# Probabilistic Method and Random Graphs

Lecture 10. The Method of Counting&Expectation

Xingwu Liu

Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

[1]The slides are mainly based on Chapter 6 of Probability and Computing.

# Recap of Lecture 9

- Threshold phenomenon in E-R random graphs
  - Sharp threshold function of connectivity: $\frac{\ln n}{n}$
  - That of the existence of major component : $\frac{1}{n}$
  - Threshold function of cycles: $\frac{1}{n}$
- W.h.p. a Hamiltonian cycle can be found in time $O(n\ln n)$ in $\mathcal{G}_{n,p}$ with $p \geq 40\frac{\ln n}{n}$
  - Independent adjacency list model

# Recap of Lecture 9

- Though elegant, E-R model is not practical
  - It can’t be both sparse and heterogenous
- Random graphs with fixed degree distribution
  - Bollobas, about 1980
- Preferential attachment model
  - Barabasi&Albert, 1999
- Rewired ring model
  - Watts&Strogatz, 1998

# Probabilistic Method
## -Elegance from graph theory

| | G1 | G2 | G3 | G4 | G5 | Score |
|---|---|---|---|---|---|---|
| G1 | | 3:1 | 5:3 | 4:2 | 0:1 | 3 |
| G2 | | | 1:3 | 3:2 | 2:1 | 2 |
| G3 | | | | 0:2 | 1:0 | 2 |
| G4 | | | | | 3:2 | 2 |
| G5 | | | | | | 1 |

- A warm-up example:
  - $n$ players against each other
  - "Top-$k$" players get prize
- But, are you sure no controversy exists?
  - Controversy: a loser defeated all prize-winners
- Unfortunately, when $k$ is small and $n$ is big, controversy does exist w.h.p.
  - $k = 1$. Controversy exists unless the prize-winner defeated all the other plays

# Proof (non-constructive)

Theorem: For small $k$ and big $n$, controversy exists w.h.p.

- $S$: random $k$-subset of players
- $A_S$: no controversy if $S$ get prize
  - i.e. no another player defeated all players in $S$
- Consider a random tournament
- $\Pr(A_S) = (1 - 2^{-k})^{n-k}$
- $\Pr(\text{no controversy}) \leq \Pr(\bigcup A_S) \leq \sum \Pr(A_S)$
$$= \binom{n}{k} (1 - 2^{-k})^{n-k}$$
$$= o\left(\frac{1}{n}\right)$$
- Find a controversial one? Just sampling

Cool?

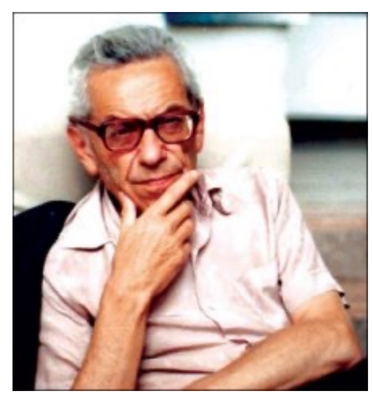
A piece of cake in probabilistic method!

# What is the Probabilistic Method?

- **Proving the existence** of an object that satisfies a certain property, **without constructing** it
- Underlying principle

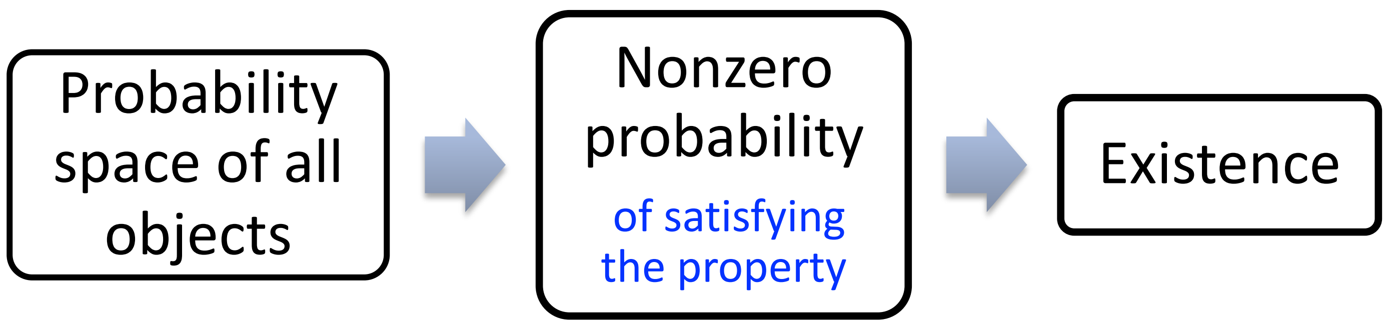
Probability space of all objects ⇒ Nonzero probability of satisfying the property ⇒ Existence

