq m \cdot 2^{1-k}$

If $m < 2^{k-1}$ the probability is strictly less than 1.

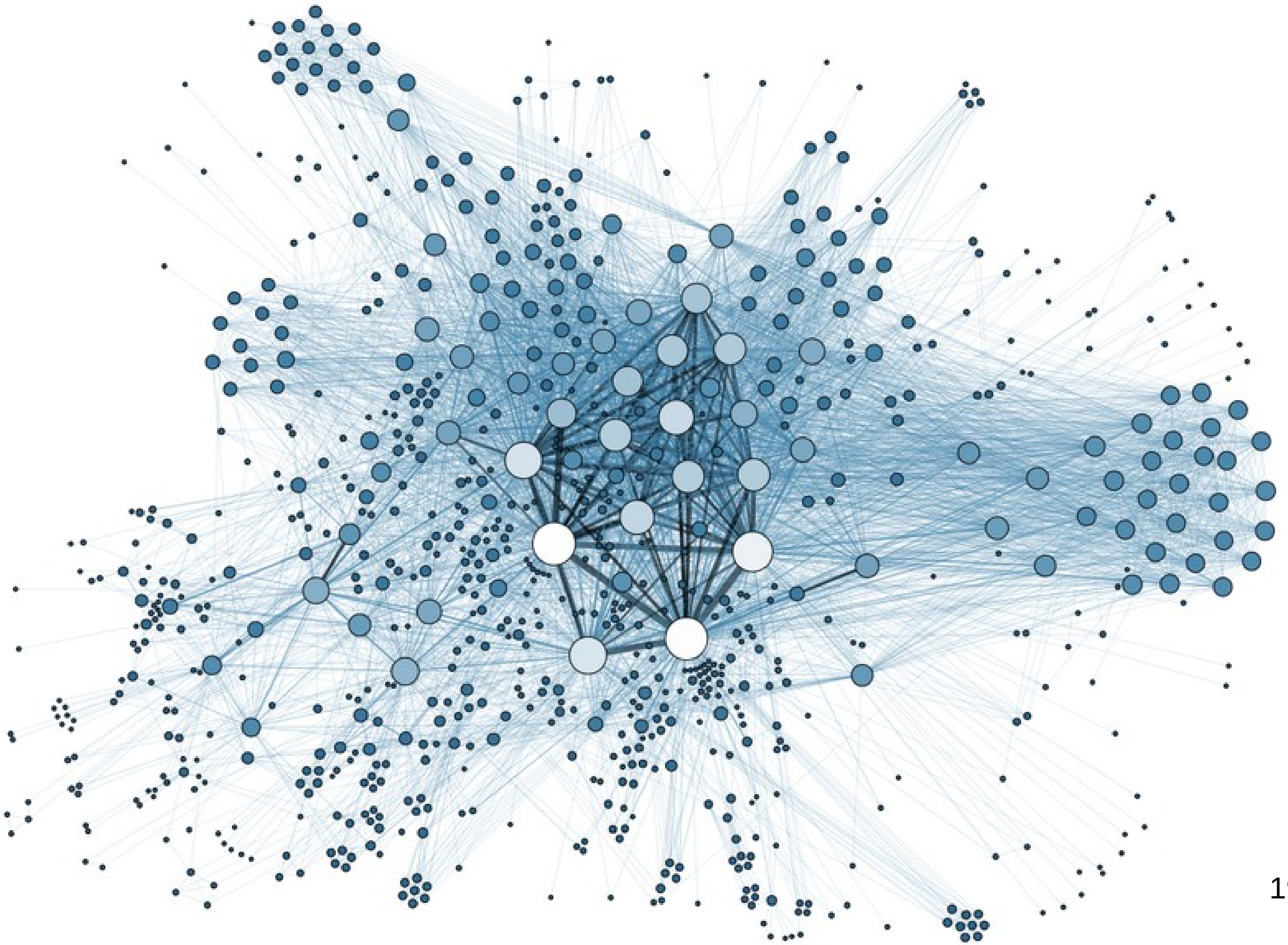
Some M is 2-colorable. $\therefore s(k) \geq 2^{k-1}$.

