

CSL 857: Randomized Algorithms

Ragesh Jaiswal
CSE, IIT Delhi

Deviation from Expectation

- Chebychev's Inequality: Let $\mathbf{X}$ be a random variable and $\mathbf{a} > 0$, then

$$\Pr[|\mathbf{X} - \mathbf{E}[\mathbf{X}]| \geq \mathbf{a}] \leq \mathbf{Var}[\mathbf{X}] / \mathbf{a}^2.$$

- Define the random variable $\mathbf{Y} = (\mathbf{X} - \mathbf{E}[\mathbf{X}])^2$.

- Lemma 1: $\mathbf{E}[\mathbf{Y}] = \mathbf{E}[\mathbf{X}^2] - \mathbf{E}[\mathbf{X}]^2 = \mathbf{Var}[\mathbf{X}]$

- Proof of Chebychev's inequality:

- $$\begin{aligned} \Pr[|\mathbf{X} - \mathbf{E}[\mathbf{X}]| \geq \mathbf{a}] &= \Pr[\mathbf{Y} \geq \mathbf{a}^2] \\ &\leq \mathbf{E}[\mathbf{Y}] / \mathbf{a}^2 && \text{(by Markov)} \\ &\leq \mathbf{Var}[\mathbf{X}] / \mathbf{a}^2 && \text{(by Lemma 1)} \end{aligned}$$

Deviation from Expectation: Example

- Birthday Problem: You uniformly sample q items with replacement from a collection of n items. What is the probability that two items are the same?
- Birthday Problem(popular version): There are q people in a room. What is the value of q such that the probability of two people having the same birthday is at least $\frac{1}{2}$. Each person's birthday is assumed to be a random day in the year.

Deviation from Expectation: Example

- Birthday Problem: You uniformly sample q items with replacement from a collection of n items. What is the probability that two items are the same?
- Let X_{ij} be an *indicator* random variable that is **1** if the i^{th} and the j^{th} person have the same birthday and **0** otherwise.
- Lemma 1: $E[X_{ij}] = 1/n$.
- Let $X = \sum_{i \neq j} X_{ij}$
- X denotes the number of distinct pairs of people with the same birthday.
- $E[X] = q(q-1)/(2n)$ (by linearity of Expectations)
- If $q \approx \sqrt{(2n)}$, then $E[X] > 1$.

Deviation from Expectation: Example

- Birthday Problem: You uniformly sample q items with replacement from a collection of n items. What is the probability that two items are the same?
- Let X_{ij} be an *indicator* random variable that is **1** if the i^{th} and the j^{th} person have the same birthday and **0** otherwise.
- Lemma 1: $E[X_{ij}] = 1/n$.
- Let $X = \sum_{i \neq j} X_{ij}$
- X denotes the number of distinct pairs of people with the same birthday.
- $E[X] = q(q-1)/(2n)$ (by linearity of Expectations)
- If $q \approx c\sqrt{(2n)}$, then $E[X] = 10$.
- Lemma 2: $\text{Var}[X_{ij}] = (n-1)/n^2$.
- Lemma 3: $\text{Var}[X] = \sum \text{Var}[X_{ij}]$.
- So $\text{Var}[X] = q(q-1)(n-1)/(2n^2) = 10 \cdot (1 - 1/n)$ for $q \approx c\sqrt{(2n)}$

Deviation from Expectation: Example

- Birthday Problem: You uniformly sample q items with replacement from a collection of n items. What is the probability that two items are the same?

- Let X_{ij} be an *indicator* random variable that is **1** if the i^{th} and the j^{th} person have the same birthday and **0** otherwise.
- Lemma 1: $E[X_{ij}] = 1/n$.
- Let $X = \sum_{i \neq j} X_{ij}$
- X denotes the number of distinct pairs of people with the same birthday.
- $E[X] = q(q-1)/(2n)$ (by linearity of Expectations)
- If $q \approx c\sqrt{(2n)}$, then $E[X] = 10$.
- Lemma 2: $\text{Var}[X_{ij}] = (n-1)/n^2$.
- Lemma 3: $\text{Var}[X] = \sum \text{Var}[X_{ij}]$.
- So $\text{Var}[X] = q(q-1)(n-1)/(2n^2) = 10 \cdot (1 - 1/n)$ for $q \approx c\sqrt{(2n)}$
- By Chebychev, we get $\Pr[X < 1] \leq \Pr[|X - E[X]| \geq 9] \leq 10/81 < 1/4$.

Birthday Problem

- Birthday Problem: You uniformly sample q items with replacement from a collection of n items. What is the probability that two items are the same?
- The previous analysis gives us that when q is $\Omega(\sqrt{2n})$, then there is better than $1/2$ chance that there is a collision.
- Question: Can we get a closed form expression for $C(n, q)$, the probability of collision, when q is small?
- Balls and bins: We throw q balls into n bins randomly. What is the probability that there is a bin that has more than one ball?
 - This is the same problem. The probability is $C(n, q)$.