- Pioneered by Erdős in 1940's



# What is the Probabilistic Method?

- **Proving the existence** of an object that satisfies a certain property, **without constructing** it
- Underlying principle




- Pioneered by Erdős in 1940's
- Naturally lead to (randomized) algorithms

# Main Probabilistic Methods

- Counting argument
- First-moment method
- Second-moment method
  - Higher-moment method
- Lovasz local lemma

Algorithms

# Counting Argument

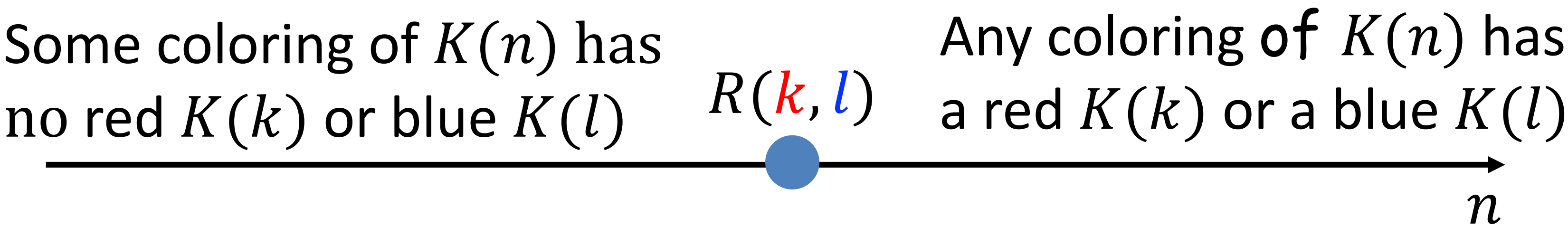
- Construct a probability space and calculate the probability

- Algorithm design: sampling

- Application
  - Tournament
  - Ramsey number: an observation by S. Szalai, 1950's

# Ramsey Number

- Given integers $k, l, n$, 2-color the edges of $K(n)$
  - Is there a red $K(k)$ or a blue $K(l)$?
  - Not guaranteed for small $n$
- Ramsey number $R(k, l)$
  - the smallest $n$ such that any 2-coloring of $K(n)$ must have a red $K(k)$ or a blue $K(l)$



# Ramsey Number is Well Defined

- Ramsey Theorem: $R(k, l)$ is finite for any $k, l$
  - Ramsey proved it in 1930 and determined $R(3,3) = 6$
  - Origin of Ramsey theory
    - The existence of rather big good substructure in a large structure
  - How much is $R(k, l)$?

# Ramsey Number is Well Defined

- Ramsey Theorem: $R(k,l)$ is finite for any $k,l$
  - Ramsey proved it in 1930 and determined $R(3,3)$
  - Origin of Ramsey theory
    - The existence of rather big good substructure in a large structure
  - How much is $R(k,l)$?
- Upper bound: $R(k,l) \leq R(k-1,l) + R(k,l-1)$
  - Proved by P. Erdős and G. Szekeres in 1935
  - The $2^{nd}$ cornerstone of Ramsey theory
  - By $R(k,2) = R(2,k) = k$, $R(k,l) \leq \binom{k+l-2}{k-1}$
  - It implies $R(k,k) \leq 4^k$
    - Best: $k^{-c\frac{\ln k}{\ln \ln k}} 4^k$ by Conlon in 2009, $k^{-c \ln k} 4^k$ by Sah in 2020

# Known bounding ranges

| r \ s | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | **1** | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| **2** | | **2** | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| **3** | | | **6** | 9 | 14 | 18 | 23 | 28 | 36 | 40–42 |
| **4** | | | | **18** | 25[5] | 36–41 | 49–61 | 59[10]–84 | 73–115 | 92–149 |
| **5** | | | | | **43–48** | 58–87 | 80–143 | 101–216 | 133–316 | 149[10]–442 |
| **6** | | | | | | **102–165** | 115[10]–298 | 134[10]–495 | 183–780 | 204–1171 |
| **7** | | | | | | | **205–540** | 217–1031 | 252–1713 | 292–2826 |
| **8** | | | | | | | | **282–1870** | 329–3583 | 343–6090 |
| **9** | | | | | | | | | **565–6588** | 581–12677 |
| **10** | | | | | | | | | | **798–23556** |