Basic Counting Argument

The Expectation Argument

Lovasz Local Lemma

1. *Dense* Partition



Theorem. Let G be a graph with an even number, $2n$, of vertices and with $m > 0$ edges. Then the set $V = V(G)$ can be divided into two disjoint n-element subsets A and B in such a way that more than $\frac{m}{2}$ edges go between A and B.

Proof. Randomly choose n vertex to form set A.
Then $B = V \setminus \mathrm{A}$.

For any edge $e = \{u, v\}$, the probability of e being lying 'across' A and B is:

$$\frac{2\binom{2n-2}{n-1}}{\binom{2n}{n}} = \frac{n}{2n-1} > \frac{1}{2}$$

$|E(G)| = m$, the expectation of the number of edges lying 'across' : $E\big(C(A,B)\big) = m \cdot \frac{n}{2n-1} > \frac{m}{2}$

There must exist a choice of A with more than half of the edges going across.

A Las Vegas algorithm for finding an partition

Let $p = \Pr\left(C(A,B) \geq \frac{m}{2}\right)$,

$$\frac{m}{2} < E\left(C(A,B)\right) = \sum_{i \leq \frac{m}{2}-1} i \cdot \Pr(C(A,B) = i) + \sum_{i \geq \frac{m}{2}} i \cdot \Pr(C(A,B) = i)$$

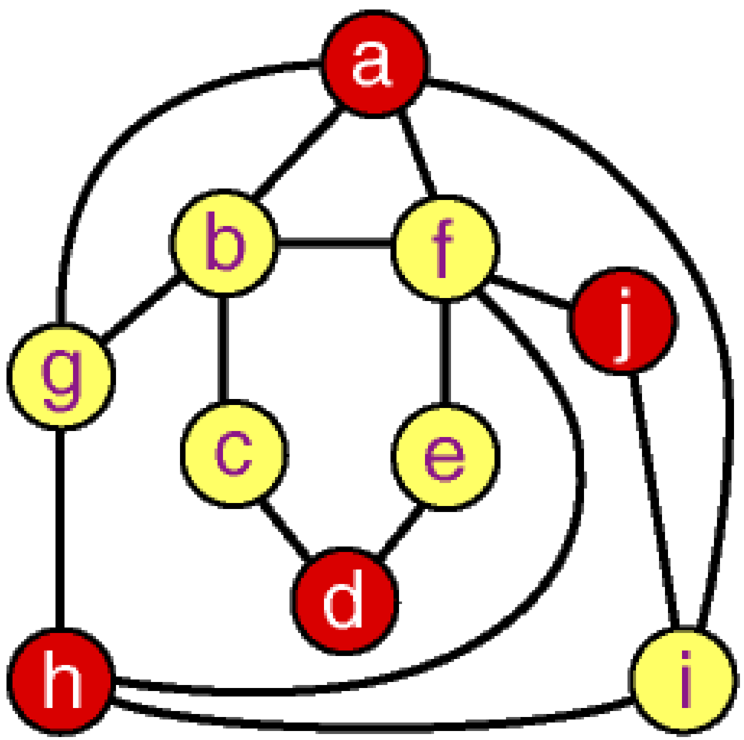
$$\leq (1-p)\left(\frac{m}{2} - 1\right) + pm$$

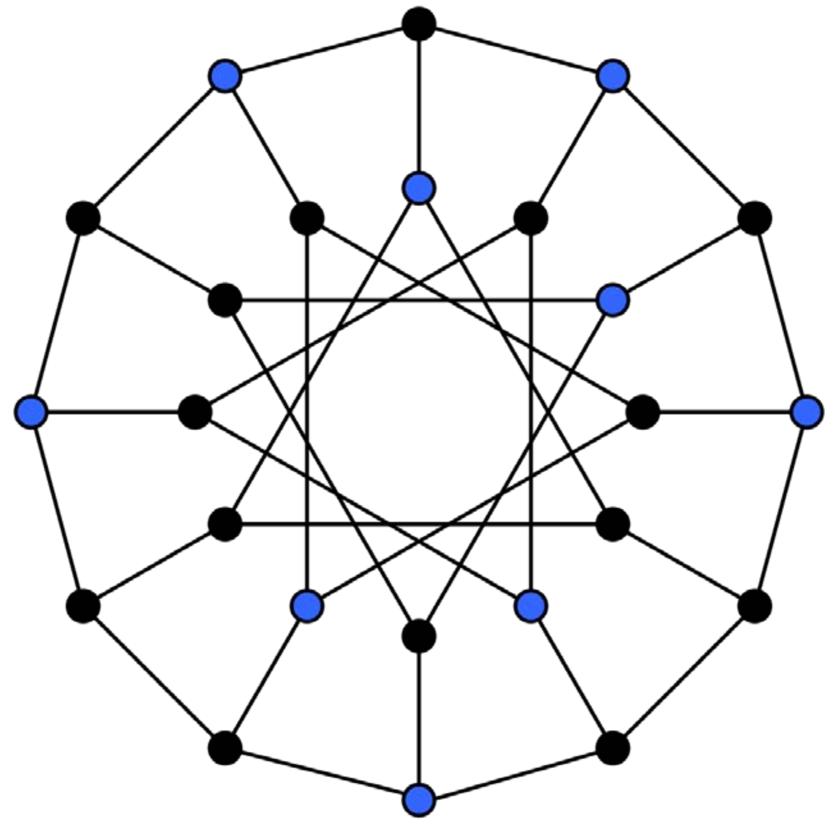
$$\therefore \; p \geq \frac{1}{\frac{m}{2} + 1}$$

