Streaming Algorithms

Meng-Tsung Tsai

02/27/2018

A Reference Book

- "The Probabilistic Method", Alon and Spencer (2004)

You may find an e-copy of this book on www.lib.nctu.edu.tw

The Probabilistic Method

The Union Bound

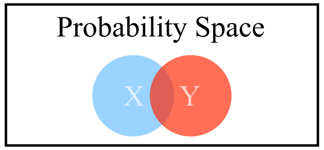
The union bound, a.k.a. Boole's inequality, is stated as follows.
For any countable set of probablistic events A_1, A_2, ..., we have

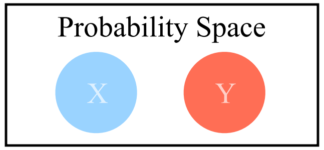
$$\Pr\left[\bigcup_i A_i\right] \leq \sum_i \Pr[A_i]$$

Note that A_i's may be dependent or independent.

An illustration of independent events

If events X and Y are independent, which of the following is a proper illustration?





The Ramsey Number

Theorem 1. Given an n-node complete graph G and any 2-coloring (red/blue) on the edges in G, then G has a k-node monochromatic clique if n is sufficiently large, as a function of k.

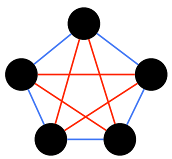
Let R(k, ℓ) be the smallest n so that for any 2-coloring G has either a k-node red clique or an ℓ-node blue clique.

The Ramsey Number

Theorem 1. Given an n-node complete graph G and any 2-coloring (red/blue) on the edges in G, then G has a k-node monochromatic clique if n is sufficiently large, as a function of k.

Let R(k, ℓ) be the smallest n so that for any 2-coloring G has either a k-node red clique or an ℓ-node blue clique.

Example. R(3, 3) = 6.



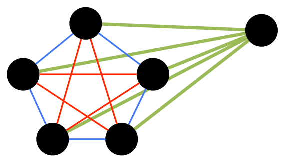
R(3, 3) > 5.

The Ramsey Number

Theorem 1. Given an n-node complete graph G and any 2-coloring (red/blue) on the edges in G, then G has a k-node monochromatic clique if n is sufficiently large, as a function of k.

Let R(k, ℓ) be the smallest n so that for any 2-coloring G has either a k-node red clique or an ℓ-node blue clique.

Example. R(3, 3) = 6.



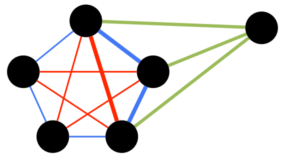
At least 3 newly-added edges have the same color.

The Ramsey Number

Theorem 1. Given an n-node complete graph G and any 2-coloring (red/blue) on the edges in G, then G has a k-node monochromatic clique if n is sufficiently large, as a function of k.

Let $R(k, \ell)$ be the smallest n so that for any 2-coloring G has either a k-node red clique or an ℓ-node blue clique.

Example. $R(3, 3) = 6$.



No matter which color is assigned to these three edges, it induces a monochromatic triangle, so $R(3, 3) \leq 6$.

The Ramsey Number

Theorem 1. Given an n-node complete graph G and any 2-coloring (red/blue) on the edges in G, then G has a k-node monochromatic clique if n is sufficiently large, as a function of k.

Let $R(k, \ell)$ be the smallest n so that for any 2-coloring G has either a k-node red clique or an ℓ-node blue clique.

Claim 1. $R(k, k) > 2^{k/2}$.

Proof of $R(k, k) > 2^{k/2}$

Proof Strategy.

Try to prove: if $n \leq 2^{k/2}$, there exists a coloring so that the resulting G has no monochromatic k-clique.

Assign a random coloring to each edge and try to show that bad events do not always happen.

Each bad event corresponds to that a certain clique is monochromatic.

Proof of $R(k, k) > 2^{k/2}$

Assign a random color, sampled uniformly from {red, blue}, to each edge.

Assign an unique ID from $\{1, 2, ..., \binom{n}{k}\}$ to each k-clique.

$\Pr[\text{the i-th clique is monochromatic}] = \frac{1}{2^{\binom{k}{2}-1}}$

Proof of $R(k, k) > 2^{k/2}$

Assign a random color, sampled uniformly from {red, blue}, to each edge.

Assign an unique ID from $\{1, 2, ..., \binom{n}{k}\}$ to each k-clique.