Birthday Problem

- Balls and bins: We throw q balls into n bins randomly. What is the probability that there is a bin that has more than one ball?
- Claim 1: $C(n, q) \leq q(q-1)/(2n)$.
 - Proof:
 - Let C_i be the event that the i^{th} ball collides with one of the previous balls.
 - Lemma: $\Pr[C_i] \leq (i-1)/n$.
 - $C(n, q) = \Pr[C_1 \cup C_2 \cup \dots \cup C_q]$
 $\leq \Pr[C_1] + \Pr[C_2] + \dots + \Pr[C_q]$
 $\leq 0 + 1/n + \dots + (q-1)/n$
 $= q(q-1)/2n$.

Birthday Problem

- Balls and bins: We throw q balls into n bins randomly. What is the probability that there is a bin that has more than one ball?
- Claim 2: $C(n, q) \geq 1 - e^{-q(q-1)/2n}$.
 - Proof:
 - Let D_i be the event there are no collisions after i balls are thrown.
 - Lemma: $\Pr[D_{i+1} | D_i] = 1 - i/n$ and $\Pr[D_1] = 1$.
 - $1 - C(n, q) = \Pr[D_q] = \Pr[D_q | D_{q-1}] \cdot \Pr[D_{q-1}]$
 $= \prod \Pr[D_{i+1} | D_i]$
 $= \prod (1 - i/n) \leq e^{-\sum i/n} = e^{-q(q-1)/2n}$

Birthday Problem

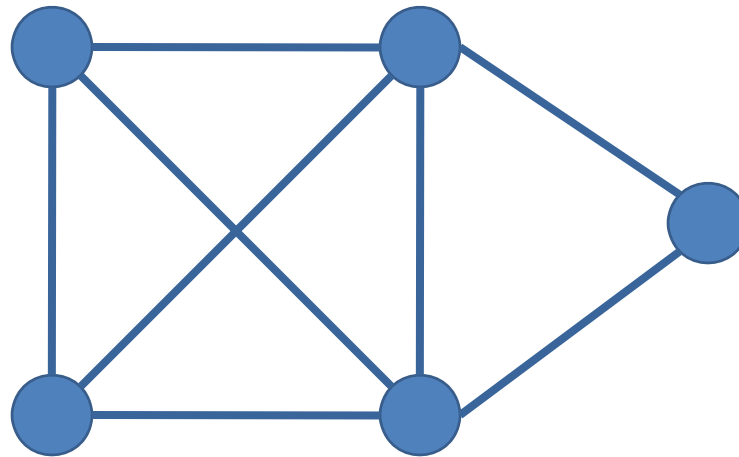
- Balls and bins: We throw q balls into n bins randomly. What is the probability that there is a bin that has more than one ball?
- Claim 2: $C(n, q) \geq 1 - e^{-q(q-1)/2n}$.
- Corollary: If $q \leq \sqrt{2n}$, then $C(n, q) \geq (1 - 1/e) \cdot q(q-1)/2n$.
 - Proof:
 - Use the fact that for $0 < x \leq 1, 1 - e^{-x} \geq (1 - 1/e)x$.

Randomized Algorithm

Karger's Min-Cut Algorithm

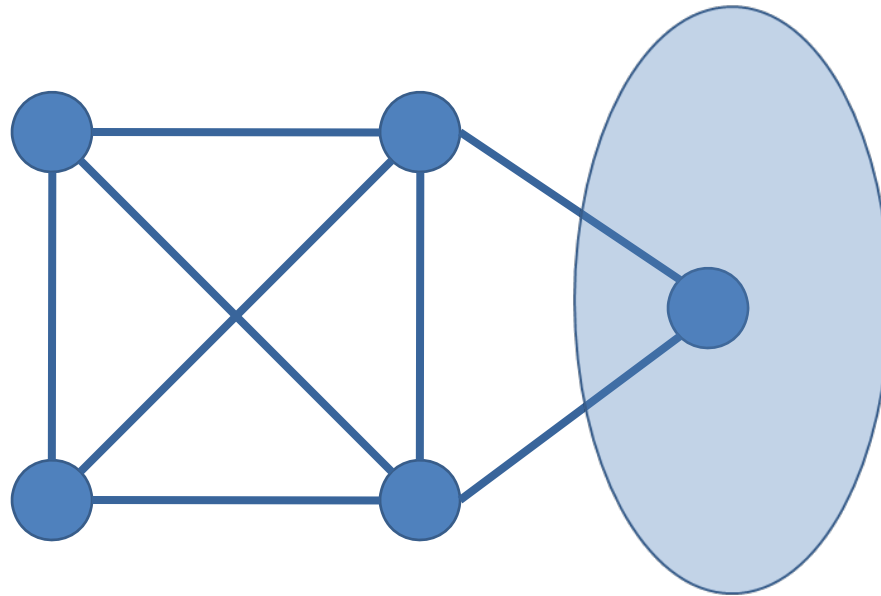
Karger's Min-Cut Algorithm

- Problem: Given a graph, find the cut of minimum size in the graph.
 - Cut: Partition of vertices into two non-empty sets. The edges between vertices in different sets are called cut edges. The size of the cut is the number of such edges.



