{ Csc 80030 | Lecture 2 }

Probabilistic Analysis & Randomized Algorithms

Hoeteck Wee · hoeteck@cs.qc.edu

PART 0

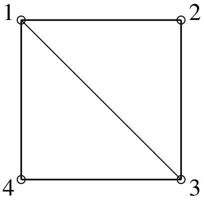
- LAST WEEK: $\binom{n}{2} \in \Omega(\sqrt{n}), \omega(n \log n)$.
- HOMEWORK: HW1 is out, due next week; hiring problem $\rightarrow$ HW2.
- TODAY: randomized min-cut; basic probability & coupon-collecting

PART 1 | randomized min-cut

Minimum cut

Definitions

1. Cut: set of edges whose removal render the graph disconnected
2. Minimum cut: cut of the smallest size (size = # edges in the cut)



- cut: $\{(1,2),(1,3),(1,4)\}$ $\langle\{1\},\{2,3,4\}\rangle$
- min-cut: $\{(1,4),(3,4)\}$ $\langle\{1,2,3\},\{4\}\rangle$
  or $\{(1,2),(2,3)\}$ $\langle\{1,3,4\},\{2\}\rangle$

Easy Fact

The minimum cut has size at most the minimum degree of any node.

Minimum cut

Definitions

1. Cut: set of edges whose removal render the graph disconnected
2. Minimum cut: cut of the smallest size (size = # edges in the cut)

MINIMUM CUT Problem

On input an undirected graph with n vertices, output a minimum cut.

APPLICATIONS

- network reliability (nodes = machines, edges = connections)
- clustering webpages (nodes = webpages, edges = hyperlinks)

Minimum cut

Definitions

1. Cut: set of edges whose removal render the graph disconnected
2. Minimum cut: cut of the smallest size (size = # edges in the cut)

MINIMUM CUT Problem

On input an undirected graph with n vertices, output a minimum cut.

ALGORITHMS

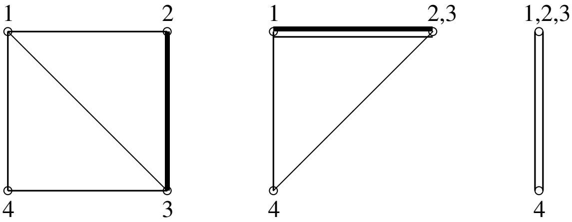
- "naive": compute s-t minimum cut n times (via Ford-Fulkerson).
- next: randomized algorithm based on edge contractions

Edge contraction

Operation EDGE CONTRACTION

Input: edge (u, v) in undirected graph

1. Merge vertices u and v.
2. Remove any self-loops and keep multi-edges.



Randomized min-cut (Karger, 1993)

RANDMINCUT Algorithm

Input: undirected graph G.

1. Repeat $n - 2$ times: contract a random edge
2. Output the edges connecting the remaining two vertices.

- Each edge contraction reduces # vertices by 1.

Analysis

FACT 1

Let C be a min-cut. If we never contract an edge in C, then C remains a cut.

- PROOF: only contract edges, so $\langle$ edges, vertices $\rangle$ on left side of C stay on left side, and the same for right side.

Analysis

FACT 1

Let C be a min-cut. If we never contract an edge in C, then C remains a cut.

- INTUITION: $\exists$ lots of edges, so we're unlikely to contract an edge in C
- GOAL: bound $\Pr[E_i]$ where E_i is "C survives the first i iterations".

Base case: $\Pr[E_1]$

1. degree of every vertex $\geq k$
2. # edges $\geq nk/2$
3. $\Pr[E_1] = 1 - \frac{k}{\#\text{ edges}} \geq 1 - \frac{k}{nk/2} = \frac{n-2}{n}$

Analysis

FACT 2

Min-cut size never decreases.

- CLAIM: Any cut C in the new graph is also a cut in the original graph.
- PROOF: Induction. Any "contracted edge" must lie on same side of C.

Analysis

Fact 2

Min-cut size never decreases.