Pr[the i-th clique is monochromatic] $= \dfrac{1}{2^{\binom{k}{2}-1}}$

Pr[some clique is monochromatic]

$\leq \sum_i$ Pr[the i-th clique is monochromatic]   (the union bound)

$= \dfrac{\binom{n}{k}}{2^{\binom{k}{2}-1}}$

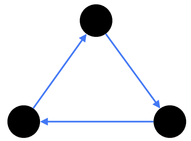
< 1 if $n \leq 2^{k/2}$

Some random coloring induces no monochromatic k-clique.

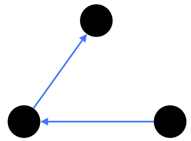
Tournament Graphs

Definition. We say a directed graph G is a tournament graph if for every pair of nodes u, v in G, either the directed edge (u, v) exists or (v, u) exists, but not both.

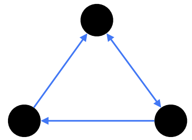
(a Yes-instance)



(a No-instance)



(a No-instance)

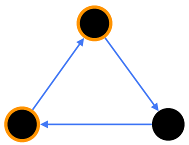


Tournament Graphs

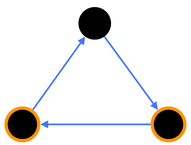
Theorem 2. Given an n-node tournament graph $G = (V, E)$, for any k-node subset $X \subseteq V$ there exists a node $v \in V \setminus X$ so that v beats (has a directed edge to) every node in X, if n is sufficiently large.

Example. $(n, k) = (3, 1)$

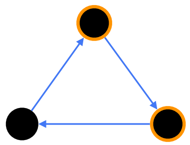
(Case 1)



(Case 2)



(Case 3)



Tournament Graphs

Theorem 2. Given an n-node tournament graph $G = (V, E)$, for any k-node subset $X \subseteq V$ there exists a node $v \in V \setminus X$ so that v beats (has a directed edge to) every node in X, if n is sufficiently large.

Proof Strategy. Assign a random orientation to each edge and try to show that bad events do not always happen.

Each bad event corresponds to that certain k nodes do not have a node in common that beats them all.

Proof of Theorem 2

Assign a random orientation, sampled uniformly from $\{\leftarrow, \rightarrow\}$, to each edge.

Assign an unique ID from $\{1, 2, ..., \binom{n}{k}\}$ to each k-node subset.

Pr[the i-th bad event happens] = $\left(1 - \frac{1}{2^k}\right)^{n-k}$

Pr[some bad event happens]

$\leq \sum_i$ Pr[the i-th bad event happens]   (the union bound)

$= \binom{n}{k}\left(1 - \frac{1}{2^k}\right)^{n-k}$

< 1 if n is sufficiently large.

Exercise 1

Problem 10 in Chap 1.

Prove that there is an absolute constant $c > 0$ with the following property. Let A be an n by n matrix with pairwise distinct entries. Then there is a permutation of the rows of A so that no column in the permuted matrix constains an increasing subsequence of length at least $c\sqrt{n}$.

Linearity of Expectation

Linearity of expectation is stated as follows. Let $X_1, X_2, ..., X_n$ be random variables, $X = \sum_i c_i X_i$ for c_i's in **R**, we have

$$E[X] = \sum_{1 \leq i \leq n} c_i E[X_i]$$

Note that X_i's may be dependent or independent.

Splitting Graphs

Theorem 3. Any n-node m-edge undirected graph G has a bipartite subgraph that has at least m/2 edges.

Proof Strategy. Show that a random partition of the n nodes induces m/2 crossing edges in expectation (on average).

Therefore, some random partition has at least m/2 crossing edges. Why?

It is impossible that every random partition of nodes has fewer crossing edges than the average.

Splitting Graphs

Place each node in {Left, Right} uniformly at random. Let X_e be the indicator varaible whether edge e is a crossing edge with respect to the random partition.

Clearly, $E[X_e] = 2 * 1/2 * 1/2 * 1 = 1/2$.

Thus, $E[\sum_{e \in G} X_e] = \sum_{e \in G} E[X_e] = m/2$.

Some random partition of nodes induces at least m/2 crossing edges.

Exercise 2

Prove that every n-node m-edge undirected simple graph G can be partitioned into O(log n) bipartite subgraphs H_1, H_2, ...; in other words, the union of H_i's forms G.

Exercise 3

Prove or disprove that every n-node m-edge undirected graph has a bipartite subgraph $H = (U \cup V, E)$ so that the following conditions are both satisfied.

(1) $||U| - |V|| = O(1)$

(2) $|E| \geq m/2$