Karger's Min-Cut Algorithm

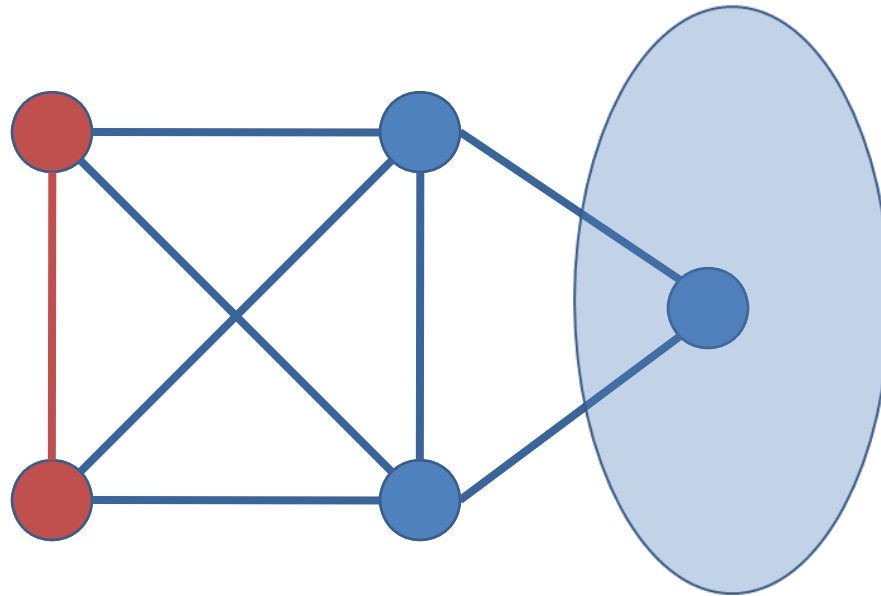
- Problem: Given a graph, find the cut of minimum size in the graph.
 - Cut: Partition of vertices into two non-empty sets. The edges between vertices in different sets are called cut edges. The size of the cut is the number of such edges.



Cut of minimum size

Karger's Min-Cut Algorithm

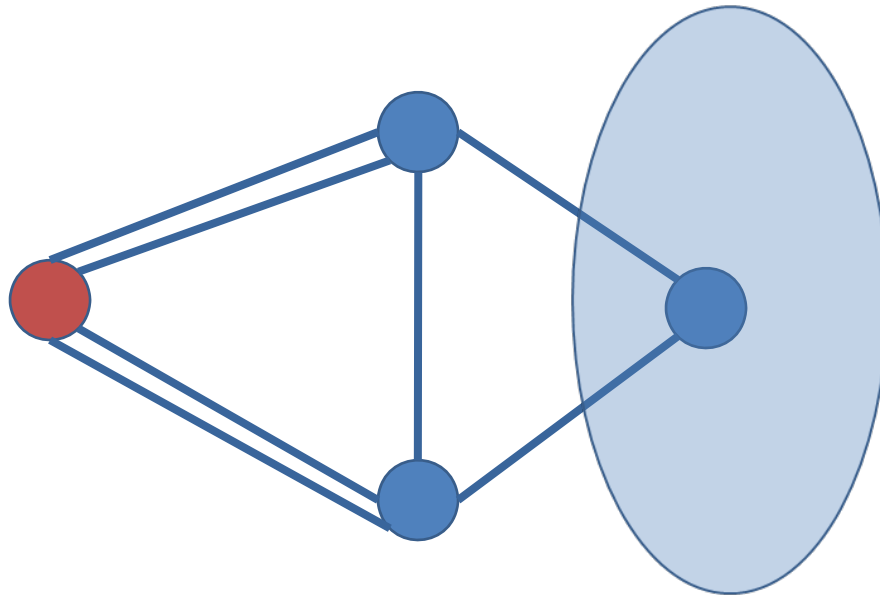
- *Collapse(e)*: Collapsing vertices across an edge $e=(u,v)$.
 - Merge vertices u and v and remove self loops. There may be multiple edges between same vertices (multi-graph).



Cut of minimum size

Karger's Min-Cut Algorithm

- *Collapse(e)*: Collapsing vertices across an edge $e=(u,v)$.
 - Merge vertices u and v and remove self loops. There may be multiple edges between same vertices (multi-graph).



Cut of minimum size

Karger's Min-Cut Algorithm

- Min-cut algorithm:
 - Repeat $\mathbf{P}$ times // For probability amplification.
 - While $|V| > 2$
 - Pick a random edge $\mathbf{e}$ in the multi-graph $\mathbf{G}$ and perform **Collapse**($\mathbf{e}$) to obtain $\mathbf{G}'$.
 - $\mathbf{G} \leftarrow \mathbf{G}'$
 - The edges across the remaining two vertices are the candidate cut edges.
 - Output the best answer from the $\mathbf{P}$ trials.

End