1989

2017
1997(49)

Vigleik Angeltveit; Brendan McKay (2017). “R(5,5)≤48". [arXiv]

# Proof of the upper bound

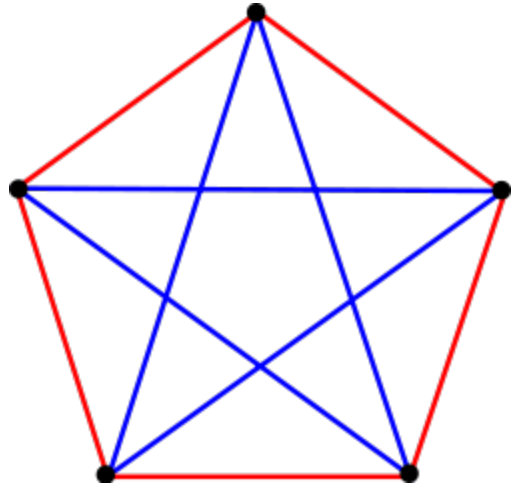
Theorem: $R(k,l) \leq R(k-1,l) + R(k,l-1)$

- 2-color the complete graph on $R(k-1,l) + R(k,l-1)$ vertices
- Pick a vertex $u$. Define subgraphs $G_r$ and $G_b$:
  - $\forall\, v \neq u, v \in \begin{cases} G_r & \text{if } (u,v) \text{ is red} \\ G_b & \text{if } (u,v) \text{ is blue} \end{cases}$
- Either $|G_r| \geq R(k-1,l)$ or $|G_b| \geq R(k,l-1)$
- Do case-by-case analysis

# Example: Ramsey Number $R(3,3)$

$R(3,3) \leq 6$

Actually, $R(3,3) > 5$



# Lower bound of $R(k,k)$

- $R(k,k) > 2^{k/2} = \sqrt{2}^k$ (Erdős, 1947)
  - Best: $[1+o(1)]\frac{k}{e}\sqrt{2}^{1+k}$ by Spencer in 1975
- For any complete graph with at most $2^{k/2}$ vertices, there is a 2-coloring without monochromatic $K(k)$

- Prove by the probabilistic method

# Prove $R(k,k) > 2^{k/2}$

- Randomly 2-color edges of $K(n)$
  - Uniform distribution on all 2-coloring
- $A_S$: the subgraph on $S$ is monochromatic
  - $S$ is a random $k$-subset of the vertices
- $\Pr(A_S) = 2^{1-\binom{k}{2}}$
- $\Pr(\bigcup_S A_S) \leq \binom{n}{k} 2^{1-\binom{k}{2}} < \frac{2^{1+\frac{k}{2}}}{k!} \frac{n^k}{2^{\frac{k^2}{2}}} < 1$ (If $n = \lfloor 2^{k/2} \rfloor$)
- $\Pr(\bigcap_S \overline{A_S}) > 0$, so there is a coloring avoiding all $A_S$

# Randomized Algorithms

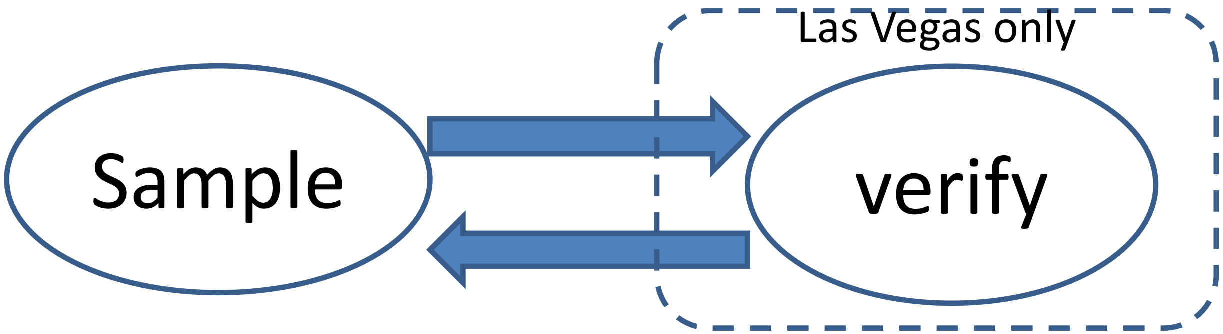
- But how to find a good coloring? By sampling!
- General approach

Sample

Monte Carlo

# Randomized Algorithms

- But how to find a good coloring? By sampling!
- General approach




- Prerequisites
  - Efficient sampling
  - Small probability of failure
  - Efficient verification (Las Vegas only)