The expected number of samples before finding a cut with value at least $m/2$ is therefore just $\frac{m}{2} + 1$.

Sample and testing.

2. Independent set





Theorem. (Tur$\grave{a}$n's theorem). For any graph G on n vertices, we have $\alpha(G) \geq \frac{n^2}{2|E(G)|+n}$.
where $\alpha(G)$ denotes the size of the largest independent set of vertices in the graph G.

Lemma. For any graph G, we have

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

Lemma. For any graph G, we have

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

- **Proof.** $V = \{1,2, \dots, n\}$

Randomly pick a permutation $\pi: \mathrm{V} \to V$,

$M = M(\pi) \subseteq V = \{v \mid \forall u \left(\{u, v\} \in E(G) \to \pi(u) > \pi(v)\right) \}$,

$M(\pi)$ is an independent set in G, $\therefore$ for any $\pi, |M(\pi)| \leq \alpha(G)$.

A_v: the event "$v \in M(\pi)$"

$$P(A_v) = \frac{1}{1 + |N_v|} = \frac{1}{\deg_{\mathrm{G}}(v) + 1}$$

$$\alpha(G) \geq E(|M|) = \sum_{v \in V} E[I_{A_v}] = \sum_{v \in V} P(A_v) = \sum_{v \in V} \frac{1}{\deg_G(v) + 1}$$

Lemma. For any graph G, we have

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

Theorem. (Turàn's theorem). For any graph G on n vertices, we have $\alpha(G) \geq \frac{n^2}{2|E(G)|+n}$.
where $\alpha(G)$ denotes the size of the largest independent set of vertices in the graph G.

$$\sum_{v \in V(G)} \frac{1}{\deg_G(v)+1}$$

will be minimal, when $d_1 = d_2 = \cdots = d_n = \frac{2|E(G)|}{n}$.

3. Maximum Satisfaction

- Logical formula:
  $(x_1 \vee \overline{x_2} \vee \overline{x_3}) \wedge (\overline{x_1} \vee \overline{x_3}) \wedge (x_1 \vee x_2 \vee x_4) \wedge (x_4 \vee \overline{x_3}) \wedge (x_4 \vee \overline{x_1})$
- SAT is NP-hard
- MAXSAT: Given a SAT formula, satisfying as many clauses as possible.

Theorem. Given a set of m clauses, let k_i be the number of literals in the ith clause for $i = 1, \ldots, m$. Let $k = \min_{1 \le i \le m} k_i$. Then there is a truth assignment that satisfies at least

$$\sum_{i=1}^{m} \left(1 - 2^{-k_i}\right) \ge m\left(1 - 2^{-k}\right).$$

- **Proof**

Assign values independently and uniformly at random to the variables.

The probability that the ith clause with k_i literals is satisfied is $1 - 2^{-k_i}$

The expected number of satisfied clauses is $\sum_{i=1}^{m} \left(1 - 2^{-k_i}\right) \ge m\left(1 - 2^{-k}\right).$

Basic Counting Argument

The Expectation Argument

Lovasz Local Lemma

- $E_1, E_2, \ldots, E_n$ is a set of bad events.
- The probability that none of the bad events occurs is

$$\Pr\left(\bigcap_{i=1}^{n} \overline{E}_i\right)$$

- Mutual independence is rare in real applications.
- What if the dependency is limited.