Iterative step: $\Pr[E_{i+1} \mid E_i]$

1. By Fact 2, min-cut size $\geq k$, so degree $\geq k$
2. # vertices $= n - i$
3. # edges $\geq (n-i)k/2$
4. $\Pr[E_{i+1} \mid E_i] = 1 - \frac{k}{\text{\# edges}} \geq 1 - \frac{k}{(n-i)k/2} = \frac{n-i-2}{n-i}$

Analysis

$$\begin{aligned}
& \Pr[\textsc{RandMinCut} \text{ outputs } C] \\
= \quad & \Pr[E_{n-2}] = \Pr[E_{n-2} \mid E_{n-1}] \cdots \Pr[E_2 \mid E_1] \cdot \Pr[E_1] \\
\geq \quad & \left(\tfrac{n-2}{n}\right)\left(\tfrac{n-3}{n-1}\right)\left(\tfrac{n-4}{n-2}\right)\left(\tfrac{n-5}{n-3}\right) \cdots \left(\tfrac{4}{6}\right)\left(\tfrac{3}{5}\right)\left(\tfrac{2}{4}\right)\left(\tfrac{1}{3}\right) \\
= \quad & \tfrac{2}{n(n-1)}
\end{aligned}$$

Question

How can we increase the probability of returning a min-cut?

- Repeat $\frac{n(n-1)}{2} \ln n$ times and output the smallest cut.
- $\Pr[\text{fails to output } C] \leq \left(1 - \frac{2}{n(n-1)}\right)^{\frac{n(n-1)}{2} \ln n} \leq \frac{1}{n}$

PART 2 | two distributions & coupon-collecting

Expectation

Definition (expectation)

The expectation $\mathrm{E}[X]$ of a discrete random variable X is given by

$$\sum_i i \Pr[X = i].$$

Fact (linearity of expectations)

Given any finite collection of r.v. $X_1, \ldots, X_n$, we have

$$\mathrm{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{E}[X_i]$$

- EXAMPLE: toss two 6-sided dice and take the sum of the two values.

Variance

Definition (variance)

The variance $\mathrm{Var}[X]$ of a random variable X is given by

$$\mathrm{E}[(X - \mathrm{E}[X])^2] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2.$$

Fact

Given any finite collection of *independent* r.v. $X_1, \ldots, X_n$, we have

$$\mathrm{Var}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathrm{Var}[X_i]$$

Bernoulli and Binomial random variables

- Consider an experiment with success probability p, and call Y the r.v.:

$$Y = \begin{cases} 1 & \text{if the experiment succeeds} \\ 0 & \text{otherwise} \end{cases}$$

  Then, Y is called a *Bernoulli* or *indicator* r.v.

- $\mathrm{E}[Y] = p \cdot 1 + (1-p) \cdot 0 = p.$
- $\mathrm{Var}[Y] = \mathrm{E}[Y^2] - (\mathrm{E}[Y])^2 = p(1-p).$

Bernoulli and Binomial random variables

- Consider an experiment with success probability p, and call Y the r.v.:

$$Y = \begin{cases} 1 & \text{if the experiment succeeds} \\ 0 & \text{otherwise} \end{cases}$$

  Then, Y is called a *Bernoulli* or *indicator* r.v.

- *Binomial* r.v. X – $B(n,p)$ – denotes # successes in n independent trials.

$$\Pr\left[X = j\right] = \binom{n}{j} p^j (1-p)^{n-j}$$

- Can write X as sum of indicator r.v. $Y_1 + Y_2 + \cdots + Y_n$
- $\mathrm{E}[X] = np$ and $\mathrm{Var}[X] = np(1-p)$.

Geometric random variable

- Perform a sequence of independent trials until the first success.
- *Geometric* r.v. X with parameter p denotes # trials until first success.

$$\Pr\left[X = n\right] = (1-p)^{n-1}p$$

- $\mathrm{E}[X] = \sum_{n=1}^{\infty} n(1-p)^{n-1}p = 1/p^2 \cdot p = 1/p$
- Fact: $\frac{1}{(1-x)^2} = 1 + 2x + 3x^2 + 4x^3 + \cdots$.
- p.g.f. $G_X(t) = \mathrm{E}[\,t^X] = \frac{pt}{1-(1-p)t}$

Coupon collector's problem

Coupon Collector's Problem

- Each box of cereals contains one of n different coupons.
- Coupon in every box is chosen independently and uniformly at random.

How many boxes must we buy to obtain at least one coupon of every type?

- Let X_i be the # boxes to go from exactly $i-1$ different coupons to i.
  1. X_i is a geometric r.v. with parameter $p_i = 1 - \frac{i-1}{n}$.
  2. Total # of boxes $X = X_1 + X_2 + \cdots + X_n$.
- $\mathrm{E}[X] = \sum_{i=1}^{n} \mathrm{E}[X_i] = \sum_{i=1}^{n} \frac{n}{n-i+1} = n \cdot \sum_{i=1}^{n} \frac{1}{i} = n(\ln n + \Theta(1))$