# First-Moment method

- Use the expectation in probabilistic reasoning
- Two types of first-moment method
  - Expectation argument

    $$\Pr(X \geq \mathbb{E}[X]) > 0, \Pr(X \leq \mathbb{E}[X]) > 0$$

  - Markov's inequality for non-negative $X$
    - $\Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$
    - When $X$ is integer-valued,

      $\Pr(X \neq 0) = \Pr(X > 0) = \Pr(X \geq 1) \leq \mathbb{E}[X]$

# First-Moment argument

- 3-CNF Boolean formula
  - $(x_1 \vee x_2 \vee x_3) \wedge \ldots \wedge (\overline{x_1} \vee \overline{x_3} \vee x_4)$
- For such a formula, at most how many clauses can be satisfied simultaneously?
  - **MAX-3SAT** is NP-hard

# First-Moment argument

- 3-CNF Boolean formula
  - $(x_1 \vee x_2 \vee x_3) \wedge \ldots \wedge (\overline{x_1} \vee \overline{x_3} \vee x_4)$
- Theorem: there is a truth assignment which satisfies $\geq \frac{7}{8}$-fraction of the clauses
- Proof: -Randomly assign truth values to each variable

  -Define r.v. $X_i$ indicating whether clause $i$ is true

  -$\mathbb{E}[X_i] = \frac{7}{8} \Rightarrow \mathbb{E}[X] = \frac{7}{8}n$ with $X = \sum_{i=1}^{n} X_i$

  -The theorem holds since $\Pr(X \geq \mathbb{E}[X]) > 0$

# Remark

- Probability of sampling a good truth assignment $\geq \frac{1}{n+1}$, leading to an efficient alg.
  - Optimum, since impossible to get a $\left(\frac{7}{8} + \varepsilon\right)$-approx.
    - J. Hastad. Some optimal inapproximability results. STOC 1997

# Proof of $\Pr\left(\sum X_i \geq \frac{7}{8}n\right) \geq \frac{1}{n+1}$

- Let $X = \sum X_i$ and $p = \Pr\left(X \geq \frac{7}{8}n\right)$
- $\frac{7}{8}n = \mathbb{E}[X]$

$$= \sum_{i<\frac{7}{8}n} i * \Pr(X = i) + \sum_{i\geq\frac{7}{8}n} i * \Pr(X = i)$$

$$\leq \left(\frac{7}{8}n - \frac{1}{8}\right)(1 - p) + np$$

$$= \frac{7}{8}n - \frac{1}{8} + \frac{n+1}{8}p$$

# Expectation argument

- Turán Theorem
  - Any graph $G$=($V$,$E$) contains an independent set of size at least $\frac{|V|}{D+1}$, where $D = \frac{2|E|}{|V|}$
- Proof: Consider the following random process for constructing an independent set $S$:
  - Initialize $S$ to be the empty set
  - Repeat: randomly choose a vertex $u$ outside the neighbood of $S$ and add $u$ to $S$
  - Return $S$

# Proof (Continued)

- $S$ is an independent set
- Vertex $u$ is selected with probability $\geq \frac{1}{d(u)+1}$
  - See the next slide
- So, $\mathbb{E}[|S|] \geq \sum \frac{1}{d(u)+1} \geq \frac{|V|}{D+1}$ due to convexity

- **Remark:** probability of sampling a good independent set is $\geq \frac{1}{2D|V|^2}$

# Proof: Pr($u$ is selected)$\geq \frac{1}{d(u)+1}$

- $u$ is selected if and only if $A$ occurs
  - $A$: when sampling first occurs in the neighborhood of $u$, $u$ rather than its neighbors is sampled
    - Neighborhood: $u$ and its then-valid neighbors
  - Denote the neighborhood by $N$, and the number of then-valid neighbors by $x$. Note that $x \leq d(u)$
- $\Pr(A) = \Pr(u$ is chosen$|$sampling occurs in $N)$

$$= \frac{1}{x+1} \geq \frac{1}{d(u)+1}$$

# References


- http://www.cse.buffalo.edu/~hungngo/classes/2011/Spring-694/lectures/sm.pdf
- http://www.cse.cuhk.edu.hk/~chi/csc5160-2007/notes/Probabilistic.ppt
- Erdős. Graph theory and probability I. 1959
- Erdős. Graph theory and probability II. 1961
- Alon&Krivelevich. Extremal and Probabilistic Combinatorics. 2006