Mutually independent of a set

- Event F is mutually independent of the events $F_1, F_2, \ldots, F_m$ if, for any subset $I \subseteq [1, n]$:

$$\Pr\left(F \mid \cap_{j \in I} F_j\right) = \Pr(F)$$

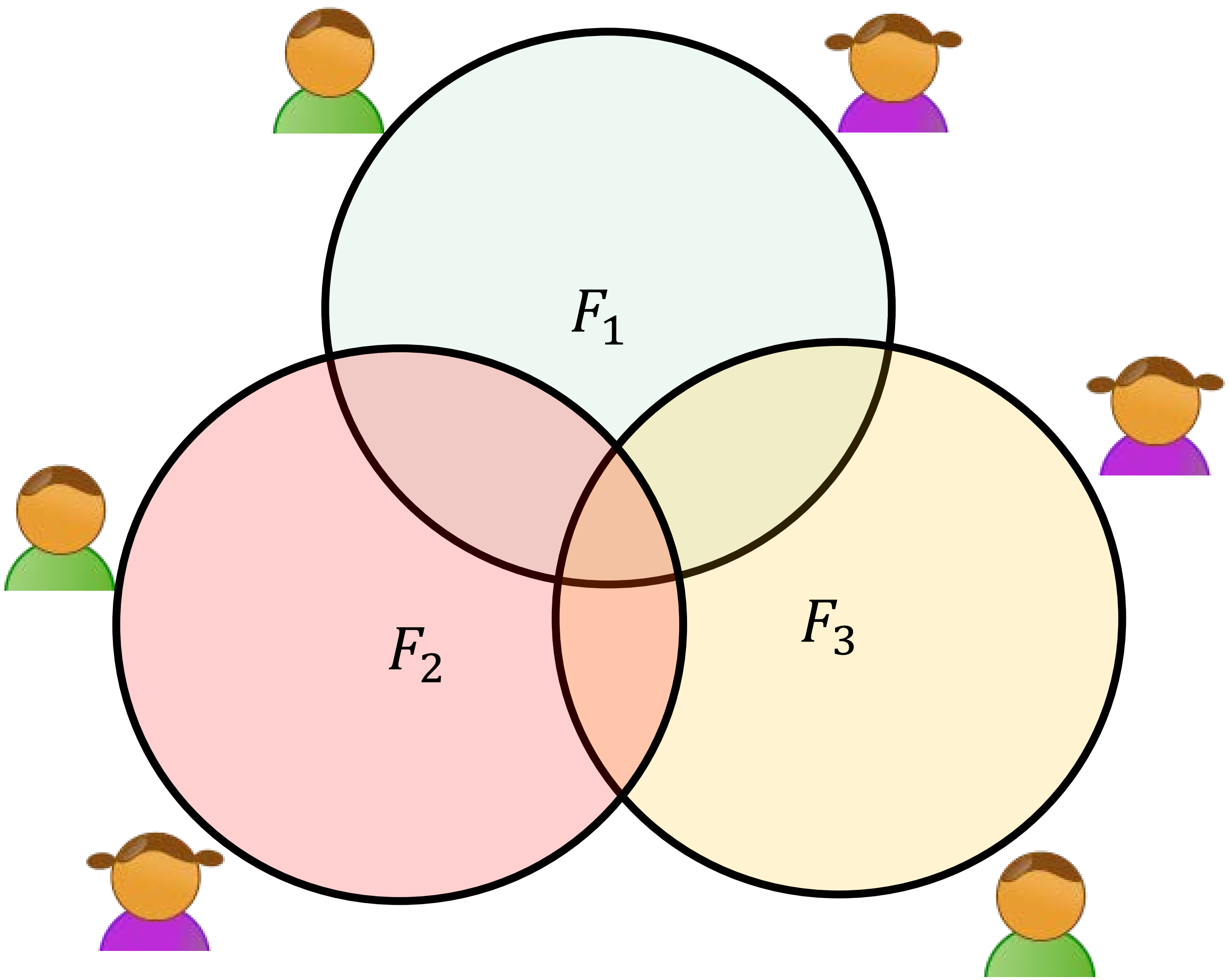
- **Dependency graph.** for a set of events $E_1, E_2, \ldots, E_n$, define graph $G = (V, E)$ such that $V = \{1,2, \ldots, n\}$ and, for $i = 1, \ldots, n$, event E_i is mutually independent of the events $\left\{E_j \mid (i, j) \notin E\right\}$.

Theorem[Lovasz Local Lemma]: Let $E_1, E_2, \dots, E_n$ be a set of events, and assume that the following hold:

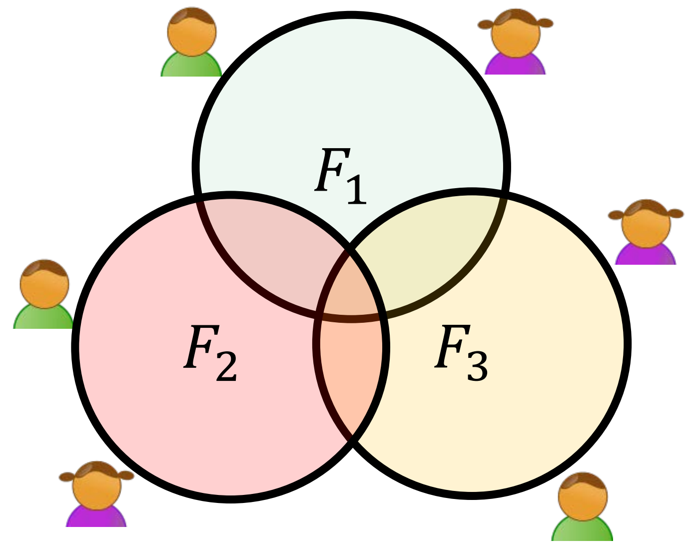
1. For all i, $\Pr(E_i) \le p$;
2. The degree of the dependency graph given by $E_1, E_2, \dots, E_n$ is bounded by d;
3. $4dp \le 1$.

Then $\Pr(\bigcap_{i=1}^{n} \bar{E}_i) > 0$.

Application 1: Edge-disjoint path



- Scenario
  - n pairs of users need to communicate using edge-disjoint paths on a given network.
  - Each pair $i = 1, \ldots, n$ can choose a path from a collection F_i of m path (i.e. $|F_i| = m$).



Theorem: If any path in F_i shares edges with no more than k paths in F_j, where $i \neq j$ and $\frac{8nk}{m} < 1$, then there is a way to choose n edge-disjoint paths connecting the n pairs.

Theorem: If any path in F_i shares edges with no more than k paths in F_j, where $i \neq j$ and $\frac{8nk}{m} \leq 1$, then there is a way to choose n edge-disjoint paths connecting the n pairs.

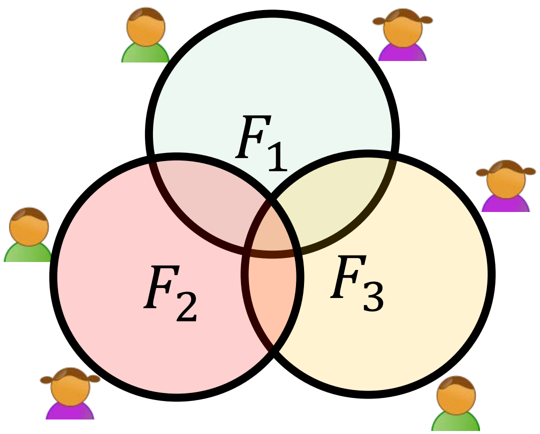
Proof. Each pair i chooses a path independently and uniformly at random from F_i.

$E_{i,j}$: the event that the path chosen by pairs i and j share at least one edge.

Obviously, $p = \Pr\left(E_{i,j}\right) \leq \frac{k}{m}$,

Dependency graph, $d < 2n$.

$$4dp < \frac{8nk}{m} \leq 1$$



$\therefore\ \Pr\left(\bigcap_{i \neq j} \overline{E_{i,j}}\right) > 0$ by Lovasz local lemma.

Application 2: Satisfiability

- If no variable in a k −SAT formula appears in more than $T = \frac{2^k}{4k}$ clauses, then the formula has a satisfying assignment.
- **Proof.**
  - E_i: the ith clause is not satisfied.
  - $p = 2^{-k}$, $d \leq k \cdot T \leq 2^{k-2}$